



Exploring Teacher's Cognition and Practices of Testing Speaking in Colombian Universities¹

Lorena Idarraga Robayo² 

Citation: Idarraga, L. (2026). Exploring Teacher's Cognition and Practices of Testing Speaking in Colombian Universities. *Colomb. Appl. Linguistic. J.*, 28(1), pp. 43-58.

Received: 24-Jan.-2025 / **Accepted:** 03-Sep.-2025

DOI: <https://doi.org/10.14483/22487085.23179>

Abstract

This study examines Colombian English teachers' knowledge, experiences, and views regarding speaking assessments. Semi-structured interviews were conducted with Colombian teachers, some of whom also shared classroom assessment samples. Thematic analysis was used for the interviews, and content analysis was applied to the assessment materials. Findings revealed a clear link between teachers' perceptions of testing and their approach to presenting it to students. Furthermore, a correlation exists between their knowledge and experience in designing, administering, and evaluating speaking tests. Lastly, there is evidence of teachers' awareness of the issues surrounding these testing practices. The shared concern about the effectiveness of current practices, along with the expressed need for further training, underscore the importance of developing targeted professional development programs to better prepare teachers in this area.

Keywords: Experiences; Language Testing; Speaking; Opinions; Knowledge

¹ This article is a result of dissertation research to achieve the title of Master in Applied Linguistic to TESOL from the University of Leicester, England.

² British Council, London, Great Britain, Email: Lorena.idarraga@hotmail.com



Explorando la cognición y las prácticas para evaluar la habilidad de habla de los docentes en las universidades colombianas

Resumen

Este estudio examina los conocimientos, experiencias y percepciones de los docentes colombianos de inglés en relación con la evaluación de la habilidad de habla. Se llevaron a cabo entrevistas semiestructuradas con profesores colombianos, algunos de los cuales también compartieron muestras de evaluaciones utilizadas en el aula. Para el análisis de las entrevistas, se empleó un análisis temático, mientras que las muestras de evaluación se examinaron mediante el análisis de contenido. Los hallazgos revelaron una relación clara entre las percepciones de los docentes sobre la evaluación y la manera en que la presentan a los estudiantes. Así mismo, se identificó una correlación entre sus conocimientos y experiencias en el diseño, aplicación y evaluación de pruebas orales. Finalmente, se evidenció que los docentes eran conscientes respecto a los desafíos asociados con estas prácticas evaluativas. La preocupación compartida sobre la efectividad de las prácticas actuales, junto con la necesidad expresada de recibir mayor formación, subrayan la importancia de desarrollar programas de desarrollo profesional específicos que fortalezcan las competencias docentes en esta área.

Palabras clave: conocimientos; experiencias; habilidad de habla; opiniones; prueba

Introduction

Although testing is a routine part of teaching, many educators find it challenging and may not view it as a favored aspect of their role. As a result, assessment practices can become disconnected from daily classroom instruction. Despite its centrality to student evaluation, there is limited research exploring how teachers perceive and integrate testing into their pedagogical practice (Reynolds, 2010; Muñoz, *et al.*, 2012; Palacio, *et al.*, 2016). This absence of cohesion can hinder a learning process that is inherently cyclical and focused on growth. Students learn and practice, but they also undergo testing, and these elements should seamlessly integrate into a unified process. Teachers must navigate numerous decisions in the classroom, including how to design and plan assessments. They need to determine what to assess; how to create the tasks; how much time to allocate; when to schedule them; how to score them, and the format; among other factors.

The decisions made in the area of testing can have a significant impact on the student's learning process, and possibly in the way these students face tests moving forward on their education; indeed, there have been numerous studies, which investigate the impact of high-stakes tests on entire education systems (Hughes, 2003; McNamara, 2000, East, 2015). Moreover, Shohamy (2001) contends that every aspect of a teacher's thought system impacts their testing practices, suggesting that teachers' beliefs are vital for understanding their classroom approaches, and for promoting fair, informed assessments that benefit both teachers and students. Despite some existing studies on assessment practices and test analysis in Colombia, a significant gap remains in understanding teachers' beliefs about testing, which constitutes a largely neglected area in educational research. This study aims to explore Colombian teachers' beliefs about testing to contribute to a more comprehensive understanding of assessment practices in the national educational context.

It could be beneficial to examine how EFL teachers design speaking tests, and what this reveal about their classroom practices, since there is evidence suggesting that teachers' beliefs about assessment strongly influence their teaching methods. (Calderhead, 1996; Clark & Peterson, 1986; Thompson, 2006). To accomplish this goal, it is essential to gain an understanding of the perspectives of current teachers regarding which elements of their experience, knowledge, and beliefs influence the design, scoring, and administration of the speaking tests they currently use. According to authors such as Díaz, Hernández and Navarrete (2012), Farrell, (2012), and Pajares (1992), teachers themselves can benefit from engaging in a reflection process about their assessment procedures. Both teachers and test developers can gain deeper insights by examining the process of developing and administering tests; approach known as reflective practice.

The goal of this study is to explore Colombian English teachers' cognition regarding speaking assessment. Specifically, it aims to answer the following research questions: (1) What do teachers know, believe, and think about testing in general, and speaking assessment in particular? (2) How does this teacher cognition influence their current classroom practices related to the assessment of speaking skills? This study seeks to illuminate the connection between teacher beliefs and assessment behavior, providing insights that can inform teacher training and assessment design.

Literature Review

Assessing speaking ability involves multiple linguistic and cognitive dimensions, such as grammar, vocabulary, listening comprehension, and non-verbal communication. Given this complexity, scholars emphasize the importance of clearly defining the construct to be assessed. A well-defined construct, whether it involves general language proficiency or specific subskills, is essential to ensure both the validity and reliability of speaking assessments (Taylor, 2011; Messick, 1996). Fulcher (2003) argues that speaking tests should reflect real-world communicative abilities, including fluency, topic management, and conversational roles, while Taylor (2011) warns that failing to define the construct can lead to construct underrepresentation or construct-irrelevant variance, both of which compromise test validity.

Once the construct is established, researchers stress the importance of task design that mirrors authentic language use. Bachman and Palmer (1996) define assessment tasks as purposeful language activities, a view supported by Pica, Kanagy, and Falodun (1993), who highlight the role of interaction and communication goals.

[Fulcher \(2003\)](#) further advocates the use of varied task types to capture the full spectrum of speaking skills. Finally, to ensure the reliability of scoring, it is necessary to develop a scoring system aligned with the construct and to prototype assessments with learners before implementation. Beyond task design, understanding how teachers perceive and implement speaking assessment is crucial.

Teacher Cognition

Teacher cognition encompasses what teachers know, believe, and think, and how these factors influence their instructional and assessment practices. [Borg \(2001\)](#) outlines four key influences on teacher cognition, namely personal schooling, professional training, contextual factors, and classroom experience. Prior language learning experiences have a strong impact on teaching strategies ([Numrich, 1996](#)). Research suggests that teachers often rely on personal, and sometimes, inconsistent beliefs about assessment ([Cizek et al., 1995](#); [Kahn, 2000](#)), with perspectives ranging from viewing assessment as a valuable tool to seeing it as a bureaucratic formality ([Brown, 2004](#)). Studies also indicate that diverse professional experiences shape assessment practices ([Cheng, et al., 2004](#)), yet many teachers lack a deep understanding of assessment principles, depending more on intuition than on formal guidelines ([Childs & Lawson, 2003](#); [Leighton et al., 2010](#)). This issue is particularly pronounced in Latin America, where research on assessment-related teacher cognition remains limited. For example, while Chilean teachers express support for communicative assessment models, effective implementation is often lacking ([Díaz, et al., 2012](#)). In Colombia, there is a notable gap in empirical research on this topic.

Teachers' Perspectives on Speaking Assessment

Research conducted by [Johnson, Mehta, and Rushton \(2015\)](#) found that teachers in England felt more engaged when they designed speaking tasks themselves. Similarly, [Devrim \(2018\)](#) highlighted Turkish EFL instructors' need for diverse and interactive testing methods, emphasizing alignment between classroom activities and assessment tasks.

The scoring system plays a crucial role in interpreting a test taker's ability. [Luoma \(2004\)](#) describes rating scales as indicators of the tester's understanding of the construct. There are two main approaches to rating scales: Measurement-Driven, based on expert consensus and generalizable across contexts (e.g., CEFR), and Performance-Data Driven which relies on discourse analysis from specific contexts ([Fulcher et al., 2012](#)). The latter, while contextually reliable, may not be efficient for large-scale testing. Having explored teachers' perspectives on the topic, it is important to situate these insights within the broader context of research.

Higher Education Testing in Colombia

In Colombia, although the National Bilingualism Program ([MEN, 2006](#)) mandates a B2 English level for higher education graduates, fewer than 5% achieve this standard. This discrepancy raises concerns about the validity of EFL testing practices which are often academically weak and potentially inflated by institutions, according to what research suggests ([López & Bernal, 2009](#)).

One contributing factor to this testing gap is the inadequate training of EFL teachers. Although some graduate programs offer assessment-related coursework, many teachers lack access to these resources, resulting in inconsistent and unsystematic assessment practices ([López & Bernal, 2009](#)). This disconnection between training and practice has significant theoretical implications: when teachers lack a clear understanding of the constructs they are assessing, they risk omitting essential dimensions of language ability, or what [Messick \(1996\)](#) coins: construct underrepresentation.

Additionally, teacher perceptions of assessment often reflect disinterest or anxiety, frequently rooted in negative personal experiences. [Reynolds \(2010\)](#) observes that many teachers perceive assessment as stressful and separate from the learning process. [López and Bernal \(2009\)](#), or even [Toro and Patarroyo \(2005\)](#) suggest that untrained teachers may conflate assessment with mere numerical grading, prioritizing surface-level linguistic features over meaningful communication. This narrow focus not only reflects a lack of construct clarity but also reinforces assessment practices that are misaligned with communicative language teaching goals.

Methodology

In this study, the interpretivist paradigm, along with qualitative research methods, is used to gain an in-depth understanding of what is happening in the classroom, offering an insider's perspective on the situation ([Richards, 2003](#)), which is essential to meet the study's objectives. Various research tools are available in qualitative research, from which interviews were chosen, because they provide richer insight into teachers' views on speaking assessments. [Heigham and Croker \(2009\)](#) note that interviews reveal beliefs, perceptions, and motivations, in contrast to other methods, such as questionnaires, which cannot capture all of this. The interviews were analyzed using thematic analysis, which involved categorizing data into themes, a process that [Dörnyei \(2007\)](#) refers to as coding. The steps in this process include transcribing the data, identifying patterns, and creating categories for later interpretation ([Brown, 2014](#)). This coding process helps organize the data into interconnected networks, making it easier to identify patterns and connections that can lead to inferences. Additionally, participants shared examples of speaking tests they use in their classrooms, which also served as valuable data. These samples illustrate the teachers' procedures for speaking assessments and provide evidence supporting the insights gathered from the interviews applied. Samples collected were examined using content analysis, a method that interprets the material considering the research questions ([Schreier, 2012](#)). This approach goes beyond analyzing the data's content by also considering the creator's perspective, the context of use, and the potential impact of the material ([White & Marsh, 2006](#)).

The study involves ten participants, all of whom are EFL university teachers with bachelor's degrees in TESOL. Eight participants hold master's degrees in TESOL-related fields, while two are MA candidates. The number of participants was chosen to allow for a deeper exploration of the data. The sample was selected purposively, and it involved teachers with, at least, three years of experience in teaching EFL at the university level. Also, participants were chosen to represent various types of universities in Colombia, including teachers from three private and three public universities. Additionally, teachers were required to have at least a bachelor's degree, and preferably a master's degree in TESOL or a related field. This ensures that the participants possess both the academic qualifications and practical experience necessary to provide well-informed insights into the EFL teaching environment at universities.

The interviews conducted in this study were semi-structured, a format described by [Heigham and Croker \(2009\)](#) as a conversation with a purpose. While the interviews allowed for natural conversation, they were guided by a structured set of topics. The interview guide was divided into three sections: the first focused on background information, the second explored the participants' perceptions of testing, including their experiences with it and their views on its impact on students' lives, while the third section delved into their current practices in speaking assessment, including test design, administration, and scoring. There was a piloting process of the interview questions; as a result, the timing was adjusted, and some follow-up questions were added. The interviews were held via Zoom, which recorded both audio and video. The participants' identities were anonymized in the transcriptions and labeled as T1, T2, etc. Four participants also shared their speaking test samples, which were analyzed through content analysis. The coding for the thematic analysis and the transcriptions of the interviews were both done manually to ensure accuracy and a detailed understanding.

Results

Thematic Analysis of Interviews

Following the thematic analysis of the interviews and utilizing the thematic network framework proposed by [Attride and Stirling \(2001\)](#), a network was developed through a structured process. The first step involved coding the material, which entailed summarizing the information into digestible segments. To facilitate this, a chart was created that summarized the various responses from participants based on the interview questions. It is important to note that not all participants received identical questions, due to the nature of the interviews; however, the topics and overarching concepts remained consistent. In the second stage of network construction, key themes began to emerge from the decoded information. The next step consisted of refining these themes into specific (non-repetitive) and broad (capable of encapsulating ideas) categories, which represented the outcome of this phase.

Table 1. Thematic Analysis Stage 1

Theme	Classification
Studies	Specific
Testing Courses	Specific
Experience University	Specific
Designing Experience	Specific
Importance of Assessment	Broad
Own experience with assessment	Broad
Own assessment practices	Broad
Challenges evaluating students	Broad
Own testing procedures	Broad
How to assess speaking	Broad
Types of speaking tests	Broad
Scoring speaking tests	Broad
Student's perceptions	Broad
Suggestions	Specific

Source: Own elaboration.

Following this stage, themes were categorized, considering their content, theoretical framework, and similarities. Each group results in a common theme that forms a global theme. During this phase, themes were organized, according to the insights they provided regarding teachers' assessment practices and teacher cognition, which ultimately became the global themes.

Table 2. Thematic Analysis Stage 2

Theme	Global theme
Studies	Teacher's Cognition
Testing Courses	Teacher's Cognition
Experience University	Teacher's Cognition
Designing Experience	Teacher's Cognition
Importance of Assessment	Teacher's Cognition
Own experience with assessment	Teacher's Cognition
Own assessment practices	Current practices in EFL Higher Education.
Challenges evaluating students	Current practices in EFL Higher Education.
Own testing procedures	Current practices in EFL Higher Education.
How to assess speaking	Current practices in EFL Higher Education.
Types of speaking tests	Current practices in EFL Higher Education.
Scoring speaking tests	Current practices in EFL Higher Education.
Student's perceptions	Teacher's Cognition
Suggestions	Teacher's suggestions to improve

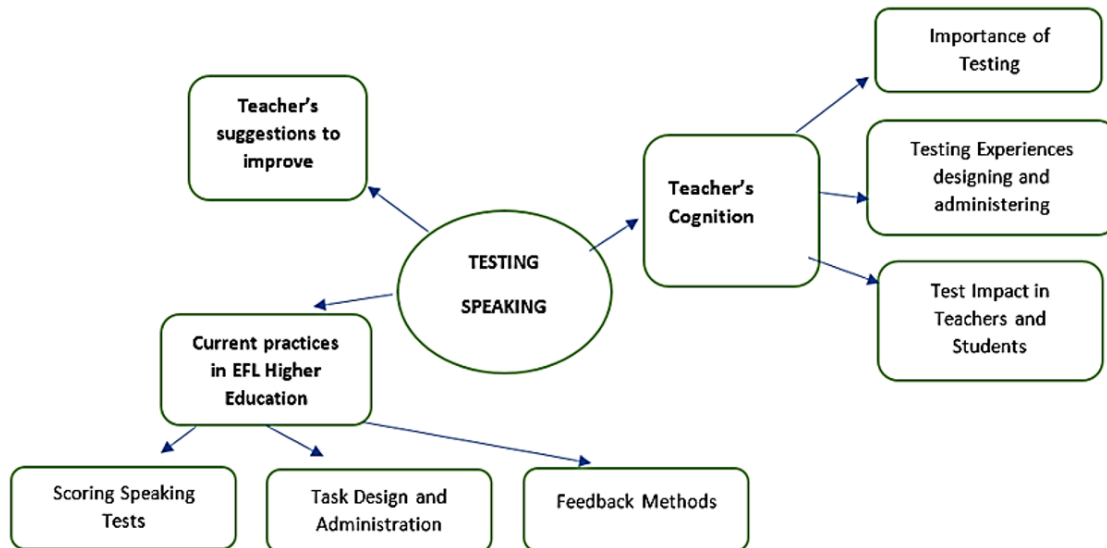
Source: Own elaboration

Subsequently, the basic themes were refined to ensure they accurately represent the collected data. During this stage, some predefined themes may be renamed or consolidated into more academic categories. For instance, themes related to teachers' practices could be divided into scoring tests, design and administration, and feedback methods, while those addressing challenges and perceptions could be grouped under teachers' experiences in

administering and designing tests, among others. This process led to the development of the next network, followed by a summary of the results for each theme.

Within this section, in accordance with the anonymity clause outlined in the consent form, participants will be identified by numbers preceded by the letter T (e.g., T1, T2, T3, etc.).

Figure 1. Thematic Network



Source: Own elaboration.

Teacher's Cognition (Importance of testing)

Participants recognized the significant role of testing in learning English as a Foreign Language (EFL); indeed, two of them described it as essential. They noted that testing informs students about their progress and provides teachers with critical insights to evaluate their teaching practices and assess whether students meet initial expectations. Two participants emphasized the role of testing in standardization, stating that it establishes fairer measures for all students and evaluates the effectiveness of EFL materials and methodologies.

However, four participants stressed the importance of a proper testing process, highlighting that effective implementation benefits all stakeholders, while poor execution can harm students. One participant pointed out that subjectivity can influence the evaluation process, as some teachers may not clearly understand what they are assessing, allowing factors such as student behaviour to affect scores. Additionally, there was a concern regarding the preference for summative evaluation over formative evaluation in Colombian institutions. This trend prioritizes numerical scores over the insights those scores might provide about the learning process. T6 remarked: "In the institution where I was working, they were more concerned about the numbers rather than the process". Overall, this discussion reflects a consensus on the critical need for effective and fair testing practices in EFL education, alongside concerns about the potential pitfalls of prioritizing scores over genuine learning outcomes.

Testing Experiences

The experiences of teachers in administering and taking language tests significantly shape their perceptions and approaches to assessment. In this study, most teachers reported generally positive experiences with language testing and did not associate it with negative emotions. However, several concerns emerged from their reflections. Two teachers highlighted the anxiety associated with speaking tests, which required them to improvise on unfamiliar topics. Another pair expressed dissatisfaction with the lack of clarity regarding evaluation criteria during their own

testing experiences, noting that there was little opportunity for teachers to explain the assessment process. One teacher recounted feeling unprepared during speaking exams, which heightened anxiety and led to unfair conditions for students who were not adequately practiced in speaking skills.

Conversely, some teachers viewed exams as challenges to overcome rather than obstacles. One teacher referred to fostering a sense of collaboration among students to face exams collectively and then reducing anxiety through a supportive classroom environment. He emphasized the importance of helping students understand the testing process, framing it as a common enemy that they could confront together.

Regarding test design, all teachers acknowledged that creating assessments was a routine part of their jobs. However, out of ten participants, only one had received formal academic training in testing. This lack of training is common, as it is often not covered in undergraduate or graduate programs. Two teachers admitted feeling uncomfortable with their testing practices, highlighting a need for improvement in this area. One teacher articulated this sentiment, noting that despite his strengths in classroom instruction and student profiling, he frequently doubts his testing approaches.

When discussing their experiences in test design, four participants shared that they had been part of an instructional design committee tasked with developing new curricula, materials, and assessments. They unanimously found this experience beneficial, as it helped them create fairer assessments, aligned with their teaching methodologies. One teacher illustrated this by explaining how their exams were intentionally connected to the lessons, reflecting a thoughtful design process.

However, some teachers perceived a disconnection between curriculum and assessment, which could undermine the validity of test scores. T4 remarked: "Some institutions give teachers the instruments of evaluation and, since they come as standardized test, they are not appropriate or coordinated with what the teacher does". Similarly, T1 offered an example from his university, where the lack of connection between the course content and assessments resulted in inconsistencies, leading to concerns about the relevance and fairness of the testing process.

Finally, teachers recognize the importance of language testing in their practice, their experiences reveal a complex relationship to assessment. Many value the role of tests in reflecting student learning and guiding instructional practices, yet express a need for better alignment between curriculum and assessment methods. Their insights underscore the necessity for ongoing professional development in test design and administration to enhance the educational experience for both teachers and students.

Impact of Testing on Teachers and Students

Regarding their testing practices, most teachers recognize the influence these practices have on their students' learning processes. They emphasize that the assessment process should hold more value than the test itself. In the context of English as a Foreign Language (EFL), they noted that standardized tests often overlook the learning process and may not be inclusive of all students. T6 addresses this issue.

Sometimes when we design a test, it can be a general one, and we don't pay attention to the fact that our students are different, some student could do bad at a test but have other skills which have to do with other learning styles, as teachers, it is something we could work on.

Moreover, nearly all participants concurred that the test should not be the most significant element of a course. They argued that teachers should focus on equipping students with the skills needed for future tasks, rather than merely preparing them for the test. They emphasized that assessment should be a continuous process rather than just an outcome. To lessen the emphasis on the test itself, some teachers utilize the flexibility available in certain institutions to assign scores that reflect more than just test performance. One participant mentioned that he observes students during class activities and addresses their weaknesses, ensuring they are better prepared for the test. Other teachers expressed that, when grading, they considered the student's process, more than the final result, and they incorporated this information into the overall score to promote fairness.

Another important aspect highlighted by all teachers was the significance of the feedback process after a test, which serves as a means to guide learning, based on the evidence from the test. Many discussed how this feedback could positively or negatively affect students' progress. For instance, T3 expressed that he feels feedback is sometimes delivered in inappropriate terms:

Teachers believe that answers need to be correct or incorrect, and that is it. And, for many different situations, this does not apply sometimes, people take that to the classroom, and we should not be looking at performance, in terms of being right or wrong, because communication is not something that we can consider like that.

Building on this idea, T1 considers how feedback can be valuable when provided correctly.

That is why I would talk about feedback, at the end of the day, is giving them a grade, but that grade has to mean something, should not be just like a grade, but it has to have more information, and it has to provide them with clues about what they have to do and recommendations in general terms for improving.

Other teachers also highlighted feedback as an area for improvement, remarking how a teacher communicates a test score can significantly impact the student. When feedback extends beyond simply presenting the score, it can serve as motivation and provide the student with insights into their progress, and how to enhance it.

Current Practices in Testing Speaking in EFL higher education

Designing and Administration of Speaking Tests

Eight teachers agreed that speaking skills were the most significant challenge in testing. They often mentioned that Colombia is a monolingual country, resulting in few opportunities to practice a foreign language. Additionally, one teacher pointed out that speaking requires accessing many linguistic resources rapidly, making it difficult for teachers to design effective lessons. This combination of factors contributes to the overall challenges educators face, when assessing speaking abilities in their students.

Participants discussed various methods for administering speaking tests, with interviews being the most common format, similar to the speaking sections of well-known standardized tests like IELTS and TOEFL. Five teachers confirmed they conduct individual interviews with each student, often incorporating prompts such as pictures for description. Some teachers, however, offered alternative approaches. One teacher described a scenario where two students engage in a conversation, while the teacher observes and provides a situational context. Another teacher assigns students to prepare a role-play or a video presentation in advance for the speaking test day. A different method involved dividing students into small groups of three or four, where they discuss a controversial topic assigned to them beforehand, allowing for preparation prior to the exam. When asked about their testing strategies, the teachers exhibited varied practices. While most relied solely on the interview format, two teachers combined interviews with conversational elements, and other two utilized discussions or video presentations. This diversity in testing approaches reflects the teachers' different philosophies and objectives in assessing students' speaking skills.

When asked for their views on the various speaking tasks in their tests, teachers generally favored those that involved interaction among students, rather than one-on-one interviews. Indeed, T6 points out that interviews can be intimidating: "For some students, it was very stressful, because they are alone, sitting in front of the teacher, and they felt judged". Moreover, T1 explains why he prefers other alternatives than the interviews:

I prefer the tasks with a social connotation where they have to do something real they have to talk to another person. It reduces anxiety a little bit because they are not just in front of a person.

In the case of T2, he underlines that for students who are shy and struggle with public speaking, having an interview with the teacher can help them to reduce their anxiety. He emphasizes the importance of offering students various options, allowing them to demonstrate their speaking skills in a way that makes them feel comfortable with.

Scoring of Speaking Tests. All participants utilize the same tool, a rubric that outlines various criteria for scoring students. Some selected criteria include grammar, vocabulary, content, and pronunciation. Around seven teachers agree on allocating the same percentage to each criterion and averaging them for a final score. However, two teachers assigned different percentages, based on what they deemed more important for that specific speaking task. T5 highlights that the rubric has three components: content and organization, which account for 40% of the final score; language focus (grammar and vocabulary), 30%; and fluency and pronunciation, 30%. These teachers mentioned that the scoring of the speaking test varies by institution and depends heavily on the teachers. Some also pointed out that, despite having a rubric, not all teachers apply it. One teacher remarked that this process lacks standardization, and even though the institution mandates the use of the rubric, subjectivity continues to impact the process. Here is T6 discussing this issue:

We have a rubric, but sometimes we do not know what to expect from our students, the communicative ability that we are trying to evaluate is not clear, so one student for a teacher can be a five³. And, for others it can be a three, so it is something that we need to work on.

In this manner, participants explained that the speaking exam was part of a five-ability exam which is similar to high-standard tests. Now, regarding the value of this five-ability test against the other class activities, or what is known as classroom assessment. The majority of teachers indicated that classroom assessment was taken into account, but only to a limited extent, resulting in exams that carry the greatest weight in the overall score. They overlapped that in two universities, the exam accounted for 60% of the final score, with classwork contributing 40%. In a different university, the distribution was equal at 50/50%. Similarly, T5 believes these proportions are unsuitable:

In one of the universities, there is a 60% of the final score given to tests, which in my opinion is high. Because since it is an English course, it should be more important what they can do with the language in class, not what they do in a test.

Interestingly, T8 indicates that these pressures significantly affect the exam, and given its impact on students, the test should be meticulously crafted and organized by a knowledgeable team of educators. Nonetheless, many teachers agreed that the final test development involved collecting questions from all teachers, which would then be compiled virtually. This approach overlooks the thorough analytical process required, which includes establishing test specifications, conducting a needs assessment, item creation, prototyping, and statistical evaluation (Fulcher, 2003).

Teachers' Suggestions to improve

Based on their field experience, teachers shared several suggestions during the interview that they believe could enhance the speaking assessment process in their context. Five teachers concentrated on improving the quality of feedback, arguing that more comprehensive feedback could significantly assist students. T3 pointed out that students often focused solely on their scores, without understanding what those numbers mean. In a similar vein, T9 expressed that, if possible, she would allocate an entire follow-up session to discuss feedback on the speaking exam. She believes this is crucial to prevent discouragement among students who work hard but may not achieve their desired results. In Colombia, such reflective sessions are uncommon due to time limitations and large class sizes, yet they could be beneficial. T9 explains: "You normally give them feedback with the rubric, but it would be very nice to keep in touch with students and help them with things where they make mistakes".

To enhance feedback methods, two teachers proposed recording students' speaking exams so they could be reviewed later, allowing for more detailed feedback for each student. Some educators expressed that they were allotted only one session for both the speaking test and feedback, which constrained time; therefore, recording and a follow-up feedback session were commonly suggested.

Another prevalent suggestion among teachers was the provision to students with various speaking tasks during tests. In relation to it, T4 points out that each student is unique, and while personalizing every task isn't feasible, offering alternatives to the interview format would be beneficial. This could involve engaging in communicative tasks

³ In Colombia the grading system works in the scale from zero to five, five being the top score.

that require students to use the language more dynamically rather than simply answering questions. T10 echoed this perspective, recommending conversational tasks that connect class content with what is being assessed: “Creating sections in the speaking test that are more realistic and better structured, ensuring students have a meaningful experience”.

Lastly, most teachers expressed the necessity for training in testing methodologies. As an illustration, T9 remarked that it could be difficult for some educators to articulate what they were assessing in their speaking exams. There is often confusion regarding their expectations for students, which stems from insufficient training in this area. Also, T1 shared this viewpoint: “Being language teachers or professors doesn’t mean you are an expert in testing, as it is a specific aspect of the wider array of services and skills we need to cultivate”. Especially, other educators felt that this lack of knowledge in testing contributes to score inaccuracies.

Content Analysis on Test Samples

Four teachers consented to share examples of the speaking tests they commonly use in their classrooms. The analysis of the test samples began by identifying features related to task types, levels of difficulty, administration formats, and scoring systems, in alignment with the study’s research questions. These test samples offer valuable insights about the design, administration, and scoring procedures employed by the participants. For clarity in this analysis, each speaking test will be referenced using the letter “S” followed by a number (S1, S2, S3, S4).

Task Design and Administration. S1 was created as a speaking test for intermediate-level students and consisted of three sections. The first section is an interview, in which the teacher poses several questions such as: When did you last travel by airplane? Did you encounter any issues at the airport? If so, what were they? What would you do if you found a wallet on the bus tomorrow? The second section requires students to select a picture and describe it, with options provided among several photos. The third section involves a conversation where two students are assigned specific roles and must act it out in front of the teacher. Here is an example:

Table 3. Extract Speaking Test S1

<i>Situation 2</i>	
Student A	Student B
You work in a company with a friend. Your friend has recently been very stressed, and he told you about some problems that he/she had with the boss. Listen to him/her for more details.	Your boss has recently been angry at you because you have arrived late and haven’t submitted a report. Talk to your friend and tell him/her about your problem .

Source: Own elaboration

Along these lines, a comparable method is applied in S2, where six different options are provided for various conversational scenarios. In this test, students are asked to discuss a topic and address specific aspects of the issue. In the following image, I provide an example on one of the options:

Figure 2. Extract Speaking Test S2

Talk to your partner about the kind of illnesses or injuries you have suffered from in the past. You can give examples of the times you have been sick and how you have handled it (the common cold, for example); you may want to mention minor or major accidents you have had and how you recovered; or if you do not mind, you may want to talk about past surgeries. Finally, discuss the kind of things you do to stay healthy, which may include some type of physical exercise, vitamins or certain kinds of food you consume or avoid, and how often you visit the doctor.

You have one minute to think about what you are going to say. You can make some notes to help you if you wish. The conversation should be between 4 and 5 minutes.

Main grammar points under evaluation: Present perfect and present perfect continuous

Vocabulary point under evaluation: Illnesses and injuries

Source: Own elaboration.

In both tests for the conversational section, students receive a prompt and have an average of two minutes to prepare before the conversation begins, during which the teacher takes notes. The other two tasks differ in their design. One is interview-based where students answer three questions that range, in difficulty levels, from easy to challenging. Moreover, it covers various topics. Here is a sample of the questions for the intermediate level:

Table 4. Extract Speaking Test S2

EASY	MEDIUM	HARD
What's there in your classroom?	What do you have for dinner? don't forget to use quantifiers	Choose one historical event in Colombian history. Where did it happen? What happened? Who participated in it?
What is your brother's/sister's name? Where is he/she now?	What are your future plans?	What were you doing yesterday at 10 pm?
What is your mom's/ dad's favourite colour? Fruit?	What did you do yesterday?	Then, ask your teacher what he or she was doing at the same time.

Source: Own elaboration.

The final sample test S4 involves a task where the teacher instructs students to record a video on one of the topics discussed in the book during the term. Students have the freedom to choose the video format they deem suitable. This assignment is collaborative, and there are technical guidelines regarding the video's length, and ensuring all group members speak during the presentation. According to the teacher, the formats may vary based on the topic, with potential options including interviews, talk shows, role plays, blogs, and video blogs.

Scoring. For all samples, the scoring system is structured using a rubric that includes various categories and a corresponding numerical value for each description. For S4, the video's categories are organization, oral presentation, content, creativity, and requirements. The descriptors vary with the score; for instance, a score of 4 in organization indicates that the content is well-structured, while a score of 1 suggests that the organization lacks clarity.

As follows, these descriptors for the content category are explained:

Table 5. Extract Rubric Test S4

Category	4	3	2	1
Content	Cover all topics thoroughly and with details and examples. The knowledge of the subject is Excellent.	Includes basic knowledge about the topic. Content seems to be good.	Includes essential information on the topic but needs some fact-checking.	Content is barely minimal and has many imprecisions in the facts.

Source: Own elaboration.

For the other tests, a similar rubric was used, which included categories such as pronunciation and fluency, content, vocabulary, grammar, and preparedness, among others. In S2, the rubric not only assigns a score but also includes descriptive terms like outstanding, satisfactory, and so on. Next, there is an example of one of the categories:

Table 6. Extract Rubric Test S2

CRITERIA	OUTSTANDING	SATISFACTORY	EVEN	NEEDS IMPROVEMENT	POOR/ NO ATTEMPT
PRONUNCIATION & FLUENCY	All sentences and sounds are clear enough to give a right message. Some mistakes don't interfere in the response (1-3 minutes) (10 points)	Sentences and sounds are clear for a message. Some mistakes don't interfere in the response, mistakes are still present. (8 points)	Sentences and sounds are not clear to the message. Mistakes interfere in the response. (more than 6) (5 points)	Sentences and sounds are not clear to the message. Mistakes interfere in the response. Message is obscured and hard to get (more than 6) (3 points)	Poor or no attempt to develop the exercise was made (0 point)

Source: Own elaboration.

For S3, teachers received an Excel spreadsheet where they could assign scores ranging from one to five for each criterion per question, with the computer automatically calculating the final score. It is important to note that the questions were assigned different weights, based on their difficulty degree: easy questions were valued at 30%, medium questions at 30%, and difficult questions at 40% in the final score. The criteria assessed included grammar, pronunciation, vocabulary, comprehension, and fluency. Here's what the rubric looks like:

Figure 3. Extract Rubric Test S3

		30%	30%	40%
Level:				
Student Name:		Q1	Q2	Q3
Group:				
General observations				
Grammar				
Pronunciation				
Vocabulary				
Comprehension				
Fluency				
Q1	0,0		0,0	0,0
Q2	0,0			
Q3	0,0			
TOTAL	0,0			

Save

Source: Own elaboration.

Discussion

At this point, the discussion section responds to two primary research questions regarding teachers' perspectives on testing, specifically, in the context of speaking assessments in Applied linguistics to English as a Foreign Language (EFL) teaching and learning.

What do teachers know, believe, and think about testing and testing speaking?

During the interview analysis, many teachers emphasized the dual role of testing: tracking student progress and allowing teachers to evaluate their teaching methods. This perspective aligns with [Hughes \(2003\)](#), who identifies positive washback, in which tests serve as reflective tools in the educational process, highlighting areas needing improvement.

In this study, responses from teachers regarding their own experiences with EFL testing varied significantly. While some viewed tests as motivational challenges, others regarded them as obstacles. [Shohamy \(2001\)](#) suggests that teachers' attitudes toward tests can influence students' perceptions, potentially leading to a negative view. For instance, T4 recalled that exams were not pivotal in his learning, but framed them as a common adversary to conquer with students. Although he believes this approach alleviates student anxiety, it may also perpetuate a negative mindset toward assessments.

To ensure tests to effectively assess student abilities, it is crucial to maintain validity and reliability. [López and Bernal \(2009\)](#) reveal that many Colombian teachers lack the training to design, administer, and score tests effectively, which often results in inconsistent practices. Comparatively, four teachers noted discrepancies in scoring the same student's speaking exams due to varying focuses among evaluators.

Additionally, teachers reported a troubling fixation on scores among students, leading to inadequate feedback practices. Most teachers indicated a need for improvement in feedback, recognizing that merely providing scores without constructive commentary hinders student development. Only a couple of teachers followed up with comprehensive feedback after speaking assessments. Despite acknowledging the importance of feedback, they struggle with the knowledge and time required to implement effective changes, echoing [Borg's \(2006\)](#) assertion that classroom experiences influence teacher cognition.

What is the influence of that teacher's cognition in their current testing practices of speaking?

Approximately 70% of teachers expressed concerns regarding the fairness of their speaking assessments, citing inconsistencies between what is taught and what is assessed. Factors such as student motivation and administrative policies were identified as contributors. Teachers like T1 and T10 aimed to align test content with classroom activities, while T2 and T4 adopted continuous assessment strategies, emphasizing holistic evaluation over single-test performance.

Such reflective practices are related to what [Farrell \(2012\)](#) conceptualizes around teachers as deliberately examining challenges to implement meaningful change. This increasing recognition of the need for fairer testing overlaps [McNamara's \(2000\)](#) call for ethical language assessment, emphasizing the importance of designing tests that accurately and justly evaluate student abilities. In the present inquiry, teachers reported adopting flexible approaches in their speaking assessments, aiming to move away from traditional testing methods toward formative assessments that emphasize students' practical language skills. For instance, one of the teachers (T5) expressed the view that speaking assessments should focus on students' functional use of language rather than solely on their scores. However, institutional demands often prioritize summative assessments, creating a gap between teachers' educational philosophies and administrative expectations.

Teachers who collaborated on curriculum design found it beneficial, promoting clarity in learning objectives and assessments. T1 noted that aligning exams with lesson goals fostered a coherent educational process, reinforcing [Hughes's \(2003\)](#) call-for teacher involvement in test design.

During interviews, teachers struggled to define what aspects of speaking ability their assessments aimed to measure, raising concerns about tests' validity. [Fulcher \(2003\)](#) emphasizes the need for speaking assessments to predict performance in real-life situations. However, many rubrics used in tests focused on general linguistic features, rather than specific spoken language competencies, paralleling findings by [Díaz et al., \(2012\)](#) who identified an overemphasis on vocabulary, instead of functional speech.

Along these lines, participants reported familiarity with test design, recognizing the complexity of measuring speaking ability within limited timeframes. Many opted for interview formats due to familiarity, even when these were not tailored to specific contexts. Particularly, T2's dual approach –combining interviews with conversational tasks– illustrated a shift toward more communicative and less anxiety-inducing assessments. Precisely, this method aligns with [Fulcher's \(2003\)](#) recommendation around diverse tasks in speaking assessments.

As evidenced in the speaking test samples (S1–S4), rubrics varied in structure and complexity, which is connected to teachers' own perceptions of inconsistency and lack of standardization. Although rubrics aimed to make scoring transparent, they often failed to provide specific insights about individual student progress. This mirrors literature on measurement-driven instruments, which may lack the detail necessary for meaningful evaluation.

This analysis reveals that EFL teachers aspire to implement fair and reflective assessment practices; yet, they face challenges such as inadequate training, institutional constraints, and unclear construct definitions. While they show a readiness to improve, both targeted support and resources are needed to align their pedagogical goals with assessment realities.

Conclusion

This study researched speaking assessment practices in Colombian higher EFL education contexts, focusing on how these ones are shaped by teachers' cognitive frameworks, including their beliefs, knowledge, and professional experiences. This study contributes to the limited body of research on speaking assessment in higher education within English as a Foreign Language (EFL) contexts, particularly, in Latin America. Findings reveal a complex interplay between teachers' backgrounds and the assessment tools they employ. Many educators rely on standardized tests such as IELTS or TOEFL, especially when lacking formal training in language assessment. This reliance often stems from their own experiences as test-takers, rather than from established principles of valid and reliable testing.

Consequently, classroom assessments may lack coherence with best practices, leading to compromised validity and fairness.

Despite this, some teachers are beginning to design more interactive and communicative speaking tasks, concurring contemporary approaches to language testing. Teachers' perceptions also play a critical role: those who see testing as a barrier tend to foster similarly negative attitudes in students. Furthermore, the absence of assessment training often results in inconsistent scoring, which undermines the reliability of results and the fairness of high-stakes decisions based on them.

Notwithstanding this study offers meaningful insights about current practices, its small sample size limits the generalizability of its conclusions. Future research with a broader participant base, and a combination of quantitative and qualitative methods, could provide deeper and more representative findings. Still, this research serves as an important starting point for further exploration of speaking assessment in EFL higher education.

Finally, results emphasize the urgent need for reform in speaking assessment practices. Stakeholders must prioritize professional development that provides teachers with the conceptual and practical knowledge required for valid, reliable, and ethical assessment. Addressing these challenges requires more than technical training; it calls for an approach that acknowledges and integrates teachers' cognitive frameworks, ensuring that any shift in assessment practices is sustainable and contextually relevant. By doing so, institutions can foster a more equitable, informed, and effective approach to evaluate speaking skills, ultimately enhancing both teaching and learning in the EFL classroom.

References

- Attride-Stirling, J. (2001). Thematic networks: An analytic tool for qualitative research. *Qualitative Research*, 1(3), 385-405. <https://doi.org/10.1177/146879410100100307>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Borg, M. (2001). Key concepts in ELT: Teachers' beliefs. *ELT Journal*, 55(2), 186-188. <https://doi.org/10.1093/elt/55.2.186>
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318. <https://doi.org/10.1080/0969594042000304601>
- Brown, H. D. (2014). *Principles of language learning and teaching* (6th ed.). Pearson Education.
- Calderhead, J. (1996). Teachers' beliefs and knowledge. In R. C. Calfee & D. C. Berliner (Eds.), *Handbook of educational psychology* (pp. 709-725). Routledge.
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3), 360-389. <https://doi.org/10.1191/0265532204lt268oa>
- Childs, R. A., & Lawson, A. (2003). What do teacher candidates know about large-scale assessments? What should they know? *Alberta Journal of Educational Research*, 49(4), 341-350.
- Cizek GJ, Fitzgerald SM, Rachor RE (1995/6) Prácticas de evaluación de los docentes: preparación, aislamiento y el fregadero de la cocina. *Educational Assessment* 3: 159-179.
- Clark, C., & Peterson, P. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 255-296). Macmillan; Collier-Macmillan.
- Claudio Díaz Larenas, P., Alarcón Hernández, P., & Mabel, O. N. (2012). The English teacher: His beliefs about English language assessment at primary, secondary, and tertiary levels. *Ikala: Revista de Lenguaje y Cultura*, 17(1), 15-26. <https://doi.org/10.17533/udea.ikala.10039>
- Devrim Höl. (2018). Current problems and proposed solutions for testing speaking: Opinions of EFL teachers. *Literacy Trek*, 4(1), 27-36. <https://doi.org/10.29081/lt.2018.4.1.27>
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.
- East, M. (2015). Coming to terms with innovative high-stakes assessment practice: Teachers' viewpoints on assessment reform. *Language Testing*, 32(1), 101-120. <https://doi.org/10.1177/0265532214556410>

- Farrell, T. S. (2012). Reflecting on reflective practice: (Re) visiting Dewey and Schön. *TESOL Journal*, 3(1), 7-16. <https://doi.org/10.1002/tesj.13>
- Fulcher, G. (2003). *Testing second language speaking*. Pearson Longman.
- Heigham, J., & Croker, R. A. (2009). *Qualitative research in applied linguistics: A practical introduction*. Palgrave Macmillan. <https://doi.org/10.1057/9780230239517>
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge University Press.
- Johnson, M., Mehta, S., & Rushton, N. (2015). Assessment, aim, and actuality: Insights from teachers in England about the validity of a new language assessment model. *Pedagogies: An International Journal*, 10(2), 1-21. <https://doi.org/10.1080/1554480X.2015.1018397>
- Kahn, E. A. (2000). A case study of assessment in a grade 10 English course. *The Journal of Educational Research*, 93(5), 276-286. <https://doi.org/10.1080/00220670009598759>
- Leighton, J. P., Gokiert, R. J., Cor, M. K., & Heffernan, C. (2010). Teacher beliefs about the cognitive diagnostic information of classroom versus large-scale tests: Implications for assessment literacy. *Assessment in Education: Principles, Policy & Practice*, 17(1), 7-21. <https://doi.org/10.1080/09695940903526880>
- 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), 55-66.
- Luoma, S. (2004). *Assessing speaking*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511733017>
- McNamara, T. F. (2000). *Language testing*. Oxford University Press.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256. <https://doi.org/10.1177/026553229601300302>
- Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *Profile: Issues in Teachers' Professional Development*, 14(1), 143-158.
- Ministerio de Educación Nacional (2006). *Estándares Básicos de Competencias en Lenguaje, Matemáticas, Ciencias y Ciudadanas. Guía sobre lo que los estudiantes deben saber y saber hacer con lo que aprenden*. Bogotá: MEN.
- Numrich, C. (1996). On becoming a language teacher: Insights from diary studies. *TESOL Quarterly*, 30(1), 131-153. <https://doi.org/10.2307/3587685>
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307-332. <https://doi.org/10.3102/00346543162003307>
- Palacio, M., Gaviria, S., & Brown, J. D. (2016). Aligning English language testing with curriculum. *Profile: Issues in Teachers' Professional Development*, 18(2), 63-74. <https://doi.org/10.15446/profile.v18n2.53302>
- Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communication tasks for second language instruction and research. In G. G. Crookes & S. M. (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 9-34). Multilingual Matters.
- Reynolds, C. R. (2010). *Measurement and assessment in education* (2nd ed.). Pearson Education International.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Longman.
- Taylor, L. (2011). *Examining speaking: Research and practice in assessing second language speaking*. Cambridge University Press.
- Thompson, A. (2006). Teachers' beliefs and conceptions: A synthesis of the research. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). IAP. <https://doi.org/10.1108/978-1-60752-874-620251010>
- Toro, C. I. A., & Patarroyo, L. M. M. (2005). Evaluación en lenguas extranjeras: Discursos y prácticas. *Ikala*, 10(1), 63-91. <https://doi.org/10.17533/udea.ikala.v10n1a06>
- Richards, J. C. (2003). *Curriculum development in language teaching*. Cambridge University Press.
- Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE Publications. <https://doi.org/10.4135/9781529682571>
- White, H. P., & Marsh, E. (2006). *Content analysis: A flexible methodology*. SAGE Publications. <https://doi.org/10.1353/lib.2006.0053>

