



# Evaluating an Online Language Assessment Course: Colombian Teachers' Voices

Frank Giraldo<sup>1</sup> 

Lenny García<sup>2</sup> 

**Citation:** Giraldo, F. and García, L. (2025). Evaluating an Online Language Assessment Course: Colombian Teachers' Voices. *Colomb. Appl. Linguistic. J.*, 27(1), pp. 112-125.

**Received:** 02-May.-2024 / **Accepted:** 07-Jan.-2025

**DOI:** <https://doi.org/10.14483/22487085.22116>

## Abstract

Language assessment literacy (LAL) has emerged as a relatively recent focus in language testing, especially when it comes to fostering this dimension among stakeholders. There are few reports on online assessment courses for English language teachers that discuss their usefulness regarding teachers' LAL. However, there is a growing call to advance LAL pedagogies and to investigate how assessment courses contribute to teachers' professional development. In this study, we employed a qualitative research design to ask six Colombian English language teachers to evaluate the contents and activities of an online assessment course specifically designed for them. Data were collected through semi-structured interviews. Our findings indicate that teachers engaged in critical reflection on the course contents, activities, and —considering both what was effective and what was not—by referencing their own assessment practices. Moreover, the findings suggest that course activities, especially quizzes, were instructionally sound, as they promoted a formative assessment approach that, in turn, contributed to development of the teachers' LAL.

**Keywords:** language assessment, language assessment courses, language assessment literacy, teachers' professional development

<sup>1</sup> Frank Giraldo is a professor at the Foreign Languages Department of Universidad de Caldas in Manizales, Colombia. [frank.giraldo@ucaldas.edu.co](mailto:frank.giraldo@ucaldas.edu.co)

<sup>2</sup> Lenny Vanessa García Rojas is an 8th semester undergraduate student of the BA in Modern Languages at Universidad de Caldas in Manizales, Colombia. [lenny.221920291@ucaldas.edu.co](mailto:lenny.221920291@ucaldas.edu.co)



# El análisis de un curso en línea de evaluación de idiomas: las voces de docentes colombianos

## Resumen

La literacidad en la evaluación de lenguas (LEL) es un tema de discusión relativamente reciente en el campo de la evaluación de idiomas. Son escasos los estudios que exploran la utilidad de los cursos de evaluación en línea para el desarrollo de la LEL de docentes de inglés. No obstante, existe un creciente llamado a problematizar las pedagogías de la LEL con el fin de comprender cómo este tipo de cursos contribuye al desarrollo profesional de estos docentes. Este estudio adoptó un diseño cualitativo y convocó a seis profesores de inglés en Colombia para que analizaran los contenidos y actividades de un curso en línea de evaluación diseñado específicamente para ellos. Con este propósito, se utilizaron entrevistas individuales semiestructuradas. Nuestros hallazgos indican que los docentes tuvieron en cuenta sus contextos de evaluación para reflexionar críticamente sobre los contenidos, las actividades y los retos del curso. Además, los hallazgos sugieren que las actividades del curso, en especial los quizzes, fueron pedagógicamente apropiadas, ya que promovieron la evaluación de enfoque formativo, lo que en últimas conllevó al desarrollo de la LEL de estos docentes.

*Palabras clave:* cursos de evaluación de lenguas, desarrollo profesional docente, evaluación de lenguas, literacidad en evaluación de lenguas.

## Introduction

In the field of language testing, research and conceptual discussions have focused on two interrelated aspects of language assessment literacy (LAL). First, there has been a sustained effort toward construct definition (Davies, 2008; Fulcher, 2012; 2022; Inbar-Lourie, 2008; 2012; Mohammadkhah *et al.*, 2022). Scholarly work in this dimension has suggested that, although LAL broadly comprises knowledge, skills, and principles for language assessment, it must be understood within the specific contexts of stakeholders (Coombe *et al.*, 2020; Csépes, 2021; Cui *et al.*, 2022). In other words, it may not be a good idea to prescribe a checklist approach for what people involved in assessment should know and do (Inbar-Lourie, 2017; Csépes, 2021). Second, researchers have discussed the LAL profiles of various stakeholders, namely teachers, students, professional language testers, among others (Butler, 2022; Kremmel & Harding, 2020; Malone, 2017). Unsurprisingly, much of this literature has concentrated on describing teachers' LAL and professional development scenarios (e.g., workshops or courses) to advance this dimension of their profession.

While research on LAL as a construct and on how it is reflected in stakeholders' profiles is ongoing, there has increasingly been a call for advancing LAL pedagogies. Thus, researchers have analyzed current language assessment courses for teachers and suggested major features for LAL training (Brindley, 2001; Levi & Inbar-Lourie, 2019; Stabler-Havener, 2018). From these studies, authors conclude that preparing language teachers to conduct professional practices is an ongoing need. Key features of assessment such as describing language ability, the design of assessment tools, and ethical concerns in assessment are commonly treated in both empirical and conceptual research.

Additionally, several research studies have reported on the design, implementation, and evaluation of language assessment training (e.g., courses) for teachers. Collectively, these studies suggest that such training increases teachers' awareness of what language assessment entails and its potential uses (Baker & Riches, 2018; Giraldo & Murcia, 2019; Lao & Yan, 2024; Montee *et al.*, 2013; Restrepo-Bolívar, 2020); become conscientious of the rigor needed to analyze and/or design assessment instruments (Cui *et al.*, 2022; Kleinsasser, 2005; Koh *et al.*, 2018); and problematize the relationship between assessment, teaching, and learning. In several of these studies (e.g., Kremmel *et al.*, 2018; Montee *et al.*, 2013), teachers provide opinions on the contents they studied, mostly from a quantitative perspective; however, there are few reports that utilize a qualitative approach to data collection and analysis. Further, there are few reports in Colombia that delve into teachers' ideas on their LAL, vis-à-vis the contents and activities in LAL courses. Against this background, the purpose of our research study<sup>3</sup> was to examine Colombian teachers' evaluation of an online language assessment course in relation to their LAL development. Specifically, we asked them for their perceptions of how course content and activities fostered their LAL. The following research question guided our qualitative inquiry: What are the perceptions of six Colombian English language teachers of an online assessment course?

We believe that this paper may be significant in three related ways. First, we consider this a contribution on how to teach language assessment, online, to in-service language teachers in Colombia. Second, our experience may be useful for educators tasked with designing assessment courses, whether in this country or elsewhere. Ultimately, with the findings and insights from our study, we seek to advance discussions of LAL pedagogies for teachers by highlighting how their voices can help to problematize this area of language testing and assessment.

## Literature Review

### *Assessment Literacy*

Stiggins (1995) was perhaps the first scholar to refer to assessment literacy (AL), defining this construct as the knowledge and skills needed to understand the purposes, skills, and methods to conduct assessment, so as to collect reliable information regarding student learning. Importantly, according to Stiggins, assessment-literate teachers are aware of the negative consequences of poor assessment practice. More recently, authors have argued that assessment

---

<sup>3</sup> This article is part of a research project called *El Diseño y la Implementación de un Curso Online Masivo y Abierto (MOOC) en Evaluación de Lenguas Extranjeras* (Designing and Implementing a Massive Open Online Course (MOOC) on Language Assessment). The project was sponsored by the Vicerrectoria de Investigaciones y Posgrados at Universidad de Caldas in Manizales, Colombia. The code for this project was 0727622.

literacy is highly contextual and profile-based ([Coombs & DeLuca, 2022](#); [Pastore, 2023](#)): In addition to knowledge and skills related to assessment, teachers' beliefs, contexts, attitudes, and emotions affect their assessment literacy. In language testing, language assessment literacy (LAL) has adopted some of the generic components of AL but has, naturally, focused on language as a central component of the construct ([Giraldo, 2018](#); [Inbar-Lourie, 2008](#)).

### *Language Assessment Literacy*

[Davies \(2008\)](#) was amongst the first scholars in language testing to bring attention to the fundamental needs in professional assessment. Davies states that those involved in language testing need knowledge of language ability and measurement; skills to design tests; and fairness and ethics as guiding principles. Other authors have used and expanded on [Davies' \(2008\)](#) ideas. For example, [Inbar-Lourie \(2008; 2012\)](#) discusses LAL from three questions: the why (or purposes) of assessment, the what (language ability), and the how (methods). Additionally, various authors have described LAL granularly, suggesting profiles and competencies ([Giraldo, 2018](#); [Kremmel & Harding, 2020](#); [Malone, 2017](#); [Rezai, 2024](#); [Taylor, 2013](#)). While not exhaustive, the following is a list of knowledge, skills, and principles that appear in scholarly discussions regarding LAL:

- *Knowledge*: of language ability frameworks and models; assessment-related concepts and theories (e.g., validity and reliability; assessment of / for / as learning, etc.); measurement theory; policies relating to language assessment; codes of ethics and practice; consequences of testing (both intended and unintended); assessment beliefs, contexts, and uses; assessment methods.
- *Skills*: test evaluation; design of various assessment instruments, both traditional and alternative; statistical calculations (basic and/or advanced); score interpretation of large-scale and classroom tests.
- *Principles*: consequential validity, democracy, ethics, fairness, transparency, and more recently, social justice (see [Randall et al., 2023](#)).

### *Teachers' Language Assessment Literacy*

Language teachers have become a central stakeholder group in LAL, with numerous studies describing their LAL profiles and needs. Specifically, scholars have argued that language teachers' contexts play a pivotal role in describing their LAL ([Coombe et al., 2020](#); [Hill, 2017](#); [Lin et al., 2024](#); [Scarino, 2013](#)). Therefore, these stakeholders should not be labeled as either literate or illiterate in assessment. Rather, their trajectories place them along a language assessment continuum, and their LAL is fostered or enacted depending on their assessment contexts and approach ([Baker, 2021](#); [Lin et al., 2024](#); [Nayernia & Mohebbi, 2023](#); [Yan & Fan, 2021](#); [Yan et al., 2018](#)). Considering this ongoing discussion on teachers' LAL, the components outlined in the previous section cannot be considered a rigid checklist for describing, evaluating, or even promoting teachers' LAL. For instance, a teacher working in an elementary-level language course may require specific knowledge, skills, and principles to develop and use a checklist to assess their learners' speaking skills in a roleplay task. What the research has consistently shown, however, is that language teachers want and need to foster their LAL ([Fulcher, 2012](#); [Vogt & Tsagari, 2014](#)). Consequently, there has been a slow but steady increase in pedagogical initiatives aimed at fostering language teachers' AL.

### *Training Language Teachers through Assessment: Towards LAL Pedagogies*

The literature on teachers' LAL training has identified several characteristics of effective professional development initiatives. Below, we review key features of workshops and courses aimed at fostering language teachers' assessment literacy.

### *LAL Needs Analysis*

Studies have shown that conducting a thorough needs analysis provides valuable feedback for designing and implementing context-sensitive LAL initiatives that are relevant and meaningful to teachers ([Arias et al., 2012](#); [Baker & Riches, 2018](#); [Giraldo, 2023](#); [Koh et al., 2018](#)). Teachers' LAL needs have been identified through both qualitative and quantitative methods and have described their situated LAL: assessment needs, challenges, course expectations, etc.

## Engaging Teachers in Test Design and Analysis

Research studies have demonstrated that when teachers are involved in analyzing and/or designing assessment instruments, they become aware of what is needed to create useful instruments and use theory in practice ([Giraldo & Murcia, 2019](#); [Kleinsasser, 2005](#); [Kremmel et al., 2018](#); [Lao & Yan, 2024](#); [Levi & Inbar-Lourie, 2019](#)). Specifically, LAL teacher educators engage teachers in analyzing tests against qualities of assessment (e.g., validity and authenticity) and in designing useful tests (based on such qualities) for a given context.

## Teachers' LAL Lifeworlds

As we mentioned earlier, it cannot be stated that language teachers are placed as either literate or illiterate in assessment. Their experiences, as both learners and teachers, inform how they conceive of and implement assessment practices. Thus, LAL initiatives that cherish teachers' assessment lives (i.e. teachers' contexts, beliefs, attitudes, reported skills, etc.) contribute to a socio-cultural approach to LAL development. In such settings, peer collaboration —through discussions, joint test development, and feedback activities—supports mutual LAL growth ([Baker & Riches, 2018](#); [González, 2021](#); [Kremmel et al., 2018](#)).

## Emerging Teachers' Reflection on LAL

Whether incidentally or explicitly, reflection on assessment becomes a major positive characteristic of LAL initiatives. As teachers learn or problematize assessment issues, they become aware of what assessment can do, how it can become fairer for language learners, and how they (teachers) have made mistakes in past assessment practices ([Arias et al., 2012](#); [Giraldo, 2023](#); [González, 2021](#); [Lao & Yan, 2024](#)). More importantly, thanks to LAL training, teachers consider improvements to assessment so that it can better help their learners ([Baker & Riches, 2018](#); [Koh et al., 2018](#)). Reflections in LAL courses can take the form of journals or discussion environments, whether online or face to face ([Kleinsasser, 2005](#); [Restrepo-Bolívar, 2020](#)).

## Contents of Initiatives for Teachers' LAL

The literature on courses and workshops for teachers' LAL generally aligns with the three central components of LAL. The contents commonly addressed in these initiatives include ([Arias et al., 2012](#); [Baker & Riches, 2018](#); [Giraldo, 2024](#)).

- Knowledge: Assessment purposes, methods, and qualities.
- Skills: Design of items and tasks for assessing receptive and productive language skills.
- Principles: Fairness and democracy.

These components are generally addressed through hands-on tasks. Whether through test analysis or design, teachers use contents (i.e., knowledge in LAL) through praxis, fostering teachers' theoretical and practical dimension of LAL.

On the other hand, the literature clearly indicates that knowledge and skills are the most prominent components in LAL courses, while principles take a secondary stage. We believe this may be because principles such as ethics and fairness are typically discussed in large-scale testing, which is not directly relevant to teachers. However, there is a growing recognition of the importance of social justice, fairness, and consequential validity as crucial aspects in language testing ([Deyger, 2019](#); [Randall et al., 2023](#)). Consequently, we believe this LAL component might become more present in LAL initiatives for teachers.

## Methodology

### Research Context

Even though the Ministry of Education in Colombia establishes general guidelines for English language teaching and even has proposed curricula with learning objectives and contents, language assessment has not received explicit policy-ingrained attention. Nevertheless, several Colombian scholars have emphasized the need to foster teachers'

LAL ([Cote & López, 2024](#); [Herrera & Macías, 2015](#)). To do this, LAL training—in the shape of official courses in undergraduate and graduate degrees or through in-service professional initiatives— should be implemented. This project was designed to respond to this need by providing LAL education for English language teachers in Colombia. Through this study, we sought to contribute to such need, and we now wish to advance LAL discussions in Colombia, especially regards instructional approaches to teaching LAL.

### *Research Design*

We adopted a qualitative research design. As [Creswell \(2014\)](#) explains, qualitative research involves documenting people's dialogue, their descriptions, and their perspectives on a central phenomenon—which, in our case, was the development of LAL. This approach allowed us to study this research phenomenon in detail in the context of an online learning environment. Our goal was to understand the nature of the participants' LAL through an interview to explain perceptions on the LAL course and its development. As [Maxwell \(2013\)](#) states, a qualitative study is coherent as it investigates the lived experiences (ideas, beliefs, challenges, etc.) of research participants given a specific topic (LAL in our study) and a focused research question.

### *Participants*

Participants in this study were six Colombian English language teachers, whose teaching experience in EFL ranged between 5 and 10 years. They worked in different settings: high schools (2 teachers), language institutes (3 teachers), and a university (1 teacher). Three of these teachers also worked as teacher trainers, i.e., teaching English and language teaching methodology to other elementary and high school teachers. The teachers came from different regions in Colombia, which added to the multi-contextual nature of the course: Bogota (1), Cauca (1), Huila (1), Valle del Cauca (1), and Risaralda (2). All six participants were teaching during the study.

They taught general English courses in their respective contexts and had to follow institutional policies for assessment—primarily concerning percentage allocations for different course components and alignment with prescribed curricula. Nevertheless, all teachers had the freedom to choose what methods to use for assessment.

### *Instructional Design*

The course was administered through Google Workspace tools (Meet, Drive, Forms, and Docs). Over a period of ten weeks, the teachers studied various language assessment topics: fundamentals of language assessment, test specifications, assessing receptive skills, assessing productive skills, assessing integrated skills, evaluating resources for language assessment, validity and reliability, authenticity, and alternative assessment.

Teachers met with the instructor online once per week for two hours. Each session involved a range of LAL LAL learning activities. Before the synchronous lessons, the teachers were asked to read a handbook organized into modules corresponding to each topic. Each module had an average of twelve pages and contained theory and practice for each course content. The handbook and the course were designed by Giraldo, who was also the course instructor, and based on another study reported in [Giraldo \(2024\)](#). The idea that the course and handbook be useful for teachers in Colombia informed such decision, especially because both were based on characterizing their LAL profile through a diagnostic stage (see Giraldo, in press).

Additionally, the teachers were expected to take a quiz with closed-ended questions about module topics and practical tasks for them to practice these contents, e.g., creating a rubric for a speaking task or reflecting on unethical/unfair practices in assessment.

The following is a list of activities, along with examples, that the course participants engaged in during the synchronous sessions.

- Homework check: checking and correcting answers from a handbook activity; sharing ideas on assessment issues, e.g., unethical/unfair practices in assessment.



- Quiz review: The instructor shared the responses to the quiz for every week with the course and provided feedback; the teachers then shared their insights into the practical task of a quiz; for instance, brainstorming the writing skills that a student would need to write a movie review.
- Discussions: After studying problematic assessment situations, the teachers shared what could go wrong and how the problem could be improved, e.g., a teacher who checked an oral presentation without a rubric.
- Analysis tasks: The teachers analyzed assessment instruments against theoretical and technical aspects of assessment; for example, validity and authenticity for the former and, for the latter, what to consider when designing a multiple-choice question.
- Design tasks: Teachers were asked to plan test specifications and then derive test items and tasks from them.

### *Data Collection and Analysis*

To collect data, each participant was interviewed individually via Google Meet. The interviews, which lasted between 40 minutes and 1 hour, included guiding questions to maintain the same structure with all participants. The questions asked participants about course contents, activities, and the perceived impact of the course on participants' professional development. Participants were given the option to conduct the interview in Spanish, English, or both, to minimize misconceptions, lack of confidence, or limited responses.

We analyzed the Google Meet transcriptions following [Saldaña's \(2016\)](#) content analysis approach. Through them, we summarized the answers in each question and constantly compared them to group salient data samples. In this initial stage, we examine inter-rater reliability by calculating Cohen's Kappa for all the initial codes ( $N = 289$ ) extracted from the interviews. Cohen's Kappa for this stage was .66, considered a good level of agreement ([Landis & Koch, 1977](#); [Riazi, 2016](#)). To reach full agreement, we discussed data samples to determine what exactly we understood in each answer given by each teacher.

We then looked for trends across participants' answers to identify the most relevant and recurring themes. These emergent trends were grouped, and during this process, we could notice some organic trends as well, i.e. those that appeared in several moments of the interview. We then discussed and refined our grouped codings, leading to the identification of major themes, which are presented in the findings below.

## **Findings**

Based on our data analysis, we identified five major findings related to the course's impact on participants' LAL. These include teachers' critical reflection on the course; teachers' context and LAL; and the challenges faced during the course. Below, we present each finding, including relevant data samples from our research participants.

### *Teachers' Critical Reflection on Course Contents*

The participants in this study had mostly positive perceptions of the course content, particularly highlighting the relationship between course contents and their LAL. For example, Teacher1 commented on the organization and practical value of the content, and how it related to her own teaching.

I found them (the contents) very interesting and useful. I think you [the course instructor] go from the conceptual part you gave us; you also gave us like the practical way, to implement those concepts you gave us. Let's say, if we talked about the structure of an assessment, for example, then it was there in the contents: how to structure it; there were examples that were very clear and true to our educational context.<sup>4</sup>

Teacher2, on the other hand, emphasized the precision and relevance of the contents, highlighting the complexity of language assessment within her LAL trajectory. Notice that her comment both reflects a positive perception and implies her conclusion about experience doing assessment.

<sup>4</sup> We translated into English all the data samples in which teachers used Spanish.

About the contents, I consider they were very precise; I consider that each topic was very concrete and the points to be addressed in each session. Something I believe is important, regarding the contents, is that assessment is extremely complex, despite the fact that we're teachers and have experience assessing.

Teacher4 shared his insights into the ethical dimension of language assessment. Although he found ethics difficult to address in his practice, he stressed its importance and recommended a stronger focus on this topic in future iterations of the course:

Topics related to ethics are complicated because it's really hard to apply them in real life. Usually, we find ourselves with different situations in which we cannot follow those statements, unfortunately, so they're like we should follow those rules to the letter or we go with the flow. So, I would go once again for questions and topics related to ethics because unfortunately is very difficult to apply to implement those in real life.

### *Course Activities for Conceptual and Social LAL Development*

As with contents, teachers had positive perceptions towards the activities they performed in class. We want to bring attention to two aspects that were apparent across the interview transcriptions. First, teachers remarked on how meaningful the activities were, especially considering their teaching context, similarly to what we described in the previous finding. In the next sample, Teacher3 commented on learning about LAL as he made connections with his assessment context.

The activities were very meaningful for the learning process, given that they allowed us to, like, I mean, the way you designed them, they allowed us to contextualize ourselves, keeping in mind our educational context.

Second, the teachers reported the positive impact of interaction activities to develop their LAL. Teacher4 explained why he liked the activities in the course and how interactive activities helped his LAL development: "I like the activities because they gave me the possibility to listen to some other teachers, talking about the assessment process. In doing so, I confirmed or I cleared on some questions that I had."

On the other hand, the interview data indicates that these teachers viewed quizzes as a useful LAL activity. The data below indicate that quizzes in this LAL course had pedagogical and, naturally, assessment functions. For example, Teacher5 reported that quizzes helped her identify areas for improvement in her LAL development: "I loved the quizzes, because I can check my mistakes or my misconceptions about something related to assessment". Teacher6, in turn, stated that he used quizzes to study with a classmate, discussing and checking on their performance (e.g., right and wrong answers to the quiz). Then, during the synchronous lessons, all the teachers received instructor feedback to further check on their LAL development.

The one that I liked the most were the quizzes. And the way feedback was given to the quizzes, because my partner, most of the times we were together watching the quizzes so we can discuss: "This [answer] was not mine. This is the one I did but you made it right and I made it wrong." So, we got discussion in the answers. But when we had the feedback, then we realized if we were doing a good analysis, or we needed to fix some of the ideas we had.

Finally, Teacher1 commented on how quizzes were a useful assessment method to monitor her progress in the course. She also highlighted that the quizzes included short hands-on items through which she could reflect on her assessment context, especially assessment practices she either was or was not conducting.

I believe those quizzes made us aware of what we were reading or learning. Besides, there were practical activities inside these quizzes, for us to look into our reality and how that was taking place, to see whether or not we were doing it [the practical activities] so that we could reflect on it.

### *The Course as a Trigger of Reflection on LAL*

A recurring theme across the interviews was the role of the course in fostering teacher reflection on assessment. The teachers remarked on the importance of assessment for professional practice, analyses of their assessment lifeworlds,



and the need to expand their LAL repertoire. In the first data sample below, for example, Teacher5 described how the course reshaped her conception of language assessment and. She also reflected on how using assessment could have a positive impact on her students.

As I told you at the beginning of this interview, I had a misconception about what is assessment or how to ask the students. And all this process to notice the importance of the ethic, validity, reliability, and to the importance of take different kinds of assessment. Not only traditional; all these help me to check, also the methodology and the materials, and the purpose of my classes and how to help my students if they have any difficulty in there, improving their communicative skills.

Teacher6, on the other hand, analyzed long-standing assessment practices and how they should broaden thanks to LAL training. He also reflected on how he can expand LAL practices to achieve validity in assessment, which he describes as assessing what needs to be assessed.

We as teachers get used to do some practices in terms of assessment. And I feel we're married to something that makes you think that you're doing it right. You get used to the type of assessment you create and you think that it is correct. But when you open the scope, when you check that is something more to think about, that is when you think, okay, I need to reflect on this and I need to start doing some other things. Not only focusing on one single aspect but to explore assessment and take advantage of what all students can do and actually assess what we need to assess in this case.

### *The Role of Context in Teachers' LAL Development and Enactment*

The teachers in this course did not study and learn about language assessment in a vacuum. As their interview responses made clear, their teaching and assessment contexts influenced the way they evaluated the course and their LAL development. For instance, Teacher3 discussed the relevance of the topics studied in the course. It is significant how Teacher3 reflected on problematic assessment situations in his context and the benefits that LAL could have when addressing these.

I think the content we studied in the course is very much connected with, well, everything we live in our educational context, so they are contents that teach us a lot about assessment. So, we come across many difficulties and realize the mistake that we believe we are assessing students well but in reality it's not the case. And so sometimes we set learning goals but if we don't have assessment instruments that allow us to collect the necessary information, then we will not achieve those goals within a given context.

Teacher2 connected the topic of score pollution (a threat to assessment validity) to a reflection in her assessment context. She explained how having to assess *effort* challenges the notion of construct validity and valid interpretations of students' assessment language performance.

In my case, the issue of score pollution made a lot of sense, because I consider myself a teacher who did not take that aspect very seriously. When judging the validity of the assessment itself, then I found myself facing that task, right? And to what extent we can provide, to quote, an "accurate" picture... but to what extent it [an assessment] can show me students' learning achievement.

### *Challenges for teachers to develop LAL in the Course*

The final theme relates to the challenges that participants encountered throughout the course. Teacher2 highlighted inherent complexity of assessment at any stage of a teacher's career. He articulated this view while also recognizing how the course offered clarity:

Assessment is extremely complex. To address it and make it concrete. Like, somehow giving contextual definitions of assessment, it is often a sensitive issue. But I think that the way that assessment was addressed, what it is and what it is not, was very clear.

On the other hand, all teachers expressed that having enough time to be and study in this course was a major challenge. Lack of time did not allow them to be on time to start the synchronous sessions or delve into reading the module for each week conscientiously before class. Teacher5, for example, described how her work schedule affected her attendance and engagement:

Be on time for me was hard or I felt that. I wasn't 100% of the course paying attention all the time. Because I finish my classes at six thirty. And I ride my bike, to arrive home as fast as I can, but most of time, I arrive late so I connected late. So I missed some information and I wasn't learning on the same way that the [other] participants.

Teacher6 also mentioned time constraints limiting his ability to engage more with the readings for each module, even though he did read through the preparatory material.

I could manage it and I think I did a good job with the modules because I get to read them but not a conscious reading. But I would love to have had more time to sit down and take notes, and highlight some parts.

## Discussion

Research on teachers' perceptions of LAL courses has shown that stakeholders welcome both contents and activities, recognizing their contribution to LAL development ([Baker & Riches, 2018](#); [Boyd & Donnarumma, 2018](#); [Koh \*et al.\*, 2018](#); [Lao & Yan, 2024](#)). Most of these studies use questionnaires to gather feedback, but they do not provide rich, thick descriptions of teachers' course evaluations. In contrast, the present study, through the use of individual interviews, enabled us to explore teachers' critical reflections on course contents and activities, their connections to contextual realities, and the challenges encountered during LAL development. Teachers' reflections might not have been possible otherwise, which is why we attribute our findings to the focal use of interviews: Deep reflection is a natural development in the use of interviews for qualitative research ([Creswell & Poth, 2018](#)).

We also believe that the teachers responded positively to the LAL content in the course because it was designed based on a diagnostic stage, during which these participants completed a questionnaire about their LAL and submitted two assessment instruments (see Giraldo, *in press*). Thus, there is a relationship between the findings in the diagnostic stage and teachers' evaluation of course contents. Regarding activities, we believed teachers enjoyed them because the course included practical tasks in the handbook, quizzes to reflect and check on LAL progress, and interactive discussions with fellow teachers. Therefore, it seems that the pedagogical approach in this course was generally successful.

A recurring theme across the findings presented in this article is teachers' reflection. Participants engaged deeply with the content and activities, and seemed to have a keen eye for what worked and what did not work in this LAL journey. As various authors have highlighted, critical reflection is a key component to engage teachers in LAL development ([Inbar-Lourie, 2012](#); [Scarino, 2013](#); [Xu & Brown, 2016](#)). For example, while Teacher2 and Teacher4 acknowledged the usefulness of course contents, they also discussed the challenges and/or difficulties of using their LAL knowledge in practice. We thus conclude that our participants used their developing LAL (which includes beliefs and contextual challenges) to evaluate our LAL pedagogy and, ultimately, their LAL progress.

Based on the above, course activities seemed to be pedagogically sound, which has been reported elsewhere ([Kremmel \*et al.\*, 2018](#); [Lao & Yan, 2024](#); [O'Loughlin, 2006](#)). In particular, the use of quizzes stood out. Before all synchronous lessons, the teachers read a module on a given topic (e.g., validity and reliability) and completed a quiz about its contents; as Teacher1 states, the quizzes also had short practical tasks. Once in the synchronous session, Giraldo and all the teachers went over all the answers and discussed their performance; in short, quizzes were the center of an instructional sequence and were used entirely for formative purposes: To put it succinctly, quizzes were assessment *as learning* (teachers' self-assessing their LAL progress) and assessment *for learning*, when teachers received feedback on strengths and aspects to improve.

Additionally, we want to highlight how the teachers in this study evaluated the course against their teaching and assessment contexts. As [Scarino \(2013; 2017\)](#) rightly argues, teachers' assessment lifeworlds (e.g., their beliefs

and practices) help shape the LAL agenda. For example, Teacher3 and Teacher6 suggested that better assessment requires making changes to their contextual practices, which might be possible to the training they received in the course. Thus, teachers' contexts become crucial to evaluate the impact of LAL training, as various studies have shown ([Baker & Riches, 2018](#); [Koh et al., 2018](#); [Kremmel et al., 2018](#); [Levi & Inbar-Lourie, 2019](#); [Montee et al., 2013](#)). In our study, participants' assessment contexts were continuously integrated into the course thanks to handbook tasks and course discussions. For instance, the task below comes from the handbook and its answers were shared in a class discussion on what makes assessments useful:

What methods do you use to collect information about your students' skills in the English language? Are they useful to collect information about how your students are developing the standards or learning objectives for a specific grade or course? Make a list of methods you use and bring it to Week 1.

Finally—and perhaps not surprisingly—the lack of time influenced engagement in the LAL course reported in this study. The issue of time is widely recognized in the literature as a key condition for effective teacher learning, both in general and in language education ([Csépes, 2021](#); [DeLuca et al., 2021](#); [Lukin et al., 2004](#)). The teachers in our study willingly participated in the LAL course, so they were not given any official time out of their work to fully engage in it. We believe that teachers found time a challenge simply because they did not have enough of it. For this reason, we echo calls in the literature for teachers to be given time to study assessment and to do related tasks.

## Conclusions

The results of this study reveal how an online language assessment course impacted six Colombian English language teachers across various dimensions of their language assessment literacy. The participants shared their insights on the course content, activities and overall experience, while also reflecting on how these components of the course helped them develop their LAL.

Regarding the course content, the teachers assessed its usefulness in relation to the course's structure and their specific context. Based on this, they indicated whether the new content could be integrated into their current practices or whether this might prove challenging. As for activities, the teachers analyzed quizzes and feedback as important opportunities for improvement prior to and during the synchronous sessions. Additionally, the synchronous sessions gave teachers the opportunity to interact with and learn from one another. Thus, the course supported a social dimension of LAL development.

Reflection was commonplace in the interviews and gave the participants a space to have a voice. Specifically, the teachers discussed how their contexts significantly influenced their evaluation of the course and the enactment of LAL—particularly in terms of their motivation to change assessment practices. By the end of the course, the teachers recognized that, although assessment remains a complex component of their professional responsibilities, beginning to develop the skills to design appropriate assessment instruments had a tangible impact on their classroom practices. The comments during the interviews also indicated some challenges these teachers faced. The teachers considered lack of time as the most challenging factor in devoting themselves to the course and learning the core language assessment concepts.

This study contributes to the growing body of literature on LAL by demonstrating the importance of providing targeted assessment education for teachers. Such training not only motivates them to interact and share with their colleagues and reflect on their practices but also has the potential to influence their immediate assessment contexts. Above all, the study allowed us to listen to these teachers to continue building a network of language assessment literacy in Colombia. Based on the findings of this study—and more broadly, the larger research project to which it belongs—we propose the following recommendations for LAL course design for English language teachers in Colombia. LAL courses should

- Be carefully designed based on a diagnostic stage in which their LAL is characterized. We suggest that such characterization include various sources of LAL, e.g., analysis of teacher-designed tests, questionnaires on LAL topics, individual interviews, among others.

- Include hands-on and reflection tasks through which teachers can examine their LAL progress *and* their LAL contexts. For this, course materials (e.g., the handbook we report) and course tasks should activate teachers' developing knowledge, skills, principles, and perceptions of assessment.
- Instill a deeply formative assessment approach to ignite teachers' LAL. Assessments in LAL courses should help teachers to reflect on their progress through self- or peer-assessment tasks, i.e. *assessment as learning*. Also, LAL instructors should explicitly use teachers' performance on tasks in the LAL course to further problematize and improve LAL, i.e. *assessment for learning*.

Finally, we emphatically encourage policymakers and institutions to provide quality spaces and time for Colombian English language teachers to foster their LAL. Serious, explicit, and ongoing training in language assessment should be encouraged, given the tendency that such training leads to evident improvements in how teachers view or enact assessment.

## References

- Arias, C. I., Maturana, L. M., & Restrepo, M. I. (2012). Evaluación de los aprendizajes en lenguas extranjeras: hacia prácticas justas y democráticas. *Lenguaje*, 40(1), 99–126. <https://doi.org/10.25100/lenguaje.v40i1.4945>
- 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>
- Boyd, E., & Donnarumma, D. (2018). Assessment literacy for teachers: A pilot study investigating the challenges, benefits and impact of assessment literacy training. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 105–126). Springer. [https://doi.org/10.1007/978-3-319-77177-9\\_7](https://doi.org/10.1007/978-3-319-77177-9_7)
- Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126–136). Cambridge University Press.
- Butler, A. (2022). Exploring the idea of “assessment with learners” in discussing language assessment literacy. *Language Teaching Research Quarterly*, 29, 92–101. <https://doi.org/10.32038/ltrq.2022.29.06>
- Coombe, C., Vafadar, H., Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia*, 10, 1-16. <https://doi.org/10.1186/s40468-020-00101-6>
- Coombs, A. & DeLuca, C. (2022). Mapping the constellation of assessment discourses: a scoping review study on assessment competence, literacy, capability, and identity. *Educational Assessment, Evaluation and Accountability*, 34, 279–301. <https://doi.org/10.1007/s11092-022-09389-9>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications.
- Creswell, J. W., & Poth, C. (2018). *Qualitative inquiry and research design: Choosing among five approaches*. SAGE Publications.
- Csépes I. (2021). The evolving concept of (language) assessment literacy: Implications for teacher education. *Central European Journal of Educational Research*, 3(1), 120–130. <https://doi.org/10.37441/CEJER/2021/3/1/9360>
- Cui, X., Liu, Y., Yu, H., & Gao, Y. (2022). Developing English teachers' language assessment literacy in an EAP reform context through test design: A case study. *System*, 109, 1-12. <https://doi.org/10.1016/j.system.2022.102866>
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/0265532208090156>
- Cote-Parra, G., & López, A. A. (2024). Examining the assessment practices of foreign language novice teachers. *Profile: Issues in Teachers' Professional Development*, 26(1), 97–113. <https://doi.org/10.15446/profile.v26n1.106384>
- DeLuca, C., Searle, M., Carbone, K., Ge, J., & LaPointe-McEwan, D. (2021). Toward a pedagogy for slow and significant learning about assessment in teacher education. *Teaching and Teacher Education*, 101, 1-12. <https://doi.org/10.1016/j.tate.2021.103316>
- Deygers, B. (2019). Fairness and social justice in English language assessment. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 541-570). Springer.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. <https://doi.org/10.1080/15434303.2011.642041>

- Fulcher, G., Panahi, A., Mohebbi, H.3 (2022). Glenn Fulcher's thirty-five years of contribution to language testing and assessment: A systematic review. *Language Teaching Research Quarterly*, 29, 20–56 <https://doi.org/10.32038/ltrq.2022.29.03>
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179–195. <https://doi.org/10.15446/profile.v20n1.62089>
- Giraldo, F., & Murcia, D. (2019). Language assessment literacy and the professional development of pre-service language teachers. *Colombian Applied Linguistics Journal*, 21(2), 243–259. <https://doi.org/10.14483/22487085.14514>
- Giraldo, F. (2023). Fostering pre-service teachers' language assessment literacy. Editorial Universidad de Caldas.
- Giraldo, F. (2024). Language assessment literacy for in-service teachers: Online course planning and implementation. Editorial Universidad de Caldas.
- Giraldo, F. (in press). Designing an assessment course for in-service English language teachers in Colombia. *Folios*, 62.
- 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>
- Herrera, L., & Macías, D. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, 17(2), 302–312. <https://doi.org/10.14483/udistrital.jour.calj.2015.2.a09>
- Hill, K. (2017). Understanding classroom-based assessment practices: A precondition for teacher assessment literacy. *Papers in Language Testing and Assessment*, 6(1), 1–17.
- 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>
- Inbar-Lourie, O. (2012). Language assessment literacy. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–9). John Wiley & Sons. <https://doi.org/10.1002/9781405198431.wbeal0605>
- Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment: Encyclopedia of language and education* (3rd ed., pp. 257–268). Springer. [https://doi.org/10.1007/978-3-319-02261-1\\_19](https://doi.org/10.1007/978-3-319-02261-1_19)
- Kleinsasser, R. C. (2005). Transforming a postgraduate level assessment course: A second language teacher educator's narrative. *Prospect*, 20(3), 77–102.
- Koh, K., Burke, L., Luke, A., Gong, W., & Tan, C. (2018). Developing the 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>
- Kremmel, B., Eberharter, K., Holzknicht, F., & Konrad, E. (2018). Fostering language assessment literacy through teacher involvement in high-stakes test development. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 173–194). Springer. [https://doi.org/10.1007/978-3-319-77177-9\\_10](https://doi.org/10.1007/978-3-319-77177-9_10)
- Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100–120. <https://doi.org/10.1080/15434303.2019.1674855>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Lao, H. & Yan, Z. (2024) Using lesson study to promote teacher assessment literacy. In E. Cheng, & B. Lander (Eds.), *Implementing a 21st century competency-based curriculum through lesson study*. (pp. 42– 55). Routledge.
- Levi, T., & Inbar-Lourie, O. (2019). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, 17(2), 168–182. <https://doi.org/10.1080/15434303.2019.1692347>
- Lin, L., Ying, G., & Guo, X. (2024). Pre-service Chinese language teachers' conceptions of assessment: A person-centered perspective. *Language Teaching Research*, 28(1), 273–295. <https://doi.org/10.1177/1362168821996529>
- Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice*, 23(2), 26–32. <https://doi.org/10.1111/j.1745-3992.2004.tb00156.x>
- Nayernia, A., & Mohebbi, H., (2023). Teachers' online language assessment literacy. In Z. Tajeddin, & C. Griffiths (Eds.), *Language education programs: Perspectives on policies and practices*. (pp. 122 –140). Springer. <https://doi.org/10.1007/978-3-031-38754-8>
- Malone, M. E. (2017). Training in language assessment. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment: Encyclopedia of language and education* (3rd ed., pp. 225–240). Springer. [https://doi.org/10.1007/978-3-319-02261-1\\_16](https://doi.org/10.1007/978-3-319-02261-1_16)
- Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3<sup>rd</sup> ed.). SAGE Publications.



- Mohammadkhah, E., Reza Kiany, G., Tajeddin, Z., ShayesteFar, P. (2022). EFL teachers' assessment literacy: A contextualized measure of assessment theories and skills. *Language Teaching Research Quarterly*, 29, 102–119. <https://doi.org/10.32038/ltrq.2022.29.07>
- Montee, M., Bach, A., Donovan, A., & Thompson, L. (2013). LCTL teachers' assessment knowledge and practices: An exploratory study. *Journal of the National Council of Less Commonly Taught Languages*, 13, 1–31.
- O'Loughlin, K. (2006). Learning about second language assessment: Insights from a postgraduate student on-line subject forum. *University of Sydney Papers in TESOL*, 1, 71–85.
- Pastore, S. (2023). Teacher assessment literacy: a systematic review. *Frontiers in Education*, 8. <https://doi.org/10.3389/educ.2023.1217167>
- Randall, J., Poe, M., Slomp, D., & Oliveri, M. (2023). Our validity looks like justice. Does yours? *Language Testing*, 41(1) 203–219. <https://doi.org/10.1177/02655322231202947>
- Restrepo-Bolívar, E. M. (2020). Monitoring preservice teachers' language assessment literacy development through journal writing. *Malaysian Journal of ELT Research*, 17(1), 38–52.
- Rezai, A. (2024). The role of teacher assessment literacy in job stress and job burnout in EFL contexts: a mixed-methods investigation. *Asian-Pacific Journal of Second and Foreign Language Education*, 9, 1-21. <https://doi.org/10.1186/s40862-023-00225-1>
- Riaz, A. M. (2016). *The Routledge encyclopedia of research methods in applied linguistics: Quantitative, qualitative, and mixed-methods research*. Routledge.
- Saldaña, J. (2016). *The coding manual for qualitative researchers*. SAGE Publications.
- 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 interpretive challenge. *Papers in Language Testing and Assessment*, 6(1), 18-40.
- Stabler-Havener, M. L. (2018). Defining, conceptualizing, problematizing, and assessing language teacher assessment literacy. *Working Papers in Applied Linguistics & TESOL*, 18(1), 1–22.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238–245.
- 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., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402. <https://doi.org/10.1080/15434303.2014.960046>
- Xu, Y., & Brown, G. (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>
- [https://cdn.ymaws.com/www.iltaonline.com/resource/resmgr/docs/ltrc\\_2021/ltrc2021\\_schedule\\_version\\_ma.pdf](https://cdn.ymaws.com/www.iltaonline.com/resource/resmgr/docs/ltrc_2021/ltrc2021_schedule_version_ma.pdf)
- Yan, X., Zhang, C., & Fan, J. (2018). "Assessment knowledge is important, but ...": How Contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, 74, 158-168. <https://doi.org/10.1016/j.system.2018.03.003>
- Yan, X., & Fan, J. (2021). "Am I qualified to be a language tester?": Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219–246. <https://doi.org/10.1177/0265532220929924>
- Zu, X & Evans, C., (2024). Enhancing the development and understanding of assessment literacy in higher education. *European Journal of Higher Education*, 14(1), 80-100. <https://doi.org/10.1080/21568235.2022.2118149>
- Yan, X. (2021, June 14-17). Towards a conceptual model of LAL development across stakeholder groups and contexts [Conference presentation]. *Language Testing Research Colloquium*, online.
- Baker, B. (2021, June). Unpacking the concept of language assessment literacy for all key stakeholders [Conference presentation]. *New Directions Colombia 2021* <https://americas.britishcouncil.org/new-directions/programme>

