

Defining, Conceptualizing, Problematizing, and Assessing Language Teacher Assessment Literacy

Michelle L. Stabler-Havener¹
Teachers College, Columbia University

ABSTRACT

Assessment literacy (AL) empowers teachers (Grabowski & Dakin, 2014) by helping them make better decisions about the development, administration, and use of assessments (Harding & Kremmel, 2016; Popham, 2009). This is crucial because when teachers make erroneous interpretations, which lead to incorrect decisions, students can suffer unintended, negative consequences (Purpura, 2016; Purpura, Brown, & Schoonen, 2015). As essential as AL is to providing quality teaching and learning opportunities, assessment education remains inadequate (Lam, 2015; Mendoza & Arandia, 2009; Vogt & Tsagari, 2014). According to Cheng (2001), teachers spend up to a third of their time in assessment related activities; however, most “do so with little or no professional training” (Bachman, 2000, p. 19-20). In light of the demand for the development of instructor AL, this paper first defines and conceptualizes language assessment literacy (LAL). Then it problematizes LAL through an investigation of empirical studies conducted to explore pre- and in-service teacher education and resources that have been used to develop teachers’ LAL. Finally, it examines how teachers’ LAL levels are evaluated through assessments such as Cambridge University Press’s Teaching Knowledge Test, the edTPA, and state teacher certification Praxis tests.

Keywords: language assessment literacy, classroom-based assessment, teacher education

INTRODUCTION

Assessment literacy (AL) empowers teachers (Grabowski & Dakin, 2014) by helping them make better decisions about the development, administration, and use of assessments (Harding & Kremmel, 2016; Popham, 2009). This is crucial because when teachers make erroneous interpretations, which lead to incorrect decisions, students can suffer unintended, negative consequences (Purpura, 2016; Purpura, Brown, & Schoonen, 2015) such as failing to promote students who deserve to be advanced to the next level or not providing adequate support to students who need additional assistance. As essential as AL is to providing quality teaching and

¹Michelle Stabler-Havener is pursuing her Ed.D. in Applied Linguistics at Teachers College, Columbia University, specializing in second language assessment. Her research interests include classroom-based assessment, scenario-based assessment, and teacher education. Correspondence should be sent to Michelle Stabler-Havener, 316 Zankel, Teacher College, Columbia University, 525 W 120th Street New York, NY 10027. Email: mls2262@tc.columbia.edu

learning opportunities, assessment education remains inadequate (Lam, 2015; Mendoza & Arandia, 2009; Vogt & Tsagari, 2014). According to Cheng (2001), teachers spend up to a third of their time in assessment related activities; however, most “do so with little or no professional training” (Bachman, 2000, p. 19-20). Also remarking on the need for increased AL, Coombe, Troudi, and Al-Hamly (2012) lament that “without a higher level of teacher assessment literacy, we will be unable to help students attain higher levels of academic achievement” (p. 20). Taylor (2009) further reminds us that the need for teachers to be assessment literate continues to grow as new test development projects are initiated at national and regional levels by countries which seek to align policy and practice with new pedagogical approaches. Recognizing the demand for the development of instructor AL, the purpose of this paper is to review the literature on teachers’ language assessment literacy (LAL) by first defining and conceptualizing LAL, then problematizing LAL by investigating empirical research, and, finally, discussing the assessment of teachers’ LAL levels.

DEFINING AL/LAL

Definitions of LAL are rooted in general education AL literature. Stiggins (1991), who first coined the term “assessment literacy,” identified an assessment literate person as one who could discern between excellent and poor-quality assessments and apply that knowledge to make informed inferences about student achievement. Since his seminal article on the topic, the language assessment community has noted that LAL should be considered separately from general AL because of the “unique complexities that are entailed in the testing and assessment of linguistic skills, knowledge, and communicative competence” (Harding & Kremmel, 2016, p. 414; see also Inbar-Lourie, 2008a; Jeong, 2013; Taylor, 2013). Take for example, Pill and Harding’s (2013) definition of LAL as “a repertoire of competencies that enable an individual to understand, evaluate and, in some cases, create *language* tests and analyze test data” (p. 382, emphasis added). Or Malone’s (2013) classroom specific definition: “*language* instructors’ familiarity with testing definitions and the application of this knowledge to classroom practices in general and specifically to issues related to assessing *language*” (p. 329, emphasis added).

Nevertheless, despite the fact that what makes LAL unique from AL is the incorporation of language in how it is conceptualized and operationalized, some in the assessment field fail to include a language component, i.e., an understanding of second language constructs, in their definition of LAL. For example, O’Loughlin (2013) describes LAL as the “acquisition of a range of skills related test production, test-score interpretation and use, and test evaluation in conjunction with the development of a critical understanding about the roles and functions of assessment within society” (p. 363). His emphasis on the need to understand that assessment administration and interpretation occur within societal contexts is laudable, but his exclusion of language seems to be an oversight. Likewise, Fulcher (2012) fails to include language in his working definition of LAL:

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

to understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals. (p. 125)

Though still the most detailed, comprehensive definition of LAL to date, it does not include knowledge of language constructs or language assessment. However, language assessment is mentioned explicitly in Fulcher's (2012) diagram of what he calls an "expanded definition" of LAL through the inclusion of "the practice of *language testing*" (p. 126, emphasis added). A discussion of Fulcher's (2012) expanded definition follows in the section entitled "Conceptualizing LAL Models Through Competencies." Although definitions provide helpful glimpses into what LAL entails, more specific conceptualizations are provided through lists of competencies that a language assessment literate instructor should possess. Educational bodies and language assessment researchers have made recommendations as to the core competencies assessment literate teachers should have; these are discussed below.

CONCEPTUALIZING GENERAL EDUCATION AL THROUGH COMPETENCIES

As with definitions of LAL, conceptualizations of LAL have been influenced by general education AL literature. A noteworthy contribution frequently referred to by researchers in the language assessment field (Brindley, 2001; Brown & Bailey, 2008; Coombe, Troudi, & Al-Hamly, 2012; Fulcher, 2012; Harding & Kremmel, 2016; Inbar-Lourie, 2013) is the American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME), and the National Education Association (NEA; 1990) "Standards for Teacher Competence in Educational Assessment of Students." According to this document, the seven assessment competencies in which teachers should be skilled are: (1) choosing and (2) developing assessment methods appropriate for making instructional decisions; (3) administering, scoring, and interpreting the results of teacher produced and externally produced assessments; (4) appropriately using assessment results to inform decisions regarding student and curriculum development; (5) devising valid grading procedures for student assessments; (6) communicating assessment results to stakeholders; and (7) identifying unethical and inappropriate assessment methods and use of assessment data. Although the AFT, NCME, and NEA Standards have not been updated, Klinger et al. (2015) consulted the AFT, NCME, and NEA Standards framework when developing the new *Classroom Assessment Standards for preK-12 Teachers*. US states such as Illinois have adapted these principles when developing their state-specific assessment documents (Zaleski, 2015). However, assessment literature reviewed for this paper did not reference the *Classroom Assessment Standards*. Reasons for this may be that it was published recently or because there are now LAL specific models for the language assessment community to reference.

CONCEPTUALIZING LAL MODELS THROUGH COMPETENCIES

The LAL competencies in this section are presented through a discussion of the following models: a five-component, professional development program model (Brindley, 2001); a skills, knowledge, and principles model (Davies, 2008; Inbar-Lourie, 2008a); a practices, principles,

and contexts model (Fulcher, 2012); and an LAL stakeholder profile model (Pill & Harding, 2013; Taylor, 2013). The section concludes with a brief reflection on some of the challenges and problems with the LAL models.

A Five-Component, Professional Development Program Model

Using the AFT, NCME, and NEA Standards as a springboard, Brindley (2001), the first language assessor to write about AL (Fulcher, 2012), proposed five components which professional assessment development programs should include. Critiquing the AFT, NCME, and NEA Standards, Brindley (2001) felt that in addition to core competencies, it should be acknowledged “that different individuals will require different levels of knowledge according to the nature and extent of their involvement in assessment issues” (p. 128). Therefore, his assessment professional development program was comprised of two core units and three optional units. One core unit dealt with social, educational, political, and ethical characteristics of assessment as they pertained to instructors’ teaching contexts and wider communities. Teacher awareness of social assessment characteristics is critical because not all contexts have the same assessment values and perspectives. For example, in some countries, formative assessment is valued over summative assessment and vice versa. Additionally, since social context affects how an assessment’s validity is perceived by stakeholders, teachers need to be aware of social issues when designing assessments and reporting their results. Teaching instructors about ethical test use is also paramount, e.g., it would not be ethical to base decisions regarding student grade level advancement on a poorly designed test that has not been proven to be valid or reliable. Regarding the political nature of language assessments, drawing teachers’ attention to ways that assessments have been used to inappropriately classify people, determine exclusive membership, and set standards for success or failure to the disadvantage of certain individuals or groups (Shohamy, 1998) may help teachers avoid bias when constructing assessments.

Concerning the remaining four components of Brindley’s (2001) professional development program, the other core unit was devoted to defining and describing language proficiency including a “critical evaluation of theoretical models of language ability such as those as proposed by Canale and Swain (1980) and Bachman and Palmer (1996)” (Brindley, 2001, p. 129). With respect to the optional units, one involved developing and evaluating language tests and an introduction to statistical analysis. Another optional unit covered criterion referenced tests and techniques associated with assessment alternatives such as portfolios, journals, self-assessments, and project work, and explored how these types of assessments are integrated into the language curriculum. The final unit mapped out a follow-up strategy for putting into practice issues raised in the other program units.

In 2008(a), Inbar-Lourie reviewed Brindley’s (2001) assessment professional development program discussed above, categorizing the themes of his instructional units according to the following: the *why* (the reason for assessment: core unit 1), the *what* (the trait or characteristics of language ability to be assessed: core unit 2), and the *how* (the method of assessment: optional units 1-3). In her discussion of the *why*, the reason for the assessment, Inbar-Lourie (2008a) agreed with Brindley’s (2001) suggestion that professional education programs include a mandatory unit on the social context of assessments. She felt this was salient given the attention that the language assessment community has drawn to the ethical responsibility of language assessors and to the role of language assessments in society

(McNamara & Roever, 2006; Shohamy, 2001). Nonetheless, she critiqued Brindley (2001) for not including the role of language assessment on civil, vocational, and educational decision-making processes. Inbar-Lourie (2008a) believed that this was an oversight because such social themes “foreground and shape perceptions of language as a social entity in terms of both the trait to be assessed and the assessment process” (p. 391). Next, in her section on the *what*, the trait to be assessed, she affirmed Brindley’s (2001) choice to require teachers to study theoretical approaches to language assessment because “language assessors are expected... to skillfully implement assessment measures that are compatible” (p. 391) with current assessment theory and research. In her discussion of the *how*, the method of assessment, Inbar-Lourie (2008a) disagreed with Brindley’s (2001) decision to make the last three units optional. As Harding and Kremmel (2016) pointed out, “all five areas identified by Brindley might be understood as mandatory” (p. 417) for language teachers’ professional development.

A Skills, Knowledge, and Principles Model

Published the same year as Inbar-Lourie’s (2008a) article, which included the review of Brindley’s (2001) professional development courses, Davies (2008) wrote a comprehensive examination of language assessment textbooks from Lado (1961) to Fulcher and Davidson (2007). In it, he identified a trend in assessment textbook content to cover *skills*, *knowledge*, and *principles* which moderately echo the *how*, *what*, and *why* of Inbar-Lourie’s (2008a) categorization of Brindley’s (2001) assessment professional development program (Harding & Kremmel, 2016). For Davies (2008), *skills* offer education in appropriate methodology and tools such as item writing, statistics, analysis, and reportage as well as software programs that assist with these endeavors. *Knowledge* provides a foundation in measurement and language proficiency as well as contextual considerations. Finally, *principles* are concerned with assessment use, impact, fairness, and ethics.

A Practices, Principles, and Contexts Model

Broadening the *skills*, *knowledge*, and *principles* model, Fulcher (2012) constructed a diagram of, what he referred to as, an expanded definition of LAL (see Figure 1). Like his working definition of LAL referred to in the “Defining AL/LAL” section, his expanded definition was based on data gathered from 278 international instructors through an online survey regarding the assessment education needs of language teachers. The diagram of Fulcher’s (2012) expanded definition implies a hierarchy, with *practices* (i.e., the knowledge, skills, and abilities involved in language testing) appearing as the foundation of LAL. The next level, *principles*, is composed of the principles, processes, and concepts that guide foundational practices. The topmost tier of Fulcher’s (2012) expanded definition, *contexts*, expresses a level of complexity and development not present in other models. It comprises the historical, social, and political philosophical frameworks in which the practices and principles are placed so that assessment stakeholders are able to discern the origins, impacts, and reasons for the practices and principles. Fulcher (2012) does not believe all elements of the expanded definition are essential for all stakeholders. Nonetheless, since his expanded definition of LAL is based on data collected

from language instructors, it could be argued that language instructors require competence in all the expanded definition descriptors.

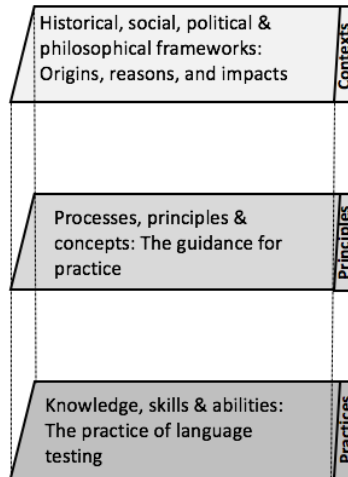


FIGURE 1
Language assessment literacy: An expanded definition (Fulcher, 2012, p. 126)

An LAL Stakeholder Profile Model

As mentioned above, Brindley (2001) and Fulcher (2012) observed that acquisition of all LAL competencies is not necessary for all stakeholders. Sensitive to this issue, Pill and Harding (2013) identified five LAL proficiency levels that are based on Bybee's (1997) work in scientific literacy and Kaiser and Willander's (2005) work in mathematical literacy. These five LAL proficiency levels encompass: (0) *illiteracy* (ignorance), (1) *nominal literacy* (basic understanding with some misconceptions), (2) *functional literacy* (sound, basic understanding), (3) *procedural and conceptual literacy* (understanding and practice of central concepts), and (4) *multidimensional literacy* (extended knowledge beyond ordinary concepts). Providing the only continuum model of LAL, Pill and Harding (2013) are to be commended for their ingenuity in applying models of scientific and mathematical literacy to the study of LAL. Nonetheless, Harding and Kremmel (2016) address potential limitations of Pill and Harding's (2013) model—generally speaking, the descriptors tend to overemphasize procedural and theoretical knowledge and, except for level 4, neglect social, ethical, and political competencies. Additionally, the continuum does not specify what level of LAL is necessary for different stakeholders. Taylor's (2013) model of LAL, which uses Pill and Harding's (2013) continuum model, addresses the latter concern.

Taylor's (2013) model of LAL takes into account eight LAL dimensions which are similar to the lists of competencies already discussed. These dimensions include *knowledge of theory, technical skills, principles and concepts, language pedagogy, sociocultural values, local practices, personal beliefs/attitudes, and scores and decision making*. One dimension that Taylor (2013) includes but other LAL models lack is teachers' *personal beliefs/attitudes*. Her inclusion of this dimension is noteworthy because, for example, researchers such as Breen et al. (1997), Leung (2004), and Scarino (2013) have mentioned that teachers' attitudes and beliefs about assessment impact their willingness to adopt new educational policies like implementing

formative assessment or state-mandated assessment. Taylor (2013) then takes the eight dimensions and maps them onto Pill and Harding's (2013) five levels of LAL proficiency. What particularly makes her model stand out from other models is its ability to differentiate between the LAL needs of different stakeholder groups.

By mapping her eight dimensions of LAL onto Pill and Harding's (2013) five LAL proficiency levels, Taylor (2013) created differential LAL profiles for four constituencies: test writers, classroom teachers, university administrators, and professional language testers. Her profile for a classroom teacher, as seen in Figure 2, is as follows (the numbers correspond to Pill and Harding's (2013) five LAL proficiency levels above): knowledge of theory (2), technical skills (3), principles and concepts (2), language pedagogy (4), sociocultural values (3), local practices (3), personal beliefs/attitudes (3), and scores and decision making (2). Taylor's (2013) model contributed a fresh perspective on LAL by addressing a gap in the literature, that is how to differentiate between various stakeholder LAL needs. However, as Harding and Kremmel (2016) note, Taylor's (2013) dimensions are speculative. For example, some might argue that classroom teachers require more advanced literacy in *local practices*, *scores and decision making*, and *personal beliefs/attitudes* than Taylor (2013) proposes. Additionally, one can only conjecture what competencies her eight dimensions might comprise because she does not define them.

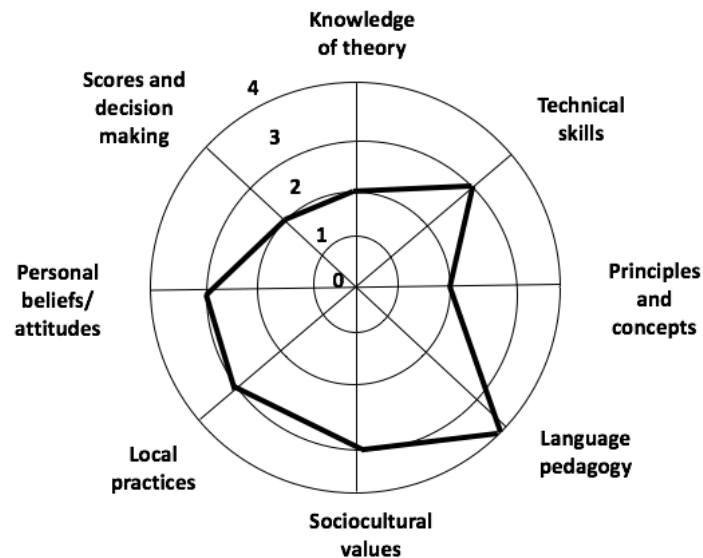


FIGURE 2
LAL profile for classroom teachers (Taylor, 2013, p. 410)

To summarize, as seen in Table 1, conceptualizations of LAL highlight the need for language teachers to be able to define and describe language proficiency and have knowledge of language assessment theory and methods. In light of this knowledge, language assessment literate individuals should be able to develop quality assessment instruments and properly analyze the evidence gathered from the administration of the assessment instruments. Additionally, LAL competence requires teachers to appropriately contextualize assessment practices and use assessment results ethically. The level of expertise a teacher is required to possess to be considered language assessment literate depends on the teacher's instructional context.

TABLE 1
Language Assessment Literacy Models

| Author(s) | Date | Recommended LAL Competencies |
|------------------|-------------|--|
| Brindley | 2001 | <ul style="list-style-type: none"> • The social context of assessment (essential) • Defining and describing proficiency (essential) • Constructing and evaluating language tests (optional) • Assessment in the language curriculum (optional) • Putting assessment into practice (optional; Brindley, 2001, pp. 129-130) |
| Inbar-Lourie | 2008 | <ul style="list-style-type: none"> • <i>Why</i>: The reason for assessment • <i>What</i>: The trait to be assessed • <i>How</i>: The method of assessment (Inbar-Lourie, 2008a, p. 390) |
| Davies | 2008 | <ul style="list-style-type: none"> • <i>Skills</i>: Item writing, test analysis, statistics, reportage and the ability to use software programs for these endeavors • <i>Knowledge</i>: Background in measurement, language description, and the contextualization of language assessments • <i>Principles</i>: language test use, impact, ethics and professionalism (Davies, 2008, p. 335) |
| Fulcher | 2012 | <ul style="list-style-type: none"> • <i>Practices</i>: Knowledge, skills and abilities (the practice of language testing) • <i>Principles</i>: Processes, principles and concepts (the guidance for practice) • <i>Contexts</i>: Historical, social, political and philosophical frameworks (origins, reasons, and impacts; Fulcher, 2012, p. 126) |
| Pill and Harding | 2013 | <ul style="list-style-type: none"> • <i>Illiteracy</i>: Ignorance of language assessment concepts and methods • <i>Nominal literacy</i>: Understanding that a specific term relates to assessment, but may indicate a misconception • <i>Functional literacy</i>: Sound understanding of basic terms and concepts • <i>Procedural and conceptual literacy</i>: Understanding central concepts of the field and using knowledge and practice • <i>Multidimensional literacy</i>: Knowledge extending beyond ordinary concepts including philosophical, historical and social dimensions of assessment (Pill & Harding, 2013, p. 383) |
| Taylor | 2013 | <ul style="list-style-type: none"> • Knowledge of theory • Technical skills • Principles and concepts • Language pedagogy • Sociocultural values • Local practices • Personal beliefs/attitudes • Scores and decision-making (Taylor, 2013, p. 410) |

Challenges and Problems with the Models

The progress that has been made in conceptualizing LAL since Brindley first wrote on it in 2001 is commendable; however, challenges and problems with LAL models remain. Though there are general similarities between the models as noted above, each model tends to articulate LAL competencies differently. Such differences can make it challenging for LAL stakeholders to clearly communicate about LAL competencies. For example, if stakeholders agree that *principles* are an important LAL competency, they need to clarify to which model they refer because Davies (2008) and Fulcher (2012) define *principles* differently. Another problem with the models is that although most of them provide general definitions of the competencies and some provide specific examples, none of them operationalize their model with detailed, descriptive specifications as to what a language assessment literate person needs to know and do in order to be considered language assessment literate.

The models' variety and lack of specificity could be considered positive by some who are assessment literate enough to select, adapt, and operationalize a model most appropriate for their context. However, there may be teacher educators and educational administrators who feel that language assessment professional development is critical for the teachers under their care, but they do not possess the know-how and/or time to select, adapt and operationalize an LAL model for their context. As Jeong (2013) found through her survey of 140 self-identified language assessment instructors, half of them came from non-testing backgrounds. Because of this, she argued that the language testing community should provide course objectives and guidelines for non-experts' use. Providing detailed specifications for an LAL model which teacher educators and educational administrators could adapt to suit their context would be a welcome addition to the LAL field. An additional issue with the LAL models described here is that most are teacher-centered given that the foundation of their development is the language assessment competencies teachers require. In order to be useful to assessment contexts beyond the classroom, other stakeholders should be included in conceptualizations of LAL.

Returning to the primary purpose of the paper, i.e., teachers' LAL, a strategy to keep learners as the primary beneficiaries of teachers' increased LAL levels would be to develop an LAL model using a learning-oriented assessment approach (LOA; Turner & Purpura, 2015). An LOA approach begins with "an examination of the learning process' followed by considerations of how assessment can serve to enhance this process (Purpura, 2004, 2009)" (Turner & Purpura, 2015, p. 255). Applying this approach to the development of an LAL model would mean that instead of basing LAL constructs on data gathered from assessment textbooks (Davies, 2008) or teacher surveys (Fulcher, 2012), LAL model construction would begin with an investigation into the learning process including an identification of the types of assessments that most beneficially promote learning. The resulting LAL model would focus on teacher LAL competencies that best meet the needs of learners.

PROBLEMATIZING LAL

In the previous section, conceptualizations of LAL were explained and reviewed. In this section, teachers' LAL will be problematized by discussing the issues, implications, and priorities that emerge from an investigation of empirical research. Topics that will be considered are pre- and in-service teacher education and resources that have been used to develop teachers'

LAL. The section concludes with a discussion on the divide that frequently separates testing experts and teachers.

Pre- and In-service Teacher Education

In order to better understand the interaction between language teaching and testing, Bailey and Brown (1996) investigated the content, structure, and students' perspectives on introductory language testing courses by creating a questionnaire, which was responded to by 84 language testing course instructors from different parts of the world. Later in 2007, Brown and Bailey (2008) distributed the same survey to discover how language testing courses may have changed since 1996. A comparison of the results of the two studies revealed that more respondents from countries outside the United States (US) volunteered to take the survey in 2007, which may be a possible indicator of the growing diversification of the field. Interestingly, the number of different textbooks being used from 1996 to 2007 decreased from 32 to 29. One participant noted that his change in textbook use was due to publications going out of print; thus, the authors suggested that publishers might want to consider updating editions of popular texts. In general, testing courses covered formal assessment (FA) practices such as item writing, analysis, facility, and discrimination, as well as basic descriptive statistics, validity and reliability. As this list of topics reveals, it would appear that pre-service and in-service teachers taking testing courses would likely receive adequate education in formal testing practices but may not be prepared to assess students using more formative, classroom-based assessment (CBA) approaches, through the use of, for instance, oral and written feedback, portfolios, student self- and peer-assessments, and so forth.

Also examining assessment education courses, Lam (2015) investigated the degree to which assessment education courses in Hong Kong develop or inhibit pre-service teachers' LAL by reviewing curricula at five teacher education institutions (TEIs). He then held focus group interviews with 40 students and individual interviews with nine language assessment instructors from one TEI. He found that assessment courses in the five TEIs which he examined could not sufficiently equip students to be language assessment literate. Major issues included the fact that some TEIs only offered general assessment courses rather than language assessment courses; moreover, three programs only offered assessment courses as electives, which is a concern because offering a course does not ensure that pre-service teachers will enroll in it. Additionally, Lam (2015) discovered that students tended to equate assessment with testing and preferred honing technical skills to learning about CBA practices. The need to bridge the theory-practice gap by pairing assessment instruction with practicum teaching also surfaced. Some implications of this study include the importance of improving the quantity and quality of TEIs' language assessment education programs. Better defining the LAL construct by providing better specified descriptors regarding what it means to be language assessment literate might assist with this endeavor. Though not very generalizable due to its small sample size and singular location, there are valuable LAL priorities which emerged from this study that correlate with results from other research.

Take for example Mendoza and Arandia's (2009) study, which explored Colombian English language teachers' perceptions and uses of classroom assessment by surveying 82 participants about their professional background, feelings regarding assessment, and their assessment practices. The authors also examined curricula from seven graduate and 27

undergraduate programs. Like Lam's (2015) study, they found that some programs only offered elective language assessment courses, which means some pre-service teachers may not receive any instruction in assessment practices. Also similar to Lam's (2015) results, Mendoza and Arandia (2009) noted a teacher preference for summative, FA practices to formative, CBA approaches. From this study, the authors concluded that it is the responsibility of undergraduate and graduate education programs to instruct pre-service teachers in how to create, use, grade, and interpret language assessments, and they argued that in-service teachers should take responsibility for their own professional development. Though ideal, the latter suggestion may not be a practical way to improve teachers' LAL because without support from administrators, teachers may not be motivated to invest in professional development or be granted time off to attend assessment workshops or conferences. An additional similarity to Lam's (2015) research, Mendoza and Arandia's (2009) study is limited by the fact that it is not very generalizable because only 82 Colombian teachers were surveyed.

Like Mendoza and Arandia (2009), Vogt and Tsagari (2014) used a survey to research language teachers' testing and assessment needs. They collected data from 853 language instructors in seven European countries. The study was based upon a questionnaire originally created and administrated by Hasselgreen, Carlsen, and Helness (2004) as well as information gathered through 63 teacher interviews. Few changes were made to the Hasselgreen et al. (2004) questionnaire; however, Vogt and Tsagari (2014) purposefully left out questions regarding large-scale testing because they felt that large-scale testing practices such as item writing were not typically part of classroom teachers' responsibilities. This may have been an oversight since the authors mentioned that teachers are often required to create their own tests. The study revealed that, in general, teachers' LAL was not well-developed. Instructors were not able to critically evaluate tests, which led them to use prepared materials without consideration of their quality. To compensate for insufficient assessment education, teachers tended to learn on the job, often modeling mentors' and colleagues' assessment practices. A principal implication of this study is that although it appears that teachers in all the countries from which data were collected would benefit from professional development in assessment practices, teachers in some countries may need assessment education in different competencies and to different degrees than teachers in other countries. For example, German teachers highlighted the need to learn skills-based assessment, unlike teachers in other countries, and Greek teachers desired advanced assessment education in contrast to their German and Cypriot counterparts, who only felt the need for a moderate amount of professional development in assessment practices.

Priorities that emerge for the field of LAL in light of the studies in this section include, as Hasselgreen (2008) notes, the fact that assessment should be an integral part of pre-service and in-service teacher education and professional development. Unfortunately, research showed that many teachers have not received assessment education. If they have, in many cases, it has been inadequate. As Brown and Bailey (2008) pointed out, even if teachers receive sufficient education in FA practices, they may not receive adequate education in CBA techniques.

Resources Used to Develop Teachers' LAL

To help meet the demand for LAL education, many resources have been generated to increase teachers' LAL levels and help them expand their understanding of assessment. For example, through a two-year longitudinal study, Wall and Alderson (1993) investigated the

effects of a new, national Sri Lankan O-Level examination which was administered to grade 11 students. The new exam was created to support new methodology being presented in textbook materials and teacher education courses. The desired outcome was that the new test and textbook materials would have a positive washback effect on lesson content, teaching methods, and assessment techniques. By comparing data gathered from classroom observations, interviews, questionnaires, and materials with that of an analysis of baseline data, the authors concluded that the new test had some positive influences on the way Sri Lankan teachers designed tests. Changes that took place were that some teachers paid more attention to assessing reading and writing than grammar and that they tested students during class using item types from the new exam. At the same time, other teachers still copied passages and questions from past exam papers to prepare students for the new exam, and the introduction of the new exam did not change the way that teachers graded class tests and assignments. Therefore, a possible key takeaway from the study is that national exams in certain contexts have a limited influence on teachers' classroom assessment practices.

In order to more directly influence teachers' classroom assessment practices, assessment frameworks have been developed and published, some of which are required for use by government mandate. For instance, Breen et al. (1997) published the results of a qualitative study in which 25 Australian pre-primary to third grade teachers were interviewed about their use of government mandated frameworks. Findings showed that novice teachers found the frameworks useful, incorporating them into their instruction. This possibly demonstrates that assessment frameworks can be helpful tools for teachers' professional development. In contrast, experienced teachers tended to feel they were already practicing the assessment competencies being introduced through the frameworks. Therefore, this study indicates that novice teachers and experienced teachers have different LAL needs, and different professional development tools may be required to address those needs.

In Leung's (2004) study, the motivation to investigate assessment frameworks came from 10 ESL elementary school teachers who were involved in a voluntary, in-service professional development program. They were interested in discovering how assessment frameworks might assist them in diagnosing and addressing the English language development needs of their students. Leung (2004) found that some teachers used the frameworks to improve their teaching skills. Other teachers were not interested in using the frameworks in their classrooms, but they desired to improve their professional knowledge base in order to help bring about large-scale systemic change at their institution. Finally, some felt comfortable with officially provided assessment frameworks and how they used them, and they did not feel that they needed further education in their use. From studies on teachers' uses of and responses to published frameworks, we can see that frameworks can assist with the development of teachers' LAL; however, "there is no neat one-to-one correspondence between external stimulus or provision of new knowledge and the desired teacher uptake" (Leung, 2004, p. 38). Therefore, an implication of these studies is that supplying teachers with professional development tools and opportunities will not necessarily improve teachers' abilities to assess their students.

Examining two different resources used to improve teachers' LAL, Malone (2008, 2013) investigated language assessment textbook publications and web-based resource development. Malone's (2008) review centered on textbook changes from the 1967-2005. She noted that initially language assessment textbooks focused on educating standardized test writers (Harris, 1969) rather than on educating teachers. All textbooks reviewed included topics such as validity, reliability, item development, statistics, and practicality, but in the 1980s, texts began to

incorporate references to formative CBAs such as portfolios. Additionally, discussions of validity were expanded due to inclusion of new theory (e.g., Messick, 1989). In the 1990s, it was observed that teachers needed practical assessment instruction which included examples. This consideration led to publications which combined theory with practice (Bachman & Palmer, 1996; Genesee & Upshur, 1996; O'Malley & Valdez Pierce, 1996). Lastly, Malone (2008) highlighted how the volume *ESOL Tests and Testing* (Stoynoff & Chapelle, 2005) specifically addressed LAL and mentioned the importance of informed decision making, text contextualization, and appropriate test selection and use. This study draws attention to the fact that the field of language assessment is dynamic, and as the field grows and develops, it is important for current research to inform assessment educational tools, such as textbooks. In another study, Malone (2013) reported on feedback provided by 44 US language teachers and 30 language testers on a web-based assessment education tool (Center for Applied Linguistics, 2017) being developed by the Center for Applied Linguistics. Her study revealed a stark contrast between the assessment education priorities of testing experts, who emphasized precision and accuracy of assessment definitions and appropriate test use, and teachers, who highlighted the need for definitions to be clear and the web-based resources to be easy to use.

The Divide Between Testing Experts and Teachers

A brief discussion of the divide between testing experts and teachers referred to by Malone (2013) will conclude this section on problematizing LAL because bridging this divide is a frequently mentioned priority in LAL literature (Brindley, 2001; Brown & Bailey, 2008; Davies, 2008; Inbar-Lourie, 2008b, 2012; Jeong, 2013; Lam, 2015; Pill & Harding, 2013; Spolsky, 2008; Taylor, 2009). Spolsky (2008) believed the isolation of testing experts was a result of the field becoming highly technical, specialized, and sophisticated. Taylor (2009) observed that although it is beneficial to publicize ethical testing standards such as making corporate test development processes more transparent and ensuring test quality through psychometric analyses, doing so may promote the impression that language assessment is highly technical and better left to testing experts. Additionally, language assessment textbooks, often written by academics concerned with large scale testing and validation, can be difficult to understand because of their technical nature and, therefore, may not be accessible to teachers or seen as applicable by teachers to their professional context (Brown & Bailey, 2008; Taylor 2009). Suggestions on how to bridge the divide between testing experts and teachers in order to help nurture teacher LAL include promoting an assessment culture, which values different sources of data as equally significant for the advancement of learning goals, rather than promoting a traditional testing culture, which perceives internal and external tests as two separate, unrelated assessments (Inbar-Lourie, 2008b). Pill and Harding (2013) conjecture that if professional assessment organizations and individual assessment professionals became more visible by engaging the media through press releases, opinion pieces, blogs, etc., such activities would help increase public awareness of assessment related issues. Moreover, by actively and intentionally connecting with language teachers to provide them with information and professional development opportunities that are initially not too technical and are clearly applicable to teachers' instructional contexts, testing experts have an important role to play in improving teachers' LAL levels. Effort on the part of testing experts to personally connect with teachers can be a mutually beneficial relationship because many experts conduct classroom-

based research and write tests that measure classroom achievement. Personal connection with teachers gives researchers and test writers valuable insights into classroom assessment contexts and practices as well as educators' needs.

ASSESSMENT OF TEACHERS' LAL LEVELS

Problematizing LAL by discussing the issues, implications, and priorities that emerge from an investigation of empirical research has brought attention to the fact that many teachers lack the competence to accurately assess their students. To determine whether teachers possess the assessment competencies required to adequately carry out their instructional responsibilities, assessments such as certification tests and performance assessments have been constructed to measure teachers' LAL levels. This section discusses the measurement of the LAL levels of teachers working outside North America and the UK as well as pre-service preK-12 teachers preparing to work in the US and in-service preK-12 teachers working in the US.

Measuring Teachers' LAL Outside North America and the UK

Although there is not an entire test devoted to measuring teachers' LAL levels, there are language teacher certification tests that include a few questions about assessment. One such test is the Teaching Knowledge Test (TKT), which was designed for teachers working outside North America and the UK by Cambridge University Press (CUP; 2017a). The TKT is a series of multiple-choice modular exams. There are three core modules: (1) *language and background to language learning and teaching*, (2) *lesson planning and use of resources for language teaching*, and (3) *managing the teaching and learning process*, and there are two specialist modules: *content and language integrated learning (CLIL)* and *young learners (YL)*; CUP, 2017a). The test is based on an operationalization of the Cambridge English Teaching Framework (Cambridge English Language Assessment, 2015) which includes *assessing language learning*. This component of the test construct is defined as "the conscious understanding of the principles of assessment as well as the necessary skills to design, mark, and give feedback on effective tests" (Cambridge English Language Assessment, 2015, p. 3). An examination of the sample practice tests (CUP, 2017b) indicates that assessment test items measure a teacher's ability to mark and give feedback on tests, and it could be argued there are items that very indirectly assess a teacher's ability to design tests; however, there is no evidence that teachers are tested on assessment principles. Thirteen percent of modules one and two and 16 percent of module three are dedicated to assessing teachers' LAL levels. In addition to sections that focus on knowledge of different types and methods of classroom-based assessment, the first three TKT modules also have a few supplemental assessment items on error correction, feedback, and knowledge of self- and peer assessment. Twenty-five percent of the YL module is dedicated to assessment related items. In contrast, just five percent of the CLIL module measures teachers' LAL levels. The inclusion of a few assessment items on each modular exam is a step toward calling attention to the importance of teacher LAL. Nevertheless, it is difficult to make claims about the level of teachers' LAL based on data gathered from so few questions, and it is unfortunate that all the elements of the *assessing language learning* component of the construct were not assessed.

Like the TKT, the online ELTeach certification program was created for teachers working outside North America and the UK. It was designed by Educational Testing Service (ETS) in collaboration with National Geographic Learning to help meet the high international demand for English language teaching professional development opportunities (Young, Freeman, Hauck, Gomez, & Papageorgiou, 2014). It is composed of two coursework modules: English-for-Teaching, which covers functional classroom English, and Professional Knowledge for ELT, which covers pedagogical instruction. The information taught in the English-for-Teaching module is assessed by the Test of English for Teaching (TEFT). According to the design framework presented by Young et al. (2014), one of three essential tasks assessed by the TEFT is *providing oral and written feedback*. According to the “Representative Examples of English-for-Teaching Teacher Tasks and Associated Claims” table of the ELTeach design framework report (Young et al., 2014, p. 21), examinees are required to provide written feedback on a student text. However, neither the “Representative Examples” table nor the rest of the article explains how oral feedback is tested. Information taught in the Professional Knowledge for ELT module is assessed by the Test of Professional Knowledge for ELT (TPK). No “Representative Examples” table is provided for the TPK nor is there information on task types for the TPK elsewhere in the article, so it is uncertain as to how this information is tested. At the same time, the ELTeach Professional Knowledge curriculum claims to cover “basic knowledge of...assessment” (Young et al., 2014, p. 26), such as how information regarding students’ performances is gathered and interpreted and how this information is used by the instructor to promote language development. The TPK purports to focus on connecting course content to teaching practice; therefore, it is expected that at least a few items on the TPK are dedicated to assessing teachers’ LAL levels. Nevertheless, the sole mention in the Young et al. (2014) article of the assessment of teachers’ LAL levels on the TPK is in the rubric (see Table 2). Only examinees whose scores fall in the highest rubric level, Band Three, can explicitly be said to have “knowledge about...assessment” (p. 27). Therefore, although assessment is one of the four core teaching practices which make up the TEFT and TPK constructs, operationally it appears that one can only make claims regarding a teacher’s ability to provide written feedback.

TABLE 2
Test of Professional Knowledge Rubric (Young et al., 2014, p. 27)

| BAND DESCRIPTORS |
|---|
| Band Three (Range 260-280): A typical test taker in Band Three demonstrates comprehensive knowledge of all content areas in the curriculum. The test taker in Band Three consistently identifies concepts and connects them to examples of English-language teaching situations presented in the Professional Knowledge program. In addition to the demonstrated knowledge of test takers in Bands One and Two, test takers in Band Three may have <ul style="list-style-type: none">• more knowledge about the application of concepts presented in the curriculum and• more knowledge about teaching techniques, classroom activities, and <i>assessment</i>. (emphasis added) |
| Band Two (Range 215-245): A typical test taker in Band Two demonstrates substantial knowledge of all content areas in the curriculum. The test taker in Band Two usually |

identifies concepts and **often** connects them to examples of English-language teaching situations presented in the Professional Knowledge program.

In addition to the demonstrated knowledge of test takers in Band One, test takers in Band Two may have

- more knowledge of the application of concepts presented in the curriculum and
 - additional knowledge in some other areas of the curriculum.
-

Band One (Range 155-195): A typical test taker in Band One demonstrates knowledge of some of the content areas in the curriculum. The test taker in Band One **sometimes** identifies concepts and **occasionally** connects them to English-language teaching situations presented in the Professional Knowledge program.

Test takers in Band One may have

- some understanding of the application of concepts presented in the curriculum and
 - more knowledge of some areas of the curriculum than others.
-

Measuring Pre-Service Teachers' LAL in the US

Thus far the LAL level assessments discussed were designed for teachers working outside North America and the UK. The following section focuses on the measurement of LAL levels of pre-service preK-12 Speakers of Other Languages (ESOL) teachers preparing to work in the US. In order to receive teaching licensure, many states require passing test scores on Praxis teacher educator tests. According to ETS (2017), Praxis tests assess the academic and content specific knowledge and skills required for teaching, and successful test completion is necessary for preK-12 teaching certification in many states in the US. In both versions (5361 and 5362) of the content-specific ESOL exam, the assessment section comprises 18 of 120 test items, that is, 15 percent of the test. The areas of assessment which both tests claim to cover are FA, CBA, assessment development, test use for decision making purposes, validity, reliability, language and cultural bias, state standards, accommodation for English language learners (ELLs), and communication of assessment results to ELLs and their guardians (ETS, 2016a, 2016b). Test 5361 also addresses national standards and self- and peer-assessment techniques, and Test 5362 includes teachers' abilities to adapt classroom assessments for ELLs. From the description of the many aspects that these tests profess to consider, it appears that the assessments provide a multi-faceted picture of a teacher's LAL level. However, even if the tests are able to address all the aspects of assessment which they purport to in 18 items, it is extremely unlikely that they are capable of doing so in much depth.

Continuing to focus on state specific certification requirements for pre-service ESOL teachers, the author reviewed the demands of her state of residence, New York (NY). There are three assessments pre-service teachers must pass in NY: (1) the Educating All Students (EAS) test, (2) the Content Specialty Test (CST), and (3) the edTPA. As described in one of its field reports (New York State Teacher Certification Examinations, New York State Education Department, 2014), one of five sections of the EAS specifically tests ELL related concerns. Only one of nine ELL test construct elements mentions a need for examinees to demonstrate knowledge of assessment practices such as the ability to select, create, evaluate, and adjust assessment practices to meet learners' needs and contribute to the achievement of learning standards (New York State Teacher Certification Examinations, 2014). Since so little of the EAS

is dedicated to assessing LAL, it is unlikely that the test provides much, if any, information about an examinee's LAL level.

The second test required for NY preK-12 teaching certification is the ESOL CST, which has an *Instructional Planning, Practices, and Assessment* section composed of 15 items or 13 percent of the test (New York State Teacher Certification Examinations, 2015). This section supposedly assesses 17 test construct elements, nine of which are assessment specific. The nine assessment construct elements are similar to those described by the Praxis ESOL tests, though they are more explicit (e.g., the Praxis test mentions the need to know state mandates, but the CST articulates which specific regulations and laws examinees are required to know). Like the Praxis test, the inclusion of at least a few assessment items in this section of the CST is commendable, but it seems unlikely, due to the limited number of test items, that a satisfactory measure of pre-service teachers' LAL levels could be ascertained. The last section of the CST, *Analysis, Synthesis, and Application*, comprises 20 percent of the total test score. In this section, examinees interpret and apply large scale testing data and classroom-based, informal assessment data (e.g., antidotal records) to identify students' strengths and weaknesses and describe how they would differentiate instruction for learners based on the data. This constructed response item has the potential to adequately measure how well examinees can interpret and use assessment results. Overall, there are a variety of assessment competencies that the EAS and CST supposedly address, but because there are few assessment-focused items, it appears unlikely that the EAS and CST are able to test examinees in such a way that comprehensively measures their LAL level.

The edTPA is the third assessment which NY requires for preK-12 ESOL teaching certification (Stanford Center for Assessment, Learning, and Equity, 2016). It is a performance assessment intended to evaluate whether pre-service teachers are prepared to adequately deal with the demands of full-time instruction. About a third of the edTPA is dedicated to assessing examinees' classroom assessment capabilities. The edTPA requires that examinees select, adapt, or design assessments that are aligned with the learning objectives of a mini-unit. Assessments created by edTPA examinees should focus on evaluating preK-12 students' abilities to explain and synthesize information. A positive aspect of the edTPA is that pre-service teachers must apply their knowledge of CBA to successfully complete the examination task. Nonetheless, it is limited by the fact that creating assessments for one mini-unit is not a satisfactory demonstration of the full gamut of assessment competencies necessary to adequately measure an examinee's LAL level.

Measuring In-Service Teachers' LAL in the US

Having discussed the assessment of pre-service teachers' LAL levels, the focus now shifts to the assessment of in-service preK-12 ESOL teachers' LAL levels. National Board Certification (National Board for Professional Teaching Standards; n.d.) provides preK-12 institutions located in the US with the opportunity to recognize, develop, and retain skilled teachers and generate ongoing institutional improvement. The certification process includes: (1) content knowledge assessment, (2) reflection on student performance, (3) video analysis of instructional practice, and (4) record of assessment impact (National Board for Professional Teaching Standards, 2017). The first component is assessed through 45 selected response items; 15 percent of the test assesses LAL competencies. The second component is assessed through

three constructed response items. One sample constructed response item asked examinees to analyze two student assessments and then create a unit plan based on learner needs. As with previous tests of teachers' assessment competency, the limited percentage of test items dedicated to assessment on the first two NBC components are unlikely to produce adequate data to make reliable decisions about teachers' LAL levels. The third component involves examinees showing how assessment is a continual part of the instructional process and includes the provision of regular, timely, and constructive feedback. The fourth component requires examinees to design assessments that gather information from multiple sources and use that information to beneficially impact learning. Designed assessments need to illustrate how examinees collaborated with colleagues, families, and the community to promote student growth and learning. Additionally, the assessments must demonstrate that examinees can assess the five language domains: reading, writing, speaking, listening, and visual literacy (e.g., the ability to make predictions about stories based on pictures). Furthermore, examinees are expected to be able to give feedback, instruct students how to self-assess to advance their learning, gather and analyze standardized testing data, and be competent in assessing for special purposes (e.g., students who are academically at risk). A strength of this fourth NBC component is its emphasis on the learner—all assessment practices are performed for the benefit of student advancement. Additionally, it appears to elicit many of the LAL competencies highlighted in the conceptualizations section, such as theoretical knowledge, technical skills, language pedagogy, understanding of local practices, and decision making. However, it does not seem to address historical and political philosophical frameworks or personal beliefs/attitudes. Still, of the certification tests and performance assessments reviewed, the NBC appears to provide the most accurate measure of teachers' LAL levels.

CONCLUSION

In order to review the literature on teachers' LAL, this paper reported on LAL definitions and conceptualizations. It then problematized LAL through an investigation of empirical studies conducted to explore pre- and in-service teacher education and resources that have been used to develop teachers' LAL. Finally, it examined how teachers' LAL levels are assessed. A limitation of this paper is that the literature upon which it is based is drawn from different countries and institutions. As previously mentioned, since teachers' LAL needs vary according to context (Davison, 2004; Vogt & Tsagari, 2014), it is difficult to generalize how LAL should be defined, conceptualized, problematized, and assessed. Nevertheless, it is critical for the language assessment community to better operationalize LAL conceptualizations through detailed, for-learning model specifications. It is also essential to address issues, implications, and priorities that emerge from LAL empirical research and to develop more comprehensive, accurate ways to assess teachers' LAL levels. Improving how LAL is assessed would help teachers' supervisors to (1) identify strengths and gaps in teachers' LAL and (2) provide professional development opportunities to address the identified limitations. Moreover, teachers themselves may not be aware of their lack of assessment knowledge. A well-developed LAL assessment instrument would draw teachers' attention to ways that they could improve their practices. Ultimately, the goal of developing more comprehensive, accurate ways to assess teachers' LAL levels is to empower teachers to become better assessors and thus to equip them to become better teachers

because “student assessment is an essential part of teaching and...good teaching cannot exist without good student assessment” (AFT, NCME, & NEA, 1990, p. 2).

REFERENCES

- American Federation of Teachers, National Council on Measurement in Education and National Education Association. (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30-32.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1-42.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.
- Bailey, K. M., & 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.
- Breen, M. P., Barratt-Pugh, C., Derewianka, B., House, H., Hudson, C., Lumley, T., & Rohl, M. (Eds.). (1997). *Profiling ESL children: How teachers interpret and use national and state assessment frameworks*. Canberra, Australia: Department of Employment, Education, Training and Youth Affairs.
- Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O’Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126–36). Cambridge, UK: Cambridge University Press.
- Brown, J. D., & Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing*, 25(3), 349–83.
- Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices*. Portsmouth, NH: Heinemann.
- Cambridge English Language Assessment. (2015). *Cambridge English teaching framework: How and why the framework was developed*. Cambridge, UK: UCLES/Cambridge University Press.
- Cambridge University Press. (2017a). *TKT: About the tests* [Informational website]. Retrieved from <http://www.cambridgeenglish.org/teaching-english/teaching-qualifications/tkt/about-tkt/>
- Cambridge University Press. (2017b). *TKT: How to prepare* [Informational website]. Retrieved from <http://www.cambridgeenglish.org/teaching-english/teaching-qualifications/tkt/prepare-for-tkt/>
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Center for Applied Linguistics. (2017). *Understanding assessment: A guide for foreign language educators*. [Educational website]. Retrieved from <http://www.cal.org/flad/tutorial/>
- Cheng, L. (2001). An investigation of ESL/EFL teachers’ classroom assessment practices. *Language Testing Update*, 29, 53-64.
- 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.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305-334.
- Educational Testing Service. (2016a). *The Praxis study companion: English to speakers of other languages* (5361). [Test preparation guide]. Retrieved from <https://www.ets.org/praxis/prepare/materials/5361>
- Educational Testing Service. (2016b). *The Praxis study companion: English to speakers of other languages* (5362). [Test preparation guide]. Retrieved from <https://www.ets.org/praxis/prepare/materials/5362>
- Educational Testing Service. (2017). *Praxis*. [Informational website]. Retrieved from <https://www.ets.org/praxis>
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London, UK: Routledge.
- Genesee, F., & Upshur, J. (1996). *Classroom-based evaluation in second language education*. New York, NY: Cambridge University Press.
- Grabowski, K. C., & Dakin, J. W. (2014). Test development literacy. In A. J. Kunnan (Ed.), *Companion to language assessment* (pp. 869-889). Oxford, UK: Wiley-Blackwell.
- Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbooks of applied linguistics [HAL]: Handbook of second language assessment* (pp. 413-427). Berlin, Germany/Boston, MA: De Gruyter Mouton.
- Harris, D. (1969). *Testing English as a second language*. New York, NY: McGraw-Hill.
- Hasselgreen, A. (2008). *Literacy in classroom assessment (CA): What does this involve?* Paper presented at the 5th Annual Conference of EALTA. Athens, Greece. [Powerpoint slides]. Retrieved from <http://www.ealta.eu.org/conferences/2008/docs/sunday/panel/Literacy%20in%20classroom%20assessment.pdf>
- Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European survey language testing and assessment needs. Part one: General findings*. [Report]. Retrieved from <http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf>
- Inbar-Lourie, O. (2008a). Constructing an assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402.
- Inbar-Lourie, O. (2008b). Language assessment culture. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education. Vol. 7: Language testing and assessment* (2nd ed., pp. 285-300). New York, NY: Springer.
- Inbar-Lourie, O. (2012). Language assessment literacy. In C. A. Chapelle (Ed.) *The Encyclopedia of Applied Linguistics*. Oxford, UK: Wiley-Blackwell.
- Inbar-Lourie, O. (2013). Language assessment literacy. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-9). Oxford, UK: Blackwell Publishing Ltd.

- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, 30(3), 345–362.
- Kaiser, G., & Willander, T. (2005). Development of mathematical literacy: Results of an empirical study. *Teaching mathematics and its applications*, 24(2-3), 48-60.
- Klinger, D., McDivitt, P., Howard, B., Rogers, T., Munoz, M., & Wylie, C. (2015). *Classroom assessment standards for preK-12 teachers*. Joint Committee on Standards for Educational Evaluation.
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. London, UK: Longman.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197.
- Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice, and change. *Language Assessment Quarterly*, 1(1), 19–41.
- Malone, M. E. (2008). Training in language assessment. In E. Shohamy (Ed.). *Language testing and assessment* (Vol. 7). In N. Hornberger (general editor), *Encyclopedia of language and education* (2nd ed). New York: Springer Science and Business Media, Inc., pp. 225-240.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329–344.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Hoboken, NJ: John Wiley & Sons.
- Mendoza, A. A. L., & Arandia, R. B. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *PROFILE*, 11(2), 55-70.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: Macmillan.
- National Board for Professional Teaching Standards. (n.d.). *About certification*. [Informational website]. Retrieved from <http://www.boardcertifiedteachers.org/about-certification>
- National Board for Professional Teaching Standards. (2017). *Guide to National Board Certification* (Version 1.9). National Board for Professional Teaching Standards.
- New York State Teacher Certification Examinations, New York State Education Department. (2014). *Field 201: Educating all students (EAS) test design and framework*. White Plains, NY: New York State Education Department and Pearson.
- New York State Teacher Certification Examinations. New York State Education Department. (2015). *Field 116: English to speakers of other languages test design and framework*. White Plains, NY: New York State Education Department and Pearson.
- O’Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363–380.
- O’Malley, J. M., & Valdez Pierce, L. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. Reading, MA: Addison-Wesley Publishing Company.
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381–402.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4–11.
- Purpura, J. E. (2004). *Assessing grammar*. Cambridge, UK: Cambridge University Press.
- Purpura, J. E. (2009). The impact of large-scale and classroom-based language assessment on the individual. In L. Taylor & C. J. Weir (Eds.), *Language testing matters: Investigating the*

- wider social and educational impact of assessment—Proceedings of the ALTE Cambridge Conference, April 2008* (pp. 301-325). Cambridge, UK: Cambridge University Press.
- Purpura, J. E. (2016). Second and foreign language assessment. *Modern Language Journal*, 100 (Supplement 2016), 290-208. (Centennial Issue)
- Purpura, J. E., Brown, J. D., & Schoonen, R. (2015). Improving the validity of quantitative measures in applied linguistics research. *Language Learning*, 65(S1), 37–75.
- 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.
- Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, 24(4), 331–345.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow, UK: Pearson Education.
- Spolsky, B. (2008). Language testing at 25: Maturity and responsibility? *Language Testing*, 25(3), 297-305.
- Stanford Center for Assessment, Learning, and Equity. (2016). Making good choices: A support guide for edTPA candidates. Palo Alto, CA: Board of Trustees of Leland Junior University.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-539.
- Stoynoff, S., & Chapelle, C. A. (2005). *ESOL tests and testing: A resource for teachers and program administrators*. Alexandria, VA: TESOL Publications.
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412.
- Turner, C. E., & Purpura, J. E. (2015). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 255-273). Berlin, Germany/Boston, MA: DeGruyter Mouton.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402.
- Wall, D., & Alderson, C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41-69.
- Young, J. W., Freeman, D., Hauck, M. C., Gomez, P. G., & Papageorgiou, S. (2014). A design framework for the ELTeach program assessments. Princeton, NJ: Educational Testing Service.
- Zaleski, D. (2015). *Guiding principles for classroom assessment*. Springfield, IL: Illinois Board of Education.