

The Impact of Assessment Training on EFL Writing Classroom Assessment: Voices of Mexican University Teachers

El impacto de la capacitación en evaluación de la escritura en el aula de inglés como lengua extranjera: voces de profesores mexicanos universitarios

Elsa Fernanda González


Universidad Autónoma de Tamaulipas, Ciudad Victoria, Mexico

The extent to which writing assessment training (WAT) impacts writing scores has been widely explored in L2 testing contexts. However, little is known of the benefits of WAT to classroom assessment of writing. This paper analyzes the impact of two WAT sessions on the classroom assessment of writing of eleven EFL Mexican university teachers. Twenty-two interview transcripts suggested an impact in three main areas: classroom teaching of writing, classroom assessment of writing, and teacher self-awareness. The category of teacher self-awareness displayed the most impact. The paper proposes a categorization of impact on writing assessment.

Keywords: EFL writing assessment, language assessment literacy, writing assessment training

El impacto de la capacitación en evaluación de la escritura sobre los puntajes que los estudiantes obtienen en dicha habilidad se ha explorado ampliamente en contextos de exámenes a gran escala en L2. Sin embargo, se sabe poco de los beneficios de dicha capacitación en el aula. Este artículo analiza el impacto de dos sesiones de capacitación en la evaluación de la escritura en el salón de once profesores universitarios mexicanos de inglés como lengua extranjera. Veintidós transcripciones de entrevistas sugirieron impacto en tres áreas principales: enseñanza de la escritura en el aula, evaluación de la escritura en el aula y autoconciencia del profesor. La categoría de autoconciencia del profesor mostró el mayor impacto. El artículo propone una categorización del impacto en la evaluación de la escritura.

Palabras clave: alfabetización en evaluación de idiomas, capacitación en evaluación de escritura, evaluación de escritura en inglés como lengua extranjera

Elsa Fernanda González  <https://orcid.org/0000-0002-3990-0239> · Email: efgonzalez@docentes.uat.edu.mx

This article is based on the dissertation completed by González (2018).

This article draws upon a research project funded by the British Council and carried out under the ARAGS 2015 program. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the British Council, its related bodies, or its partners.

How to cite this article (APA, 7th ed.): González, E. F. (2021). The impact of assessment training on EFL writing classroom assessment: Voices of Mexican university teachers. *Profile: Issues in Teachers' Professional Development*, 23(1), 107–124. <https://doi.org/10.15446/profile.v23n1.85019>

This article was received on December 9, 2019 and accepted on October 5, 2020.

This is an Open Access article distributed under the terms of the Creative Commons license Attribution-NonCommercial-NoDerivatives 4.0 International License. Consultation is possible at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Introduction

In many English as a foreign language (EFL) contexts, the process of language assessment is part of an instructor's daily job (Lam, 2014; López-Mendoza & Bernal-Arandia, 2009), including the assessment of writing. Teachers' language assessment literacy (LAL), then, becomes of vital importance to ensure the quality of the process. Assessment training, a common LAL tool, has benefited from extensive research of its impact on rater scoring. However, little is known of how training may benefit the classroom assessment of writing (CAW). This paper analyzes the impact that two sessions of writing assessment training (WAT) had on the assessment practices of eleven Mexican EFL university teachers.

Literature Review

The assessment of EFL students' writing is crucial for its development. It is through assessment, in its different modalities, that a text may be improved (Crusan, 2010). It is a reflective process for which teachers need to know how to create fair assessments that provide information about their students' writing ability (Crusan et al., 2016, p. 43). Assessment literate teachers are aware of the potential consequences of inaccurate assessment (Stiggins, 1999) to the quality of texts and the development of writing skills. Teachers may also need to be aware of their involvement in a process that is influenced by contextual factors such as economy, social bonds, and the cultural practices of the institution (Chen et al., 2013; Yan et al., 2017). Managing the interaction with these factors in benefit of CAW becomes a necessary ability of those assessing students. Therefore, conducting writing assessment requires stakeholders to be assessment literate (Crusan, 2010, Crusan et al., 2016; Fulcher, 2012; Taylor, 2009; Weigle, 2007), that is, to have assessment knowledge to be capable of choosing how and when to use this knowledge (Coombe et al., 2012; Malone, 2013).

LAL has gained importance due to three factors: the worldwide use of large-scale test results (Inbar-Lourie, 2008), the role of tests in the globalization of languages, and

the need for teachers to implement assessment (Fulcher, 2012). It is a social practice (Inbar-Lourie, 2008) in which stakeholders' sociocultural perspectives (Scarino, 2013) are embedded in their interpretations and preconceptions of students' language performance. LAL goes beyond the knowledge of assessment, its practice in the classroom, and the involvement of social and contextual factors. It may also involve teachers' reflection processes of their assessment practices in the classroom as a means to reconceptualize their assessment knowledge and practice. Xu and Brown (2016), for instance, proposed the Teacher Assessment Literacy in Practice (TALiP) pyramid which had the purpose of portraying the interrelationship among theoretical knowledge of assessment, sociocultural perspectives, and the development of preservice and in-service teachers. They argued that the ultimate level of LAL is "Teacher Assessor (re)construction" since it is through professors' constant metacognitive activity combined with classroom experience that may lead to the improvement of assessment processes (p. 162).

Depending on stakeholders' roles and contextual factors, researchers have pointed out the need to expand the concept of LAL by grouping the degree of knowledge and skills required into three different levels (Taylor, 2013): (a) core LAL (researchers, test developers), (b) intermediary (language teachers, course instructors), and (c) peripheral levels (policy makers) which are constantly related to stakeholders' sociocultural values (Baker & Riches, 2018). These levels may face difficulties that hinder the development of LAL and writing assessment literacy (WAL) including stakeholder's lack of time, perceived anxiousness, fear, and a lack of WAL opportunities (Falvey & Cheng, 1995 [as cited in Coombe et al., 2012]; González & Vega-López, 2018; Lam, 2014; López-Mendoza & Bernal-Arandia, 2009).

Language Assessment Perceptions

Crusan et al. (2016) identified L2 tertiary teachers' ($N=702$) perspectives of writing assessment with a 54-item electronic survey. Most of the respondents (80%) had received

WAT throughout their postgraduate courses or as part of their teaching certification courses while others reported no previous assessment training experience ($n = 130$). Teachers had contradicting views of the nature of assessment and negative perceptions of training: 66% considered writing assessment always inaccurate, 60% considered it a subjective process while 80% considered rater training did not encourage them to improve their assessment.

Teachers have different views of assessment training courses and their contribution to assessment practice (González & Vega-López, 2018; Jeong, 2013; Koh et al., 2017; Lam, 2014; Malone, 2013; Nier et al., 2013). For instance, Nier et al. (2013) analyzed the perceptions of eighty EFL teachers to an online assessment course concluding that most of them considered it useful for their future assessment practice but more samples were needed to further understand the process of assessment. Similarly, in Mexico, González and Vega-López (2018) found that elementary school teachers held positive views of an online productive skill assessment course, but more time was needed to analyze its contents.

Teachers also have diverse perceptions of LAL needs (Fulcher, 2012; Lam, 2014; Vogt & Tsagari, 2014; Volante & Fazio, 2007). Vogt and Tsagari (2014) found that 42.4% of the surveyed teachers in primary, secondary, and tertiary levels in Europe claimed to have not received assessment training. Most of the teachers preferred receiving training to improve assessment of productive, receptive, and integrated skills as well as statistical analysis for language assessment. Training has become a widely used strategy to enhance LAL. However, knowledge of its impact varies.

Assessment Training Impact on Writing Assessment

Weigle (1994, 1998), Shohamy et al. (1992), and Elder et al. (2005, 2007) focused on the effects that rater training had on assessors and the scores provided in L2 assessment contexts. Elder et al. (2005) found that after eight experienced raters participated in an online training course, scored diagnostic writing papers, and received group and

individual feedback of their scoring performance, they viewed training as useful. This allowed raters to become aware of their own rating behavior and more consistent in their scoring; therefore, suggesting that feedback may have an impact on assessment procedures.

Much of the research has focused on the teachers' assessment knowledge, their experiences in assessment training, their LAL needs, their views of assessment, and the viewed impact of online assessment courses. Additionally, raters' views of assessment and the impact that training has on their rating in large scale tests has been widely explored. Research has yet to understand the relationship among LAL, classroom assessment, and Latin American assessors in foreign language contexts. Exploring the level of impact that training, as a mode of LAL, may produce in EFL instructors' CAW, may contribute to closing this gap. Considering that studies which focus on language assessment training have been conducted in Asia (Koh et al., 2017), North America (Malone, 2013, Nier et al., 2013; Weigle, 1998, 2007), Europe (Fulcher, 2012; Vogt & Tsagari, 2014), and Australia (Knoch, 2011), the Latin American context has remained underexplored. Therefore, this paper seeks to explore the impact that WAT had on teachers' CAW in the Mexican university EFL setting by answering the research question, to what extent does WAT impact EFL teachers' reported classroom assessment of students' writing skills?

Method

The study intended to provide the researcher's interpretation of a phenomenon in a particular research context at a specific period of time (Paltridge & Phakiti, 2010), without seeking generalization. Thus, a qualitative, interpretative-constructivist perspective (Creswell, 2015; Johnson et al., 2007) was adopted.

Research Context

The eleven EFL university teachers were in service at one of three public universities (Institution A, B, and C) or a language institute (Institution D) in the north-eastern

region of Mexico. Teachers among institutions A, B, and C assess their students following distinct procedures and criteria, without considering writing. At Institute D, teachers are required to assess the four language skills. Teachers are provided with the writing prompts and the analytic rubric to assess writing. Therefore, three of the participants (Institution D) assessed writing regularly at their institutions.

Participants

Following a convenience sampling method (Dörnyei, 2007), eleven teachers (TPs) were recruited: seven women

(22–52 years) and four men (24–45 years old). The least experienced was Olivia with one year whereas the most experienced was Maria with more than 20 years of experience (Table 1). Two men and three women TPs had a BA (applied linguistics, English language, or human resources) combined with a teaching certification (ICELT); two female participants and one male participant had an MA (applied linguistics, TESOL, bilingual education) and one male teacher was an applied linguistics undergraduate student. In an interest to maintain the anonymity of participants, pseudonyms were used to refer to each TP.

Table 1. Background Characteristics of Participating Teachers

TP	Gen	Age	Months TE	Academic background	Institution of work	Teach/ Assess Writing?	Use rubrics?	Assessment training	Rubric use training
Martin	M	28	84	BA Applied linguistics TKT/ICELT	PU	Always	Always	Yes	Yes
Elena	F	26	96	BA Administration	PU	Often	Often	Yes	Yes
Julio	M	41	84	BA Human resources TKT/ICELT	LI	Often	Always	Yes	Yes
Olivia	F	24	4	BA International affairs	PU	Often	Rarely	No	No
Alberto	M	40	108	MA Bilingual education	PU	Sometimes	Never	No	No
Karina	F	26	48	MA Applied linguistics	PU	Sometimes	Always	Yes	No
Jessica	F	22	48	BA Applied linguistics TKT/ICELT	LI	Always	Always	Yes	Yes
Maria	F	52	240	MA TESOL	PU	Often	Sometimes	No	Yes
Antonia	F	35	144	BA English language TKT/ICELT	PU	Sometimes	Hardly Ever	Yes	Yes
Silvia	F	44	120	BA English language TKT/ICELT	PU	Often	Sometimes	No	No
Octavio	M	23	24	BA Applied linguistics Student	LI	Always	Often	Yes	Yes

Note. LI= Language institute; PU = Public university; TE = Teaching experience.

TPs had in common the desire to improve their assessment practice, the type of students they worked with, and they were all required by their language institutions to assess language skills. Teachers who had these characteristics and agreed to participate were considered. Four participants had never experienced assessment training while the rest had minimal experience with training.

Data Collection Instruments

Semistructured Interviews

Two semistructured interviews were conducted with the intention of identifying possible changes in instructors' CAW. Interview protocols were followed (Creswell, 2013) while conducting the 22 interviews.

Interview 1 (see Appendix A) focused on TPs' assessment prior to WAT. It included 13 questions in Spanish or English providing interviewees with a comfortable environment in their natural language (Cohen et al., 2011). Both languages were used to avoid transcript translation diminish data objectivity (Pavlenko, 2007). TPs were given the option of choosing the language of their preference.

Interview 2 (see Appendix B) explored the changes that training had encouraged. Additionally, it intended to obtain data to compare with the information obtained from Interview 1. Interview 2 was also available in Spanish or English for participants to choose.

Writing Assessment Training (WAT)

Two sessions of WAT were provided to TPs by the researcher, one prior training (WAT1) and the second (WAT2) from six to eight months after session one. WAT1 focused on the analysis of the nature of EFL writing, its assessment and writing assessment practice using holistic and analytic rubrics. WAT2 gave priority to the importance of using a rubric as a classroom tool for assessment and to provide formative feedback to

students to enhance "assessing for learning" (Stiggins, 1999). It included opportunities for teachers to reflect on the role of assessment in their classrooms, and their current assessment processes to analyze their improvement. The contents of sessions were adapted by the researcher from the *CEFR Manual for Language Examinations* (Council of Europe, 2009a, 2009b), the *ALTE Manual for Language Test Development and Examining* (Council of Europe, 2011) and the principles suggested by Bachman and Palmer (2010).

Data Collection Procedures

Stage 1. The 11 participants signed an informed consent. The researcher individually contacted teachers via email, telephone, or social media to schedule each interview and the first WAT. Then, the pretraining interview was conducted and audiotaped with participants' consent.

Stage 2. WAT1 was led by the researcher. For approximately two hours and a half, participants interacted and provided their previous experiences assessing writing as well as their views of writing, and the importance of its assessment in their classrooms. WAT1 was provided at one of the participating institutions.

Stage 3. WAT2 was delivered by the researcher. During three and half hours teachers were encouraged to share their reflections of the changes they noticed in their assessment of writing since they experienced WAT1. Participants were encouraged to participate freely and to share experiences in a group-led discussion. Then, in accordance with the researcher, participants were scheduled to participate in the post-training interview.

Stage 4. Interview 2 was conducted two to three months' post WAT2 session with the purpose of providing teacher time to reflect on information discussed and to implement possible changes in their assessment processes. As in Interview 1, consent for its recording was obtained from participants.

Data Analysis

Interview transcript analysis followed a grounded theory approach (Strauss & Corbin, 1994). Theory was generated with the researcher's interpretations of the voice of the participant to understand their individual actions (Strauss & Corbin, 1994, p. 274), and the underlying reasoning grounded in participants' rationale for their actions and practice (Taber, 2000). Data analysis relied on themes and categories as they emerged in data.

Emerging themes were identified in data and grouped in main themes then clustered into subthemes. Finally, categories were signaled with the purpose of noting relationships among variables and the context in which participants were immersed. Each category was coded (Creswell, 2015). Emerging themes, subthemes, and categories were compared among participants to find similarities and differences. Finally, these were

compared within the pre-training and post-training data to obtain the impact of WAT.

Certain researcher bias should be considered since the main researcher of the study conducted the two WAT sessions, the two semistructured interviews, and was responsible for data analysis. To diminish this, the results were shared with an expert researcher in the field of applied linguistics. This external researcher reviewed the data and came up with her own analysis. Finally, results of both analyses were discussed and agreed upon.

Results

Data suggested that assessment training had an impact (Table 2) on (a) the teaching of writing in the EFL classroom, (b) the teachers' regular classroom assessment procedures, and (c) their self-awareness as an EFL teacher or as a classroom assessor.

Table 2. Impact of Assessment Training

Main theme	Subtheme	Teacher participant
Impact on classroom teaching	Writing activities	Silvia, Antonia, Olivia, Maria
	Feedback techniques	Silvia, Alberto, Antonia, Maria
Impact on classroom assessment	Assessment procedures	Martin, Karina, Octavio
	Scoring tools	Martin, Karina, Elena, Octavio, Olivia
Impact on teacher self-awareness	The nature of writing	Olivia, Maria
	The teaching writing	Silvia, Antonia, Octavio, Olivia, Maria
	Assessment procedures	Martin, Silvia, Alberto, Karina, Maria
	Writer stance	Silvia, Alberto, Antonia, Olivia
	Student stance	Olivia, Maria
No impact was experienced		Julio, Jessica

The Teaching of Writing in the EFL Classroom

Two subthemes emerged within this main theme: (a) Writing Activities and (b) Feedback Techniques.

Writing Activities

Silvia reported she continued giving importance to the assessment of other skills mainly because of time issues and EFL program demands. After training she reflected on the importance that writing has for language students and their need of it in their future professional lives. During Interview 2, she explained that she now includes more writing activities in her lessons and provides more feedback to students' texts 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. (Silvia)

Maria reported receiving the most impact after experiencing assessment training. During Interview 2, she reported to have more interest in writing and its treatment in her classroom. 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. 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 based 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 training: (a) Implementation of writing techniques, (b) Increase of writing activities such as the case of Silvia, and (c) Focus of activities on students' needs carried out by Maria as portrayed in the previous interview excerpt.

Feedback Techniques

Three categories were identified from teachers' feedback activities after training: (a) Implementation of varied feedback techniques, (b) Increase of feedback provision, and (c) Improvement of feedback provision. For instance, Alberto explained his feedback had changed post-training. He was more careful and precise in the comments he provided to his students. He became aware of the importance of feedback in students' development of skills and in their assessment. He pointed out:

I only read the text and provided comments...I believe that in the new methods, to assess students' feedback 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. (Interview 2)

Maria, in addition to increasing the number of writing activities done in the classroom, modified her feedback focus by paying attention to the genre and the structure of the text: "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."

Maria paid more attention to the type of feedback she gave to her students specifically on language accuracy.

Classroom Assessment of Writing

Two subthemes were identified in this second emerging main theme: (a) classroom procedures to assess writing and (b) PTs' use of scoring tools to assess writing.

Classroom Procedures to Assess Writing

Martin suggested he had (a) reoriented his assessment purpose; Karina reported she had (b), increased her provision of assessment feedback while Octavio explained (c), an increase of the use of assessment techniques.

Martin reported that he modified his leniency in his assessment of students' texts in addition to his use of holistic rubrics to manage his time. 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 the text without looking at their writing...especially when I am expecting something from them I was perhaps demanding a higher proficiency....The sessions helped me notice that. (Interview 2)

Karina reported that prior to WAT she would directly provide the corrections to the text and score it according to her personal judgment without feedback. She pointed out that post WAT she had been able to implement change to her regular procedure by explaining to students prior to 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. (Interview 2)

Octavio explained, during Interview 2, that he did not consider encouraging students' self-assessment of writing. Post-training, he had managed to encourage 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 self-assessment. 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 and 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 their self and teacher evaluation.

After Octavio's implementation, his 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.

Use of Scoring Tools

From this second subtheme of the emerging theme Classroom Assessment Procedures, four categories emerged: (a) innovation of current scoring tool (exemplified with Martin), (b) implementation of scoring tool (portrayed by Karina), (c) adaptation of current scoring tool (pointed out by Elena) and (d) a combination of scoring tools (reported by Octavio).

Martin reported to have changed his scoring tool post WAT. 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 global score to students' work and for managing his time more wisely.

Karina explained that, prior to training, she used an analytic scoring tool that she later found did not suit the capabilities of her students. She described how she adapted a rubric 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 and self-assessment of their texts:

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?” (Interview 2)

During Interview 2, Elena explained she did not change her assessment methods, but she considered WAT had allowed her to adapt a different tool to her classroom needs. She considered that this cannot be done in every class because writing assessment, assessment tools, and assessment purposes depend on the institutional EFL program’s goals and students’ proficiency.

In relation to analytic and holistic scoring tools, Octavio combined the use of both types of rubrics not just to assess students’ written work but also as a tool to provide feedback. This change in conceptualization of rubric and its use allowed this participant to improve his assessment procedures in his classroom, which resulted in students’ understanding of the task and a smoother assessment of writing skills. Finally, he added that WAT had also allowed him to feel more confident in his assessment procedures and to have an objective explanation to a specific grade given to students’ texts, therefore setting aside his personal views.

Teacher Self-Awareness

Five subthemes were identified that corresponded to this main theme: (a) The Nature of Writing, (b) Teaching of Writing, (c) Assessment Practices, (d) Writer Stance, and (e) Students’ Stances.

The Nature of Writing

Data revealed two categories in regard to teachers’ reflection in this subtheme: (a) importance of writing for a language student (Maria, Antonia) and (b) the social role of learning to write (Octavio). Maria 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 in Interview 2:

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 [writing] and find a way to do it and give it a little time for feedback.

Antonia added that WAT had allowed her to recall the importance of writing for students’ future professional life. Octavio explained he had reflected on how writing is an activity that is best learnt as a social activity: “I was able to see how writing works better when shared with someone.”

Teaching of Writing

This subtheme aimed to explain teachers’ reflections about the process of writing instruction. It reflected emerging categories such as (a) improvement of teaching skills, (b) future inclusion of writing, (c) future inclusion of feedback, and (d) future inclusion of process writing. These categories were reported by Silvia, Antonia, Octavio, Olivia, and Maria.

Silvia was now aware that assessment could be standardized with the use of an assessment tool. But she needed to give more importance first to its teaching then move on to its assessment. She found her lack of organizational skills affected her lack of change in the classroom. Antonia explained that her assessment process in the classroom was not changed but she was still seeking to increase the assessment of writing in the classroom. She added that she increased her writing activities in the classroom even though it required large amounts of time.

Writing Assessment Procedures

Data suggested that WAT allowed teachers to reflect on the procedures they followed to assess their students’ work. TPS reported to have analyzed how to: (a) update their assessment techniques (Julio), (b) update their assessment procedure (Martin and Elena), (c) begin planning future assessment (Karina and Alberto), (d)

make teaching/assessment purposes agree (Maria), and (e) consider students' needs (Antonia). Alberto explained he had noticed the need to implement tools to assess his students' writing. He specified he did not change his CAW post assessment training. However, the sessions allowed him to reflect on his present assessment procedure. Karina explained that training had helped her reflect on her classroom procedures, how she conducted them, and therefore plan how she could improve her assessment to make it an easier task for her and more reflective for her students. She explained:

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 analyze 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. (Interview 2)

Julio continued assessing students' written work with their portfolio work and a monthly exam that included a writing component. However, he reflected on how his actual assessment methods could be improved, 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:

I'm thinking a little bit more on changing the way we evaluate students; I consider of course the portfolio...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"...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. (Interview 2)

Olivia explained she felt more confident and secure of scores provided when students, parents, or the administration required further explanations.

She manifested she had found a way of combining the institutional requirements with her own assessment beliefs.

Writer Stance

This fourth subtheme describes participants' conceptualization of themselves as writers and how it changed post WAT. TPS analyzed and reflected on their performance as writers emphasizing their (a) weaknesses as a writer (Silvia and Antonia), (b) improvement of teacher writing to improve student writing (Antonia), and (c) strengths as a writer (Alberto and Olivia). In this sense, Silvia reflected on the need for her to write to therefore transmit to students the skills needed to develop a text. She became aware of her weaknesses and her needs as a novice writer and explained:

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 [in], I'm not good at writing, so to be able to teach you need to know how to do it. (Interview 2)

Antonia explained that, after the sessions, she had been able to analyze herself and conclude that she had weaknesses that needed to be improved upon, and if improved upon, there would be a possibility of providing more quality feedback and writing assessment in the classroom. Silvia and Antonia, during Interview 2, stated that their reflection on their writing weaknesses led them to visualize their needs of further training and began planning to seek for other courses or workshops to attend.

Student Stance

This fifth subtheme identified was TPS' views from the perspective of a student writer. TPS reported to have reflected on themselves as writers who are constantly being evaluated in their programs of study and in their working environment. Participants Olivia and Maria

were impacted in their (a) stance as a student in their understanding of assessment knowledge and became aware of (b) their performance as a student while being a BA or master's student. For instance, Olivia explained that WAT had helped her understand the considerations when producing a text and when being assessed by her BA professors. Secondly, it allowed her to better understand the use of scoring tools, and/or to adapt them to her and her students' needs. Olivia explained she had changed her perspective as a student and as a teacher about assessment. Before taking the WAT, 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...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. (Interview 2)

With this excerpt, it may be inferred that, as a student, Olivia 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 the use of scoring tools. Corresponding to this view, Maria 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 gained a deeper understanding of her writing performance and her professors' assessment.

Discussion and Conclusions

This study analyzed the extent to which WAT impacted EFL teachers' reported CAW. Nine of the eleven TPs described changes in their regular assessment

procedures (Table 2). Those that reported so (Martin, Karina, Octavio, and Olivia) explained that minor changes conducted included a redefinition of assessment purposes, student participation in assessment process (Leung & Mohan, 2004), and an improvement of the assessment process. Time was found a constraint: TPs needed more time to process information and implement assessment change.

WAT was found to strongly impact TPs meta-cognitive skills, reflecting on themselves as writers, as teachers, and assessors who take an active part in their institution's assessment procedures. Change was found in the self-awareness of the nature of writing, the importance of teaching writing, the importance of writing assessment, and teachers' stances as a writer and as an EFL teacher (Koh et al., 2017; Lam, 2014; Scarino, 2013). Data suggested that while experiencing WAT with a group of peers that faced the same contextual difficulties, TPs became aware of the importance of the socialization of assessment (Koh et al., 2017; Lam, 2014; Scarino, 2013, 2017), encouraged the understanding of their own knowledge of assessment, and made way for the understanding of new knowledge (Scarino, 2013, 2017) thus triggering their self-awareness skills.

WAT and Impact on Teachers' Assessment Behaviors and Perceptions

Results of this study emphasized that the impact of WAT may be minor in teachers' classroom assessment (Koh et al., 2017), but beneficial for other aspects of teachers' LAL such as their conceptualizations and interpretations of assessment, as suggested by Scarino (2013, 2017). This finding may suggest that training impacted positively participants' awareness of the importance of assessment rather than on the improvement of assessment procedures. Positive impact on classroom assessment as an effect of training may be hard to obtain and is rarely measured (Jin &

Jie, 2017). However, the analysis of this impact may enlighten the understanding of the benefits of WAT. These may have important implications for the EFL teacher. For instance, training as a tool leading to the socialization of assessment procedures among relevant stakeholders may help improve teachers' confidence in their knowledge and therefore lead to improved classroom assessment. Since time is an issue that impacts the assessment process, it is crucial for teachers to consider writing assessment as an ongoing process that seeks improvement constantly with the support of decision makers to improve the CAW. Finally, TPS' perceptions may suggest that training may also be beneficial when accounted for as an ongoing process instead of a one-time opportunity.

This study may correspond to the study led by Koh et al. (2017) in which the implementation of follow-up measures, such as permanent training or reflective sessions, supported by educational institution authorities, may provide teachers with the opportunities to improve their practice. According to Alberto and Antonia, if training were more constant, their assessment in the classroom may change.

Results of this paper may support Crusan et al.'s (2016) results where participants (80%) considered rater training did not encourage them to improve their assessment. In this study, nine of eleven TPS confirmed that innovation in their assessment post WAT did not occur, confirming that impact on assessment is quite complex.

WAT and Impact on Teachers' Future Academic Development

Results suggested that WAT may trigger future training sessions. Silvia, Antonia, and Maria described

their intention of seeking other writing assessment courses. Experiencing WAT triggered their reflection regarding their strengths and weaknesses as teachers and as assessors. Therefore, it can be argued that training may be an initial step that may trigger further assessment literacy and/or professional development.

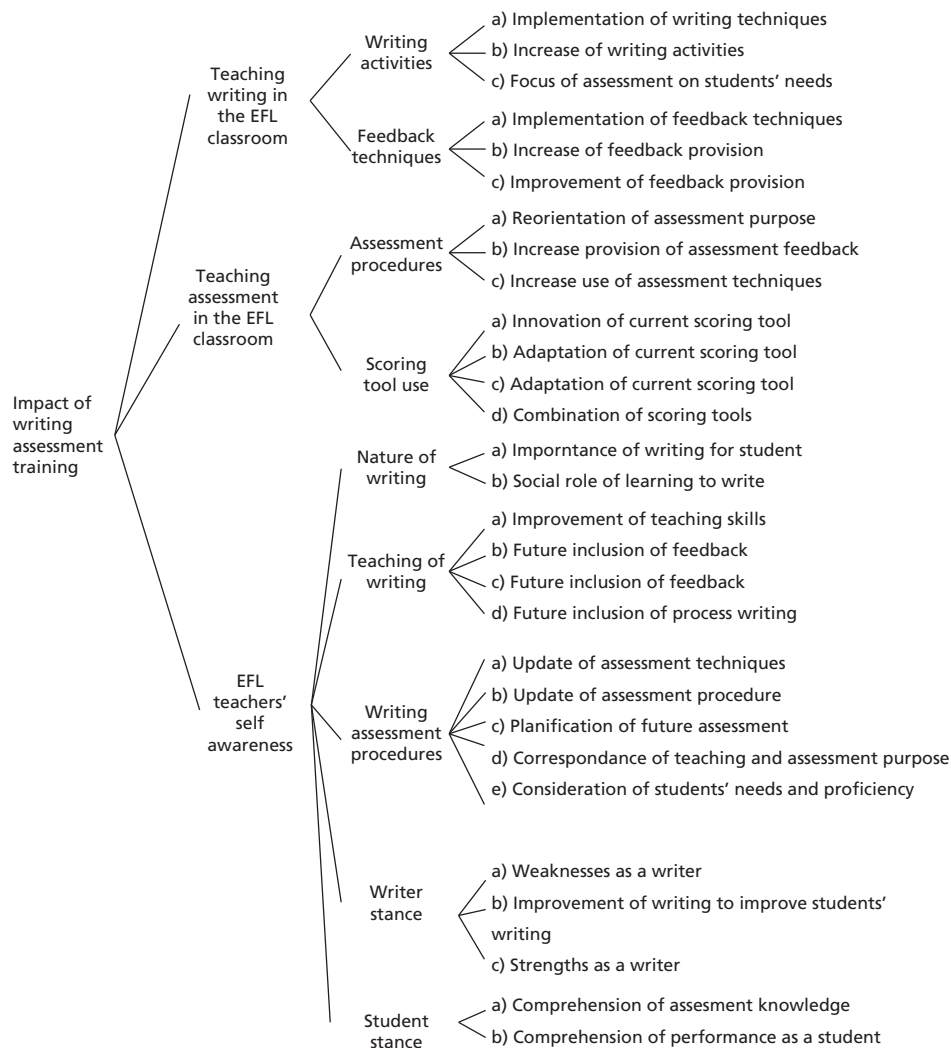
The Writing Assessment Training Impact Categorization

Data led to the initial construction of the writing assessment training impact categorization (WATIC; Figure 1), which categorizes WAT impact in this specific Mexican EFL context. It acknowledges the importance of contextual factors (Crusan, 2010; Yan et al., 2017) such as institutional policies or the nature of the EFL program being taught.

The WATIC is a multi-level assessment impact construct compiled from the themes, subthemes, and categories that emerged from the data. The first level includes the three major areas of impact: Writing in the EFL Classroom, Classroom Assessment of EFL Writing, and Teacher Self-Awareness (Level 1). Each subtheme represents the actions that TPS reported and that represented, in the TPS' perception and my own, the effect of training in their writing assessment practice. Each subtheme portrays from two to five categories (Level 3).

The WATIC may exemplify the conceptual framework of the TALIP proposed by Xu and Brown (2016), specifically at the top level of the construct "assessor identity (re)construction" during which teachers reconstruct their identity and stances as assessor teachers. This may be portrayed in the third level of the WATIC where teacher self-awareness is projected as the level that enhanced teacher metacognition.

Figure 1. Writing Assessment Training Impact Categorization (WATIC)



Limitations

The construction of the WATIC included the qualitative views of 11 active EFL teachers. Although qualitative insight may not pursue the generalization of knowledge (Cohen et al., 2011; Dörnyei, 2007), data from more participants could provide a wider perspective of TPs experiences allowing for categories to be added or deleted. Additionally, the construction of the WATIC is an initial attempt to portray impact, so further research could be conducted for its validation. Validating the specific categories that attempt to

describe the effects of WAT are crucial since this would allow the exploration of different assessment contexts to obtain a more objective categorization.

An important limitation of the study was that the researcher was the trainer and the interviewer of the study. Thus, data could be influenced by the TPs' desire of performing how they believe the researcher expects them to (Dörnyei, 2007, p. 53). To diminish this, future research could consider the triangulation of data with different qualitative instruments such as assessment observation and document analysis.

Implications

TPs' perceptions may portray the difficulty of reaching positive impact on classroom assessment as a result of WAT. They suggest that teacher self-reflection is a complex process that takes time, thus EFL teachers would benefit from multiple sessions that can encourage the reflection of assessment procedures. Additionally, the WATIC may guide teachers in the identification of their assessment strengths and the decision making to improve their weaknesses. The WATIC may raise awareness among teacher trainers, language managers, and heads of educational/language institutions of the staff needs for LAL. It may also aid the identification of the potential benefits of providing teachers with WAT that could lead to training sessions that are cost and time feasible.

References

- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford University Press.
- Baker, B. A., & Riches, C. (2018). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing*, 35(4), 557–581. <https://doi.org/10.1177/0265532217716732>
- Chen, Q., Kettle, M. A., Klenowski, V., & 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. <https://doi.org/10.1080/02602938.2012.726963>
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed). Routledge.
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges and recommendations. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyhoff (Eds.), *The Cambridge guide to second language assessment* (pp. 20–29). Cambridge University Press.
- Council of Europe. (2009a). *The manual for language test development and examination*.
- Council of Europe. (2009b). *Relating language examinations to the common European framework of reference for languages: Learning, teaching, assessment (CEFR): A manual*. <https://bit.ly/3fsDXNR>
- Council of Europe. (2011). *Manual for language test development and examining: For use with the CEFR*. <https://bit.ly/3ftyojR>
- Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.
- Creswell, J. W. (2015). *A concise introduction to mixed methods research*. Sage Publications.
- Crusan, D. (2010). *Assessment in the second language writing classroom*. The University of Michigan Press. <https://doi.org/10.3998/mpub.770334>
- Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing*, 28, 43–56. <https://doi.org/10.1016/j.asw.2016.03.001>
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. Oxford University Press.
- 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. <https://doi.org/10.1177/0265532207071511>
- Elder, C., Knoch, U., Barkhuizen, G., & von Randow, J. (2005). Individual feedback to enhance rater training: Does it work? *Language Assessment Quarterly*, 2(3), 175–196. https://doi.org/10.1207/s15434311laq0203_1
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- González, E. F. (2018). *The impact of assessment training on English as a foreign language university professors classroom writing assessment: Reported practice and perceptions* [Doctoral dissertation, University of Southampton]. <https://bit.ly/3em1al7>
- González, E. F., & Vega-López, N. A. (2018). Exploring Mexican EFL elementary school teachers' perceptions of online language assessment training. *Journal of Applied Linguistics and Language Research*, 5(1), 225–241.

- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402. <https://doi.org/10.1177/0265532208090158>
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, 30(3), 345–362. <https://doi.org/10.1177/0265532213480334>
- Jin, Y., & Jie, W. (2017, July 17–21). *Do workshops really work? Evaluating the effectiveness of training in language assessment literacy* [Paper presentation]. 39th Language Testing Research Colloquium, Bogotá, Colombia.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112–133. <https://doi.org/10.1177/1558689806298224>
- Knoch, U. (2011). Investigating the effectiveness of individualized feedback to rating behavior: A longitudinal study. *Language Testing*, 28(2), 179–200. <https://doi.org/10.1177/0265532210384252>
- Koh, K., Burke, L. E. C., Luke, A., Gong, W., & Tan, C. (2017). Developing assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, 22(3), 264–288. <https://doi.org/10.1177/1362168816684366>
- Lam, R. (2014). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197. <https://doi.org/10.1177/0265532214554321>
- Leung, C., & Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts: assessment as discourse and assessment of discourse. *Language Testing*, 21(3), 335–359. <https://doi.org/10.1191/0265532204lt2870a>
- López-Mendoza, A. A., & Bernal-Arandia, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile: Issues in Teacher' Professional Development*, 11(2), 55–70.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329–344. <https://doi.org/10.1177/0265532213480129>
- Nier, V. C., Donovan, A. E., & Malone, M. E. (2013, October). Promoting assessment literacy for language instructors through an online course [Poster presentation]. East Coast Organization of Language Testers Conference, Washington, D.C.
- Paltridge, B., & Phakiti, A. (Eds.). (2010). *Continuum companion to research methods in applied linguistics*. Continuum International.
- Pavlenko, A. (2007). Autobiographic narratives as data in applied linguistics. *Applied Linguistics*, 28(2), 163–188. <https://doi.org/10.1093/applin/amm008>
- 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. <https://doi.org/10.1177/0265532213480128>
- 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.
- Shohamy, E., Gordon, C. M., & 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. <https://doi.org/10.1111/j.1540-4781.1992.tb02574.x>
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 18(1), 23–27. <https://doi.org/10.1111/j.1745-3992.1999.tb00004.x>
- Strauss, A. L., & Corbin, J. (1994). Grounded theory methodology: An overview. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (1st ed., pp. 273–285). Sage Publications.
- Taber, K. S. (2000). Case studies and generalizability: Grounded theory and research in science education. *International Journal of Science Education*, 22(5), 469–487. <https://doi.org/10.1080/095006900289732>
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36. <https://doi.org/10.1017/S0267190509090035>

- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412. <https://doi.org/10.1177/0265532213480338>
- Vogt, K., & Tzagari, D. (2014). Assessment literacy for foreign language teachers: Findings of a European university. *Language Assessment Quarterly*, 11(4), 374–402. <https://doi.org/10.1080/15434303.2014.960046>
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749–770. <https://doi.org/10.2307/20466661>
- Weigle, S. C. (1994). Effects of training on raters of ESL compositions. *Language Testing*, 11(2), 197–223. <https://doi.org/10.1177/026553229401100206>
- Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing*, 15(2), 263–287. <https://doi.org/10.1177/026553229801500205>
- Weigle, S. C. (2007). Teaching writing teachers about assessment. *Journal of Second Language Writing*, 16(3), 194–209. <https://doi.org/10.1016/j.jslw.2007.07.004>
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Yan, X., Fan, J., & Zhang, C. (2017, July 17–21). *Understanding language assessment literacy profiles of different stakeholder groups in China: The importance of contextual and experiential factors* [Paper presentation]. 39th Language Testing Research Colloquium, Bogotá, Colombia.

About the Author

Elsa Fernanda González holds a PhD in Applied Linguistics from the University of Southampton. She is a recipient of the 2015 Research Assessment Award from the British Council and the 2019 Latin American Research Grant awarded by Michigan Language Assessment. She currently coordinates the BA in applied linguistics program at Universidad Autónoma de Tamaulipas.

Appendix A: Outline of Teacher Interview 1

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 semesterly 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?

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 the rubric provided improved your scoring of the writing samples?

The training session.

8. Do you consider that 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?

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 B: Outline of Teacher Interview 2 (Post-Training Sessions)

1. Do you continue assessing writing in your EFL classroom?
2. If so, how do you do it?
3. What changes, if applicable, have you implemented in your assessment of writing after the training session?
Why or why not?
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? Why or why not?
8. How has the training session helped you in your writing assessment practice? Why or why not?
9. What changes would you make in the training sessions?
10. How do you feel about writing after taking the training sessions?
11. How do you feel about writing assessment after taking the training session?
12. Do you have any additional comments?