

Characterizing Teachers' Language Assessment Literacy to Design an Online Assessment Course

La literacidad en evaluación de docentes de idiomas para el diseño de un curso de evaluación en línea

Frank Giraldo

Angela Patricia Velásquez-Hoyos

Luis Herney Villegas

Departamento de Lenguas Extranjeras, Universidad de Caldas, Manizales, Colombia

This article presents a qualitative case study that characterizes the language assessment literacy (LAL) of five Colombian English-language teachers to inform the design of a context-sensitive LAL course. We analyzed data from questionnaires, interviews, and institutional documents. Findings revealed that teachers' assessment practices were partially influenced by the communicative approach. On the other hand, teachers' challenges in assessment include assessing irrelevant criteria and the impact of administrative factors (e.g., national policies), which cause misalignments in assessment. This study emphasizes the need for an LAL course that addresses the design of communicative assessment tasks and methods and that aligns language assessment practices with institutional requirements.

Keywords: curriculum, evaluation, foreign language instruction, professional development, teacher education

Este artículo reporta un estudio de caso cualitativo que caracterizó la literacidad en evaluación de lenguas (LEL) de cinco profesores de inglés colombianos, con el fin de diseñar un curso de LEL contextualizado. Analizamos datos de cuestionarios, entrevistas y documentos institucionales. Los hallazgos revelaron que las prácticas docentes de evaluación se ven influenciadas por el enfoque comunicativo. Por otro lado, existen retos como evaluar criterios irrelevantes para la competencia comunicativa y el impacto de factores administrativos (p. ej., políticas nacionales), lo que puede generar incoherencias en la evaluación. Nuestro estudio enfatiza el diseño de un curso de evaluación de lenguas que incluya tareas y métodos comunicativos e integre los requisitos institucionales.

Palabras clave: formación docente, desarrollo profesional, enseñanza de una lengua extranjera, evaluación, programa de estudios

Frank Giraldo  <https://orcid.org/0000-0001-5221-8245> • Email: frank.giraldo@ucaldas.edu.co

Angela Patricia Velásquez-Hoyos  <https://orcid.org/0000-0002-9743-5769> • Email: angela.velasquez@ucaldas.edu.co

Luis Herney Villegas  <https://orcid.org/0000-0001-5313-9370> • Email: luis.villegas@ucaldas.edu.co

This project was sponsored by the Vicerrectoría de Investigaciones y Posgrados of Universidad de Caldas in Manizales, Colombia (project code: PRY-181).

How to cite this article (APA, 7th ed.): Giraldo, F., Velásquez-Hoyos, A. P., & Villegas, L. H. (2026). Characterizing teachers' language assessment literacy to design an online assessment course. *Profile: Issues in Teachers' Professional Development*, 28(2), 49–65. <https://doi.org/10.15446/profile.v28n2.124263>

This article was received on November 28, 2025 and accepted on May 7, 2026.

This is an Open Access text 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

Given the impact of language testing on individuals and society in general (Fulcher, 2015; McNamara & Roever, 2006), there is interest in studying stakeholders' language assessment literacy (LAL): the knowledge, skills, and principles for understanding and engaging with language testing and assessment. Recently, scholars have encouraged the development of pedagogies that foster the LAL of language teachers, given their continuous engagement in assessment (Baker & Taylor, 2024; Fulcher, 2020).

Thus far, available reports on LAL courses have indicated that, as teachers participate in assessment initiatives, they develop greater awareness of what language assessment is (Arias et al., 2012; Restrepo Bolívar, 2020). Additionally, they become more knowledgeable and/or skilled in designing assessment instruments for different language skills (Baker & Riches, 2018; Levi & Inbar-Lourie, 2020). In general, studies reporting on teachers' LAL describe course design superficially, while paying most attention to course delivery and impact; this is sensible, given the need to advance LAL pedagogies, as we mentioned above. However, limited research has characterized teachers' LAL—especially through case studies—to inform the design of context-sensitive LAL courses. The few case studies that exist tend to focus on describing teachers' LAL trajectories, that is, how their knowledge, skills, and principles in language assessment have developed (Gan & Lam, 2025; Yan & Fan, 2021; Yan et al., 2018). We agree with Zhang et al. (2025) that teachers' LAL is contextually shaped; thus, we acknowledge that the LAL of the participants in our study is partly influenced by institutional and national policies.

Considering the previous overview, our purpose is to report on how we characterized the LAL of five Colombian English-language teachers through a multiple-case study. Based on findings from our examination of these five cases, we report pedagogical implications for designing an LAL course for these teachers. Thus,

we believe that our paper may further problematize teachers' LAL, while fueling discussions on pedagogies for planning LAL courses for these stakeholders. The research questions that guided our inquiry were:

- What characterizes the contextual LAL of five Colombian English language teachers?
- How can these teachers' LAL characteristics inform the design of a tailor-made assessment course?

Literature Review

LAL has been described through components, models, and profiles (Kremmel & Harding, 2020; Taylor, 2013), which are useful for conceptualizing LAL across stakeholder groups while also acknowledging its high contextual variability (Baker & Taylor, 2024; Taylor, 2013). Across LAL definitions, three major components emerge: *knowledge* (e.g., language ability models and testing concepts); *skills* (e.g., the design of assessment instruments); and *principles* (e.g., social justice; Randall et al., 2024). Our study is rooted in a componential definition of LAL, first introduced by Davies (2008) and used in numerous conceptual and empirical articles (Fulcher, 2012; Inbar-Lourie, 2008; Taylor, 2013). Notwithstanding this tripartite LAL characterization, there is consensus that the components are differential rather than balanced, depending on the stakeholders' assessment tasks and contexts (Kremmel & Harding, 2020; Tsagari & Armostis, 2025).

In LAL research, teachers have remained a key stakeholder group. Some studies have indicated limited attention to teachers' LAL at the preservice level (Sultana, 2019; Vogt & Tsagari, 2014); therefore, in-service teachers have expressed a need to foster their LAL (Fulcher, 2012). Recently, researchers have argued that teachers' LAL cannot be described in deficit terms (Baker, 2021). Teachers' expertise, experience, and knowledge of local assessment policies and practices are central components of their LAL (Berry et al., 2019; Scarino, 2013). Thus, rather than one literacy, scholars have called for *literacies* in context (Inbar-Lourie, 2024).

In Latin American studies that have addressed the importance of contextualizing LAL, various authors highlight the need to understand LAL as locally situated and shaped by specific contextual conditions (Giraldo & Yan, 2023, 2025). Villa Larenas and Brunfaut (2023) argue that “a rich exploration of [teachers'] contextual working conditions, local assessment culture, access to language assessment knowledge, and assessment belief system appears necessary to understand their LAL and develop suitable LAL training programmes” (p. 489). This idea aligns with other studies from Latin America, such as those by Giraldo and Yan (2023, 2025), which suggest that LAL courses are more effective when grounded in teachers' contexts, feedback, and needs. Thus, our literature review builds on these regional contributions to justify the need for a localized characterization of LAL.

The Use of Case Studies for Teachers' LAL

As we mentioned earlier, a few qualitative case studies have reported teachers' LAL trajectories. Overall, these studies highlight the role of various factors in LAL development and enactment, including institutional and national policies such as standardized tests, testing-related tasks, and professional development scenarios (Gan & Lam, 2025; Yan et al., 2018). Notably, Cui et al. (2022) and Gan and Lam (2025) emphasize that teachers' agency leads to improved LAL when they willingly participate in LAL courses.

Other case studies have shown that teachers' engagement in collaborative work is conducive to LAL development. This development can occur as they discuss assessment practices and challenges, and through hands-on tasks in which they give and receive feedback from colleagues (Spiliotopoulos et al., 2024).

These case studies provide evidence that the three components of teachers' LAL (i.e., knowledge, skills, and principles) have distinct trajectories shaped by external factors (e.g., a newly introduced language test)

and personal decision-making. Additionally, the studies tend to remark on teachers' knowledge of and skills in language assessment. This trend is warranted, as research highlights teachers' need for a clear relationship between theory and practice in language assessment (Fulcher, 2012; Giraldo, 2024; Vogt & Tsagari, 2014). However, the principles component of teachers' LAL remains under-researched and underexamined, as shown in other studies (Moraes & Furtoso, 2026; Moraes et al., 2025).

Method

We adopted a qualitative approach to gain insight into the participants' LAL. We examined their assessment beliefs and practices, as well as the language education documents of their institutions. Based on this diagnosis, we planned to implement a professional development program to address the participants' LAL.

This is a multiple case study, as the five participating teachers came from diverse educational contexts, including public and private universities, rural and urban public schools, and a technical institution. Such diversity and the depth inherent in case studies provided a thorough understanding of the teachers' LAL—within their institutions—thereby enabling the creation of assessment courses attuned to their settings (Simons, 2012; Yin, 2014).

Participants

The participants in this study were five in-service Colombian English-language teachers. They came from different regions, educational institutions, and academic backgrounds. To protect the participants' privacy and ensure that they would share their experiences and insights freely, each teacher chose a pseudonym.

Participants were selected through convenience sampling. Initially, 107 teachers were contacted via email, as they had previously participated in assessment courses led by Giraldo. Such prior participation could have a meaningful impact on our study, as the teachers had already gained theoretical insights into assessment

and could share them to enrich the course. Of the 107 teachers, five submitted all the requested documents: a consent form, two assessment instruments, and documents related to their school's language curriculum. These five participants were not selected through a maximum variation strategy; however, the five cases come from diverse contexts as described in the within-case analysis and findings section.

Data Collection Instruments and Validation

We collected data using three instruments to ensure a comprehensive diagnosis of the participants' LAL. An online questionnaire was administered through Google Forms (see Appendix A). It collected information such as teachers' highest level of education, institution where they worked, education in assessment, and their teaching and assessment methodologies. Further, each participant shared either a syllabus or a language curriculum from their institution, as well as two assessment instruments (e.g., rubrics) that they had designed and implemented in their classrooms. These documents provided insights into the teachers' pedagogical approaches to teaching and assessment, as well as the alignment of their practices with institutional policies. To analyze these documents, we used a document analysis grid (see Appendix B).

Afterward, we used a semi-structured interview protocol (in Spanish; see Appendix C) with teachers to gather their opinions on assessment, how language should be assessed, the purposes of assessment, and related topics. We also asked them about expectations for the LAL course. The interviews lasted between 30 and 50 minutes and provided information on the use of documents to guide assessment, institutional initiatives to promote LAL, evaluation practices related to assessment, and the strengths and weaknesses of assessment in the teachers' institutions.

To enhance the validity of the data, we sent the questionnaire, interview protocol, and document anal-

ysis grid to two external professors for content expert review and to identify potential biases. They provided suggestions to improve the instruments' overall design. Additionally, the research instruments were piloted with three teachers to facilitate further adjustment.

Data Analysis

After data collection, we created a spreadsheet to store the information from all instruments. Then we read all the data for each instrument separately to identify codes. Later, we met to discuss the main tendencies in each data collection method and developed themes, which were operationalized individually and then shared as a group; thus, we identified 16 categories. To assess inter-rater reliability, we used Fleiss' Kappa because we were three raters observing 16 categories (Egbert & Baker, 2019). There was full agreement among the three raters in 11 categories, corresponding to an absolute agreement rate of 68.75%. For the categories we disagreed on, we discussed them thoroughly to reach a consensus.

In another stage, we consolidated major themes across instruments that emerged from the identification of underlying codes. Once again, to reach a consensus, we identified the themes separately and then discussed them as a group. In this stage, five overarching themes were drawn and condensed in the spreadsheet. The data were analyzed inductively, with information organized from the bottom up to identify codes and categories that were ultimately grouped into major themes (Creswell, 2013). Through this collaborative analysis, we identified three major themes, which we then used as baseline data to design the assessment course for the five teachers. In the findings section, we present each teacher's case first, followed by cross-case findings.

Member Checking

We created a profile for each participant, including analyses and findings for their cases, as described in the next section. Then we contacted each teacher via

email and shared their report with them. We did so to identify any misinformation or mistakes in their descriptions, to obtain feedback, and to answer any questions they might have. In sum, this stage was useful to portray teachers' LAL accurately.

Within-Case Analysis and Findings

Sofia's Case

At the time of the study, Sofia had been an English teacher for four years and was teaching at a state-funded institution that offers free vocational education. During her BA, Sofia took a language assessment course because of her interest in the topic. She was interested in our project because, as she stated, "In my everyday teaching, I often don't put into practice what I learned."

Sofia explained that she used an eclectic teaching approach: communicative language teaching and the presentation-practice-production technique. She has taught technical English, which can be considered English for specific purposes. In curriculum documents, we found no explicit evidence of approaches to language teaching in Sofia's institution. This implies that Sofia is allowed to use the language approach that she considers appropriate.

Sofia's institution provides teachers with a curriculum matrix outlining learning outcomes, vocabulary topics, and assessment criteria. We noticed that this document does not include assessment-related information, but does include grammar and vocabulary topics. According to this document, teachers can assess using "instruments and tasks, including checklists, questionnaires, rubrics, and roleplays." In the interview, Sofia mentioned she could provide feedback on this document, particularly on sequencing and topics.

Sofia assessed her students' skills through workshops, oral presentations, group work, class participation, online games, and roleplays. She mentioned that she believes language assessment is the use of

"tools to know whether they are understanding you." For her, language assessment should focus on communication rather than form, which she claimed was not reflected in her teaching context, which emphasized technical vocabulary.

Sofia believes her assessment skills include designing rubrics or checklists, providing clear feedback, and modifying her assessments based on results. This suggests that she is free to administer and adjust her assessments as she considers suitable. Sofia also reported that, unfortunately, her institution does not offer instruction in language teaching, let alone assessment; thus, she studies autonomously or through participation in projects. We interpret the lack of LAL education as an area for improvement at Sofia's institution, especially given her willingness to invest in her LAL.

Sofia shared two assessment instruments: a reading comprehension test and a student self-assessment checklist. The test focused on developing learner awareness of vocabulary in context, which may evidence an alignment with communicative principles. The checklist, on the other hand, allowed students to monitor their progress through various linguistic descriptors, such as the appropriate use of grammar rules. Both instruments were well-structured, provided clear instructions for students, and, given their focus on linguistic knowledge, aligned with institutional guidelines.

In conclusion, Sofia's LAL is shaped by two aspects: institutional policies emphasizing linguistic aspects and her assessment beliefs. This influence can be reflected in her instruments, which assess grammar and vocabulary, and in some of her assessment practices, which she described as communicative. Furthermore, our analysis of Sofia's case led us to identify that although her assessment instruments align with institutional documents, she adjusts her assessment practices. As implications for the course, we believe Sofia's case can be useful for exploring connections among teaching, learning, and assessment, as well as for addressing vocabulary and grammar in communicative tests.

Maleja's Case

Maleja has been teaching English for 10 years. At the time of this study, she taught general English at a high school in a rural area in Colombia, using a communicative approach and gamification. Maleja holds an MA in English Language Teaching, and during her studies, she took a course called *Language Testing and Evaluation*, which focused on “rubrics to assess learners’ performance in different skills.” She agreed to be in our study because she wanted to be informed about trends in language education. In the interview, she stated that she wanted to learn “new ways to grade [and] assess students . . . and [how to create] effective rubrics and [implement] feedback strategies.”

In the institutional documents Maleja provided, we found that the core construct to be assessed was communicative competence, with an explicit mention of grammar and vocabulary. Communicative competence was described in language learning standards for each grade level, which Maleja used to plan assessments. The purposes of assessment, as stated in the documents, were to evaluate students’ progress and to support test preparation. Regarding the latter, she explicitly includes test tasks that resemble those used in Colombia’s national tests.

Maleja believes that language assessment is a process for checking learners’ ability to use language in context, for example, through role-plays and games. For this process, she emphasized that summative and formative assessments should occur. Regarding her assessment skills, Maleja stated that she is good at using technology in class and designing practical assessments for students’ grammar use. She also highlighted: “I don’t pass to other topics . . . until I don’t feel like they got it; I don’t continue with another thing.”

Maleja shared two assessment instruments: a reading and grammar test that resembled the national test and a rubric to assess language and research skills among 11th graders (the highest grade in Colombian high schools). The 20 items in the reading and grammar

test were multiple-choice questions, 13 of which were isolated grammar items, two about understanding signs, and five about literal ideas in a dialogue. The rubric included language-based criteria such as fluency, intonation, and pronunciation, as well as research-based criteria such as the connection between results and conclusions in an oral report.

Upon analysis, we found that the rubric aligned with the curriculum Maleja used for teaching and assessment, specifically regarding learners’ ability to orally explain a given problem and its solution. The reading test, on the other hand, may not align with the syllabus because it assessed literal meaning about everyday events, whereas the relevant syllabus unit read that students should develop the ability to “recognize cause-and-effect relationships in texts about academic topics.” We also note that the task in the rubric was for realistic language use, that is, to give a research report; additionally, the instrument clearly delineated what was to be measured in the report.

To sum up, Maleja’s assessment approach integrates three aspects: the school curriculum, her considerations for assessment design and use, and what is ultimately assessed. Although she has studied language assessment, she wishes to advance her LAL, which we think might be possible by addressing topics such as the design of communicative grammar tests and by connecting reading tests to their corresponding learning objectives in a syllabus.

Prudente's Case

Prudente has been a professor of English for six years at a private university in a major Colombian city. She taught various subjects in a language education program—such as literature, English writing, and assessment—using a project-based learning approach. Prudente noted that this approach enabled her to implement “accurate and reliable formative and summative assessment.” The documents that she used to plan her courses were the course syllabus and the

Common European Framework of Reference; however, the syllabus does not specify how students are graded. In the questionnaire, she reported having taken two assessment courses in her undergraduate degree, one in general assessment and one in language assessment.

Prudente seeks to improve her assessments by ensuring they are fair, ethical, reliable, and transparent. Prudente expresses that she might benefit from studying the design of multiple-choice items and test layout. Additionally, she believes that traditional and alternative assessments of English should collect both summative and formative evidence. Such evidence must then be used to identify students' language-learning processes, including their strengths and areas for improvement. She also believes that "formative assessment is essential." She asserts that there must be harmony between traditional and alternative methods. This is evident when she reports that sometimes she does not "get authentic information from alternative assessments."

Based on the interview and questionnaire, we conclude that Prudente's LAL centers on assessing her students' communicative skills and is influenced by her practices and principles. As her institutional document does not specify grading criteria for tasks, this may imply that Prudente can be autonomous in her grading, responding to students' needs. As she stated, she seeks to assess in a fair, ethical, reliable, and transparent manner. Therefore, we think her case can reflect how teachers' beliefs impact their assessment practices and the design of context-specific assessment tasks.

Johnny Rotten's Case

Johnny Rotten is an English language teacher with 25 years of experience at a public high school in a small town. He took a course in assessment as part of his undergraduate degree, though he reported not remembering much information from it. Additionally, he completed an assessment course during his MA studies. He indicated that he wanted to be part of this

study because he is highly interested in assessment and wants to improve his assessment skills and practices.

For designing and teaching his classes, Johnny implements task-based instruction and the official documents provided by the Colombian Ministry of Education (e.g., the English suggested curriculum and the textbook *English, please!*). In addition, he uses the guidelines given by the *Risaralda Bilingüe Program*.¹ For his English for general purposes classes, he considers his students' contexts and interests when planning, teaching, and assessing. He reported that this is also mentioned in the institutional documents.

After revising the document, we noticed that Johnny's curriculum includes a suggested assessment for learning that proposes rubrics, quizzes, and instruments for peer and self-assessment. However, the document does not clearly specify the types of assessment, the percentages, or the purpose. In the interview, Johnny explained that although he can make changes to the document, the institution does not provide any assessment instruments, so teachers must design their own.

In the interview, Johnny identifies some strengths and weaknesses related to assessment. For example, he mentioned that at his school, teachers meet regularly to report on students who are successful and unsuccessful, but they never discuss students' learning process or assessment methods to help students improve. He also revealed that the institution does not provide instruction in assessment but offers materials such as books, headphones, and other items. He is most concerned with the results in standardized tests such as *Pruebas Saber*.²

Finally, the assessment instruments he shared included a quiz and a makeup test. Both instruments

¹ It is a regional initiative aimed at improving English language proficiency among students and teachers in public schools through training, resources, and curriculum development.

² *Pruebas Saber* are a series of standardized tests administered in Colombia to assess the academic performance of students in primary school, middle school, and high school in content areas.

provided simple instructions in English and did not include an example to guide the students' answers. These instruments tended to focus more on vocabulary, grammar, and reading comprehension.

In summary, Johnny conceives of language assessment as a process that needs to be continually revised to address the challenges faced by students, teachers, and institutions. Through this revision, as Johnny argued, teachers can identify areas of improvement and focus on processes rather than on results. While he uses task-based instruction and aligns his teaching with official documents, this teaching approach should be better reflected in the assessment instruments we examined. Johnny's case highlights the importance of addressing the connection between teaching, learning, and assessment, as well as the need to design tasks with a communicative approach to language testing.

Teacher C's Case

Teacher C is an English language teacher with nine years of experience, working at a high school. His knowledge of assessment comes from a course he took in his MA in English Language Teaching. His main interest in this research project is to update his assessment and professional knowledge. In his teaching context, he teaches English by incorporating business-related content, as the school emphasizes entrepreneurship.

He reports that he employs an eclectic approach focused on meaningful learning. In the institutional documents we analyzed, we did not find any direct reference to specific language-teaching methodologies. However, we inferred that Teacher C's language teaching is guided by communicative competence and by the standards outlined in Booklet 22 of the Ministry of Education, which emphasize listening, reading, speaking, and writing skills.

Regarding language assessment, the English syllabus provides no information about theoretical or methodological references to assessment. However,

it uses a holistic approach for all students, addressing aspects beyond language, such as punctuality. The document also outlines that the main purpose of assessment is to observe and collect evidence of progress, using a qualitative grading scale: superior, high, basic, and low. Nevertheless, it does not specify types of assessment, assessment tasks, or percentages. Teacher C added that at his school, assessment serves administrative purposes, evaluating not only language skills but also factors such as attendance and participation. He mentioned that he can provide feedback on this institutional document.

Teacher C conceives assessment as measuring the impact of teaching on students' understanding of class content. In the interview, he stated that he aims to align assessments with the course plan and integrate the four language skills. His goal is to create assessments that reflect course objectives integrated with real-life language use. He also noted that the institution does not provide instruction on assessment but does offer it in teaching and curriculum.

He indicated that his main skill is creating a flexible and comfortable assessment environment. He does so to prevent anxiety, ensuring that students' performance reflects their language skills rather than unrelated factors. He feels he does not have enough time to properly assess his students and identifies this as an area for improvement in his assessment practices. Further, he acknowledged the need to prioritize and adjust his assessments to better suit his students' needs.

He shared a writing task, which aligned with the English curriculum. He also shared an exam with Spanish instructions that assessed listening, reading, and writing. Both instruments reflected Teacher C's intention to assess language through communicative activities.

To conclude, Teacher C perceives assessment as a tool to gauge the impact of teaching, focusing on students' understanding of content, and aligning assessments with course objectives and real-world language use. Finally, his strengths lie in creating a supportive

classroom environment, but there is room to refine his assessment methods to better suit his students' context and skills.

Cross-Case Findings

The Communicative Approach as a Backbone for Teaching and Assessment

After analyzing data from the four data collection methods, it became apparent that the communicative approach permeated teachers' assessment practices in different ways and was operationalized through methodologies such as task-based instruction. Such an approach was most prominent in curriculum documents and the teachers' reported practices. Firstly, we found that communicative competence was the main goal for teaching and assessment. For example, the following excerpt comes from the syllabus provided by Johnny Rotten, where communicative competence is mentioned as a goal for teaching and assessment: "Promote communicative competence and sub-competences: sociolinguistic, pragmatic, and linguistic. . . . The syllabus has tasks as a central component, which are integrated with communicative competence and relevant learning standards in the English language."

The communicative approach was also evident in teachers' views of language use through communicative tasks. In the interview, Teacher C reported that he believes assessment should address language "as close as possible so students can use it in real-life situations." In the questionnaire, Maleja and Prudente reported using rubrics to assess communicative skills through oral presentations.

Content Validity in Teachers' Approach to Language Assessment

Another assessment pillar we found in the data was content validity, which refers to the extent to which assessments cover relevant curriculum content. We

found evidence of content validity, as teachers aligned their assessments with their lessons and the curriculum, specifically the syllabi for the courses they taught. For example, Sofia, in the interview, reported the following about test design: "I take into consideration those assessment criteria listed in the [curriculum] matrix, so that [students] meet those criteria and that they are effectively putting what we study into practice."

Our analysis of assessments also provided evidence of content validity. In a progress test of listening skills, grammar, and vocabulary in context, Teacher C aligned the test items with the learning objective: "To recognize vocabulary related to greetings and answering basic personal information questions." His assessment instrument included these two items:

Listening. Listen to four people introducing themselves and write down the expressions they use to greet.

- a. _____
- b. _____
- c. _____
- d. _____

Writing. Answer.

What is your name?

What is your last name?

How are you in this moment?

Salient Challenges of Construct Irrelevance in Language Assessment

After our data analysis, we found evidence of construct irrelevance (i.e., assessing something unrelated to language ability) in participants' approaches to assessment in various ways. These issues were evident in the inclusion of factors that may unintentionally introduce bias or measure skills beyond language constructs. The issues were salient in institutional documents such as the English curriculum, syllabi, and assessment instruments designed by the participants. For instance, Teacher C's and Prudente's English curriculum clearly state that there should be specific elements that teachers must assess in their language

classes and that do not have to do with the language itself, as indicated in the following excerpts from the two institutional documents:

Assessment criteria: Collaboration and solidarity, sense of belonging, purpose, punctuality, aesthetics. The student shows aesthetics and organization in the presentation of work, activities, and the guide notebook. (Teacher C, English curricular plan)

Assessment criteria: Students' attitude: Responsibility, creativity, and initiative; autonomy and interaction with members of the specific educational community. (Prudente, English syllabus)

Construct irrelevance was also evident in the assessment instruments designed by the teachers. For example, some assessment tasks incorporated content or skills that were not directly related to language, such as requiring knowledge of specific terminology from other areas in a general English class. For example, Teacher C included this in his writing rubric: "Your text is aesthetic and well-presented." Similarly, Maleja includes volume in a rubric to assess speaking skills: "Volume: The intensity which the student instills in their voice during the presentation." This can be problematic, as someone with a low voice can speak the language well, or vice versa.

The data analysis revealed significant issues of construct irrelevance in language assessment practices. The inclusion of extraneous factors, such as punctuality, aesthetics, and attitude, may detract from the primary goal of assessing language proficiency. Institutional documents, including curricula and syllabi, established assessment criteria unrelated to linguistic ability.

To conclude, based on LAL's three components and the data analysis and cross-case findings, we drew three conclusions. The cases provide evidence of *knowledge* of the communicative approach in language planning (curriculum), teaching, and assessment. Most importantly, such knowledge was reported through what can be considered the *skills* side of LAL (e.g., the use

of communicative criteria in rubrics). Furthermore, LAL skills were evident in teachers' use of assessment instruments clearly aligned with the learning objectives of a given course. Finally, the presence of construct irrelevance in the teachers' LAL may pose threats to LAL principles (e.g., ethics and fairness). To illustrate, if teachers have to grade responsibility and creativity in an English language class, this may introduce a lack of ethics (e.g., having to grade something outside of language skills) or biases (a problem of fairness) by overgrading or downgrading students' language level based on irrelevant knowledge or skills (e.g., punctuality and aesthetics).

Discussion

Teachers' LAL develops through their assessment trajectories and related factors, such as engaging in assessment-related projects or attending professional development opportunities (Gan & Lam, 2025; Yan et al., 2018). In our study, participants' LAL trajectories are also influenced by institutional documentation, particularly policies and constraints. Most importantly for LAL development, we argue, was these teachers' stated motivation to continue developing their LAL. Language teachers' interest in assessment has been reported elsewhere and, perhaps not surprisingly, is a major driver of participation in LAL initiatives (Fulcher, 2012).

Our study also highlights the richness and utility of case studies in advancing LAL (Baker, 2024; Giraldo, 2024). Analyzing the five cases, individually and together, allowed us to arrive at findings that, we believe, point to the specific needs of all participants. Such careful analysis can be useful to devise appropriate LAL courses (Bøhn & Tsagari, 2021; Giraldo & Yan, 2023). Additionally, each case uniquely informs what to include and how to teach an assessment course (see Conclusions section), which we believe is the starting point for a pedagogy that integrates LAL (Fulcher, 2020; Giraldo & Yan, 2025).

Our cross-case analysis uncovered three dominant themes: the prominence of the communicative approach, evidence of content validity, and persistent construct irrelevance for assessment. First, all cases suggested that while teachers embraced the communicative approach, operationalizing it in assessment instruments was challenging; a finding that echoes calls for LAL pedagogies that bridge principles and practice (Fulcher, 2020; Giraldo & Yan, 2025). Second, content validity emerged as a strength in teachers' practices; however, it was frequently undermined by construct irrelevance, a theme that has received limited attention in prior research (Gan & Lam, 2025). Institutional documents included non-linguistic criteria (e.g., punctuality, aesthetics), and test items assessed extraneous knowledge. Instead of addressing these as individual deficits, we interpret them as tensions embedded in institutional policies that should be considered when contextualizing LAL (Inbar-Lourie, 2017). Thus, we argue that, based on this diagnosis, future LAL research must directly address construct (ir)relevance and validity threats as situated, institutional issues rather than as teachers' knowledge gaps.

Research on teachers' LAL has also shown discrepancies between what they believe language assessment should do and what they ultimately do (Díaz Larenas et al., 2012; López Mendoza & Bernal Arandia, 2009). In our study, however, we highlight that such discrepancies (e.g., believing that language assessment should be communicative while assessing linguistically) are not necessarily a teacher's problem but rather a mismatch between teachers' beliefs and practices and what they are required to do where they work (see Sofia's case, for example). Thus, as Inbar-Lourie (2017) indicates, these may be considered tensions in teachers' context and, we argue, not necessarily a problem with teachers' LAL.

Conclusions and Implications for Course Planning

This study characterized the LAL of five Colombian English language teachers to inform the design

of a context-sensitive assessment course. The findings reveal a complex interaction among personal convictions, professional practices, and institutional constraints. Besides, the findings highlight that while teachers hold strong beliefs about the communicative and formative purposes of language assessment, their practices are often shaped by institutional policies, resource limitations, and external demands.

Consequently, this research provides insights into the diverse trajectories of LAL development and emphasizes the need for context-sensitive LAL courses that address teachers' individual and institutional needs. Supporting teachers in aligning their assessment beliefs and practices requires not only LAL but also fostering institutional environments that prioritize meaningful and communicative assessment practices.

As synthesized in the discussion section, the three-case themes: the communicative approach, content validity, and construct irrelevance, revealed shared strengths and systemic constraints across participants. In our cross-case analysis, we also identified individual LAL factors that emerged and derived specific, actionable implications for course planning. For instance, Sofia required support in connecting communicative goals with vocabulary and grammar assessment. Maleja needed guidance on designing communicative grammar tests and aligning reading tests with learning objectives. Prudente needed to refine the design of multiple-choice items and the test layout. Johnny Rotten required explicit connections between task-based instruction and communicative assessment tasks. Teacher C needed strategies to manage time constraints and adjust assessment rigor to student readiness. Thus, while shared themes emerged, each teacher's LAL profile demanded specific attention.

Drawing from our findings, we propose that the assessment course for these teachers should incorporate the study of communicative language testing and its relation to teaching and learning, including an examination of teachers' beliefs and practices in this

area. Within this approach to assessment, the course should focus on instrument design that specifically assesses linguistic criteria, while setting aside unrelated aspects or irrelevant constructs (e.g., aesthetics). This would mean delving into teachers' beliefs and practices regarding language assessment.

Given the contextual nature of LAL, the course should include discussion tasks in which teachers share their individual LAL profiles to problematize language assessment at their institutions: policies, practices, strengths, and aspects to improve. Ultimately, such discussions could give teachers insights to develop a strong assessment culture in the contexts where they work. Finally, the course should foster a collaborative learning environment where teachers can share experiences, challenges, and solutions, thereby building a community of practice for LAL development. Appendix D shows the contents and tasks we decided to include in the assessment course for the five teachers in our study, along with commentary on the lessons.

Regarding limitations, even though this was a multiple-case study with a small sample size, our purpose is not to generalize teachers' LAL experiences, but to offer a contextualized account that teachers and teacher educators in other contexts could adapt. Additionally, this study did not include classroom observations as a data-collection instrument; therefore, we recommend that future research incorporate observational data to gain deeper insights into the institutional realities that teachers face as they enact their LAL after participating in LAL courses.

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. (2021, July 2). *Unpacking the concept of language assessment literacy for all key stakeholders* [Conference session]. New Directions in English Language Assessment Conference, Bogotá, Colombia.
- Baker, B. (2024). Introduction. In B. Baker & L. Taylor (Eds.), *Language assessment literacy and competence volume 2: Case studies from around the world* (pp. 1–7). Cambridge University Press & Assessment.
- 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>
- Baker, B., & Taylor, L. (Eds.). (2024). *Language assessment literacy and competence volume 2: Case studies from around the world*. Cambridge University Press & Assessment.
- Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal*, 73(2), 113–123. <https://doi.org/10.1093/elt/ccy055>
- Bøhn, H., & Tsagari, D. (2021). Teacher educators' conceptions of language assessment literacy in Norway. *Journal of Language Teaching and Research*, 12(2), 222–233. <https://doi.org/10.17507/jltr.1202.02>
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). SAGE Publications.
- Cui, Y., 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, 102866. <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>
- Díaz Larenas, C., Alarcón Hernández, P., & Ortiz Navarrete, M. (2012). El profesor de inglés: sus creencias sobre la evaluación de la lengua inglesa en los niveles primario, secundario y terciario. *Íkala, Revista de Lenguaje y Cultura*, 17(1), 15–26. <https://doi.org/10.17533/udea.ikala.10039>
- Egbert, J., & Baker, P. (Eds.). (2019). *Using corpus methods to triangulate linguistic analysis*. Taylor & Francis. <https://doi.org/10.4324/9781315112466>

- 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. (2015). *Re-examining language assessment: A philosophical and social inquiry*. Routledge. <https://doi.org/10.4324/9781315695518>
- Fulcher, G. (2020). Operationalizing assessment literacy. In D. Tsagari (Ed.), *Language assessment literacy: From theory to practice* (pp. 8–28). Cambridge Scholars.
- Gan, L., & Lam, R. (2025). Language assessment literacy development of a novice university English teacher in the Chinese context. *RELC Journal*, 56(2), 346–361. <https://doi.org/10.1177/00336882241231304>
- Giraldo, F. (2024). *Language assessment literacy for in-service teachers: Online course planning and implementation*. Editorial Universidad de Caldas. <https://doi.org/10.2307/jj.13027271>
- Giraldo, F., & Yan, X. (2023). Planning an online assessment course for English language teachers in Latin America. *Profile: Issues in Teachers' Professional Development*, 25(2), 147–164. <https://doi.org/10.15446/profile.v25n2.104703>
- Giraldo, F., & Yan, X. (2025). Evaluating an online assessment course: Teachers' voices on their language assessment literacy. *Profile: Issues in Teachers' Professional Development*, 27(1), 97–114. <https://doi.org/10.15446/profile.v27n1.113304>
- 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. (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. 257–270). Springer. https://doi.org/10.1007/978-3-319-02261-1_19
- Inbar-Lourie, O. (2024). 'A course of our own': Reflections following a language assessment literacy (LAL) course. In B. Baker & L. Taylor (Eds.), *Language assessment literacy and competence volume 1: Research and reflections from the field* (pp. 231–238). Cambridge University Press & Assessment.
- 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>
- Levi, T., & Inbar-Lourie, O. (2020). 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>
- 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 Teachers' Professional Development*, 11(2), 55–70.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Wiley.
- Moraes, I. T., & Furtoso, V. A. B. (2026). Who will teach the teacher educator? Findings and implications of an online workshop for fostering Brazilian teacher educators' language assessment literacy. In F. Giraldo & X. Yan (Eds.), *Fostering language assessment literacy: Materials, methods, and emerging issues* (pp. 145–165). Routledge. <https://doi.org/10.4324/9781003483915-11>
- Moraes, I. T., Furtoso, V. A. B., & Giraldo, F. (2025). Developing pre-service teachers' language assessment literacy: An analysis of classroom tasks. *Ilha do Desterro*, 77, 1–20. <https://doi.org/10.5007/2175-8026.2024.e96898>
- Randall, J., Poe, M., Slomp, D., & Oliveri, M. E. (2024). Our validity looks like justice. Does yours? *Language Testing*, 41(1), 203–219. <https://doi.org/10.1177/02655322231202947>
- Restrepo Bolivar, E. M. (2020). Monitoring preservice teachers' language assessment literacy development through journal writing. *Malaysian Journal of ELT Research*, 17(1), 38–52.
- 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>

- Simons, H. (2012). *Case study research in practice*. SAGE Publications.
- Spiliotopoulos, V., Van Viegen, S., Early, M., & Lam James, C. (2024). Teacher professional learning communities and language assessment reform in western Canada. In B. Baker & L. Taylor (Eds.), *Language assessment literacy and competence volume 2: Case studies from around the world* (pp. 8–19). Cambridge University Press & Assessment.
- Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9(1), 2–14. <https://doi.org/10.1186/s40468-019-0077-8>
- 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>
- Tsagari, D., & Armostis, S. (2025). Contextualizing language assessment literacy: A comparative study of teacher beliefs, practices, and training needs in Norway and Cyprus. *Education Sciences*, 15(7), 927. <https://doi.org/10.3390/educsci15070927>
- Villa Larenas, S., & Brunfaut, T. (2023). But who trains the language teacher educator who trains the language teacher? An empirical investigation of Chilean EFL teacher educators' language assessment literacy. *Language Testing*, 40(3), 463–492. <https://doi.org/10.1177/02655322221134218>
- 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>
- 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>
- Yan, X., Zhang, C., & Fan, J. 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>
- Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). SAGE Publications.
- Zhang, J., Yu, G., & Browne, W. J. (2025). Teachers' language assessment literacy: Exploring its construct and contextual factors. *Studies in Educational Evaluation*, 87, 101525. <https://doi.org/10.1016/j.stueduc.2025.101525>

About the Authors

Frank Giraldo holds an MA in Teaching English as a Second Language from the University of Illinois at Urbana-Champaign. He is a professor at the Foreign Languages Department of Universidad de Caldas in Manizales, Colombia. His main academic and research interest is the interface between language assessment literacy and teachers' professional development.

Roles: conception of the research project; collection, analysis, and interpretation of data; major contribution of investigation materials; and drawing up the different text versions of the paper.

Angela Patricia Velásquez-Hoyos holds an MA in English Didactics from the Universidad de Caldas in Colombia. She is a professor in the Department of Foreign Languages at Universidad de Caldas. Her research interests include critical pedagogy in English language teaching, materials design, and interculturality.

Roles: collection, analysis, and interpretation of data; major contribution of investigation materials; and drawing up the different text versions of the paper.

Luis Herney Villegas holds an MA in English Didactics from the Universidad de Caldas in Colombia. He is a professor of English in the Department of Foreign Languages at Universidad de Caldas. His research interests are teaching English to students with special educational needs, international exams, and assessment.

Roles: collection, analysis, and interpretation of data; major contribution of investigation materials; and drawing up the different text versions of the paper.

Appendix A: Sample Items From the General Information Questionnaire

- During your formal studies, did you receive any training in English language assessment?
- If applicable, mention and describe the spaces or contexts in which you have received training on topics related to assessment.
- Why are you interested in participating in this training project?
- Describe the methodology or methodologies you use to conduct your English classes.
- What documentation is used in your institution to plan and organize English courses?
- Mention and describe the type of English courses you teach.
- How do you assess your students' academic progress in your English classes?

Note: The original questionnaire was in Spanish. This translation was made for publication purposes.

Appendix B: Document Analysis Grid

About assessment in curriculum documents (if applicable):

Why (Purposes)

What (Constructs)

How (Methods)

Methodological considerations / Approaches to language teaching and learning:

Theoretical and methodological references (about assessment):

For example: language planning as per the MEN in Colombia + policies:

Percentages + assessment criteria:

Type: Summative and/or formative

Appendix C: Interview Protocol

1. What do you like most about the institution where you work?
2. What do you understand by English language assessment? What do you think is the purpose of assessing this language?
3. Do you use any documents for your English assessment practices? What do you use those documents for? Can you contribute to these documents, and do you make any contributions to them?
4. In your opinion, how should your students' English be assessed?
5. When designing your assessments, what do you take into account? What aspects do you consider when assessing?
6. Does the institution offer teacher training opportunities on assessment?
7. Does the institution have resources to support assessment in the English classroom? Which ones, and how are they used?
8. What are your strengths regarding English language assessment in the institution where you work?
9. Do you think there is anything that could be improved?
10. Do you conduct any type of analysis of your assessment practices?
11. What recommendations do you have for the online course we are going to offer? What would you like to learn about English language assessment?

Appendix D: Topics and Tasks for an Online LAL Course

Topics	Tasks
1. Communicative Language Assessment (CLA)	Design communicative tasks, criteria, and items.
Commentary: Connect the communicative approach to CLA; assessment implications for the curriculum, teaching, materials, tasks, test items, criteria, and student involvement.	
2. CLA in the Curriculum, Lessons, and Instruments	Align curriculum documents, assessment instruments, and assessment for learning.
Commentary: Explore the influence of CLA on syllabi, lessons, and tests.	
3. Authenticity in Language Planning, Teaching, and CLA	Write communicative language learning objectives. Design authentic tests.
Commentary: Study authenticity within a multicultural and multilingual framework; authenticity in the curriculum, instruction, learning, and assessment.	
4. Construct (Ir)relevance and Score Pollution in CLA	Evaluate the presence (or lack thereof) of content and construct validity, and score pollution and its impact.
Commentary: Problematicize criteria unrelated to language and discuss score pollution and its impact.	
5. Principles for an Assessment Culture	Write a small-scale proposal to advance a language assessment culture in their school(s).
Commentary: Share ideas on how teachers could raise the <i>status quo</i> of language assessment in their institutions.	