
Training in Language Assessment

Margaret E. Malone

Abstract

This chapter examines the major issues involved in providing appropriate training and professional development for language instructors to improve their knowledge and skills and make informed decisions throughout all aspects of the assessment process. First, the chapter reviews the major approaches to conducting language assessment within the context of educational policies and societal beliefs over time. In reflecting on the changing contexts and approaches, the chapter identifies the underlying philosophies of training in assessment and how such philosophies align with approaches to both assessment practice and how teachers have learned how to conduct assessment. The chapter also investigates different approaches to training. Traditional language testing textbooks and their content, as well as the recent increase in the availability of such textbooks, are highlighted. The chapter then turns from traditional textbooks to traditional professional development and addresses ways that such professional development has changed, including the availability of distance learning and other online resources, as well as how such approaches provide opportunities for innovation and improved understanding of language assessment. Finally, the chapter addresses the ongoing challenges in training for teacher professional development, including lack of resources and frequent lack of agreement between language testers and language teachers regarding the most essential topics for teachers to understand in learning about, developing, selecting, and using language assessments with their students.

M.E. Malone (✉)

Assessment and Evaluation Language Resource Center, Department of Linguistics, Georgetown University, Washington, DC, USA

e-mail: malonem@georgetown.edu; mmalone@cal.org; meg@cal.org

Keywords

 Language assessment literacy • Teacher professional development • Assessment

Contents

Introduction	226
Early Contributors	227
Current Trends	229
Major Contributors	229
Text-Based Materials	230
Non-Text-Based Assessment Training	234
Works in Progress	236
Problems and Difficulties	236
Future Directions	237
Cross-References	238
Related Articles in the Encyclopedia of Language and Education	238
References	238

Introduction

This chapter updates and addresses some of the major issues in training language instructors to make informed decisions in all aspects of the assessment process; in this context, the “assessment process” refers to developing, scoring, interpreting, and improving classroom-based assessments developed by language instructors as well as selecting, interpreting, and sharing results of large-scale tests developed by professional testing organizations (Stoyhoff and Chapelle 2005; Bachman and Palmer 1996). Within the context of providing training in language assessment, this chapter explores “language assessment literacy” (Taylor 2013; Inbar-Louie et al. 2013; Malone 2013; Stiggins 1997; Stoyhoff and Chapelle 2005; Boyles 2005), discusses expanded definitions of assessment literacy, and reviews the available resources for training in language assessment, as well as work that still needs to be done.

As pressure for language instructors and educational institutions to provide information on students’ progress has increased since the 1880s and skyrocketed in the past decade (Llosa 2011; Brindley 1997), attention has focused on the testing that takes place within the context of language teaching and learning. The 2001 passage of *No Child Left Behind* (NCLB) in the United States mandates annual assessment of the English language proficiency of all English language learners enrolled in elementary and secondary programs and emphasizes the need to track and monitor student outcomes and progress in both English language and content areas (Alicea 2005). Although Europe and other countries do not mandate the use of the Common European Framework of Reference for Languages (CEFR), in that member nations are not required to adopt it or its aligned tests, by emphasizing language teaching and learning (Little 2012), the CEFR exerts great influence on the teaching and assessment of language (Davies et al. 1999) in Europe and beyond, thus demonstrating one way that language assessment has increased in importance in many places in the world.

Despite the growth of standards-based education, standards for teacher certification, and an increase in tests administered, there is no clear framework of what is required or even needed for language instructors to reliably and validly develop, select, use, and interpret tests or the extent to which these standards are used for classroom assessment (Llosa 2011). Therefore, the issue is how to identify the best approaches for support and training for those who “have to do the real work of language teaching” (Carroll 1991, p. 26) when they assess their students.

In addition to the practical and pedagogical concerns about teacher assessment knowledge and skills, the political arena also influences how, when, and why students are assessed. With the arrival of NCLB in the United States and the CEFR in Europe and beyond, assessment of language learners’ progress has only strengthened in political, practical, and pedagogical importance. This chapter examines how the underlying philosophies of training in assessment have changed over time, in response to societal and educational changes in policy and practice. It also examines how different approaches for training in language assessment, from textbooks to distance learning, have altered such training. Finally, it examines ongoing challenges and future directions for increasing the “assessment literacy” of language instructors for the improvement of language learning and teaching.

Early Contributors

Like education, language assessment is a microcosm of what is happening in larger society. This part of the chapter describes the three early periods of language testing (1800s–1980s) and discusses how each period’s philosophies were reflected in available assessment training. Spolsky (1977) has divided language testing from the 1800s through the 1980s into three major periods: prescientific, psychometric, and sociolinguistic.¹ The prescientific approach, as practiced in the United States and Europe, relied mainly on the judgments of instructors as they assessed a translation, composition, or oral performance or another open-ended task presented to students. The very term “prescientific” judges this approach “unscientific”; the lack of science as applied to language testing during this period resulted in debates as to the reliability of written and oral exams administered to large groups of students and rated by different instructors. The literature does not reveal any systematized, required training for instructors on how to develop the questions for these tests, guidelines for rating the test results, or available training for the instructors in rating the examination performances.² As far back as 1888, debates ensued as to the reliability of these written (or oral) exams, administered to large groups of students and rated by different instructors with varying understanding of expected outcomes

¹Spolsky (1981), Barnwell (1996), and others have alternative names for these periods; this paper uses the original terms.

²While some large-scale tests for admittance to universities or professions included oversight by committees, there is no evidence of such oversight for classroom assessment.

(Spolsky 1995). Despite these criticisms, it is important to note that such exams, including professional exams for admittance to, for example, the Indian Civil Service Exam supplemented patronage for candidates to the civil service. In other words, early language tests, though their developers and raters may have lacked rigorous formal training in language assessment, were often viewed as a more democratic means of admitting students to university and the workplace than simply using personal connections (Spolsky 1995).

By contrast, the second period, termed as the psychometric period, emphasized statistics and measurement and moved away from open-ended test questions to test items focusing on discrete aspects of language, such as vocabulary, grammar, pronunciation, and spelling. The format for testing also changed from the first to the second period, while in the prescientific period, students may have responded to prompts for a written essay or oral response and test items in the psychometric period included more, but shorter, questions. It was at this time that item types such as multiple choice, true/false, and similar short questions gained popularity in testing. The popularity of this approach was thus reflected in course offerings at institutions of higher education; Jonic (1968), as cited by Spolsky (1995), reports that, by 1920, courses in educational measurement were being offered by most US state universities, although such educational measurement approaches had not yet spread to language learning.

Therefore, the shift from fewer test items with long responses that took time to score to more test items with short, easy to score test items, was underway. While this new phase in language testing addressed some of the criticisms of the prescientific phase, it introduced new challenges. Despite Jonic's (1968) reference to the development and availability of educational measurement courses, there is no indication that such courses were uniformly required of teachers; therefore, the change was not accompanied by a similar change in approach to language testing courses. During this period, the work of testing and teaching was divided; testing organizations developed large-scale tests to measure student progress, and teachers provided instruction to students (Stoynoff and Chappelle 2005). Therefore, a gulf developed between instructors and test developers.

By the 1970s, changes in society, educational measurement, and theories of language learning resulted in a shift toward the sociolinguistic period.³ During this period, the focus shifted from discrete-point testing toward tests to measure meaningful communication (Ommagio 1986). A great deal of literature is devoted to how language instructors should (and should not) be trained to assess according to variations of this approach (Bachman and Savignon 1986; Lantolf and Frawley 1985). One of the most popular approaches to assessing communicative competence during this period in the United States was the *ACTFL Proficiency Guidelines*, while later in Europe, work began on what would become the *Common European Framework of Reference*. By the early 1980s, training in various approaches to assessing

³Canale and Swain (1982) and others refer to this as "communicative competence" or "the proficiency approach" (Barnwell 1996).

communicative competence became available, and language instructors could seek and receive training in various approaches. As this period in testing spread into the 1980s, educational reform in the United States and efforts by the Council of Europe to reform language teaching prodded the sociolinguistic movement toward measuring outcomes based on shared standards for language learning (Stoynoff and Chapelle 2005).

However, the gap in skills held by teachers and test developers that developed during the psychometric period tightened during the sociolinguistic period and narrowed further with the introduction and incorporation of standards in the language classroom. With the 1980s and 1990s, a new era of language testing, with roots in the education reform movements in Europe and the United States, emerged.

Current Trends

Spolsky (1995) and others have described thoroughly the three early periods in modern language testing. Following and overlapping the sociolinguistic period, the literature shows an increased emphasis on authentic, performance (or outcomes-based) assessment to reflect what students need to do with the language in real-life settings (Wiggins 1994) as well as an increased importance on shared, common standards with which to assess students. During this time, methods of collecting information from students gained popularity, such as portfolios of student work and student self-assessment, and increased emphasis on the authenticity of the task the student was to perform with respect to language use in daily life (Moore 1994). In the 2000s, emphasis on testing, including language testing, has steadily increased. The release of the CEFR in Europe and beyond and the passage of NCLB, as well as the introduction of the Common Core State Standards Initiative in the United States, have only magnified the importance of testing worldwide. The connection between assessment, standards, and politics highlights the importance of training language instructors so that they can adequately assess their students' progress toward local, national, and/or international goals and standards.

Major Contributors

Any history of language testing will readily name a number of influences on language assessment; it is more difficult to pinpoint at what point changes in the language testing arena begin to influence the pre- and in-service training of classroom teachers because of the gradual nature of the change. The impetus for the three periods described in the previous section began with primarily large-scale assessments, such as admission to university and professions; the rate at which results and lessons learned from large-scale assessments trickle down to instructors and into preservice teacher texts is unclear and undocumented. This emphasis is reflected not only in the volume of assessments available throughout the world but also in the number of texts available for training instructors in assessment. Reviewing the three

periods is important to contextualize how training for language assessment has evolved over the past two centuries. During the prescientific period, the assessment role fell largely on individual instructors, while during the psychometric period, test development was largely in the hands of expert psychometricians, and thus language teachers did not receive much, if any, training in language test development. However, the sociolinguistic period represented a time when language teachers began to become increasingly involved in language testing. The impact of the sociolinguistic period is evidenced by the titles and content of texts developed on language testing over a 40-year period. In this section, I will address two major contributions to training in language assessment: traditional text-based materials and technology-mediated materials and information that became available in the 1990s and beyond.

Text-Based Materials

There are several ways to examine language testing textbooks, including length, content, and quantity of available textbooks. Cohen (1994) references seven other textbooks on language testing available at the time of printing and points out that there were not as many available in the edition published 15 years earlier. This gap shows the crux of the issue of training in language assessment; during the psychometric period, “large-scale standardized instruments [were] prepared by professional testing services to assist institutions in the selection, placement and evaluation of students” (Harris 1969, p. 1), and the focus was on training professionals to develop items for standardized tests rather than training language instructors to assess their students. Examining the bibliographies of over 560 language testing texts, the author initially selected ten published from 1967 to 2005 to contrast on page lengths and number of citations listed in Google Scholar and then three more published or revised from 2005 onward. Table 1 shows these results.

While this table includes only a very small sample of textbooks available in language testing from the late 1960s until present, it shows differences and similarities over time. For example, while Valette and Harris were contemporaries, the lengths of their textbooks were different, and Valette had nearly four times as many references as Harris. In 2005, Harris has twice as many citations on Google Scholar as Valette; 10 years later, his Google Scholar citations dwarf hers. In addition to the contrasts between specific texts, there are definite changes over time. First, text length increased over time, as knowledge about language testing grew, and, similarly, the number of references included in texts increased. It is also interesting to note the contrast between the number of Google Scholar citations for each text in 2005 and 10 years later is remarkable. This growth first speaks to the increased power of the Internet in general and Google Scholar in particular of tracking citations and secondly shows how much more frequently all sources are cited even 10 years later. Table 1 also shows how the numbers of pages and the number of references have increased over time

Table 1 Distinctions in page lengths and number of references in language testing books

Author/text	Date of publication	Page length	Number of references	Citations on Google Scholar (2005)	Citations on Google Scholar (2015)
Harris, D. <i>Testing English as a Second Language</i>	1969	146	7	40	679
Valette, R. <i>Directions in Foreign Language testing</i>	1967	200	26	18	21
Oller, J. W. <i>Language Tests at School</i>	1979	421	370	140	1,209
Cohen, A.D. <i>Testing Language Ability in the Classroom</i>	1980	132	172	56	153
Henning, G. <i>A Guide to Language Testing</i>	1987	158	117	37	506
Hughes <i>Testing for Language teachers</i>	2003	154	66	343	3,030
Bachman, L. <i>Fundamental Considerations in Language testing</i>	1990	359	751	751	6,477
Weir, C. <i>Understanding and Developing language tests</i>	1995	170	83	65	583
Brown, H.D. <i>Language Assessment: Principles and Classroom Practice</i>	2004	160	302	9	1,511
Stoynoff and Chapelle <i>ESOL Tests and Testing: A Resource for Teachers and Program Administrators</i>	2005	204		1	45
Bachman and Palmer. <i>Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World</i>	2010	493	193	n/a	281
Fulcher, G. <i>Practical Language Testing</i>	2010	346	377	n/a	97
Carr, N. <i>Designing and Analyzing Language Tests</i>	2011	361 (plus CD appendix)		n/a	31

In preparing this chapter, the author examined over 100 language testing publications, including books, articles in peer-reviewed journals, and guidelines. In addition to the gap between page length and number of citations that exists between various texts, there is also a difference between earlier and later editions of texts, as

Table 2 Changes in Hughes', Cohen's, and Bachman and Palmer's textbooks

Author and book title	Date of publication	Page length	Number of references
Hughes <i>Testing for Language Teachers</i>	1989	154	66
Hughes <i>Testing for Language Teachers</i>	2004	217	186
Cohen <i>Testing Language Ability in the Classroom</i>	1980	132	172
Cohen <i>Testing Language Ability in the Classroom</i>	1994	362	433
Bachman and Palmer <i>Language Testing in Practice: Designing and Developing Useful Language Tests</i>	1996	370	88
Bachman and Palmer <i>Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World</i>	2010	493	193

Cohen points out. Therefore, Table 2 shows the differences in Hughes', Cohen's, and Bachman and Palmer's textbooks over time.

The differences in length and references mirror additions of content to the text. While all texts referenced above include steady reminders of reliability, validity, and practicality, the 1990 and onward versions include more references to assessments such as portfolios and other practices that became widespread in the 1980s. In addition, Hughes added a chapter on assessing children because of the increased emphasis on testing this age group (Hughes 2004). Cohen (1994) and Bachman and Palmer (2010) more than doubled the number of references, suggesting that teachers required more information in the 15 years that passed between publications. Bachman and Palmer also adapted their title from *Language Testing in Practice: Designing and Developing Useful Language Tests* to *Language Testing in Practice: Developing Language Assessments and Justifying their Use in the Real World*. The shift in title shows the emphasis on assessment rather than testing and the growing emphasis of "real-world" use of assessment. As assessments change, the textbooks used in teacher training must change as well.

Just as new language testing textbooks began focusing on classroom teachers' practical needs, additional text-based resources emerged in the 1990s and have continued to be used in the field. While early textbooks often combined theoretical explanations with samples from actual assessment practices, the 1990s saw an explosion of textbooks that could supplement existing ones by supplying examples that could be readily included in the classroom or a "how to" on classroom assessment.

O'Malley and Valdez Pierce's (1996) *Authentic Assessment for English Language Learners: Practical Approaches for Teachers* represented a new approach to language testing textbooks; it combines theory and practice in an accessible volume for classroom teachers. Its rubrics, checklists, and practical advice on applications can easily be incorporated into the classroom. At a similar time, Brown (1998) produced

a volume with 18 different activities with input from three to eight international contributors for each activity type. Others in language testing also worked to model and explain solid theories of language testing coupled with practice; Bachman and Palmer (1996) and Genesee and Upshur (1996) published textbooks on language testing with an emphasis in both their titles and tone toward classroom teacher use. Unlike traditional language testing textbooks, both volumes emphasized the specific issues and problems faced by classroom teachers and aimed to combine a theoretically strong approach to language testing with practical help. For example, Genesee and Upshur (1996) include conferencing and portfolios, both approaches that gained popularity in the 1990s, as well as tables that describe the benefits of portfolios.

In the spirit of combining the information of a language testing textbook and the practicality of a “how-to” manual for teachers, Davidson and Lynch (2002) have produced *Testcraft: A Teacher’s Guide to Writing and Using Language Tests*. Their approach emphasizes the importance of developing solid test specifications based on language testing research. At the same time, they tackle practical issues of teamwork in the test development process and ways to approach inevitable conflicts, as well as including scenarios applicable to situations their readers may encounter. Few language testing texts address the importance of teamwork and the challenges inherent in working with colleagues who have differing viewpoints about the purposes and uses of the test as well as suggest approaches for addressing not just the content of such issues but also working with colleagues.

Stoynoff and Chapelle (2005) published *ESOL Tests and Testing*, a volume which includes reviews of common English language tests, as well as chapters on the “basics” that language instructors should know before using any test. Stoynoff and Chapelle stress the importance of making informed decisions in all aspects of the testing process, and the structure of the volume supports this approach. The reviews are embedded in the book, rather than appearing at the beginning or the end, and this sequence emphasizes the importance of contextualization in test selection. This volume points to the issue of “assessment literacy” in language instructors and the need to provide practical and usable resources to language instructors to ensure that tests are selected and used properly.

Bachman and Palmer (2010) updated their original 1990 book, and it is widely used. In addition, the slight change to the title emphasizes the use of testing in “real-world” situations and the decisions made on the basis of language tests that can have an impact on students, instructors, and programs. This focus on the real world reflects the changes in language testing textbooks over the past three decades; the shift from providing basic information on assessment to demonstrating ways to integrate authenticity into assessment is striking. In addition, Carr’s (2011) book includes a CD to help users apply the information in the text, with a specific emphasis on using statistics. Such approaches show that language testing texts are working to meet the needs of their users through contextualization and additional resources such as computer-based activities beyond a written text that allow users to practice what they have learned.

While the above provides only a glimpse into the kinds of text-based materials offered to classroom teachers, the very existence of such materials points to the

importance of assessment for language instructors, as well as an understanding on the part of textbook authors and publishers that theoretical texts were insufficient to explain testing to language instructors. It is also important to note that encyclopedias such as this one also provide a resource for language professionals to explore in depth a variety of issues in language assessment.

Non-Text-Based Assessment Training

In addition to training provided by written texts employed during a formal university or graduate level class or independently, other formats have become available for training language instructors on assessment. This section outlines some self-paced self-instructional materials and web-based instructional materials for instructors.

Self-Instructional Materials

Professional development workshops are frequent approaches to help instructors in all subjects supplement their formal training and improve their classroom effectiveness. With the proficiency movement in the United States in the 1980s, language instructors could participate (for a cost) in a 4-day training on oral proficiency assessment, a format previously restricted primarily to government employees.

As technologies became more accessible and less costly, tape-recorded materials, accompanied by tapes, could begin to replace live, face-to-face workshops; Kenyon and Stansfield (1993) and Kenyon (1997) investigated one new format: allowing potential language raters to participate in training through use of a kit rather than a live training workshop. Such self-instructional approaches allowed instructors to seek on their own (or upon advice from supervisors or other colleagues) new methods of language assessment to use in their classroom. Similarly, ETS developed self-training kits for raters of the SPEAK test; these kits included tapes and ancillary materials. These new formats allowed instructors who had not received training in new approaches during their education or for whom the approaches came after their formal education was completed to learn about and apply new testing methods.

As use of computers and the Internet grew throughout the 1990s, computer-based approaches gained in popularity throughout education. So, too, did access to more information on language assessment training.

Since 1995, Fulcher has hosted the *Resources in Language Testing* webpage (<http://languagetesting.info>, accessed 12/5/2015), which includes references, relevant organizations, and streaming video of well-known language testers responding to frequently asked questions in language testing on topics such as reliability, validity, test impact, item writing, and statistics. This page contains a plethora of useful information. Recently, he has added podcasts to accompany articles published in *Language Testing*, one of the two major journals devoted to language assessment. The addition of podcasts to supplement such academic articles demonstrates the growing need in academic journals, as in academic texts, of users to go beyond the written word and to use multiple forms of communication to describe and explain language testing to different users.

As the CEFR gains popularity in Europe, uses of it continue to grow. Among other useful resources is a “passport” to demonstrate student progress on the CEFR that students and instructors can complete to show student growth. These resources are available on the web and can be downloaded for use in schools. The Council of Europe has a website that provides resources on both the CEFR and assessment in general (<http://www.coe.int/t/dg4/education/elp/>, accessed 11/30/2015), including ways to develop an online portfolio to document language outcomes. The Centre for Canadian Language Benchmarks provides resources for learners and assessors on its website, including guidelines and resources for test development. Many European-based resources include information for language learners in addition to instructors; such resources are less plentiful for US-based resources. Two examples of learner-oriented resources in the United States are housed at the National Council of State Supervisors of Foreign Languages (NCSSFL) and CAL. NCSSFL developed a first paper-based and now online self-assessment system for US K-16 learners inspired by the CEFR efforts. This resource (http://www.ncssfl.org/LinguaFolio/index.php?linguafolio_index accessed 12/20/2015) is designed to help learners develop and track their progress toward language proficiency goals and requires registration. On a different note, in developing a new, computer-based Arabic oral proficiency assessment, CAL worked with learners to design a five-module online resource that describes different aspects of Arabic oral language proficiency, including both examples of student performances at different proficiency levels and clips of student interviews that describe how these students attained proficiency in Arabic (<http://www.cal.org/aop/>, accessed December 15, 2015).

In addition to resources for students, some organizations also provide support for teachers. In the late 1990s, the Center for Advanced Research on Language Acquisition (CARLA) of the University of Minnesota has developed a seven-module, online Virtual Assessment Center (VAC) to provide both resources, background information and guidance on second language classroom assessment (<http://www.carla.umn.edu/assessment/vac/index.html>, accessed 11/30/2015). The VAC includes an annotated bibliography of assessment resources, as well as a virtual item bank. The virtual item bank provides model items for teachers and is accompanied by item-writing tips. The VAC represents an early effort not only to help classroom language instructors develop good items and assessments for their students but also to understand the principles of assessment that undergird the process. Although the VAC is a valuable resource, the annotated bibliography has not been updated since the early 2000s. Perhaps one of the most challenging aspects of online resources is keeping them current; updating such resources regularly represents a significant commitment. If such resources are not reviewed regularly, they fall out of date quickly.

Swender et al. (2006) reported on a web-based survey of assessment uses and needs of 1,600 foreign language instructors in the United States. In addition to highlighting tests currently being used and needed for language instructors, the survey also highlighted a lack of understanding of many testing concepts, such as appropriate test use, by those who responded. As a result of this survey and other reports, in 2009, the Center for Applied Linguistics updated its foreign language test directory and developed a tutorial for users in test selection. In developing the

tutorial and soliciting feedback from a variety of stakeholders, Malone (2013) found a dichotomy between the perceived needs of such a tutorial by language instructors and by language testers. Language instructors stressed the needs for a succinct, understandable tutorial, while many language testing specialized and emphasized on the importance of explaining complex language testing concepts, such as assessment use and validity arguments, to such language instructors. The directory is updated biannually and the tutorial will be reviewed and updated by 2018. In 2015, the tutorial and directory received 63,000 unique views, thus highlighting the need for such online instruments.

Works in Progress

Many of the current projects described are simultaneously works in progress and represent ongoing efforts to enhance both practice and understanding of assessment by language instructors. The addition of online tutorials, podcasts, videos, and e-portfolios across the world demonstrates the continued interest in and need for these resources. A recent edition of *Language Testing* was devoted to the issue of language assessment literacy; this special issue highlighted many facets of language assessment literacy from how language assessment is viewed in the parliament (Pill and Harding 2013) to the identity of the language tester (Jeong 2013) to the contrast between information valued by language testers and instructors (Malone 2013). In reviewing the wide range of topics addressed by this issue, it is clear that a variety of stakeholders could benefit from information about language assessment and that the audience for such information has expanded both beyond simply language testers and language teachers. As the field progresses, it is likely that still more online resources will become available; a likely issue to arise is how to evaluate the efficacy of the different resources to ensure that users not only use high-quality resources that reflect best practices but also that the resources they access are appropriate for their own needs. Although the university in general and teacher preparation programs in particular have been the traditional focus of language assessment, online resources represent an important way to provide both ongoing professional development to in-service teachers as well as basic information about language assessment to those outside the field. In addition to online resources, the International Language Testing Association (ILTA) provides funding for two or three workshops to be held annually in parts of the world where language assessment literacy could be improved or where such efforts are scarce.

Problems and Difficulties

Although the landscape for including more stakeholders in the language assessment process and educating these stakeholders about language assessment is hopeful, it is nonetheless an ongoing and daunting task. The amount of resources available in print and online continues to grow, and a language instructor inexperienced in

language assessment might not understand how to select from among the many resources in the world.

This original chapter was released in 2008, and many more resources, from textbooks to online resources, have been released and are being used internationally. In 2008, the major challenge identified was determining who is and who should be trained in language assessment, how and to what extent such individuals are trained, and what the expected outcomes of such training should be. To some degree, raising the issue of training in language assessment is as important as any information contained therein; in 2008, the focus was on the training of language instructors. Although there is still no consensus on the assessment literacy needs of language instructors and, indeed, there are differences in perspective as to what language instructors believe, they need to know about assessment and what language testers believe that language instructors need to know (Malone 2013). It is important that the area of inquiry on language assessment literacy has expanded during the past 7 years beyond discussing what language instructors need to know. Pill and Harding (2013) investigated the assessment literacy needs from a parliamentary perspective; Jeong (2013) explored the language assessment literacy needs of testers and non-testers, and O'Loughlin (2013) explored the needs of university test users. However, one major gap that remains is how to best educate students about language assessment. Although three examples have been used in this chapter, the fact remains that students of language also need to understand why and how they are being assessed and how the results of their assessments will be used.

Future Directions

As this chapter indicates, progress has been made to increase language assessment literacy efforts. The focus on language instructors and their understanding of assessment has expanded to include additional test users such as parents and administrators. The studies cited in this article represent the understanding that language assessment results are used for a number of far-reaching goals, from language students and teachers to our representatives in government. While this progress is helpful, additional work is needed. Pill and Harding (2013) mentioned above show that governments need education on assessment literacy; their efforts should be extrapolated to other governments. It is important to note, too, that students and test takers have not yet emerged as a focus of language assessment literacy research and this group is most affected by language tests and their results.

For continued progress to take place, it will be important to continue and expand work that explores both stakeholder perceptions of language assessments, the extent to which these perceptions are accurate, and how to mediate these expectations to help improve assessment literacy. As stakeholder language assessment literacy grows, it will become crucial for stakeholders to use this information to hold themselves and the developers of the tests they use accountable for the ways the tests are used and the decisions made on the basis of these tests. Finally, language learners themselves and their families must be included in this work. Learners and

their families benefit most and least from language assessment; thus, they must fully understand the tests they take and the implications the results have for them.

Cross-References

- ▶ [Ethics, Professionalism, Rights, and Codes](#)
- ▶ [Language Assessment Literacy](#)
- ▶ [Washback, Impact, and Consequences Revisited](#)

Related Articles in the Encyclopedia of Language and Education

- Nike Arnold: [Technology and Second Language Teacher Professional Development](#),
In Volume: Language, Education and Technology
- Klaus Brandl: [Task-Based Instruction and Teacher Training](#). In Volume: Second and Foreign Language Education
- Linda von Hoene: [The Professional Development of Foreign Language Instructors in Postsecondary Education](#). In Volume: Second and Foreign Language Education
- Margaret E. Malone: [Developing Instructor Proficiency in \(Oral\) Language Assessment](#). In Volume: Second and Foreign Language Education
- Oleg Tarnopolsky: [Nonnative Speaking Teachers of English as a Foreign Language](#).
In Volume: Second and Foreign Language Education

References

- Alicea, I. (2005). NCLB requirements prompt changes in ELL assessment. *The ELL Outlook*, 4(5), 19.
- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Bachman, L. F., & Savignon, S. J. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *The modern language journal*, 70(4), 380–390.
- Boyles, P. (2005). Assessment literacy. In M. Rosenbusch (Ed.), *National assessment summit papers* (pp. 11–15). Ames: Iowa State University.
- Brindley, G. (1997). Assessment and the language Teacher: Trends and transitions. *Language Teacher Online*, 2.
- Brown, J. D. (1998). *New ways of classroom assessment. New ways in TESOL series II*. Innovative Classroom Techniques. TESOL, Alexandria, VA.
- Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford University Press.
- Carroll, B. J. (1991). Resistance to change. In J. C. Alderson & B. North (Eds.), *Language testing in the 1990s* (pp. 22–27). London: Modern English Publications and the British Council.
- Cohen, A. (1994). *Assessing language ability in the classroom*. Boston: Heinle and Heinle.
- Davidson, F., & Lynch, B. (2002). *Testcraft*. New Haven: Yale University Press.

- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamra, T. (1999). *Dictionary of language testing*. Cambridge: University of Cambridge Local Examinations Syndicate and University of Cambridge.
- Genesee, F., & Upshur, J. A. (1996). *Classroom-based evaluation in second language education* (Cambridge Language Education). Cambridge/New York: Cambridge University Press.
- Harris, D. P. (1969). *Testing english as a second language*. Washington, DC: ERIC Clearinghouse.
- Hughes, A. (2004). *Testing for language teachers*. New York: Cambridge University Press.
- Inbar-Lourie, O., Scarino, A., Malone, M. E., Jeong, H., O'Loughlin, K., Pill, J., & Harding, L. (2013). The special issue on language assessment literacy. *Language Testing*, 30(3), 301–307.
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers?. *Language Testing*, 30(3), 345–362.
- Kenyon, D., & Stansfield, C. W. (1993). Evaluating the Efficacy of Rater Self-Training.
- Kenyon, D. M. (1997). Further research on the efficacy of rater self-training. In A. Huhta, V. Kohonen, L. Kurki-Suonio, & S. Luoma (Eds.), *Current developments and alternatives in language assessment: Proceedings of LTRC 96* (pp. 257–273). Jyväskylä: University of Jyväskylä.
- Lantolf, J. P., & Frawley, W. (1985). Oral-proficiency testing: A critical analysis. *The Modern Language Journal*, 69(4), 337–345.
- Little, D. (2012). Elements of L2 proficiency: The CEFR's action-oriented approach and some of its implications. In E. Tschirner (Ed.), *Aligning frameworks of reference in language testing: The ACTFL proficiency guidelines and the common European framework of reference for languages*. Tübingen: Stauffenburg Verlag.
- Llosa, L. (2011). Standards-based classroom assessments of English proficiency: A review of issues, current developments, and future directions for research. *Language Testing*, 28(3), 367–382.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329–344.
- Moore, Z. (1994). The portfolio and testing culture. In C. Hancock (Ed.), *Teaching, testing and assessment: Making the connection* (pp. 163–182). Lincolnwood: National Textbook Company.
- 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: Addison-Wesley Publishing.
- Ommagio, A. (1986). *Teaching language in context*. USA: Heinke & Heinle Publishers.
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381–402.
- Spolsky, B. (1977). Language testing: An art or science. In *Proceedings of the fourth international congress of applied linguistics* (pp. 7–28). Stuttgart: Hochschulverlag.
- Spolsky, B. (1995). Measured words: The development of objective language testing. In D. Kenyon, & C. W. Stansfield (1993) (Eds.), *Evaluating the efficacy of rater self-training*. Oxford University Press.
- Stiggins, G. (1997). *Student centered classroom assessment*. Upple Saddle River: Prentice Hall.
- Stoyhoff, S., & Chapelle, C. A. (2005). *ESOL tests and testing: A resource for teachers and program administrators*. Alexandria: TESOL Publications.
- Swender, E., Abbott, M., Vicars, R., & Malone, M. (2006). *The assessment of performance and proficiency in testing: a report*. Presentation at the convention of the American Council of Teachers of Foreign Languages, Nashville, TN.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412.
- Wiggins, G. (1994). Toward more authentic assessment of language performances. In C. Hancock (Ed.), *Teaching, testing and assessment: Making the connection* (pp. 69–86). Lincolnwood: National Textbook Company.