How
ISSN: 0120-5927
Asociación Colombiana de Profesores de Inglés

Giraldo, Frank
Language Assessment Practices and Beliefs: Implications for Language Assessment Literacy
How, vol. 26, no. 1, 2019, January-June, pp. 35-61
Asociación Colombiana de Profesores de Inglés

DOI: https://doi.org/10.19183/how.26.1.481

Available in: https://www.redalyc.org/articulo.oa?id=499462432003

https://doi.org/10.19183/how.26.1.481

Language Assessment Practices and Beliefs: Implications for Language Assessment Literacy*

Creencias y prácticas en la evaluación de lenguas: implicaciones para la literacidad en evaluación de lenguas

Frank Giraldo

He holds a BA in English language teaching from Universidad Tecnológica de Pereira; an MA in English didactics from Universidad de Caldas, Colombia; and an MA in the teaching of English as a second language from University of Illinois at Urbana-Champaign (USA). His interests include language assessment, curriculum design, and the professional development of English language teachers.
frank.giraldo@ucaldas.edu.co

Abstract

This study reports the contextual Language Assessment Literacy (LAL) of five Colombian English language teachers. Two semi-structured interviews and reflective journals were used for data collection. The findings show that the teachers used varied traditional and alternative assessment instruments, assessed language and non-language constructs, used assessment information to improve teaching and learning, evaluated assessment results,

1 Received: August 19, 2018. Accepted: February 4, 2019
How to cite this article (APA 6th Edition):
Giraldo, F. (2019). Language assessment practices and beliefs: Implications for language assessment literacy. *HOW, 26*(1), 35-61. https://doi.org/10.19183/how.26.1.481

and engaged students in quantitative peer assessment. As for beliefs, data show that students' success and failure in assessment were connected to past experiences, and that assessment was appropriate given a number of features. Participants' answers about LAL show a complex and multifaceted construct. Taken together, the findings serve as baseline data to further professional development in language assessment.

Keywords: evaluation, language assessment, literacy, language teaching, teacher knowledge.

Resumen

Este estudio reporta la Literacidad en Evaluación de Lenguas (LEL) en contexto de cinco docentes de inglés. Se usaron dos entrevistas semiestructuradas y diarios de reflexión como instrumentos de recolección de datos. Los hallazgos muestran que los docentes usan instrumentos tradicionales y alternativos de evaluación, evalúan constructos de lengua y otros constructos e incluyen a sus estudiantes en evaluación par cuantitativa. En cuanto a creencias, los datos muestran que el éxito de los estudiantes, o falta de él, en la evaluación se conecta a experiencias pasadas, y que la evaluación es apropiada según un número de condiciones. Las respuestas de los participantes sobre LEL dan cuenta de un constructo complejo y multifacético. En conjunto, los hallazgos proveen información para el desarrollo profesional docente en evaluación de lenguas.

Palabras clave: enseñanza de lenguas, evaluación, literacidad en evaluación de lenguas, conocimiento docente.

Introduction

Language Assessment Literacy (henceforth LAL) is a major area in language testing; as such, scholars highlight that the construct needs more research to understand it as it relates to different stakeholders. For example, several authors argue that not only should language teachers be assessment literate but that those who make decisions based on assessment data (i.e. school administrators and even politicians) should have some knowledge of language assessment (Stiggins, 1991; Taylor, 2009). Because of the power tests have on teachers, students, institutions, and society at large (Fulcher, 2012), language teachers

and other stakeholders are expected to be skillful in interpreting, designing, implementing, and evaluating language assessment, as well as to be critical towards the implications of their assessment-based actions (Scarino, 2013). Consequently, language teachers and teachers in general have been central in assessment literacy discussions (Giraldo, 2018; Popham, 2011). As Taylor (2009) comments, language teachers should have knowledge and skills in test design, development, and evaluation for large-scale and classroom-based assessments.

Inherent in Taylor's argument is the scope of LAL for language teachers. The author highlights LAL to be related to both large-scale and classroom-based assessment. Additionally, other authors contend that assessment literacy requires knowledge of statistics (Brookhart, 2011; Davies, 2008), skills in test and item construction (Fulcher, 2012; Giraldo, 2018), knowledge of language and language education issues such as second language learning theories, approaches to communicative language testing, and even the relation between culture and language in assessment (Davies, 2008; Inbar-Lourie, 2008; Scarino, 2013).

In the case of language teachers, Scarino (2013) has made the call that the field needs to embrace the local realities of teachers and how they come to shape their assessment literacy. This author argues that teacher beliefs, practices, attitudes, and experiences –what she calls their interpretive frameworks– should be part of LAL as a construct. Thus, while core knowledge of assessment and skills for assessment are indeed necessary, understanding teachers' contexts is likewise pertinent. Given the complexity of the concept and its ongoing discussions, Inbar-Lourie (2017) encourages more research of local realities in LAL to understand the intricacies of the matter and ignite discussions that can feed the field of language assessment.

Based on this background, this article reports the findings of a qualitative case study which looked into the language assessment practices and beliefs of five Colombian English language teachers. This exploratory study elicited information about a group of teachers' LAL so that it could serve as baseline data for professional development opportunities. This was, then, a needs analysis exercise for LAL.

As opposed to most studies in LAL, which have used large populations and questionnaires predetermined by experts (see Fulcher, 2012, for example),

this study took an interpretive approach with a small group to see what five English language teachers do and think about language assessment in a particular context. As the findings below suggest, the information from this study may provide a fine-grained meaning of LAL and the richness of case studies as a diagnostic stage for professional development programs in language assessment.

Theoretical Framework

In language assessment, there seems to be a consensus as to three core components of LAL. Based on a study of language testing textbooks, Davies (2008) explained that LAL entails knowledge, skills, and principles. Knowledge refers to a background in educational measurement, knowledge of language and linguistic description, language teaching approaches, as well as knowledge of socio-cultural aspects related to assessment. Skills include item construction and analysis, use of statistics, and technology for language testing. Lastly, Davies stated that principles include the validity of assessment, the consequences of testing on stakeholders (e.g. teachers and students), and a sense of ethics and professionalism in the field.

Now found in a common definition, Davies' (2008) components have been used in other lists and taxonomies for LAL. For example, Inbar-Lourie (2008) argued that LAL should also include knowledge of the influence a first language and its culture can have on language learning; norms of English as an international language; the linguistic profile of multilingual learners; and current approaches to language teaching and testing, namely task-based assessment.

Specifically, for teachers, LAL also includes knowledge, skills, and principles that should be part of their assessment repertoire, as Fulcher (2012) argued. This author (2012, p. 125) offered the following ongoing definition of LAL for language teachers, in which the depth and scope of the concept can be elucidated:

> The knowledge, skills and abilities required to design, develop, maintain or evaluate, large-scale standardized and/or classroom based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order understand

why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals.

As can be observed, the LAL proposed for language teachers is a complex multi-layered enterprise (Inbar-Lourie, 2013). It places teachers at the forefront of sound theoretical, practical, and pedagogical practices for language assessment. To add to the layers of LAL, Giraldo (2018) proposed a list of sixty-six descriptors for nine categories subsumed under the three core components of LAL, as follows:

- *Knowledge*: Of applied linguistics; theory and concepts; own language assessment context.
- *Skills*: Instructional skills; design skills for language assessments; skills in educational measurement; technological skills.
- *Principles*: Awareness of and actions towards critical issues in language assessment.

While the core components from Davies (2008) are constantly cited in the literature, Scarino (2013) claimed that this core knowledge base is not sufficient to account for language teachers' LAL. Thus, she contended that the field needs to understand teachers' beliefs, practices, and experiences to articulate the meaning of LAL for this particular group. Consequently, LAL for language teachers includes knowledge, skills, principles, *and* "the assessment life-worlds of teachers" (Scarino, 2013, p. 30). These life-worlds include their practices, beliefs, and their own knowledge.

Given this conceptual discussion, Taylor (2013) discussed four stakeholder profiles and corresponding components in LAL. The profiles are of test writers, classroom teachers, university administrators, and professional language testers. For each of these groups, Taylor delineates the core contents they are supposed to have in increasing levels of depth. As regards language teachers, Taylor (2013) explained that language pedagogy is highest in the priorities for teachers, while sociocultural values, local practices, personal beliefs/attitudes, and technical skills are second in the profile. Lastly, scores and decision making, knowledge of theory, and principles and concepts rank at an intermediary level of LAL. Not surprisingly, Taylor (2013) invited the field to scrutinize these profiles through reflection and research.

Related Research

This section overviews research conducted around the particular LAL of language teachers. The review is based on practices, beliefs, LAL as a construct, LAL needs, and experiences of professional development in LAL.

A trend in practices by language teachers is the overuse of traditional assessment methods and assessment of micro-skills. This trend is evident in the studies by Frodden, Restrepo, and Maturana (2009), which reported that teachers tend to use quizzes as these were practical assessment instruments. Similar findings were reported in López and Bernal (2009), Cheng, Rogers, and Hu (2004), and Diaz, Alarcon, and Ortiz (2012). Overall, these studies indicate that while teachers express that they use a communicative approach to language testing, their actual practices are rather limited in that they emphasize micro-skills, namely vocabulary and grammar, and tend to disregard speaking and writing in their assessment.

The research by Rea-Dickins (2001) and McNamara and Hill (2011) identified four stages for assessment practices. The first stage involves planning, where teachers get students ready for assessment. In the second stage, teachers present the rationale, instruction, and means to conduct assessments; this stage also includes the actual development of assessment as it engages teachers in scaffolding and students in providing feedback. Stage three refers to teachers going over the results of assessment on an individual or group basis (i.e. with peers). Lastly, the final stage includes providing formal feedback and reporting and documenting assessment results.

Additionally, other research studies have focused on beliefs about language assessment. The results from these studies highlight the belief that assessment should provide feedback to improve teaching and learning (Brown, 2004; Muñoz, Palacio, & Escobar, 2012), and that language assessment should be communicative and based on both summative and formative methods (Arias & Maturana, 2005; Muñoz et al., 2012). Interestingly, these studies highlight that while teachers have these strongly-held beliefs, their practices indicate otherwise; for example, in López and Bernal (2009) and Muñoz et al. (2012), teachers used a summative approach to assessment, even though they think assessment should serve a formative purpose.

Another research focus of LAL has been the perceived needs of language teachers. The studies with this focus point to the fact that teachers need mostly a practical approach to language assessment, but they also expect a blend of practice with theory and principles. Thus, findings of these studies show that, overall, language teachers express needs in all areas of language assessment. To illustrate this, Fulcher (2012), for instance, used a questionnaire to find out the language assessment needs of language teachers from several countries. According to the findings in this study, teachers needed a comprehensive treatment of theory, techniques, principles and statistics for language assessment. In a similar study, Vogt and Tsagari (2014) used questionnaires and interviews to ask language teachers in Europe about their knowledge as well as their training needs in language assessment. Findings in this study indicated that the language teachers were, in general, not well trained in language assessment. Hence, they reported they needed training in test construction for both traditional instruments as well as alternative ones (e.g. portfolios).

Particularly in Colombia, there is scarce research explicitly targeting LAL for language teachers. Giraldo and Murcia (2018) conducted a study with pre-service teachers in a Colombian language teaching program. Through questionnaires and interviews, the authors asked participants (pre-service teachers, professors, and an education expert) what they would expect to have in a language assessment course for pre-service teachers. The answers reiterated what has appeared elsewhere: The need to have a course that combines theory and practice, with a strong emphasis on the latter. Additionally, language assessment within general frameworks such as Task-Based Instruction and Content and Language Integrated Learning (CLIL) also emerged as prominent in the data. Interestingly, the participants in this study also made it clear that they would like to have a course that addresses Colombian policies for *general* assessment, i.e. the Decreto 1290 (Decree 1290).

Lastly, research studies have observed the impact of professional development programs on language teachers' LAL. The impact of these studies occurs at a practical, theoretical, or critical level. For example, in Arias, Maturana, and Restrepo (2012), the researchers engaged English language teachers in collaborative action research geared towards improving assessment practices. As the authors report, the teachers' assessment became more valid in light of models

of communicative ability, and keener towards democratic and fair assessment practices. Whereas Arias et al.'s study had an impact on assessment practices, Nier, Donovan, and Malone's (2009) blended-learning assessment course helped instructors of less commonly taught languages increase their understanding of assessment and generate discussions of their practice. Lastly, the research by Walters (2010) highlighted how a group of ESL teachers became critical towards standards for language learning. As the author argued, this criticality should be part of teachers' LAL.

The Problem

As a need to cater to teachers' professional development, authors such as González (2007) have argued for a context-sensitive approach. In this regard, the institute where the current study took place started a process to examine the language assessment practices of its language teachers. To gather contextual data on language assessment, this current study focused on the life-worlds (Scarino, 2013) of five Colombian English language teachers and analyzed their practices and beliefs to elucidate some shape of LAL for these particular teachers. Thus, the present case study sought to collect baseline data on LAL for proposing professional development opportunities, as well as to analyze such data in light of LAL theory. The study was then informed by these three questions:

> What language assessment practices do the five Colombian English teachers have?
>
> What beliefs about language assessment do these teachers have?
>
> What implications for language assessment literacy can be derived from these teachers' practices and beliefs?

Context and participants. I conducted this study in a language institute of a public Colombian university. The institute teaches English to undergraduate students (teenagers, young adults, and other or older adults) enrolled in different university programs. The English courses are based on general interest themes (e.g. *sports and recreation*, *university life*, among others), language functions, and listening, speaking, reading, and writing. Language assessment at the institute is divided into 60% of skills development, whereby teachers assess the four language skills through the means they consider pertinent. The remaining 40%

of language ability is assessed through an achievement test the teachers design and administer at the end of each course.

The five teachers (one female and four male) in the study have worked for several years at this language institute. Tita was the pseudonym that the female participant chose for the study, while Mooncat, Vincent, Professor X, and Kant were the pseudonyms that the four males selected. The participants' ages ranged from 25 to 50 years old, and their experience teaching at the institute ranged from four to 29 years. All the teachers, except Vincent, had had some training in language assessment. Table 1 provides details about each participant.

Table 1. Relevant characteristics of the five English teachers in the study.

Theme	Tita	Kant	Mooncat	Vincent	Professor X
Experience in years (teaching EFL)	8	7	29	6	4
Has taught	Kindergarten Adolescents Adults	Children Adolescents Adults	Adolescents Adults	Adolescents Adults	Children Adolescents Adults
Training in (language) assessment	Online course: Modules Formative Summative assessment	Workshop for assessing writing and speaking Admin. of TOEFL iBT Psychometrics course	Module in a specialization course	None	Workshop: "testing as you teach"

Research Methodology

This research was a qualitative case study as it examined the contextual language assessment practices and beliefs of the five teachers. The approach I used was naturalistic (Cohen, Manion, & Morrison, 1998) as I inquired into teacher thinking and action in order to understand LAL from the participants' worldviews.

McKay and Gass (2005) describe qualitative research as providing rich and detailed descriptions; describing participants and their contexts as naturally as possible, without intervening in any way; including few participants given the depth of description; building research from an *emic* perspective, which means categories arise and are not pre-determined by the researcher; narrowing data patterns in a cyclical manner; permitting certain research ideologies (such as *a priori* categories for data analysis); and framing itself on open research questions.

Case studies provide data that explain participants' context and yield rich explanations for its complexities. In essence, case studies are qualitative and interpretive. On the other hand, given their specific nature, one disadvantage of case studies is that the findings will not necessarily generalize to other contexts; therefore, rather than generalizing, I focused on the usefulness of the findings in my study.

Data collection and analysis. Because I was not living in Colombia during the time of the study, I collected data online. I used *Google's YouTube Live* to conduct two online interviews. This technology allows participants to have video-recorded evidence of their talk and store it safely so only interested parties can have access. The first interview was about the five teachers' general assessment practices and beliefs, and the second interview was based on their practices and beliefs towards the design and implementation of the aforementioned achievement test.

The participants completed an online reflective journal through *Google Docs*. This allowed me to study participants' answers and ask further questions for clarification and illustration (e.g. How did this happen? What topics did the test include?). The journal (with a total of eight entries) asked the five teachers to describe a weekly assessment of their choice and reflect upon it. Appendix A includes the questions for the interviews and Appendix B has the prompts for the journal entries.

For data analysis, a grounded approach was used (Glasser & Strauss, 1967). I scrutinized the teachers' answers on both instruments and identified patterns across questions, across instruments, and across teachers. For example, the teachers reported they included assessment of the four skills in their practices in both the interview and journal; this became a code in data analysis. Additionally, they

also reported they assessed non-language ability factors such as eye contact, design of PowerPoint slides, and confidence. These factors became a second code. Lastly, these two codes were grouped so as to arrive at a finding; in this case, the finding was the practice of assessing language and non-language constructs, under the major category *Practices*. Thus, the major categories that emerged from the data were Practices, Beliefs, Knowledge, Skills, and Principles.

Findings

Findings from this study are grouped in the aforementioned five major categories. Each category embraces related findings, and each finding includes evidence coming from both data collection instruments. Below, the first two sections include findings related to the five teachers' practices and beliefs in language assessment. The last three sections report findings that provide implications for LAL, as seen from the language assessment realities of the participants. Taken together, the findings identify areas for improvement in language assessment.

Practices in language assessment. This category includes what teachers did for assessing the English language during normal classroom sessions and at determined moments, whether through a quiz, an oral presentation, or a final achievement test. The data come from sample interview answers and journal entries that reflect group consensus. A common practice among all five teachers was the use of both traditional and alternative methods for assessment. Quizzes and the final achievement test are part of the traditional methods; integrated-skills tasks, debates, and others are part of the alternative methods. Below are a journal entry and an answer from an interview.

> Professor X: Journal entry 3, question 1
>
> This week I worked with getkahoot.com. I designed 10 multiple choice questions related to the grammar topic (past perfect); then I created a quiz on getkahoot.com with those questions.
>
> *Vincent: Interview 1, question 2*
>
> Mostly, I do, I assess them orally. We do a lot of role-plays, presentations, things like that… we do debates.

The data suggest that teachers were resourceful in collecting information about language ability, rather than relying entirely on tests for doing so. The use of these instruments made their practices more substantiated as more evidence on student learning was collected.

In terms of constructs, data show that the participants assessed the four skills (listening, speaking, reading, and writing), two micro-skills (vocabulary and grammar) as well as non-language constructs such as confidence, physical performance, and the design of slides in PowerPoint presentations. When asked about the reasons why they included non-language constructs in their assessment, the teachers reported that these were part of communicative competence, and they helped convey messages clearly.

Mooncat: Journal entry 8, question 4

> In oral presentations, I include physical performance because I consider it to be part of the communicative competence of the students. Something like their illocutionary abilities. The design is assessed since it is important to consider the student's ability to elaborate good presentations that make comprehension of the message easier.

As the sample shows, other aspects beyond language were assessed. This practice shows that teachers were interested in providing students with opportunities to display general skills that go beyond language ability, e.g. design of presentations.

Another clearly articulated practice among the five teachers was the interface between teaching and assessment. The five teachers described their assessment practices as they were connected to teaching. For instance, Professor X (Journal entry 2, question 1) commented that:

> The last assessment activity I implemented was a role-play. First, we studied modals verbs (should, have to, must, etc.) and we did some controlled exercises; then, learners formed small groups and each group received a problematic situation. They needed to create a drama based on the problem and include at least one modal verb to give a piece of advice or a possible solution. They had some time to prepare the role-play and then they presented it to the whole class. The rest of the groups had to listen and write down what the problem was.

The sample shows that assessment was considered as a means to language learning. Classroom activities revolved around assessment and they sought to help learners to succeed in assessment performance, as Professor X did when he gave students time to prepare.

Data analysis also provides evidence to ascertain that these five teachers gave students feedback on their language progress; this was done orally or in written form. Notice that the sample below also reflects assessment as connected to teaching.

Kant: Journal entry 5, question 1

> Ls were grouped based on their preferences and were then given time to come up with as many crazy, funny, strange, and interesting ideas as possible. Afterwards, Ls were given feedback on their creative drafts and each group handed in a definitive proposal. Throughout the process, each group was provided with feedback on length, grammar mistakes, and sociocultural aspects (For instance: use of idioms or slang words).

That teachers gave feedback provides more support to reveal that they used assessment for improving language ability rather than just measuring it. Kant, for example, made sure he provided feedback on several occasions so that the writing assessment showed students' best performance.

The five teachers reported that they evaluated assessment practices (i.e. checked their quality) in cases where the results were unexpected. Specifically, teachers conducted an instrument evaluation differently, from analyzing grades to providing washback on teaching.

Mooncat: Interview 1, question 2

> When the results for example of a test are dramatic. In the course I had this semester I applied a test to 22 students and about 16 failed the test. In those cases, you think, well, something is wrong here, with the test or with the way I have been teaching or explaining the topics.

Mooncat's answer supports the idea that assessment did more than just measure. It impacted teachers and led them to reflect on the quality of instruments and even their teaching.

One last practice that emerged as consistent in these five teachers' approaches to language assessment was the use of quantitative peer-assessment. The teachers reported they did this by engaging students in giving each other grades on their performance.

Tita: Interview 1, question 4

> At the end of each task, I interchange the worksheets so they can grade one of their partners.

Interestingly, with peer assessment, the teachers used assessment for measurement rather than for formative purposes; however, this did not happen in other practices shown above. Thus, responsibility for providing grades was partially bestowed upon students, and this represented the teachers' approach to engaging students in assessment.

Beliefs in language assessment. The next category of findings pertains to the prominent beliefs about language assessment that these five teachers held. This section specifically highlights two beliefs among the participants in this study. The first commonly held belief was that success or failure in an assessment occur because of previous teaching or learning experiences.

Tita: Journal entry 7, question 3

> Learners previously had the possibility not only to read this type of text and get familiar with this style of writing throughout the semester, but also to write one register of experience with the help of their partners, by doing collaborative writing.

Mooncat: Journal entry 2, question 3

> The main difficulty arose because some students did not attend the previous classes when the topic of the structure of a paragraph had been studied.

Additionally, the samples above confirm that language assessment and teaching had a symbiotic nature. They fed each other and contributed to success, as Tita's entry describes. Conversely, external factors, such as lack of attendance, negatively impacted this relationship.

Additionally, these five teachers believed assessment was good when it provided washback on learning and teaching; was authentic, valid, and practical; and appealed to students' interests and affect.

Professor X: Interview 1, question 5

> For learners to know how they are doing and to know what they can improve; we [teachers] receive some insight about what we do and how we do it affects them a lot.

Tita: Interview 1–question 6

> I think the tasks need to be aligned to real-life situations; it is demonstrated that if it not connected to real life situations, it is not a good assessment.

Kant: Interview 1, question 6

> An assessment is good if what you're assessing is valid in the sense of having a direct relation with what was covered or studied throughout, along the course.

Vincent: Interview 1, question 6

> It needs to be interesting for them. The more fun they have, the better because they're enjoying and it will be memorable, I believe.

Together, the sample data above suggest that these teachers' beliefs towards language assessment were not negative but rather empowering for teaching and learning. Furthermore, the beliefs positioned assessment as a core element in the language classroom.

The last section of findings in this research report integrates with the core components of LAL, as the literature and research have discussed them. Thus, the findings are categorized according to knowledge, skills, and principles in language assessment. As with practices and beliefs, the data below reflect the group's views.

Knowledge of language assessment. The five teachers seemed to be aware of the meaning of validity as applied to classroom assessment. This meaning related to the connection between an assessment instrument and what had happened before in the course.

Kant–Interview 1–Question 2

> I think: Did I actually go through the whole process of thinking what we have done in class, whether they have actually been exposed to the sort of input, the sort of instructions. Is it really valid?

Another finding related to these teachers' knowledge in language assessment reflected how they had learned test design. The teachers reported that such knowledge came from studying sample tests and their own experience.

Vincent: Interview 2, question 5

> They [advisors] send examples so I look at those. I have learned from advice given by advisors: how to make a good multiple-choice task.

This sample suggests that language assessment knowledge of the technical kind (e.g. how to design a test) came from analyzing others' instruments. Knowledge, then, came from emerging opportunities rather than formal training.

In the last journal entry, all teachers were asked what they thought they knew about language assessment. Their answers varied in the scope of knowledge (or lack thereof) they say they had. Table 2 summarizes the five teachers' answers, which were taken from journal entry 8–question 1, and were not modified in any way.

Table 2. Reported Knowledge of Language Assessment

Tita	I want to mention that during this time (while writing the journal entries) I have realized there are many theoretical gaps I have about language assessment. I am not theoretically well documented.
Mooncat	Ideas such as practicality, reliability and validity Assessing is not only testing, but helping the students through their complete process of learning
Professor X	Some types of assessment such as formative, summative, achievement and self-assessment, although not very deeply Terms such as validity, reliability, practicality and washback which should be principles for any type of assessment I ju st know basic things that I have learned during the process of designing tests about the design or types of questions, but in an informal way
Vincent	*No answer provided*
Kant	Difference between testing, assessment, and evaluation. Constructs such as validity, reliability, practicality, authenticity and washback. The possibilities of assessment within the scope of summative, formative, diagnostic, performance-based.

Perhaps, not surprisingly, the answers in Table 2 attest the differential profiles that these language teachers had for language assessment. Recall, for example, that Vincent had had no training in language assessment, while Mooncat and Kant had been engaged in LAL initiatives. Also, as Professor X claimed, learning about language assessment happened on the go.

Skills for language assessment. Among the skills the teachers report, Mooncat and Tita highlighted their approach to dealing with students' affective dimensions. Professor X and Vincent commented on their assessment approach, and Kant notes validity in classroom assessment as his skill.

Mooncat: Journal entry 8, question 2

I think the most important skill is the good T-S relationship I establish with the students.

Tita: Journal entry 8, question 3

When assessing learners, I always take into account the emotional component of assessing and I try to make them feel confident.

Vincent: Journal entry 8, question 2

I think my language assessment is diverse, I tend to use a variety of ways to assess language proficiency and development of students.

Professor X: Journal entry 8, question 2

I don't think I have any skills in language assessment. However, I do try to incorporate formative assessment during my courses.

Kant: Journal entry 8, question 2

I make sure that the questions I propose have a direct relation with the content covered during our classes.

These five data samples further indicate how different language assessment was, given every teacher's life-world and interpretive framework. The skills reported included interpersonal, psychological, methodological, and technical dimensions. Truly, the data display a wide variation in the skills these teachers explained they had.

Principles in language assessment. Finally, in relation to principles for language assessment, the teachers mentioned concepts such as validity, reliability, and washback in the form of feedback. The data provide evidence of some of the principles the teachers reported in their LAL.

Vincent: Journal entry 8, question 3

I think my assessment practices measure what they need to measure, they are relevant to what students learn during the course and have a clear purpose. They are also reliable as students' results are often very consistent.

Vincent remarked on something that underlay the five teachers' practices; that is, language assessment needed to be valid and relevant for students, which made assessment useful. Principles, the data show, illuminated these teachers' practices and reflected beliefs of what good language assessment implied.

Needs in language assessment literacy. The last category for findings in this study entails the specific needs for training in language assessment that these teachers had. Table 3, which summarizes their needs, includes answers from interview one, question nine.

Table 3. Training Needs in Language Assessment

Tita	I am very kind and I give them thousands of opportunities. I don't know if I need to improve it. Probably I need to be stricter with students. I want to be able to design more instruments for assessment. To have the ability to use all of them, using different types of tasks.
Mooncat	The difficulty in designing tests. I think sometimes it's difficult. That's the main thing I would like to improve: The design.
Professor X	I would like to be more process-oriented. Probably I would like to learn more about assessment, more about the theory.
Vincent	Maybe giving instructions, of the activity, the task, as I talk to them about them.
Kant	Listening assessment would be one; more personalized and specific training in terms of language assessment. A more advanced course that would allow me to be an expert in assessment.

As Table 3 shows, all five teachers wished to improve different aspects in language assessment. Specifically, design of instruments was apparent in Tita and Mooncat; general approaches for assessment can be inferred from the answers given by Professor X and Kant; and specific details about the teaching-assessment relationship emerge in Vincent's answer. In conclusion, the answers imply the need for general, differentiated training for language assessment among these teachers.

Discussion and Implications

As Mckay and Gass (2005) state, qualitative research provides rich detailed data, and this study has not been an exception. In fact, because of space constraints, only the most apparent findings emerging from the data have been presented. Notwithstanding the wealth of information, the findings can be analyzed against research and conceptual discussions in LAL.

Firstly, the practices these five teachers had contrast with those presented in Arias and Maturana (2005), Cheng et al. (2004), Frodden et al. (2004), and López and Bernal (2009) in that the teachers of the present study use both summative and formative assessment instruments and include all four skills in their repertoire. The inclusion of non-language constructs such as confidence and voice projection, strictly speaking, may be considered sources of construct-irrelevant variance (Messick, 1989), something that is perceived as a threat to validity in language testing. However, the assessment ecology of these five teachers provides room and rationales for these constructs to be included in their approach to assessment. As Brookhart (2003) highlights, context for classroom assessment is construct relevant.

Second of all, the knowledge that the five teachers reported aligns with what the literature has discussed in terms of concepts such as validity and assessment methods (Inbar-Lourie, 2008; Stiggins et al., 2004). However, there is no evidence to ascertain that the teachers were knowledgeable of measurement and language description (Davies, 2008) or language teaching methodologies (Inbar-Lourie, 2008), among others. Therefore, this can be considered a limitation of conducting qualitative case studies. However, I must state that the data collection instruments did not seek to measure knowledge of language assessment (through, for example, pre-determined categories as most studies do) but rather to elicit knowledge as the teachers themselves conceived it. Thus, reported knowledge of language assessment from teachers' perspectives, in the case of these five teachers, did not necessarily reflect discussions of LAL. It reflected, rather, what they cherished as important knowledge in their particular assessment life-worlds.

Additionally, the teachers reported that their knowledge of test construction came from their own experiences. This finding confirms the importance

that Scarino (2013) adheres to teachers' contexts and their impact on LAL. Test construction, as reported in the literature (Brown & Bailey, 2008; Davies, 2008) comes from language testing textbooks and experts. However, it should not be argued that these teachers somehow lacked assessment literacy or that they were fully literate. As Inbar-Lourie (2013) and Taylor (2013) suggest, there is no solidified content knowledge to describe and evaluate the depth and width of LAL among various stakeholders.

Third, this study reports on skills that have not been documented in the general literature for LAL. The fact that teachers stated that they had affective skills for assessment (Tita, for example) further highlights the strong influence of teachers' contexts, as Scarino (2013) elaborated. Davies (2008) explained that skills include item construction and use of statistics, which the teachers in this study did not comment on. In fact, as Tita and Mooncat commented, test construction was a perceived need in their language assessment. Recall that Taylor's proposal includes technical skills for teachers (see literature review above). However, her proposal does not state anything about other types of skills. In closing, language teachers' specific skills may add to the discussions of what LAL has come to mean, a growing discussion in the field (Inbar-Lourie, 2013). In Giraldo's (2018) review, there is no allusion to affective skills for language assessment, yet they were meaningful in these teachers' LAL.

Fourth, the way the five teachers conceptualized their principles for assessment differed from general discussions in LAL. In the literature, two prominent principles are ethics and fairness (Davies, 2008; Fulcher, 2012). However, in the present study the teachers viewed concepts such as validity, reliability and washback as principles. This is not surprising, considering that language testing textbooks treat these as principles. Accordingly, the concepts may seem slippery in the literature. Most importantly, the teachers viewed feedback as an integral principle for their practice. Thus, while providing feedback may be considered a practice in language assessment (McNamara & Hill, 2011; Rea-Dickins, 2001), these teachers envisioned it as a principle that undergirds their unique approach to assessment.

In closing, the information from this study can indeed serve as a needs analysis to recommend professional development experiences. Based on the findings in this research, a course that focuses on theoretical and practical

aspects of testing, including both summative and formative assessment types, should prove useful for the five teachers in this study. To substantiate the course, it should consider contextual factors as elements that can foster, or given the case, impede the development of LAL. These ideas correlate with what Fulcher (2012), Scarino (2013), and Vogt and Tsagari (2014) report: In essence, the five teachers would benefit from a language testing and assessment course that combines knowledge, skills, and principles within their particular contexts. Such a combination may help the teachers to consolidate their strengths and empower them to increase their LAL.

Conclusions

The present study described the practices and beliefs that five Colombian English language teachers held in their language assessment approach. The practices included a multi-method, multi-construct view of language assessment, a close relationship between assessment and teaching, the use of assessment data to improve teaching and learning, evaluation of assessment after specific results, and the use of quantitative peer assessment. As for beliefs, the findings yielded a coordination between assessment success and failure on the one hand, and previous teaching and learning experiences on the other; what is more, the five teachers believed that good language assessment is valid, reliable, sensitive to students' affect, and that it provides feedback to improve learning. An analysis of data showed that the classical components of language assessment literacy –that is knowledge, skills, and principles– are praxized and conceptualized in what could be complementary ways to those highlighted in the field.

How teachers reported their assessment literacy arises from their assessment life-worlds. Then, this research provides support for Scarino's (2013) call to understand teachers' interpretive frameworks in the hope to better articulate the meaning of LAL for this population. Finally, the research highlights the complexity of the knowledge base in LAL as viewed from language teachers' perspectives, a complexity that gives insight into professional development opportunities in language assessment. Based on the findings in this study, a program to foster LAL among these teachers should draw them nearer to the knowledge dimension as reported in the literature, while contrasting it with their own knowledge base. Most importantly, these five teachers might benefit

from having an LAL course where design of assessments is a priority. Finally, the teachers might be interested in learning about the way the field conceives principles for language assessment. In synthesis, such a program could provide wholesome learning through LAL.

References

Arias, C., & Maturana, L. (2005). Evaluación en lenguas extranjeras: discursos y prácticas. Íkala: Revista de Lenguaje y Cultura, 10(1) 63 – 91.

Arias, C., Maturana, L., & Restrepo, A. (2012). Evaluación de los aprendizajes en lenguas extranjeras: hacia prácticas justas y democráticas. *Lenguaje, 40*(1), 99-126.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice,* 22, 5-12.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30, 3-12.

Brown, G. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*(3), 301-318.

Brown, J. D., & Bailey, K. (2008). Language testing courses: What are they in 2007? *Language Testing, 25*(3), 349-383.

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: purposes, methods, and procedures. *Language Testing, 21*(3) 360–389.

Cohen, L., Manion, L., & Morrison, K. (1998). *Research methods in education*. London, UK: Routledge.

Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing, 25*(3), 327–347.

Díaz, C., Alarcón, P., & Ortiz, 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.

Frodden, M., Restrepo, M., & Maturana, L. (2004). Analysis of assessment instruments used in foreign language teaching. Íkala: *Revista de Lenguaje y Cultura, 9*(15), 171 – 201.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, *9*(2), 113–132.

Giraldo, F. & Murcia, D. (2018). Language assessment literacy for pre-service teachers: Course expectations from different stakeholders. *GiST Education and Learning Research Journal*, 16, 56-75.

Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development, 20*(1), 179-195.

Glasser, B. J., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New Jersey, USA: Aldine Transaction.

González, A. (2007). Professional development of EFL teachers in Colombia: Between colonial and local practices. Íkala, *Revista de Lenguaje y Cultura, 12*(18), 309-332.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, *25*(3), 385-402.

Inbar-Lourie, O. (2013). Guest editorial to the special issue on language assessment literacy. *Language Testing, 30*(3) 301–307.

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, S. May, & I. Or (Eds.), *Language Testing and Assessment (third edition), Encyclopedia of Language and Education* (pp. 257-268). Cham, Switzerland: Springer.

Lopez, A., & Bernal, 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.

Mackey, A., & Gass, S. (2005). *Second language research methodology and design*. London: Lawrence Erlbaum Associates, Inc.

McNamara, T., & Hill, K. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing, 29*(3), 395–420.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13-103). New York: American Council on Education and Macmillan.

Muñoz, A., 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.

Nier, V. C., Donovan, A. E., & Malone, M. E. (2009). Increasing assessment literacy among LCTL instructors through blended learning. *Journal of the National Council of Less Commonly Taught Languages*, 9, 105–136.

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *Teacher Educator*, 46, 265-273.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18*(4), 429–462.

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.

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, *72*(7), 534–539.

Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right—using it well*. Portland, OR: Assessment Training Institute.

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36.

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing, 30*(3), 403-412.

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, *11*(4), 374-402.

Walters, F. S. (2010). Cultivating assessment literacy: Standards evaluation through language-test specification reverse engineering. *Language Assessment Quarterly*, 7, 317–342.

Appendix A.
Questions from Semi-Structured Interview

Interview One

1. Have you taken any language testing courses? What can you tell me about it/them?

2. Now please tell me about your language assessment practices; please describe how you assess your students' English language.

3. Are there any other assessment instruments that you design? If so, which ones and how do you design them?

4. Do you involve your students in your language assessment, for example through self-assessment? If so, how?

5. In your opinion, do you think assessing students is necessary? Why (not)?

6. Please tell me what you think are the characteristics of good language assessment.

7. In your opinion, do you think English teachers should have any principles in their language assessment?

8. Overall, how do you feel about your assessment practices?

9. Is there anything you'd like to improve?

Interview Two

1. What was the purpose of this achievement test?

2. What did you assess with this test?

3. How did you design this achievement test? What steps did you take?

How do you feel about the design of this test?

How did you learn how to design the items and tasks in this test?

4. How did you administer it? How did it go (the administration)?

5. How did you score/grade this test?

6. Did you do anything with the results of this test?

Appendix B.
Online Teacher Journal

Dear teacher,

Recalling the last week you taught, think about an assessment activity you used and reflect upon it. You may use the guiding questions below and include as many other comments as you think are necessary.

In terms of language assessment:

1. What did you do?

2. What went well and what did not go so well?

3. What do you think about what happened? For example:

- If something went well, what do you think about it (what went well)? Why do you think it went well?

- If something did not go well, what do you think about it (what did not go well)? Why do you think it did not go well?