Language Assessment Literacy and Teachers' Professional Development: A Review of the Literature

La literacidad en evaluación de lenguas y el desarrollo profesional docente: una revisión de la literatura

Frank Giraldo

Universidad de Caldas, Manizales, Colombia

In this literature review, I analyze the features and impacts of 14 programs which promoted teachers' language assessment literacy. I used content analysis to build a coding scheme with data-driven and concept-driven categories to synthesize and then analyze trends in the 14 research studies. Regarding core features, findings suggest that the programs were geared towards practical tasks in which teachers used theory critically. Also, the studies show that teachers expanded their conception of language assessment, became aware of how to design professional instruments, and considered wider constructs for assessment. Based on these findings, I include implications for the construct of language assessment literacy and recommendations for those who educate language teachers.

Keywords: language assessment, language assessment literacy, language testing, professional development programs, teacher professional development

En esta revisión literaria, analizo las características e impactos de catorce programas para literacidad en evaluación de lenguas de docentes de idiomas. A través del análisis de contenidos, diseñé un esquema de códigos con categorías basadas en datos y conceptos para examinar tendencias en los catorce estudios. Los hallazgos sugieren que los programas se basaron mayormente en actividades prácticas para que los docentes usasen la teoría de manera crítica. Los estudios indican que los docentes ampliaron su concepción sobre evaluación de lenguas, se hicieron conscientes de cómo diseñar instrumentos de manera profesional y expandieron los constructos de evaluación. Desde estos hallazgos, discuto unas implicaciones sobre la literacidad en evaluación de lenguas y unas recomendaciones para la formación de docentes.

Palabras claves: desarrollo profesional docente, evaluación de lenguas, literacidad en evaluación de lenguas, programas de desarrollo profesional

Introduction

In the field of language testing, much is discussed about language assessment literacy (LAL) for different stakeholders (Kremmel & Harding, 2020; O'Loughlin, 2013; Pill & Harding, 2013; Taylor, 2013). For discussions on this matter, authors have focused on the conceptual dimension of LAL, that is, drawing the construct (see for example Davies, 2008; Inbar-Lourie, 2008, 2012, 2017; Taylor, 2013). Thus, LAL is now generally known as the combination and use of knowledge, skills, and principles for conducting language assessment in the various educational contexts where it is needed.

Another focus of scholarly work in LAL is empirical research. This research has predominantly focused on teachers' LAL, although work has been done with other stakeholders (for example, O'Loughlin, 2013, with staff members at a university; Pill & Harding, 2013, with policymakers). Regarding teachers, there has been an emphasis on practices, beliefs, and needs around language assessment and the contexts where teachers conduct it (Berry et al., 2019; Crusan et al., 2016; Hill, 2017; Hill & McNamara, 2011; López-Mendoza & Bernal-Arandia, 2009). The research has been robust and provided descriptions of what LAL means for these stakeholders, and specifically what they need to improve in their LAL—The research has suggested teachers want special attention to practical matters, for instance, design of assessment instruments (Fulcher, 2012; Vogt & Tsagari, 2014; Yastıbaş & Takkaç, 2018). Finally, other scholars have focused on reviewing resources for advancing teachers' LAL (Davies, 2008; Giraldo, 2021; Inbar-Lourie, 2017; Malone, 2017). They have explained that textbooks have been a fundamental element to foster LAL, but other initiatives exist as, for instance, open online resources (Giraldo, 2021; Malone, 2017). Additionally, the authors have explained that these materials remark theoretical and practical aspects and, more recently, the social and ethical dimensions of language testing (e.g., Davies, 2008).

Even though resources and initiatives for helping teachers to improve their LAL exist, there is, as of now, no synthesis on their characteristics and impact on teachers' professional development. Therefore, in this literature review, I provide a critical account of programs for advancing teachers' LAL. To make the review useful, I focused my analysis on courses and workshops in which language teachers specifically studied issues related to language assessment. Reviewing existing programs for teachers' LAL is necessary given that initiatives for teacher education in this area should be encouraged at both the preservice and in-service levels. Thus, a critical account of these programs may shed light on what seems effective to advance teachers' professional development through LAL. I start this paper with what LAL means for language teachers and a review of trends in teachers' LAL research. Then, I report findings related to the methodology used in fourteen LAL courses and workshops, their contents, and the impact they have had on teachers' professional development. Based on these findings, I then discuss some implications and recommendations for the nature and implementation of professional development programs in LAL.

What is Language Assessment Literacy for Language Teachers?

LAL is the theoretical framework underlying the present research study, and I use this construct to present the findings and corresponding discussion. Hence, a definition of LAL is warranted.

Through a review of language testing textbooks, Davies (2008) explained that LAL consists of *knowledge* of theories and models of language proficiency, *skills* for design and educational measurement, and *principles* ethics and the impact of language testing. Davies's is a characterization that is generally accepted in the field. I use Davies's proposed components for LAL as a theoretical framework because it has been steadily used to discuss and problematize LAL either as a concept

or set of competencies (Fulcher, 2012; Giraldo, 2018; Inbar-Lourie, 2013b; 2017; Kremmel & Harding, 2020; Stabler-Havener, 2018). Additionally, LAL as a theoretical framework is appropriate enough to analyze professional development initiatives for teachers' LAL, because these programs can target and/or impact teachers' knowledge, skills, or principles for language assessment. A specific type of positive impact on teachers can be traced to one of LAL's components; for example, if teachers improve their *design* of peer assessment instruments, this can primarily mean a positive impact on the *skills* side of LAL. Finally, I choose LAL's three components for this paper because they are amenable to qualitative content analysis as a research method—The framework is flexible but can be used for systematic data reduction that leads to major trends in the literature.

It is important to note that, although LAL's three overarching components have been constant in the literature, they are still going through refinement (Giraldo, 2020; Inbar-Lourie, 2013a). In the case of language teachers, the LAL construct is intricate and still gaining research attention (for examples, see Coombe et al., 2020; Vogt et al., 2020). Notwithstanding the growing discussions, trends in LAL for language teachers are clear. Table 1 groups examples of elements within each component of LAL for teachers: knowledge, skills, and principles. Table 1 also gathers ideas from various authors in a 19-year span (Boyles, 2005; Brindley, 2001; Davies, 2008; Fulcher, 2012; Giraldo, 2018; Inbar-Lourie, 2008, 2013a, 2013b; Kremmel & Harding, 2020; Scarino, 2013; Stabler-Havener, 2018; Taylor, 2013; Vogt et al., 2020).

The elements in Table 1 represent core aspects with which scholars have contributed to the meaning of LAL for language teachers. As can be seen from the list, expectations of teachers' practice in language assessment are high and, consequently, position LAL as a high-impact dimension of their professional development. Empirical evidence from descriptive research studies, on the other hand, has shown that many, but not all, of the issues in Table 1 can be traced as needs that language teachers report. The next section, then, focuses on studies researching the intricacies of teachers' LAL.

What Has the Research on Teachers' LAL Suggested?

Research studies delving into teachers' LAL have provided thick descriptions of, particularly, practices, beliefs, and needs. In terms of practices, the research

Table 1. Examples of Knowledge, Skills, and Principles in Language Assessment for Teachers

Knowledge	Skills	Principles
• models describing language ability • frameworks for doing language assessment, e.g., criterion-referenced • purposes and theoretical concepts • relevant theories in second language acquisition and language teaching pedagogy • personal assessment contexts, which include practices and beliefs • critical issues such as ethics, fairness, and impact	*Design of assessment instruments*: • assessment specifications • items and tasks • rubrics with criteria • alternative assessment instruments, e.g., peer-assessment protocols • statistics and score interpretation *Test critique*: • identification of poorly designed items • evaluation of assessment instruments against qualities such as validity and reliability	• ethical use of assessment data • fair treatment of students • democratic practices in which students can share their voice • transparent practices to inform the nature of assessment systems and their consequences • critical stance towards unfair or unethical uses of language assessment • awareness of consequences (both intended and unintended) of assessment systems

has shown that language teachers tend to emphasize linguistic competence over others; tend to use and rely mostly on traditional assessment procedures (e.g., a final test); and avoid the use of alternative assessments such as peer-assessment (Babaii & Asadnia, 2019; López-Mendoza & Bernal-Arandia, 2009; Sultana, 2019). However, these stakeholders believe that assessment should serve formative purposes and be meaningful to impact teaching and learning (Díaz-Larenas et al., 2012; López-Mendoza & Bernal-Arandia, 2009).

As for the needs in LAL that teachers report, studies show that they feel underprepared and, therefore, want training in all areas, that is, knowledge, skills, and secondary attention to principles (Giraldo & Murcia, 2018; Vogt & Tsagari, 2014). Fulcher (2012) reports that teachers want special training in the design of assessment instruments, what the author calls the practice of language testing. As the author discusses, when it comes to theory and statistics, teachers seem to want clarity and practical examples rather than abstract notions. This sentiment is echoed in Jeong (2013), who states that teachers tend to see language assessment from a practice-based perspective. An interesting finding that has emerged in this LAL research is that teachers learn from more experienced peers to compensate for their lack of preservice or in-service training in LAL (Babaii & Asadnia, 2019; Tsagari & Vogt, 2017; Vogt & Tsagari, 2014).

Overall, research regarding teachers' LAL has made it clear that professional development in this area is expected and encouraged. As I mentioned earlier, LAL can be fostered through textbooks and online resources; however, it is not easy to track the impact of these materials on teachers' professional development if relevant research reports are not published. Thus, the purpose in the present study was to synthesize and analyze features and findings from 14 published research studies which describe professional development initiatives to support teachers' LAL.

Method

This literature review is grounded in a qualitative approach to research as it sought to interpret robust descriptive information from research studies on professional development through LAL. Particularly, I relied on a method called content analysis (Schreier, 2012) to read through, categorize, synthesize, and analyze codes that were data- (e.g., *purposes of the programs*) and concept-driven, that is, LAL's knowledge, skills, and principles. Such information was gathered from research reports on workshops and courses for language assessment. I reviewed the studies guided by these two questions:

- What are the characteristics of professional development programs for teachers' LAL?
- What impact do these programs have on teachers' professional development?

To find relevant studies, I searched major specialized journals in language testing and assessment (*Language Testing* or *Language Assessment Quarterly*) and also major journals in language teaching (e.g., *TESOL Quarterly*). My search included Latin American, North American, European, and Asian journals. Finally, I used the *Directory of Open Access Journals* to search for more papers. In all of these websites, I typed keywords such as *language assessment literacy* or *language testing program*.

The Literature Corpus

The corpus for this review consisted of 14 research studies in which various language teachers (preservice, in-service, student teachers) were engaged in learning about theories and practices in language testing and assessment. To choose studies fit-for-purpose in the review, I used three selection criteria:

- The study had to exclusively describe a course or workshop about language assessment rather than one in general applied linguistics or language teaching. This criterion was necessary because

several scholars have questioned the usefulness of courses which touch upon assessment in passing rather than giving extensive attention to it (Babaii & Asadnia, 2019; Vogt & Tsagari, 2014).

- The participants in the course or workshop had to be language teachers. In this case, language teachers include those who are preservice, in-service, and those student teachers in graduate programs (see O'Loughlin, 2006, for example). This criterion was necessary to collect more studies and, therefore, provide aggregated evidence for the findings in this report.

- The study describes, either explicitly or implicitly, the contents, methodology, and impact of the course/workshop on teachers' LAL. This was the key criterion for the research questions in this review.

Data Analysis

To arrive at the findings for this review, I first iteratively analyzed data inside each study independently, and then across studies, to form what Schreier (2012) calls a "coding frame." For this, I employed a matrix through which I collected the following information: *purpose*, *context*, and *participants*; *methodology for teaching language assessment*; *contents of the workshop or course*; *findings* (impact on teachers); and *other insights*, for example, recommendations emerging from each study.

With these codes, or categories as Schreier (2012) calls them, I then proceeded to group information across all 14 studies and observe instances of data to illustrate each category. For example, language assessment contents such as *item analysis* and *task analysis* were data-driven categories for which I associated examples from all studies. I then grouped these categories into one called *assessment design*, and then this and other grouped categories (e.g., *task development*) formed a synthesized category, that is, a finding—*rigorous design of assessments*. I explain and discuss this and other findings next.

Findings and Discussion

In line with the research questions for this review, I will first present the findings regarding the nature of the professional development programs. Then, I will explain findings which highlight major impacts the programs had on the participating teachers. After each finding, I provide a discussion based on LAL as the theoretical framework and empirical research reported elsewhere in this paper.

On the Context, Purpose, and Methodology of These Language Assessment Programs

The first finding relates to the contexts, objectives, and professional development approaches in the 14 programs. Although these elements varied in their focus, naturally the programs had a common goal, which was to help teachers improve different aspects of their LAL. Table 2 lists down basic features about the context of these programs, their purpose, and methodologies.

Based on the corpus of 14 studies, four trends are evident. The initiatives reported in the literature are aimed at helping language teachers in general, with most emphasis placed on those who teach the English language. However, the presence of teachers from different languages suggests that training in this area is necessary regardless of the language taught (for example, Montee et al., 2013 and Koh et al., 2018). Also, there is an emphasis on in-service language teachers, but studies with preservice teachers are starting to appear, with Giraldo and Murcia (2019), Jaramillo-Delgado and Gil-Bedoya (2019) and Restrepo-Bolívar (2020) being examples of this trend. This is positive, given that authors have emphasized the burning need for preservice teacher training in language testing and assessment (Hill, 2017; Lam, 2015; López-Mendoza & Bernal-Arandia, 2009; Vogt & Tsagari, 2014). The fact that most studies included in-service teachers attests to the need for providing these teachers with continuous

Table 2. Context, Purposes, and Methodologies of LAL Programs

Context		
Graduate	**In-service**	**Preservice**
2 studies: Kleinsasser (2005); O'Loughlin (2006)	9 studies: Arias et al. (2012); Baker & Riches (2017); Boyd & Donnarumma (2018); Koh et al. (2018); Kremmel et al. (2018); Levi & Inbar-Lourie (2019); Montee et al. (2013); Nier et al. (2009); Walters (2010)	4 studies: Giraldo & Murcia (2019); Jaramillo-Delgado & Gil-Bedoya (2019); Restrepo-Bolívar (2020); Walters (2010)
Purposes		
Critique or design of assessments		**Development of LAL at large**
9 studies: Arias et al. (2012); Kleinsasser (2005); Koh et al. (2018); Kremmel et al. (2018); Levi & Inbar-Lourie (2019); Montee et al. (2013); Nier et al. (2009); O'Loughlin (2006); Walters (2010)		5 studies: Baker & Riches (2017); Boyd & Donnarumma (2018); Jaramillo-Delgado & Gil-Bedoya (2019); Giraldo & Murcia (2019); Restrepo-Bolívar (2020)
Methodologies		
Critiquing and designing assessments: Face to face		**Critiquing and designing assessments: Blended**
11 studies: Arias et al. (2012); Baker & Riches (2017); Boyd & Donnarumma (2018); Giraldo & Murcia (2019); Jaramillo-Delgado & Gil-Bedoya (2019); Kleinsasser (2005); Koh et al. (2018); Kremmel et al. (2018); Levi & Inbar-Lourie (2019); Restrepo-Bolívar (2020); Walters (2010)		3 studies: Montee et al. (2013); Nier et al. (2009); O'Loughlin (2006)

professional development in LAL; however, the number of LAL initiatives for preservice teachers should be higher, so that they are more professionally prepared for the inevitable task of in-service language assessment.

Another trend in the corpus regards the purposes for training in LAL. Naturally, all studies were meant to help teachers improve their LAL. The clear tendency, however, is that improving assessment design is a key objective of these programs. Twelve out of 14 studies explicitly aimed at helping teachers improve their LAL either through analysis or design of assessment instruments; Kleinsasser (2005) and Restrepo-Bolívar (2020) do mention design in their studies but not as the approach

to teaching LAL. The trend of design as a main purpose in these programs reflects one expectation that teachers have about language assessment: They want practical, hands-on training (Fulcher, 2012; Giraldo & Murcia, 2018). This finding emphasizes teachers' need of a focus on the *skills* side of LAL. The implication seems to be that if teachers study knowledge and principles in LAL, they should do so within a practice-based framework.

The approaches to teaching language assessment in these programs were variegated. However, and as just commented, design-based learning is fundamental in these studies. The programs, whether face-to-face or online, use hands-on design and critique of assessments

for helping teachers improve the *skills* component of LAL but *knowledge* emerges in these tasks, as I will explain later. This is in line with what authors have discussed about LAL for teachers, that is, that the practical side of assessment is pivotal (Fulcher, 2012; Giraldo, 2018; Inbar-Lourie, 2008). As Boyd and Donnarumma (2018) argue, "teachers need a proper course in assessment design to allow them the time to absorb what is, after all, an expert subject" (p. 120). As the data corroborate, the *skills* side of LAL, and particularly design, seem to be pivotal to help teachers further their LAL. Clearly, the studies indicate that engaging teachers in designing assessment instruments impacts them positively at the technical and theoretical levels.

Lastly, one issue that may merit attention in these studies is the time devoted for training. The range is wide, from one-off workshops lasting three hours (Boyd & Donnarumma, 2018), week-long or semester-long programs (Baker & Riches, 2017; O'Loughlin, 2006, respectively), to sustained training for two or three years (Koh et al., 2018; Kremmel et al., 2018). As the authors in these programs suggest, initiatives with longer times should be encouraged for teachers' LAL (Baker & Riches, 2017; Boyd & Donnarumma, 2018; Kremmel et al., 2018). Since LAL is so much needed, as language testing experts and language teachers agree, the long-term impact of short professional development programs, if such exist, should be questioned or researched. Programs like the ones in Arias et al. (2012) and Koh et al. (2018) attest to the need for sustainable initiatives that, in addition to training in language assessment, accompany teachers in the implementation and scrutiny of the assessments they design. While short programs seem to raise teachers' awareness of what language assessment implies, actual impact on teaching and student learning seem to come with sustained programs (months and even years, such as Koh et al., 2018) that connect assessment to the contexts where teachers work. Clearly, the longer the LAL program, the more beneficial it might be for all stakeholders involved.

On the Contents of These Language Assessment Programs

Whereas the contents found in all 14 programs naturally varied, the data suggest clear tendencies at a theoretical and a practical level of language assessment. The data also show that there are contents which are not addressed in the majority of the studies, and there may be reasons for this.

The information in Table 3 corroborates the finding explained above: In this review, language assessment programs for teachers' LAL prioritize the critique and, most importantly, the design of language assessments. There is evidence in all 14 studies that teachers are engaged in studying and creating instruments with either items (e.g., multiple-choice questions) or tasks (a rubric for a speaking assessment) for traditional or alternative assessment. Thus, it can be suggested that these programs have responded to the need that teachers have expressed (see the relevant section on teachers' LAL research). In line with this finding and with a focus on preservice teachers, Giraldo and Murcia (2019) state that "language assessment courses for pre-service teachers should emphasise highly structured design tasks because they trigger conscientious decisions fueled by seasoned theoretical frameworks" (p. 255). Based on the data from the corpus, the programs seemed to have understood language assessment as design-driven rather than merely conceptual, which should have implications for language teacher education in various contexts, for instance, pre- and in-service: The design of assessment instruments should be a top priority for teachers' LAL.

Qualities for language assessment is another type of content that is common across most studies (11 out of 14). Of these, *validity* and *reliability* are the qualities that occur most in the studies, with *authenticity* and *practicality* coming second in the data. In scholarly discussions, these qualities are included in the *knowledge* dimension of LAL, which underscores them as an essential part of the fundamental knowledge base

Table 3. Language Assessment (LA) Contents in the Programs

Study	Knowledge			Skills		Principles
	Meaning of LA	Purposes in LA	Qualities of LA: validity, reliability, etc.	LA methods: critique and/or design	Assessing language skills	Ethics, fairness, impact, etc.
Kleinsasser (2005)		✓	✓	✓	✓	✓
O'Loughlin (2006)		✓	✓	✓	✓	
Nier et al. (2009)		✓	✓	✓	✓	
Walters (2010)				✓		
Arias et al. (2012)		✓	✓	✓	✓	✓
Montee et al. (2013)		✓	✓	✓	✓	✓
Baker & Riches (2017)			✓	✓	✓	
Boyd & Donnarumma (2018)	✓	✓	✓	✓	✓	✓
Kremmel et al. (2018)			✓	✓	✓	
Koh et al. (2018)			✓	✓	✓	
Jaramillo-Delgado & Gil-Bedoya (2019)			✓	✓		
Giraldo & Murcia (2019)	✓		✓	✓	✓	
Levi & Inbar-Lourie (2019)	✓	✓		✓		✓
Restrepo-Bolívar (2020)	✓	✓		✓		✓

that is needed for language assessment (Inbar-Lourie, 2008, 2012). Besides, as various authors in these studies report (Giraldo & Murcia, 2019; Kleinsasser, 2005), the participating teachers used these qualities to critique and design instruments for language assessment, which suggests that theory, apparently, was not studied in isolation but through the analysis assessments. A major implication from these data may be that theory should connect to design so teachers can make sense of it in the assessments they critique or analyze. Finally, three studies did not explicitly address key qualities such as reliability and validity (Levi & Inbar-Lourie, 2019; Restrepo-Bolívar, 2020; Walters, 2010), which are staple in language testing and assessment given their overarching impact. Thus, these two qualities should be central topics in LAL programs and studied accordingly given the specific nature of each initiative.

Another common content that stands out in Table 3 is, perhaps naturally, the assessment of language skills: listening, speaking, reading, and writing. The studies that do not mention skills assessment (Jaramillo-Delgado & Gil-Bedoya, 2019; Levi & Inbar-Lourie, 2019; Restrepo-Bolívar, 2020; Walters, 2010) perhaps did cover these topics, but they are not explicitly or implicitly reported in the articles. In the case of Walters' study, the program focused on a rather particular set of strategies for analyzing items: specifications writing and reverse specifications. Overall, the focus on assessing skills aligns with studies which report that teachers want to learn or improve how they assess language

skills (for example, see Giraldo & Murcia, 2018; Vogt & Tsagari, 2014). In all 14 studies there is an absence of construct-based discussions which are prominent in language testing, namely the assessment of multilingual competence. This may happen given that teachers work in contexts where there is a majority target language to be assessed.

As for the absence, in some cases, of the meaning of and purposes for language assessment, it may be the case that these topics were in fact studied in the programs. However, these topics are not reported in all studies because they may be taken for granted or because the nature of the study did not need to address them at length. For example, Kremmel et al. (2018) state that their study was not meant for classroom assessment, in which case *instructional purposes* for assessment may be irrelevant. However, language testing is predominantly done from a purpose-based angle: Assessments respond to purposes for them to be useful. Thus, purposes should be explicitly studied so that knowledge, skills, and principles in language assessment correspond to them.

Lastly, only six out of 14 programs included or addressed, at least explicitly, principles for language assessment. The principles included in the studies were *ethics*, *fairness*, *democracy*, and *transparency*, with a major focus in Arias et al. (2012) and, to a lesser extent, Levi and Inbar-Lourie (2019). The other studies (Boyd & Donnarumma, 2018; Kleinsasser, 2005; Montee et al., 2013; Restrepo-Bolívar, 2020) merely mention these contents. This finding contrasts with overall discussions of LAL, which state that principles such as ethics and fairness are a fundamental piece of this puzzle, for all stakeholders, teachers included (Davies, 2008; Inbar-Lourie, 2017; Kremmel & Harding, 2020). Interestingly, the finding does seem to align with teacher-reported needs. Descriptive studies such as Fulcher (2012) and Giraldo and Murcia (2018) show that teachers rank *principles* as a low priority for LAL. In one of the studies in this review—Kremmel et al. (2018)—teachers reported

that they learned little about principles such as ethicality, but as the authors comment, this was not the focus of their LAL study. However, in the study by Arias et al. (2012), teachers implemented fair, transparent, and democratic practices in their assessment approach. Thus, it seems that principles for LAL should be further studied in these programs, and their impact on teachers elucidated, especially because such principles have become central regarding the role and impact of language assessment in society (Fulcher, 2012; Giraldo, 2018; Inbar-Lourie, 2017).

On the Impact of These Programs on Teachers' Professional Development

In this section, I will provide evidence to answer the second research question that guided this review. The impact that these programs had on teachers' LAL can be explained in three aspects: *Heightened conception of language assessment*, *rigorous design of assessments*, and *broader constructs for assessment*. Next, I will explain and discuss these impacts.

Heightened Conception of Language Assessment

Most studies in this review (12 out of 14) report that teachers' conception of language assessment went beyond merely using tests and reporting grades to using assessment to improve learning and teaching. According to the reports, the teachers in the studies explained that they viewed language assessment as a powerful tool to impact student learning. Montee et al. (2013), for example, state that the teachers in their program developed "an increased awareness of and appreciation for assessment as a tool for guiding and improving language instruction" (p. 23). Similarly, in Restrepo-Bolívar's (2020) study, one of the preservice teachers viewed assessment as "a process in which the teacher gathers relevant information about the student's weaknesses and strengths in the learning process to make decisions about the instruction and students' learning" (p. 45).

Kremmel et al. (2018) and Walters (2010) do not report teachers' change regarding their conception of language assessment as a whole but do mention they became aware of issues in test development and design. Arguably, these areas may cause a change in perspective. In conclusion, the studies suggest that, even with short workshops (e.g., Boyd & Donnarumma, 2018), language assessment programs seem to exercise a positive impact on teachers' perspective towards what language assessment represents in instructional contexts. Some studies have shown that teachers have a limited view of language assessment and equate it with testing only (e.g., Díaz-Larenas et al., 2012), and this may be attributed to lack of training in LAL, especially in preservice teacher education (Lam, 2015; López-Mendoza & Bernal-Arandia, 2009; Vogt & Tsagari, 2014). Studies such as López-Mendoza and Bernal-Arandia (2009) have indicated that teachers with limited training in language assessment tend to see this area negatively and equate it with grading only. Thus, the call to provide early and continuous education in LAL is necessary so that teachers can use assessment for positive impact on teaching and learning.

Rigorous Design of Assessments

The design of language instruments is another prominent positive impact the programs had on the participating teachers. As shown in previous findings, workshops for language assessment prioritize a design-based course. Particularly, the studies report that teachers become aware of the necessary procedures to create high-quality assessments and, as they do so, they intertwine knowledge of theory to either critique or improve their design. In other words, design is not a procedural task but one in which theory and practice converge. Arias et al. (2012) explain this trend: "Inter-rater reliability was possible thanks to the existence of instruments and formats that included a complete rubric, with explicit instructions, criteria and construct" (p. 118, translated from Spanish). Similarly, Kremmel et al. (2018, p. 187)

report the following based on the answers from the 56 participating teachers: "The item writing stage appears to have been particularly beneficial for their learning about validity (89%), item writing (88%), reliability (86%), selecting tests for their classroom use (79%) and authenticity (77%)."

The answers in both studies suggest that, as teachers are trained in designing assessments, the task itself triggers theoretical constructs from their LAL. This is perhaps why Fulcher (2012) places the practical aspect of language assessment as fundamental for teachers; the studies in this review seem to align well with this idea and, perhaps most importantly, the needs teachers report in diagnostic studies for LAL (for instance, Vogt & Tsagari, 2014, and others). In conclusion, LAL programs that prioritize design benefit not only the *skills* side of assessment but also *knowledge*, and more importantly, the needs that teachers have reported consistently. What remains open for further investigation is how *principles* in LAL intertwine with *skills* and *knowledge*.

Broader Constructs for Language Assessment

A last outstanding positive impact these programs had on teachers regards the *what* of language assessment: constructs. The studies report that, through these initiatives, teachers moved from assessing minor linguistic skills such as grammar and vocabulary to assessing language ability more holistically. The trend in the studies is that teachers become more aware of assessing listening, speaking, reading, and writing, and this may be attributed to the programs' emphasis on these skills. As Baker and Riches (2017) report, there was "a broadening of the teachers' understanding of the construct of language ability relative to what they had previously held" (p. 566). In this study, teachers thought assessing only grammar and vocabulary was enough but understood the need to assess reading skills as well. Restrepo-Bolívar (2020) also reports how her students understood language ability more intricately. As the author explains, a participant in her

study: "moved from considering mere development of knowledge and skills to focus on language use as the language to be assessed, which is consistent with what current views state about the ultimate goal of teaching and learning a language" (p. 46).

This impact on teachers reflects a need to which scholars have referred in LAL: that of understanding language ability as an intricate construct (Inbar-Lourie, 2008; Stabler-Havener, 2018). Thus, these LAL programs may contribute to assessment practices that are more on par with current understandings of language ability models, which is a crucial component of LAL (Brindley, 2001; Davies, 2008). However, as I stated in the findings regarding content, teacher educators and the teachers themselves do not seem to bring up multilingual assessment as an issue in these courses, despite the growing consensus in language testing that this phenomenon is a crucial discussion. Like *principles* in LAL, professional development programs for teachers' LAL may lead to interesting findings when they explore the construct of multilingualism and the design of multilingual assessments.

Conclusions

My purpose with this literature review was to elucidate the nature of language assessment programs and their impact on language teachers' LAL. Data from all 14 studies suggest that training is conducted with language teachers in various educational settings and languages. Particularly, the studies remark the need to advance teachers' LAL through methodologies that use the critique and design of assessments as central tasks, that is, the *skills* component of LAL; such tasks lead to careful design of instruments, as the studies show. Besides, it appears that, with such a practice-based methodology, teachers learn more about conceptual aspects and expand the language ability constructs they assess, two aspects which are part of the *knowledge* side in LAL. Lastly, the *principles* component in LAL and core issues in language testing (e.g., multilingual

assessment) were not prominent in these studies, so further research may be needed to fully explicate the pertinence of these topics in teachers' LAL.

Taken together, the studies are in line with the needs that teachers have reported in the available LAL literature. Further, they exercised a highly positive impact on teachers, even in cases in which training was limited due to time constraints. A final call is that professional development programs for teachers' LAL should become more prominent in the literature so we can learn from others' experiences and then provide high-quality teacher education, which in turn should lead to positive consequences for those involved in language learning.

Implications and Recommendations

Based on my review of the literature, professional development programs for teachers' LAL started to become commonplace in the late 2010s. This is why, out of 14 studies, eight come from 2017 or later. Thus, the small corpus in this review may be considered a limitation. However, despite the limited number of papers and range of years among studies, the trends are clear and point to areas of consensus regarding programs in LAL for teachers. A related limitation is that each program had a particular impact on teachers, which made it difficult to synthesize and confirm other trends. As studies for teachers' professional development in LAL appear, more generalities and specificities might surface in the literature. Finally, the analysis of the data in the corpus for this review depended on my view entirely. Other analyses and conclusions may be possible given different research orientations and purposes.

The studies in this review suggest that as teachers engage in the development of assessments, they also use their theoretical knowledge in their LAL. This is what Davies (2008) conceptualized as a knowledge + skills angle on language testing. Thus, programs designed to foster teachers' LAL should definitely place major

emphasis on the analysis *and* design of assessments; theory-only courses may not be as successful to impact teachers' LAL or even be based on their actual needs for training. Another issue that requires attention is the role of principles in teachers' LAL education. Few studies in this review explicitly addressed them extensively. Thus, professional development programs should include principles such as ethics, fairness, and transparency as contents for teachers' LAL and careful observation as to how these principles can be meaningful for teachers' educational contexts. The feedback from research may be useful in LAL discussions to confirm the need for principles such as ethics and fairness, as authors have argued, or to challenge their presence in these discussions.

Based on the studies in this literature review, and the related conceptual review, LAL should become a core dimension of language teacher education. It may be a disservice not to include courses for language assessment in language teaching curricula, especially because learning about language assessment may lead teachers to become aware of its critical role on three fronts: current understandings of what it means to know and use a language; the impact of language assessment on teaching and learning; and the use of rigorously designed assessments to account for student learning.

References

Arias, C. I., Maturana, L. M., & Restrepo, M. I. (2012). Evaluación de los aprendizajes en lenguas extranjeras: hacia prácticas justas y democráticas [Assessment in foreign language learning: Towards fair and democratic practices]. *Lenguaje*, *40*(1), 99–126. https://doi.org/10.25100/lenguaje.v40i1.4945

Babaii, E., & Asadnia, F. (2019). A long walk to language assessment literacy: EFL teachers' reflection on language assessment research and practice. *Reflective Practice: International and Multidisciplinary Perspectives*, *20*(6), 745–760. https://doi.org/10.1080/14623943.2019.1688779

Baker, B. A., & Riches, C. (2017). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing*, *35*(4), 557–581. https://doi.org/10.1177/0265532217716732

Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal*, *73*(2), 113–123. https://doi.org/10.1093/elt/ccy055

Boyd, E., & Donnarumma, D. (2018). Assessment literacy for teachers: A pilot study investigating the challenges, benefits and impact of assessment literacy training. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 105–126). Springer. https://doi.org/10.1007/978-3-319-77177-9_7

Boyles, P. (2005). Assessment literacy. In M. H. Rosenbusch (Ed.), *New visions in action: National assessment summit papers* (pp. 18–23). Iowa State University.

Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126–136). Cambridge University Press.

Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia*, *10*(3), 1–16. https://doi.org/10.1186/s40468-020-00101-6

Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing*, *28*, 43–56. https://doi.org/10.1016/j.asw.2016.03.001

Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, *25*(3), 327–347. https://doi.org/10.1177/0265532208090156

Díaz-Larenas, C., Alarcón-Hernández, P., & Ortiz-Navarrete, M. (2012). El profesor de inglés: sus creencias sobre la evaluación de la lengua inglesa en los niveles primario, secundario y terciario [The English teacher: His/Her beliefs about English language assessment at primary, secondary and tertiary levels]. *Íkala, Revista de Lenguaje y Cultura*, *17*(1), 15–26.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, *9*(2), 113–132. https://doi.org/10.1080/15434303.2011.642041

Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, *20*(1), 179–195. https://doi.org/10.15446/profile.v20n1.62089

Giraldo, F. (2020). A post-positivist and interpretive approach to researching teachers' language assessment literacy. *Profile: Issues in Teachers' Professional Development*, *22*(1), 189–200. https://doi.org/10.15446/profile.v22n1.78188

Giraldo, F. (2021). A reflection on initiatives for teachers' professional development through language assessment literacy. *Profile: Issues in Teachers' Professional Development*, *23*(1), 197–213. https://doi.org/10.15446/profile.v23n1.83094

Giraldo, F., & Murcia, D. (2018). Language assessment literacy for pre-service teachers: Course expectations from different stakeholders. *GIST: Education and Research Learning Journal*, (16), 56–77. https://doi.org/10.26817/16925777.425

Giraldo, F., & Murcia, D. (2019). Language assessment literacy and the professional development of pre-service language teachers. *Colombian Applied Linguistics Journal*, *21*(2), 243–259. https://doi.org/10.14483/22487085.14514

Hill, K. (2017). Understanding classroom-based assessment practices: A precondition for teacher assessment literacy. *Papers in Language Testing and Assessment*, *6*(1), 1–17.

Hill, K., & McNamara, T. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, *29*(3), 395–420. https://doi.org/10.1177/0265532211428317

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, *25*(3), 385–402. https://doi.org/10.1177/0265532208090158

Inbar-Lourie, O. (2012). Language assessment literacy. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2923–2931). John Wiley & Sons. https://doi.org/10.1002/9781405198431.wbeal0605

Inbar-Lourie, O. (2013a). Guest Editorial to the special issue on language assessment literacy. *Language Testing*, *30*(3), 301–307. https://doi.org/10.1177/0265532213480126

Inbar-Lourie, O. (2013b, November). *Language assessment literacy: What are the ingredients?* [Paper presentation]. 4th CBLA SIG Symposium, University of Cyprus, Cyprus.

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment: Encyclopedia of language and education* (3rd ed., pp. 257–268). Springer. https://doi.org/10.1007/978-3-319-02261-1_19

Jaramillo-Delgado, J., & Gil-Bedoya, A. M. (2019). Pre-service English language teachers' use of reflective journals in an assessment and testing course. *Funlam Journal of Students' Research*, (4), 210–218. https://doi.org/10.21501/25007858.3010

Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, *30*(3), 345–362. https://doi.org/10.1177/0265532213480334

Kleinsasser, R. C. (2005). Transforming a postgraduate level assessment course: A second language teacher educator's narrative. *Prospect*, *20*(3), 77–102.

Koh, K., Burke, L., Luke, A., Gong, W., & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, *22*(3), 264–288. https://doi.org/10.1177/1362168816684366

Kremmel, B., Eberharter, K., Holzknecht, F, & Konrad, E. (2018). Fostering language assessment literacy through teacher involvement in high-stakes test development. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 173–194). Springer. https://doi.org/10.1007/978-3-319-77177-9_10

Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, *17*(1), 100–120. https://doi.org/10.1080/15434303.2019.1674855

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, *32*(2), 169–197. https://doi.org/10.1177/0265532214554321

Levi, T., & Inbar-Lourie, O. (2019). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, *17*(2), 168–182. https://doi.org/10.1080/15434303.2019.1692347

López-Mendoza, A. A., & Bernal-Arandia, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile: Issues in Teachers' Professional Development*, *11*(2), 55–70.

Malone, M. E. (2017). Training in language assessment. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment: Encyclopedia of language and education* (3ʳᵈ ed., pp. 225–240). Springer. https://doi.org/10.1007/978-3-319-02261-1_16

Montee, M., Bach, A., Donovan, A., & Thompson, L. (2013). LCTL teachers' assessment knowledge and practices: An exploratory study. *Journal of the National Council of Less Commonly Taught Languages*, *13*, 1–31.

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*, *7*, 95–118.

O'Loughlin, K. (2006). Learning about second language assessment: Insights from a postgraduate student on-line subject forum. *University of Sydney Papers in TESOL*, *1*, 71–85.

O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, *30*(3), 363–380. https://doi.org/10.1177/0265532213480336

Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, *30*(3), 381–402. https://doi.org/10.1177/0265532213480337

Restrepo-Bolívar, E. M. (2020). Monitoring preservice teachers' language assessment literacy development through journal writing. *Malaysian Journal of ELT Research*, *17*(1), 38–52.

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, *30*(3), 309–327. https://doi.org/10.1177/0265532213480128

Schreier, M. (2012). *Qualitative content analysis in practice*. Sage.

Stabler-Havener, M. L. (2018). Defining, conceptualizing, problematizing, and assessing language teacher assessment literacy. *Working Papers in Applied Linguistics & TESOL*, *18*(1), 1–22.

Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, *9*(1), 2–14. https://doi.org/10.1186/s40468-019-0077-8

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, *30*(3), 403–412. https://doi.org/10.1177/0265532213480338

Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, *6*(1), 41–63.

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, *11*(4), 374–402. https://doi.org/10.1080/15434303.2014.960046

Vogt, K., Tsagari, D., & Spanoudis, G. (2020). What do teachers think they want? A comparative study of in-service language teachers' beliefs on LAL training needs. *Language Assessment Quarterly*, *17*(4), 386–409. https://doi.org/10.1080/15434303.2020.1781128

Walters, F. S. (2010). Cultivating assessment literacy: Standards evaluation through language-test specification reverse engineering. *Language Assessment Quarterly*, *7*(4), 317–342. https://doi.org/10.1080/15434303.2010.516042

Yastıbaş, A., & Takkaç, M. (2018). Understanding language assessment literacy: Developing language assessments. *Journal of Language and Linguistic Studies*, *14*(1), 178–193.

About the Author

Frank Giraldo holds an MA in English Didactics from Universidad de Caldas (Colombia), and an MA in TESOL from University of Illinois at Urbana-Champaign (The USA). He works as a teacher educator in the Modern Languages Program and in the MA in English Didactics at Universidad de Caldas, Colombia.