

Language Assessment Literacy: Implications for Language Teachers

Literacidad en evaluación de lenguas: implicaciones para docentes

Frank Giraldo*

Universidad Tecnológica de Pereira, Pereira, Colombia

Recently, the applied linguistics field has examined the knowledge, skills, and principles needed for assessment, defined as language assessment literacy. Two major issues in language assessment literacy have been addressed but not fully resolved—what exactly language assessment literacy is and how it differs among stakeholders (e.g., students and teachers). This reflective article reviews assessment literacy from general education experts and language education scholars and shows how the meaning of language assessment literacy has expanded. To add to the discussion of this construct, the article focuses on the specific language assessment literacy for language teachers and proposes a core list of assessment knowledge, skills, and principles for these stakeholders.

Key words: Assessment, language assessment literacy, language teacher education, language testing.

Recientemente, el campo de la lingüística aplicada ha evaluado el conocimiento, habilidades y principios necesarios para la evaluación, definidos como la literacidad en evaluación de lenguas. Dos temas centrales han sido discutidos, pero no resueltos —exactamente qué es literacidad en evaluación de lenguas y cómo se diferencia entre poblaciones interesadas (p. ej. estudiantes y profesores)—. Este artículo de reflexión hace una revisión de la literacidad en evaluación vista por expertos en educación y enseñanza de lenguas, y muestra una expansión del significado de literacidad en evaluación de lenguas. Como adición a la discusión de este constructo, el artículo se enfoca en la literacidad en evaluación de lenguas específica para docentes de lenguas y propone una lista central sobre el conocimiento, las habilidades y los principios de evaluación para este grupo.

Palabras clave: educación de profesores de lenguas, evaluación, evaluación de lenguas, literacidad en evaluación de lenguas.

* E-mail: icaros@utp.edu.co

How to cite this article (APA, 6th ed.): 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>.

This article was received on January 19, 2017, and accepted on May 22, 2017.

This is an Open Access article distributed under the terms of the Creative Commons license Attribution-NonCommercial-NoDerivatives 4.0 International License. Consultation is possible at <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

The “39th Language Testing Research Colloquium: Language Assessment Literacy Across Stakeholder Boundaries”¹ held in Bogotá (Colombia) in July, 2017, explored the issue of language assessment literacy (LAL) for various stakeholders. The colloquium was guided by the consensus that LAL is a competency engaging different parties, from teachers to school administrators. The fact that such colloquium was mostly devoted to this topic speaks of the relevance that LAL has gained in language education and language testing. The purpose of this reflective paper is to contribute to ongoing discussions in LAL and seeks to illustrate what this construct implies for language teachers.

In general terms, LAL refers to knowledge, skills, and principles in language testing (Davies, 2008; Fulcher, 2012; Malone, 2008). These three components have in fact remained constant in theoretical and research discussions about LAL. However, its scope and boundaries have been questioned (Inbar-Lourie, 2013a; Taylor, 2013). Specifically, scholars are wondering what specific knowledge, skills, and principles are needed to define the term. What is clear is that knowledge of language, language use, and language pedagogy differentiate LAL from assessment literacy, the generic term in general education (Brookhart, 2001; Popham, 2009).

Another crucial discussion, the core of the aforementioned colloquium, refers to the people involved in LAL. Taylor (2009) argues that not only should language teachers be involved in knowledge of language assessment; other stakeholders such as school principals, parents, and politicians should know about language assessment and its implications (i.e., decisions based on scores). Based on the available research, Taylor (2013) identifies four stakeholder profiles in LAL: test writers, classroom teachers, university administrators, and professional language testers (more on these profiles in

the literature review section). Since several stakeholders should be engaged in language assessment, the picture of what exactly LAL means becomes even more complex (Inbar-Lourie, 2013a; Taylor, 2013). Thus, a general consensus in LAL is that research needs to be ongoing and welcomed (Fulcher, 2012; Coombe, Troudi, & Al-Hamly; 2012; Taylor, 2013).

Notwithstanding the need to involve others in LAL, language teachers remain central stakeholders whose teaching contexts should be considered to further define LAL (Scarino, 2013). López and Bernal (2009) and more recently Herrera and Macías (2015) have made the call that (Colombian) language teachers should improve their LAL. The authors have argued that LAL is needed among in-service language teachers, and that pre-service language teaching programs should raise the bar to provide quality LAL opportunities for teacher development. This is justified not only in language education but education in general, where scholars have argued for assessment literacy among teachers (Brookhart, 2011; Popham, 2009; Schafer, 1993). While the call for better LAL among language teachers is indeed necessary, the field must ask what it is exactly that LAL entails. A careful reconsideration of LAL is therefore the central theme of this paper.

The paper consists of a literature review that starts with a discussion of López and Bernal’s (2009) and Herrera and Macías’ (2015) argumentation; later, it overviews general assessment literacy and its change over time in education. Then, the bulk of the paper explores LAL from two themes: its meaning and scope, and stakeholder profiles. This theoretical exploration will serve as a basis to present a core list of LAL for language teachers. Such list is derived from conceptual discussions and research insights into knowledge, skills, and principles related to language assessment for teachers. Thus, the list is meant to fuel discussion in LAL, particularly for language teachers, and suggest what the implications of LAL for these stakeholders can be.

¹ Organized by the International Language Testing Association, ILTA and Universidad de Los Andes, Bogotá.

Literature Review

Background

The research by López and Bernal (2009) indicates that there are different practices of assessment among language teachers. Those with language assessment training used assessment to improve teaching and learning, whereas those with no training used it as a way to solely obtain grades. Thus, López and Bernal report that teachers without training placed grades and assessment on the same level, which the researchers perceive as a limited approach to language assessment. Additionally, the teachers in this research implemented more summative than formative methods.

In terms of professional development, López and Bernal inform that while graduate programs do have language assessment courses, few in-service language teachers gain access to MA degrees in Colombia. Because of this situation, the researchers argue that pre-service language teaching programs should offer more language assessment training. A majority (20 out of 27) of the undergraduate programs the researchers analyzed did not have any language assessment courses; the picture becomes more complicated when the authors explain that out of 27 programs, only two public universities offered assessment courses, as opposed to five courses offered by private universities.

Similar concerning results of language assessment practices can also be found in Arias and Maturana (2005); Frodden, Restrepo, and Maturana, (2004); and Muñoz, Palacio, and Escobar (2012), all studies conducted in Colombia. What is more, such findings have also been present in other parts of the world such as Chile (Díaz, Alarcón, & Ortiz, 2012), China (Cheng, Rogers, & Hu, 2004), and Canada (Volante & Fazio, 2007). In their conclusions, López and Bernal (2009) urge teachers to improve the validity, reliability, and fairness of their language assessment practices, and to implement assessment that is conducive to enhancing

teaching and learning. Addressing language teaching programs, the researchers find it central that

all prospective teachers take at least a course in language testing before they start teaching, and should strive to better themselves through in-service training, conferences, workshops and so forth to create a language assessment culture for improvement in language education. (López & Bernal, 2009, p. 66)

Herrera and Macías (2015) start their article by stating that “teachers are . . . expected to have a working knowledge of *all aspects of assessment* to support their instruction and to effectively respond to the needs and expectations of students, parents, and the school community (p. 303, my emphasis). Teachers with an appropriate level of LAL, according to Herrera and Macías, connect instruction and assessment, criticize large-scale tests, and design and choose from an available repertoire of assessments. Echoing López and Bernal, Herrera and Macías urge language education programs to provide more and better opportunities for LAL so that language teachers can focus on the spectrum that language assessment really entails—and not only focus on tests as instruments to measure learning. The authors then propose that for LAL experiences, questionnaires can be used to tap into teachers’ knowledge and skills in language assessment. However, as the researchers clarify, such instrument alone is not sufficient to describe and/or offer information to improve LAL among teachers.

Both articles claim that LAL is needed among pre- and in-service teachers. If language teachers are effectively trained in LAL, as these authors suggest, assessment for formative purposes—that is to enhance teaching and learning (Davison & Leung, 2009)—can become essential in language education. While the call of these four authors is one with which I agree I believe we need to take a deeper, more critical look towards what assessment literacy and specifically LAL involve. With this in mind, the next section of this article reviews the generalities of assessment literacy and specifics of LAL.

Assessment Literacy: Generalities in Education

The literature in assessment literacy reports on an expansion of the knowledge and skills that teachers and other stakeholders are expected to have—although the focus has been on assessment literacy for teachers. Historically, assessment literacy has expanded teachers' toolbox to monitor, record, improve, and report on student learning. There has also been increasing attention as to how assessment has consequences on teaching, learning, and school curricula (Brookhart, 2011; Popham, 2009, 2011); this attention has led to a belief that teachers should have a critical stance towards how assessment impacts stakeholders (Popham, 2009).

The first allusion to assessment literacy in education was proposed by the American Federation of Teachers, National Council on Measurement in Education, and National Education Association (1990) in their *Standards for Teacher Competence in Educational Assessment of Students*. They believed these guidelines were needed to help teachers become aware of assessment in and out of classroom contexts. The guidelines can be categorized into two strands. The first deals with instruction; teachers should be able to choose, design, and evaluate valid assessments for positive effects on learning, teaching, and schools. The second strand has to do with uses of tests and test results; teachers are expected to know when assessments are being used inappropriately, and to know how to communicate results well to various stakeholders. Later, Stiggins (1995) used the term assessment literacy to include knowledge and skills that stakeholders such as teachers and school administrators should have about assessment.

In addition to the standards above, Popham (2009) explains that assessment literacy includes knowledge of reliability and threats to it, tests' content validity, fairness, design of closed-ended and open-ended test

tasks, use of alternative assessments such as portfolios, formative assessment, student test preparation, and assessment of English language learners. Popham argues that assessment literacy is needed so teachers become aware of the power that tests, especially external, can have on education.

Furthermore, Brookhart (2011), who argues that the standards above are not comprehensive enough for classroom teachers, believes assessment literacy has to do with knowledge of how students learn in a specific subject; connection between assessment, curriculum, and instruction; design of scoring schemes that are clear for stakeholders; administration of externally-produced tests; and use of feedback to improve learning.

Other areas that have received attention in assessment literacy involve the use of basic statistics for educational measurement (Popham, 2011; White, 2009), student motivation (White, 2009), and the use of multiple methods in assessment (Rudner & Schafer, 2002). Similarly, the use of technology has been proposed as part of teachers' assessment literacy (Rudner & Schafer, 2002).

The previous section has shown a steady historical increase in the knowledge, skills, and principles related to the assessment literacy that teachers are expected to have. While the meaning of LAL shares similarities with assessment literacy, LAL is unique in specific ways. The next section of the paper pinpoints what has been carried over from general assessment literacy, and what has made LAL a construct on its own. For this purpose, this paper addresses two related, ongoing debates in LAL: the need to pursue a knowledge base in the field and the realization that LAL means different things to different people. After these two debates, the section will focus on a recent addition to the meaning of LAL by Scarino (2013), who argues that LAL should also involve teachers' contexts of teaching.

Language Assessment Literacy: Generalities and Specifics

Overall, conceptual discussions and research findings in LAL have provided insights for a concept that is far from being defined in limited terms. In a review of language testing textbooks, Davies (2008) places the field within three components: knowledge, skills, and principles related to the assessment of language ability. While emphasis has been given to the first two components, there is an increase in the need to instill language testing with principles such as fairness (non-discriminatory testing practices) and ethics (appropriate use of assessment data) (Kunnan, 2003). In fact, research has indicated that this trend is stable because language testing textbooks focus on knowledge and skills (Bailey & Brown, 1996; Brown & Bailey, 2008) more than they do on principles. In fact, the trend is also evident in language testing courses (Jeong, 2013; Jin, 2010), which include some but not sufficient attention to principles as well as to consequences of assessment. Thus, Davies's global view of LAL is generally accepted by authors (Inbar-Lourie, 2008; Fulcher, 2012; Taylor, 2009).

Fulcher (2012) used a questionnaire to find out the LAL needs among language teachers from around the world ($N = 278$). Based on the survey results, his definition emphasizes the interplay among Davies's three major components of LAL, as they impact practice and society at large. Fulcher also argues that teachers need to view language assessment from its historical development. Fulcher's (2010) book, *Practical Language Testing*, is an operationalization of this definition of LAL. What is particularly interesting about the author's definition is that it refers to both large-scale and classroom tests, which suggest LAL for language teachers is not limited to classroom assessment. Besides, Fulcher strongly suggests that LAL require that teachers be critical toward language assessment practices, and there exists a general con-

sensus in the field regarding that suggestion (Coombe et al., 2012; Inbar-Lourie, 2012; Taylor, 2009).

The previous section shows that LAL shares components with assessment literacy. However, language as a construct for assessment is what differentiates LAL from its generic term. Thus, in Davies (2008), LAL includes knowledge of *language* and *language methodologies* such as *communicative language teaching*. Inbar-Lourie (2008, 2012) calls language *the what* in LAL (after Brindley, 2001). Additionally, Inbar-Lourie (2008) argues that LAL includes knowledge of *multilingual learners* and *content-based language teaching*.

Skills and principles in LAL are therefore directly related to assessing language. Specifically, skills needed for test design (e.g., item-writing), use and interpretation of statistics, and test evaluation are part of LAL because they are used to assess language ability (Davies, 2008; Fulcher, 2012; Inbar-Lourie, 2013a). Concerning principles, they are viewed the same in LAL as in assessment literacy; that is, principles refer to codes of practice for ethics, fairness, and consequences of assessment.

One way to picture the wide scope of LAL is by looking at Inbar-Lourie's (2013b) *ingredients* of LAL for language teachers. She argues that LAL is "a unique complex entity", similar yet different from general assessment literacy for teachers. According to the author, the ingredients of LAL for language teachers are:

1. Understanding of the social role of assessment and the responsibility of the language tester. Understanding of the political [and] social forces involved, test power and consequences. (p. 27)
2. Knowledge on how to write, administer and analyze tests; report test results and ensure test quality. (p. 32)
3. Understanding of large scale test data. (p. 33)
4. Proficiency in *Language Classroom Assessment*. (p. 36)
5. Mastering language acquisition and learning theories and relating to them in the assessment process. (p. 39)

6. Matching assessment with *language* teaching approaches. Knowledge about current language teaching approaches and pedagogies. (p. 41)
7. Awareness of the dilemmas that underlie assessment: formative vs. summative; internal external; validity and reliability issues particularly with reference to authentic language use. (p. 45)
8. LAL is individualized, the product of the knowledge, experience, perceptions, and beliefs that LANGUAGE TEACHERS bring to the teaching and assessment process (based on Scarino, 2013). (p. 46)

Given the array of elements in LAL, it is not surprising that scholars in language testing are still debating the boundaries of the concept (Fulcher, 2012; Jeong, 2013; Malone, 2013; Scarino, 2013; Taylor, 2013). Inbar-Lourie (2013a) wonders what the essentials for LAL actually are, and invites discussions and research to expand and clarify LAL and its uniqueness. What further fuels the debate around the meaning and scope of LAL is how it relates to different stakeholders.

LAL and Different Stakeholders

Taylor (2009) contends that given the impact assessment can have other people besides teachers should possess knowledge of language assessment. Pill and Harding's (2013) study testifies the need to have others involved in LAL. Their study found that there were misconceptions and a lack of language assessment knowledge at the Australian House of Representatives Standing Committee on Health and Ageing. This political body was responsible for determining which doctors could be granted entrance to Australia, based on the results of two tests: The International English Language Testing Service (IELTS) and the Occupational English Test (OET). Additionally, the study by O'Loughlin (2013) reports the LAL needs (e.g., score interpretation) of the administrative staff at an Australian university using IELTS for admission of international students. Finally, the study by Malone (2013) reports that language instructors and language

testers had differing views and needs as regards the contents of an online language testing tutorial. While the former group expected the tutorial to be clear and include practical matters, the latter expected comprehensiveness of concepts. These three studies certainly provide convincing evidence that several stakeholders—and not only teachers—should be recipients of LAL.

To define the level of LAL among different stakeholders, Taylor (2013) proposes a figure that places them at different levels. Thus, researchers and test makers are at the core of the figure, language teachers and course instructors are placed at an intermediary level, and policy makers and the general public are on a peripheral level of LAL. Additionally, this author outlines the profiles for four different stakeholder groups; namely, test writers, classroom teachers, university administrators, and professional language testers. These four profiles are described against eight dimensions: “knowledge of theory, technical skills, principles and concepts, language pedagogy, socio-cultural values, local practices, personal beliefs and attitudes, and scores and decision making” (Taylor, 2013, p. 410). Taylor (2013) presents her proposal as open to debate and invites the field to inspection and operationalization of the suggested levels and profiles of LAL.

In conclusion, as commented elsewhere, scholarly discussions and research in LAL have indicated that this concept has come to have different shades of meaning for various people directly or indirectly involved in language assessment. While it is certain that others should be engaged in LAL, language teachers remain central in the efforts to deliver professional development opportunities in LAL (Boyles, 2006; Brindley, 2001; Fulcher, 2012; Nier, Donovan, & Malone, 2009; Taylor, 2009). Accordingly, I now move on to exploring LAL for language teachers and the implications that this construct may have for them.

LAL for Language Teachers

Scarino (2013) argues that in addition to knowledge, skills, and principles in LAL, it is pertinent to include teachers' interpretive frameworks. That is, discussions in LAL need to acknowledge that language teachers have particular teaching contexts, practices, beliefs, attitudes, and theories, all of which shape their own LAL. Recognition of language teachers' interpretive frameworks is particularly important in fostering professional development, as Scarino suggests. Knowledge, skills, and principles in language assessment coexist with teachers' ways of thinking and acting upon the act of assessment. Thus, Scarino explains that, in the case of language teachers, the components of their LAL influence each other, a notion briefly addressed by other authors (Fulcher, 2012; Taylor, 2009).

LAL discussions and research, even for language teachers, have provided a top-down perspective. Thus, the knowledge-base of LAL has been described from language testing textbooks (Davies, 2008), language testing courses (Bailey & Brown, 1996; Brown & Bailey, 2008; Jeong, 2013; Jin, 2010), and even pre-determined by language testing scholars themselves. For example, Fulcher (2012) and Vogt and Tzagari (2014) use questionnaires with pre-determined categories to find out needs among language teachers. However, what has not been clearly addressed in the literature is how language teachers engage in or display LAL. In tandem with Scarino's (2013) proposal, I believe there are particularities to LAL that should come from the bottom up, or language teachers' assessment practices.

Rea-Dickins' (2001) and McNamara and Hill's (2011) research studies do not overtly refer to teachers' LAL. However, their research scope certainly deals with areas that, according to the literature, are part of a language teacher's knowledge, skills, and principles for assessment viewed from a formative lens. Based on a purely qualitative approach using observations and

interviews, these two studies provide descriptions of language assessment stages. In Rea-Dickins (2001), there are four stages to language assessment in the classroom: planning, implementation, monitoring, and recording and dissemination. In the first stage, language teachers select the purposes and tools to assess and prepare students for assessments. In stage two, teachers introduce the why, what, and how of assessment, and also provide scaffold while assessment unfolds, ask learners to monitor themselves and others, and provide immediate feedback to students. During stage three, teachers bring together their observations and analyze them with peers, with the hope to provide delayed feedback to improve learning and teaching. In the last stage, teachers formally report their analyses to whomever they need to. In McNamara and Hill (2011), the stages are called planning, framing, conducting, and using assessment data. They are, essentially, the same as those in Rea-Dickins (2001) as the stages refer to the same assessment activities. From these last two studies, I believe we can add more layers to what LAL can entail—LAL includes the ability to effectively plan, execute, evaluate, and report assessment processes and data.

Lastly, other studies report findings of skills that should be part of teachers' LAL. In Walters' (2010) study, English as a second language (ESL) teachers became aware of a process for test and item analysis called standards reversed engineering (after Davidson & Lynch, 2001), through which they could derive test specifications and critique state-mandated standards for ESL. The study by Vogt and Tzagari (2014) with European language teachers identified that participants mostly needed skills to critique external tests. The researchers report that "the lack of ability to critically evaluate tests represents a risk for the teachers to take over tests unquestioningly without considering their quality" (p. 391). Lastly, even though not explicitly using the term LAL, the study by Arias, Maturana,

and Restrepo (2012) helped language teachers instill transparency and democracy in their practices. The researchers conceptualized *transparency* as making students aware of testing modes, rubrics, grades, and others; and *democracy* in language assessment as negotiation and the use of multiple methods and moments to assess learners. In summary, the knowledge and use of reversed engineering and test specifications, skills for critiquing existing tests and ESL standards, and transparency and democracy as assessment principles should all be part of language teachers' LAL.

Given all these possible additions to the construct under examination, LAL is still not clearly delimited for language teachers, and in fact appears to be far-reaching. For instance, if located on a spectrum, Inbar-Lourie's (2013b) ingredients of LAL can range from specific skills (e.g., item-writing) to complex issues such as the relationship between second language acquisition theories, language teaching approaches, and language assessment. Amidst all these ingredients and components, I believe we need to have a way to reconcile and streamline the implications of LAL for language teachers. To this end, in the next section I propose a core list of LAL that brings together thinking and research around LAL.

A Core List of LAL for Language Teachers

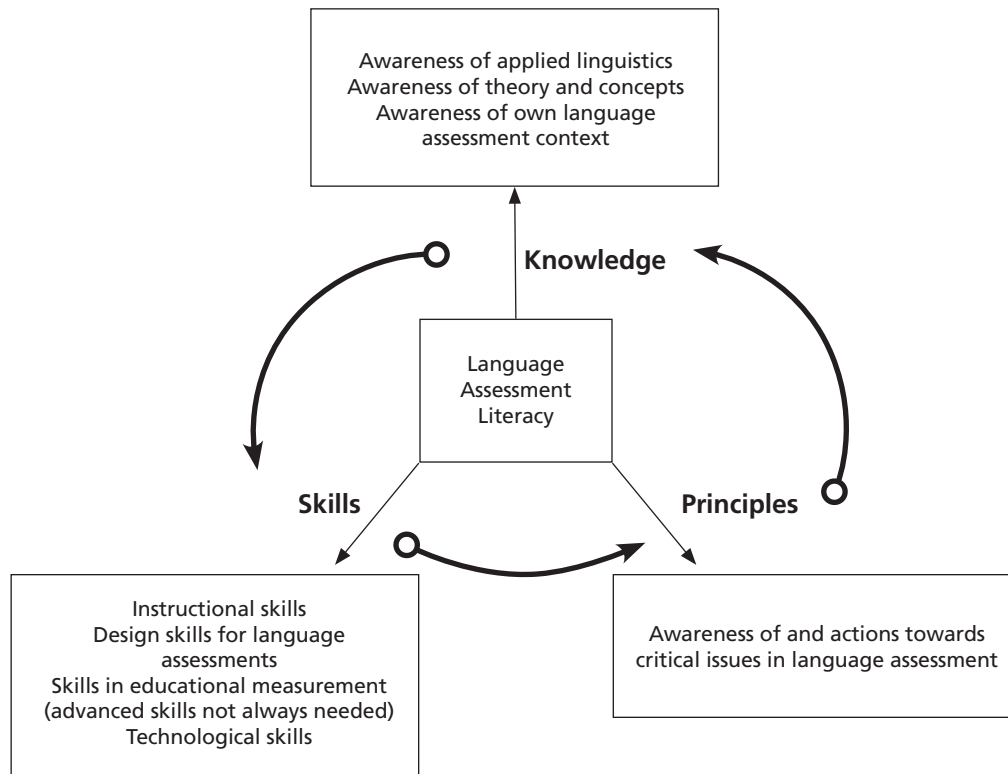
The proposed list is based upon three central components, introduced by Davies (2008), each with corresponding dimensions. *Knowledge* (three dimensions) reflects theoretical considerations such as the meaning of validity and reliability, two classical discussions in language testing. This component ranks high in the list as it deals with language and language use, the uniqueness of LAL (Inbar-Lourie, 2013a). Following this, within knowledge I include Davies' (2008) and Inbar-Lourie's (2008) suggestion

that knowledge of major issues in applied linguistics should be part of language teachers' LAL; for example, communicative approaches to language testing. Finally, this component includes teachers' knowledge of their own contexts for language assessment, an inclusion that I derive from Scarino's (2013) proposal.

Following in the list are *skills* (five dimensions), which first and foremost include instructional skills. I base this addition to LAL largely on the studies by McNamara and Hill (2011) and Rea-Dickins (2001) into assessment practices. Following, design refers to test and item construction for the four language skills and their integration in assessments (Fulcher, 2012; Taylor, 2009). Germane to educational assessment are measurement skills, which I include based on Davies (2008) and Fulcher (2010, 2012). In the case of language teachers, I agree with Popham (2011) that while advanced statistical expertise is not needed, teachers should know quantitative methods that can illuminate their assessment practice. Lastly, technological skills come from Davies (2008) and Inbar-Lourie (2012).

The last component of the list refers to *language assessment principles*. I derive this part from various authors (Arias et al., 2012; Coombe et al., 2012; Malone, 2013; Taylor, 2009; etc.). It has been discussed that large-scale tests are consequential and powerful (Shohamy, 2001), so ethics and fairness should be present in language assessment. In Taylor's (2013) proposed profile for language teachers, the author argues that this group may not be as concerned about ethics and fairness as language testing professionals must. However, I believe language teachers need to realize that these two principles are in fact codes for the professional practice of those involved in language assessment (ILTA, 2000). Most importantly, scholars in LAL argue that teachers need to become critical towards assessment practices (Fulcher, 2012; Scarino, 2013). Thus, transparency and democracy appear in

Figure 1. A Core List of Language Assessment Literacy Dimensions: Knowledge, Skills, and Principles



this last component thanks to the research by Arias et al. (2012).

Figure 1 summarizes the core list while Table 1 shows the complete list with an illustrative descriptor for each dimension.

Some similarities between this list and the work by Newfields (2006) are possible. The author proposes a series of statements for items that I also include in my list. For example, the “ability to interpret statistical raw data in terms of common measures of centrality (mean, mode, median) and deviation (*SD*, quartiles)” (p. 51) is similar to the following skill in the present list: the ability to interpret data from large-scale tests,

namely means, modes, medians, bell curves, SEMs, reliability and correlation coefficients, and so on.

Newfields’ inventory, however, is not presented hierarchically (i.e., by ranking high core components such as language and language teaching) and is based on content validity perceptions from college students, high school language teachers, and test developers. The present list is based on conceptual reviews of the literature in LAL and personal experience in language assessment courses through information from well-known language testing textbooks such as Bachman (2004), Fulcher (2010), and others.

Table 1. Descriptors for Knowledge, Skills, and Principles in Eight Dimensions of LAL for Language Teachers

Knowledge	
<i>Awareness of applied linguistics</i>	
1	<i>Compares</i> approaches for language teaching and assessment; e.g., communicative language testing; task-based assessment.
2	<i>Explains</i> major issues in applied linguistics; e.g., bilingualism, language policy and planning, pragmatics, sociolinguistics, etc.
3	<i>Analyzes</i> trends in second language acquisition and their impact on language assessment; e.g., motivation, cross-linguistic influence, learner strategies.
4	<i>Integrates</i> theories related to language and language use; e.g., models of language ability, discourse analysis, and grammar teaching.
<i>Awareness of theory and concepts</i>	
5	<i>Illustrates</i> history of language testing and assessment, and its impact on current practices and society.
6	<i>Interprets</i> reliability in language assessment and its implications: dependability, classical test theory, item analysis, threats, calculating reliability of tests and items, inter- and intra-rater reliability, etc.
7	<i>Interprets</i> validity in assessment and its implications: construct, content, and criterion validities, construct validity as unitary, Messick's (1989) consequential validity; validity as argument.
8	<i>Calculates</i> statistics procedures for investigating validity such as Pearson Product Moment Correlation (PPMC).
9	<i>Interprets</i> major qualities for language assessment practices (apart from reliability and validity), and their implications for language assessment: authenticity, practicality, interactiveness, fairness, ethics, and impact (including washback).
10	<i>Computes basic statistical analyses:</i> mean, mode, median, range, standard deviation, score distribution, etc.
11	<i>Differentiates</i> concepts related to assessment paradigms: traditional versus alternative; norm-referenced and criterion-referenced testing.
12	<i>Differentiates</i> major purposes and related decision-making for language testing: placement, achievement, proficiency, etc.
13	<i>Explains</i> major steps in developing tests: test purpose, construct definition, content specifications, test specifications, etc.
14	<i>Examines</i> the meaning and implications of critical language testing: power, ethics, and fairness.
15	<i>Judges</i> the consequences (intended or unintended) stemming from assessments in his/her context.
16	<i>Evaluates</i> the kind of washback that assessments can have on learning, teaching, curricula, and institutions.
17	<i>Contrasts</i> assessment methods, with their advantages and disadvantages; tests, portfolios, performance assessment, self- and peer-assessment, role-plays, among others.
18	<i>Articulates</i> the nature, purpose, and design of scoring rubrics; for example, holistic and analytic.
19	<i>Recognizes</i> what feedback implies within a formative assessment paradigm.

Awareness of own language assessment context

20	<i>Explains</i> own beliefs, attitudes, context, and needs for assessment.
21	<i>Evaluates</i> the test and assessment policies that influence his/her teaching.
22	<i>Assesses</i> the existing tensions that influence language assessment in his/her school.
23	<i>Illustrates</i> the general guidelines and policies that drive language learning and assessment in his/her context; for example, type of language curriculum.
24	<i>Criticizes</i> the kind of washback assessments usually have on his/her teaching context.

Skills**Instructional skills
has the ability to:**

25	align curriculum objectives, instruction, and assessment.
26	plan, implement, monitor, record, and report student language development.
27	provide feedback on students' assessment performance (norm- and criterion-referenced).
28	collect formal data (e.g., through tests) and informal data (while observing in class) of students' language development.
29	improve instruction based on assessment results and feedback.
30	utilize alternative means for assessment; for example, portfolios.
31	use language assessment methods appropriately: to monitor language learning and nothing else.
32	provide motivating assessment experiences, giving encouraging feedback, or setting up self-assessment scenarios.
33	communicate norm- and criterion-referenced test results to a variety of audiences: students, parents, school directors, etc.
34	use multiple methods of assessment to make decisions based on substantive information.
35	incorporate technologies in assessing students.

**Design skills for language assessments
has the ability to:**

36	clearly identify and state the purpose for language assessment.
37	clearly define the language construct(s) a test will give information about.
38	design assessments that are valid not only in terms of course contents but also course tasks.
39	construct test specifications (or blueprints) to design parallel forms of a test.
40	write test syllabuses to inform test users of test formats, where applicable.
41	design assessments that are reliable, authentic, fair, ethical, practical, and interactive.
42	write selected-response items such as multiple-choice, true-false, and matching.
43	improve test items after item analysis, focusing on items that are either too difficult, too easy, or unclear.
44	design constructed-response items (for speaking and writing), along with rubrics for assessment.
45	design rubrics for alternative assessments such as portfolios and peer-assessment.
46	provide security to ensure that unwanted access to tests is deterred.
47	design training workshops for raters, whenever necessary.

Skills in educational measurement (advanced skills not always needed)
has the ability to:

48	interpret data from large-scale tests, namely descriptive statistics such as means, modes, medians, bell curves, etc.; has the ability to calculate descriptive statistics.
49	infer students' strengths and weaknesses based on data.
50	criticize external tests and their qualities based on their psychometric characteristics.
51	interpret data related to test design, such as item difficulty and item discrimination.
52	calculate reliability and validity indices by using appropriate methods such as Kappa, PPMC, and others.
53	investigate facility and discrimination indices statistically.

Technological skills
has the ability to:

54	use software such as Statistical Package for the Social Sciences.
55	run operations on Excel; for example, descriptive statistics and reliability correlations.
56	use internet resources such as online tutorials and adapt contents for his/her particular language assessment needs.

Principles

Awareness of and actions towards critical issues in language assessment

57	Clearly informs the inferences and decisions that derive from scores in assessments.
58	Uses assessment results for feedback to influence language learning, not other construct-irrelevant sources (e.g., personal bias towards a student).
59	Treats all students, or users of language assessment, with respect.
60	Uses tests, test processes, and test scores ethically.
61	Provides assessment practices that are fair and non-discriminatory.
62	Critiques the impact and power standardized tests can have and has a stance towards them.
63	Observes guidelines for ethics used at the institution in regard to language assessment.
64	Criticizes external tests based on their quality and impact.
65	Implements transparent language assessment practices; informs students of the what, how, and why of assessment.
66	Implements democratic language assessment practices, by giving students opportunities to share their voices about assessment.

Recommendations

The proposed list can be used by language teachers in five ways. First, they can utilize the descriptors as a Yes/No checklist to evaluate their own language assessment contexts, paying attention to what they do well and what they need to strengthen. Second, they can use the descriptors to observe each other's LAL and

provide feedback on knowledge, skills, and principles. For example, applicable abilities in the 25 to 35 range can be turned into an observation protocol for stages in classroom language assessment. Third, teachers can identify topics to know more about what is in LAL and seek for training opportunities such as professional development teams or study groups; in such groups,

teachers may want to understand large-scale tests, so they would need to read about *educational measurement* and develop corresponding data interpretation skills (e.g., What does a mode tell me about test results?). Fourth, if teachers design tests, they may need to see what skills in this list are appropriate for their enterprise. Lastly, teachers can use this list for an overview of language assessment literacy: a large and still developing construct in applied linguistics. Overall, teachers are encouraged to use this list however they feel useful for their purposes.

Besides language teachers, this list may prove useful for teacher educators in both pre- and in-service programs. For pre-service teachers, educators can use it to introduce future language teachers to the field of language assessment; the list may be used as a pre-test and post-test to language testing courses and provide the pre-service teachers with the chance to observe how much they have learned in a language testing course. Regarding in-service teacher education, tutors can turn the list into a needs assessment or a diagnostic test in order to plan programs in language assessment; the pre-test/post-test treatment can be used in in-service teacher development.

A caution that I feel necessary to address is that the list includes parts of a greater whole. The dimensions in the list should not be seen separately but have been separated here for the sake of clarity and organization. Rather, they should be envisioned as complementary, first and foremost, depending on teachers' contexts. For example, teachers who are required to design language tests with considerable impact may need strong design skills, some knowledge of educational measurement, and awareness of theory and concepts. The combination of these skills and this knowledge should help them bring about quality products.

Limitations

There are four limitations in this core LAL list that deserve discussion. To start, this list is not meant to

be an authoritative account of what LAL actually is for language teachers; it does, however, bring together thinking from scholars and researchers in assessment literacy and most specifically in LAL. What is more, the list has a personal bias. I have developed it based on my understanding of the literature and my own experience as a test writer and student of language testing. Additionally, the 66 descriptors may not do justice to the width and depth of LAL but only comprise a fraction of what the construct implies in theory and practice for language teachers; I may have overlooked key skills, knowledge, or principles that are indeed part of teachers' LAL. In this same vein, there are descriptors that can include other more detailed skills. For example, in descriptor one, one sub-component is knowledge of issues within task-based assessment, namely the discussion of task-centered and construct-centered assessment in test design (Bachman, 2002).

Lastly, this list includes statistical procedures (e.g., descriptor 55) teachers need not concern themselves with, according to some authors (Brookhart, 2003; Popham, 2009). However, the idea that teachers do not need knowledge of statistics (at least at a basic level) may underestimate their potential. In the study by Palacio, Gaviria, and Brown (2016), the participating English language teachers used statistical procedures such as correlations and reliability analyses to improve the quality of the tests they designed.

Notwithstanding these limitations, I invite readers to examine the arguments and proposal I present to advance the knowledge base necessary to operationalize the meaning and implications of LAL for language teachers.

Conclusions

Language teachers throughout the world make decisions based on assessment data. In turn, such decisions impact teaching and learning. Given this scenario, there is a need for language teachers to have solid assessment literacy. Likewise, language teaching programs should

be more profoundly engaged in providing quality LAL—and not do so through elective courses which may not do much (Siegel & Wissehr, 2011) or merely mention assessment in passing. More importantly, programs and opportunities for in-service teachers are also central to improving the state of LAL. While the call is indeed necessary, the field of language education should carefully reflect upon the nature and scope of LAL, as it is indeed an expanding notion.

Historically, the meaning of assessment literacy has extended to include issues such as technology and even student motivation. While the meaning of LAL has been rather stable, the actual scope of each component (knowledge, skills, and principles) is still expanding. This expansion has become all the more prominent due to the call that several stakeholders (e.g., university administrators and politicians) must be included in the LAL equation.

While the contents of and people involved in LAL are still the focus of scholarly work and commentary, this paper has presented a comprehensive list to operationalize LAL for language teachers, an essential stakeholder group. Such list is proposed as a way to highlight the knowledge, skills, and principles that, according to the literature and research, language teachers should have when assessing language. The paper has discussed five ways in which language teachers can use the list; besides, the paper has discussed its limitations, ending with a call to further discussion. Even though it cannot be prescribed that *all* language teachers have such a repertoire, as Taylor (2013) explains, the overarching categories—that is *knowledge, skills, and principles*—still apply, whether we discuss assessment in the language classroom or out of it (Fulcher, 2012).

While assessment literacy may be far-reaching, the importance of such literacy for the language teacher cannot be underestimated, and it should be complemented by what their contexts have to offer so that such construct is better operationalized. The effect of what LAL truly means should be language teachers who display

knowledge, skills, and principles that are consonant with language teaching and language learning. High quality assessment is done by language teachers who plan, design, implement, monitor, record, evaluate, provide, and improve opportunities for the overarching goal in the language classroom and beyond; that is, the development of students' language ability. Lastly, because it is an expanding controversy in language education, the meaning and implications of LAL are still in fruitful development.

References

- American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. Washington, DC: Authors. Retrieved from <http://buos.org/standards-teacher-competence-educational-assessment-students>.
- Arias, C., & Maturana, L. (2005). Evaluación en lenguas extranjeras: discursos y prácticas [Assessment in language teaching: Discourse and practices]. *Íkala, Revista de Lenguaje y Cultura*, 10(1) 63-91.
- Arias, C., Maturana, L., & Restrepo, M. (2012). Evaluación de los aprendizajes en lenguas extranjeras: hacia prácticas justas y democráticas [Evaluation in foreign language learning: Towards fair and democratic practices]. *Lenguaje*, 40(1), 99-126.
- Bachman, L. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19(4), 453-476. <https://doi.org/10.1191/0265532202lt2400a>.
- Bachman, L. (2004). *Statistical analyses for language assessment*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9780511667350>.
- Bailey, K., & Brown, J. D. (1996). Language testing courses: What are they? In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 236-256). Clevedon, UK: Multilingual Matters.
- Boyles, P. (2006). Assessment literacy. In M. Rosenbusch (Ed.), *New visions in action: National Assessment Summit papers* (pp. 18-23). Ames, US: Iowa State University.

- Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, . . . K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126-136). Cambridge, UK: Cambridge University Press.
- Brookhart, S. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5-12. <https://doi.org/10.1111/j.1745-3992.2003.tb00139.x>.
- Brookhart, S. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>.
- Brown, J. D., & Bailey, K. (2008). Language testing courses: What are they in 2007? *Language Testing*, 25(3), 349-383. <https://doi.org/10.1177/0265532208090157>.
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3) 360-389. <https://doi.org/10.1191/0265532204lt2880a>.
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges, and recommendations. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyhoff (Eds.), *The Cambridge guide to second language assessment* (pp. 20-29). Cambridge, UK: Cambridge University Press.
- Davidson, F., & Lynch, B. (2001). *Testcraft: A teacher's guide to writing and using language test specifications*. New Haven, US: Yale University Press.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347. <https://doi.org/10.1177/0265532208090156>.
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, (43)3, 393-415. <https://doi.org/10.1002/j.1545-7249.2009.tb00242.x>.
- 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 [The English teacher: His beliefs about English language assessment at primary, secondary and tertiary levels]. *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(1), 171-201.
- Fulcher, G. (2010). *Practical language testing*. London, UK: Hodder Education.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. <https://doi.org/10.1080/15434303.2011.642041>.
- Herrera, L., & Macías, D. (2015). A call for language assessment literacy in the education and development of Teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, 17(2), 302-312. <https://doi.org/10.14483/udistrital.jour.calj.2015.2.a09>.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402. <https://doi.org/10.1177/0265532208090158>.
- Inbar-Lourie, O. (2012). Language assessment literacy. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-9). Oxford, UK: John Wiley & Sons. <https://doi.org/10.1002/9781405198431.wbealo605>.
- 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 presented at the 4th CBLA SIG Symposium Programme, University of Cyprus.
- International Language Testing Association. (2000). *Code of ethics for ILTA*. Retrieved from http://c.ymcdn.com/sites/www.iltaonline.com/resource/resmgr/docs/ilta_code_english.pdf.
- 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>.

- Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555-584. <https://doi.org/10.1177/0265532209351431>.
- Kunnan, A. (2003). Test fairness. In M. Milanovic & C. Weir (Eds.), *Selected papers from the European Year of Languages Conference, Barcelona* (pp. 27-48). Cambridge, UK: Cambridge University Press.
- López, 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.
- Malone, M. (2008). Training in language assessment. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education: Language testing and assessment* (2nd ed., Vol. 7, pp. 225-233). New York, US: Springer. https://doi.org/10.1007/978-0-387-30424-3_178.
- Malone, M. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329-344. <https://doi.org/10.1177/0265532213480129>.
- 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. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, US: 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.
- Newfields, T. (2006, May). Teacher development and assessment literacy. In T. Newfields, I. Gledall, M. Kawate-Mierzejewska, Y. Ishida, M. Chapman, & P. Ross (Eds.), *Authentic communication: Proceedings of the 5th Annual JALT Pan-SIG Conference* (pp. 48-73). Shizuoka, JP: Tokai University College of Marine Science.
- Nier, V., Donovan, A., & Malone, M. (2009). Increasing assessment literacy among LCTL instructors through blended learning. *Journal of the National Council of Less Commonly Taught Languages*, 9, 105-136.
- 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>.
- Palacio, M., Gaviria, S., & Brown, J. D. (2016). Aligning English language testing with curriculum. *Profile: Issues in Teachers' Professional Development*, 18(2), 63-77. <https://doi.org/10.15446/profile.v18n2.53302>.
- 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>.
- Popham, W. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4-11. <https://doi.org/10.1080/00405840802577536>.
- Popham, W. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265-273. <https://doi.org/10.1080/08878730.2011.605048>.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462.
- Rudner, L., & Schafer, W. (Eds.). (2002). *What teachers need to know about assessment*. Washington, DC: National Education Association.
- 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>.
- Schafer, W. (1993). Assessment literacy for teachers. *Theory Into Practice*, 32(2), 118-126. <https://doi.org/10.1080/00405849309543585>.
- Shohamy, E. (2001). *The power of tests*. Harlow, UK: Pearson Education.
- Siegel, M., & Wissehr, C. (2011). Preparing for the plunge: Pre-service teachers' assessment literacy. *Journal of Science Teacher Education*, 22(4), 371-391. <https://doi.org/10.1007/s10972-011-9231-6>.

- Stiggins, R. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238-245.
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36. <https://doi.org/10.1017/S0267190509090035>.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412. <https://doi.org/10.1177/0265532213480338>.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374-402. <https://doi.org/10.1080/15434303.2014.960046>.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749-770. <https://doi.org/10.2307/20466661>.
- Walters, F. (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>.
- White, E. (2009). Are you assessment literate? Some fundamental questions regarding effective classroom-based assessment. *OnCue Journal*, 3(1), 3-25.

About the Author

Frank Giraldo has an MA in the teaching of English as a second language from University of Illinois at Urbana-Champaign (USA). He also holds an MA in English didactics from Universidad de Caldas, Colombia. He is an academic consultant at the Instituto de Lenguas Extranjeras (ILEX) at Universidad Tecnológica de Pereira, Colombia. His interests are language testing and assessment, curriculum development, and professional development of language teachers.