Annual Review of Applied Linguistics (2009) **29**, 21–36. Printed in the USA.
Copyright © 2009 Cambridge University Press 0267-1905/09 $16.00
doi:10.1017/S0267190509090035

2. DEVELOPING ASSESSMENT LITERACY

Lynda Taylor

Language testing and assessment have moved center stage in recent years, whether for educational, employment, or sociopolitical reasons. More and more people are involved in developing tests and using test score outcomes, though often without a background or training in assessment to equip them adequately for this role. Simultaneously, increasing professionalization of the field has led to the generation of standards, ethical codes, and guidelines for good testing practice. Although these can help make assessment practices more transparent and accessible to a wider constituency, they also risk promoting a view of language testing as highly technical and specialized–best left to experts. These trends have implications for both policy and practice. This article reviews efforts to promote understanding of assessment within the field of applied linguistics and within education and society more broadly. The role of professional associations, academic institutions, and commercial organizations in developing assessment literacy is considered, as well as the contribution of published material and other types of training resources. This article reflects on how the international language testing community can encourage the sharing of the core knowledge, skills, and understanding that underpin good quality assessment as widely and accessibly for the benefit of all.

The Growth of Testing and Assessment Worldwide

There can be no doubt that testing and assessment are growing in importance and influence around the world today. Countries with established educational systems that already include public examinations are considering how these can be reengineered to reflect new approaches to teaching and learning as well as to meet the changing needs of contemporary society. Other nations are seeking to establish a national education system for the first time in their history, or to upgrade an existing system to extend educational opportunities to larger sections of their population. In such contexts there is often a policy imperative to design new assessment components into the educational infrastructure at local, regional, or national level in order to promote teaching and learning and to provide systems for accountability. In today's globalized environment there is sometimes an additional aspiration to align local or indigenous assessment frameworks with internationally accepted standards or

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

recognized benchmarks. A further dimension of globalization over the past decade has been the unprecedented movement of vast numbers of people to and from various parts of the world. Though the reasons for this mass migration are complex and varied (including travel for tourism, access to educational or professional opportunities, economic or environmental migration, and conflict or political unrest), assessment tools such as skills-based competency tests and citizenship tests, as well as language tests, are being integrated into the sociopolitical and economic infrastructures that support and regulate these trends.

Against this wider backdrop, the testing and assessment of language skills (where the language is a second, third, or additional language) have undoubtedly moved from periphery to centre stage in recent years. Language tests, and the scores they generate, are increasingly used across contemporary society worldwide: within education, from primary age to higher education; in employment contexts, from the professional registration of health professionals to health and safety issues in the catering or construction industry; and, more controversially, in migration and citizenship policy and practice around the world. As Spolsky observed, "testing has become big business" (Spolsky, 2008, p. 297), and not simply in commercial terms. Shohamy (2008) also noted "the societal role that language tests play, the power that they hold, and their central functions in education, politics and society" (p. xiv). The No Child Left Behind act in the United States (see Menken & Shohamy, 2008), the Every Child Matters legislation in the United Kingdom, and the recent educational reform program in Hong Kong all testify to the central role that language assessment practices now occupy in society.

The Growing Numbers Involved in Language Assessment

These trends mean that there are not simply more people taking tests; in addition, growing numbers are directly involved in selecting or developing tests, and/or in using test scores for decision-making purposes. In practice, they often find themselves doing this without much background or training in assessment, or indeed in linguistics and language education, to equip them for the role. They include classroom teachers tasked with designing and/or delivering standardized tests to evaluate their pupils' progress, tests that are sometimes then used to hold teachers and schools accountable for that progress, or the apparent lack of it. A similar burden of expectation is laid on staff in university admissions, in professional registration bodies, and in immigration agencies to know what tests measure and what test scores mean, and to understand how to integrate them into their complex and often high-stakes decision-making processes. Public and policymaker expectations can be unrealistic about what a single score represents, as well as what information it does or does not provide, and this is exacerbated by a general lack of understanding about the imprecision or error inherent in any measurement. Spolsky (2008) complained that an "oversimplified view of the ease of producing meaningful measurement is held by the general public and educational administrators alike" (p. 300).

Applied linguists are among those regularly asked to help shape policy and practice in language planning, curriculum and syllabus design, and pedagogy and

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

teacher education, and they are increasingly called upon to advise on developing or revising language assessment systems. This consultancy role tends to assume that they (along with classroom teachers, university admissions staff, and immigration officials) possess a sound understanding of the relevant theoretical and practical issues. This may be true in relation to language teaching and learning, but it is much less so for matters of testing and assessment. Language education programs at graduate level typically devote little time or attention to assessment theory and practice, perhaps just a short (often optional) module; and although there is no shortage of books on language testing and assessment available today, many of these are perceived to be (and often are) highly technical or too specialized for language educators seeking to understand basic principles and practice in assessment.

The Professionalization of Language Testing and Assessment

The field of language testing and assessment has itself undergone a process of increasing professionalization over recent years, in response to calls from both within and outside the field (Bachman, 1990). A growing raft of quality standards, ethical codes, and guidelines for good testing practice has been embraced or generated by language testers, and in many cases these touch upon matters that extend well beyond a test's purely technical qualities. Examples include the Code of Fair Testing Practices in Education (2004) and the AERA/APA/NCME Standards (1999), both of which concern assessment in general. There also exist language testing specific codes such as the ALTE Code of Practice (1994), the ILTA Code of Ethics (2001), the EALTA Guidelines for Good Practice (2006) and the ILTA Guidelines for Practice (2007). Professional associations of language testers and testing organizations were established during the 1990s, at national, regional, and international level, including the Japan Language Testing Association (JLTA), the Association of Language Testers in Europe (ALTE) and the International Language Testing Association (ILTA). At the same time, there has been growing awareness of the consequences of testing, intended and unintended, positive and negative, as illustrated by the burgeoning research literature on the theory and practice of language testing washback and impact (see, among others, Alderson & Wall, 1996; Cheng, 2005; Cheng, Watanabe, & Curtis, 2004; Green, 2007; Hawkey, 2006; Kunnan, 2000; Wall, 2005).

Kunnan (2004, 2008) reflected on the ethical milieu that has emerged for language testers, prompted by various factors in the professional field such as demands for accountability and responsiveness to clients, increased use of language tests and types of delivery methods, use of new measurement and other analytical techniques, and expanded concepts of validity (i.e., Messick, 1989, 1996). He suggested the language testing community has responded by developing explicit standards and codes for its community and practice, and that its professional ethic continues to evolve informed by the wider literature in ethics and moral philosophy.

One might imagine that the articulation of standards, ethical codes, and guidelines for good practice makes the business of designing and developing language tests more transparent, accessible, and manageable to anyone charged with this task, including those nonspecialist testing stakeholders mentioned above. To

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

some extent this may be true, but it also risks promoting a view of language testing as rather technical and specialized–and best left to professional experts, the only people properly equipped and qualified to attain industry standards. This has clear policy and practice implications among those test stakeholders who find themselves increasingly involved, either directly (e.g., classroom teachers and university lecturers) or indirectly (e.g., college admissions officers, immigration officials, and parents), in producing and using tests and test scores. Although such stakeholders may be keen to improve their assessment knowledge and skills, they may feel discouraged from doing so by perceptions of the field's growing professionalization. Both Davies (2008a) and Spolsky (2008) expressed a concern that this process of professionalization over the last quarter century has led to the isolation of the language testing field from other academic disciplines. This may be equally true for its relationship with the wider world of education and society. We might even speculate that the language testing community's preoccupation with technical and measurement qualities in testing from the 1960s to 1980s, followed in turn by a similarly inward-looking focus on developing its own professional ethic during the 1990s, means that it has singularly failed to encourage wider public understanding of assessment principles and practices, that is, it has neglected to engage in the critical task of developing assessment literacy.

The Concept of Assessment Literacy

The term *assessment literacy* has been coined by several writers in recent years to describe what the constituency of language teachers and instructors needs to know about assessment matters (Inbar-Lourie, 2008; Malone, 2008; Stiggins, 1997; Stoynoff & Chapelle, 2005). Clearly, language teachers and instructors need some measure of assessment training if they are engaged in selecting, administering, interpreting, and sharing results of large-scale tests produced by professional testing organizations, or in developing, scoring, interpreting, and improving classroom-based assessments. The requirement for adequate assessment training among language test developers and researchers is also self-evident, whether they are working in academically oriented contexts such as university departments of language or education, or in more commercially oriented contexts such as examination boards or testing agencies, or in the hybrid professional contexts of organizations such as the Educational Testing Service (ETS) or University of Cambridge ESOL Examinations (Cambridge ESOL). But the concept of assessment literacy could be expanded to describe the level of knowledge, skills, and understanding of assessment principles and practice that is increasingly required by other test stakeholder groups, depending on their needs and context (e.g., among educational advisors or government officials, policy planners and decision makers, the media, and the general public).

Since the 1960s there has been a steady stream of textbooks and course materials offering general training in testing theory and practice, from basic to very advanced. Examples include, in order of their publication date, Lado (1961), Carroll (1961), Vallette (1967), Harris (1969), Heaton (1975), Allen and Davies (1977), Oller (1979), Carroll (1980), Madsen (1983), Carroll and Hall (1985), Henning (1987),

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

Hughes (1989, 2003), Weir (1990), Davies (1990), Bachman (1990), Weir (1993), Cohen (1994), Alderson, Clapham, and Wall (1995), Bachman and Palmer (1996), McNamara (2000), Davidson and Lynch (2002), Brown and Hudson (2002), Weir (2005), McNamara and Roever (2006), Downing and Haladyna (2006), Fulcher and Davidson (2007). Other titles have focused more narrowly on testing specific language skills, subskills, or domains: the testing of reading (Alderson, 2000), listening (Buck, 2001), writing (Shaw & Weir, 2007; Weigle, 2002), and speaking (Fulcher, 2003; Luoma, 2004); the testing of vocabulary (Read, 2000) and grammar (Purpura, 2004); performance testing (McNamara, 1996); specific purpose testing (Douglas, 2000); computer-based testing (Chapelle & Douglas, 2006); and the testing of young learners (McKay, 2006). Most of these publications were written primarily for an academic or specialized readership, and though some were specifically aimed at language teachers, few could be described as accessible to a general, nonspecialist audience, which might include parents, politicians, or policymakers. The textbooks listed above typically set out to offer practical advice in technical aspects of test production and score interpretation, or instruction in theoretical aspects of language testing and assessment, or a combination of the two. A few notable exceptions (e.g., McNamara, 2000) offer a general introduction to the sociopolitical and educational nature of assessment, together with practical guidance on the appropriate selection and use of tests in these contexts.

The increasingly central role of testing and assessment worldwide, however, and the concerns raised from within the language testing community recently about the potential misuse or abuse of tests, suggest that an appropriate level of assessment literacy needs to be nurtured not just among engineers and technicians who are actively involved in test development or research activities, or even among applied linguists and language teachers involved in delivering language education, but much more broadly in the public domain if a better understanding of the function and values of assessment tools and their outcomes is to be realized throughout society. Assessment literacy needs to be developed, therefore, across an increasingly diverse set of test stakeholder groups or constituencies including: personnel in both existing and newly established and emerging national examination boards, academics and students engaged in language testing research, language teachers or instructors, advisors and decision makers in language planning and education policy, parents, politicians, and the general public.

This is perhaps not dissimilar to the notion of computer literacy that emerged during the 1990s when computer knowledge, skills, and expertise, previously restricted to a fairly narrow, specialist sector of the population, gradually extended throughout the wider population as computers themselves became easier to access and more widely used in daily life. Higher levels of assessment literacy, that is, a sound knowledge and understanding of the principles and practice of assessment, are essential if applied linguists and other language specialists (e.g., teachers) are to engage effectively in matters of language assessment policy and practice, and if the benefits and the limitations of assessment tools are to be communicated to those in wider society (e.g., policymakers, parents).

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

The Components of Assessment Literacy

Views on the types of knowledge and understanding underpinning assessment literacy have changed over time. Davies (2008a), for example, reviewed the teaching of language testing over the past 50 years and observed two main trends: (1) a steady expansion in the volume of assessment training materials required due to the growing professionalism of the field (as we have seen from the sample selection above) and (2) an expansion in the profession's own view of the knowledge and skills needed by its members. He noted a steady curricular expansion from basic technical testing expertise (the how-to), to knowledge about language and measurement (the what), and, most recently, to understanding of key principles with regard, for example, to validity and ethics (also reflected in the list of titles). Davies characterized the present paradigm as that of Skills + Knowledge + Principles:

> Skills provide the training in necessary and appropriate methodology, including item-writing, statistics, test analysis, and increasingly software programmes for test delivery, analysis and reportage. Knowledge offers relevant background in measurement and language description as well as in context setting. Principles concern the proper use of language tests, their fairness and impact, including questions of ethics and professionalism. (Davies, 2008a, p. 335)

Davies's article reviews several seminal textbooks used to teach language testing between 1960 and the present day (including some of those referenced above), showing how they exemplify changing trends and perceptions in training focus. The focus in his article is clearly on the training needed by those within the field of language testing, a mainly academic community that occasionally extends to embrace language teachers and classroom based assessment, but rarely concerns itself with the need to educate other test stakeholder groups in assessment matters.

In the mid-1990s Bailey and Brown (1995) reported on a survey of language testing courses in the context of language teacher training programs around the world. They examined various course characteristics, including topics covered and core textbooks used. A decade later, when they replicated their study, they found "the presence of a stable knowledge base that is evolving and expanding rather than shifting radically" (Bailey & Brown, 2008, p. 371). They also found evidence of other types of advanced language testing courses being provided to prospective language teachers in some graduate programs. It remains unclear, however, how far the content of such courses is likely to equip teachers to manage the complex sociopolitical and educational issues that are implicated in assessment today.

McNamara and Roever (2006) cautioned against a potentially narrow view of the training needed by test developers and, presumably, other test users: "In terms of academic training, we stress the importance of a well-rounded training for language testers that goes beyond applied psychometrics . . . a training that includes a critical view of testing and social consequences" (p. 255). Their view parallels the earlier

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

concern raised by Davies (2008a) and Spolsky (2008) that language testing, if viewed in purely professional or technical terms, risks being isolated from other key areas of life, and that this carries with it an inherent danger given the central role assessment now occupies in education and society.

Like McNamara and Roever, Inbar-Lourie (2008) took a broader view of how assessment literacy can be conceptualized and how it might be promoted more widely. She described an attempt to articulate the possible components of a language assessment literacy knowledge base as the foundation for establishing a core knowledge framework for assessment training. What seems innovative in this approach is that it was initiated from largely outside the professional testing culture and community, and it is informed by social-constructivist perspectives that encourage a more holistic, integrated, and dynamic understanding of assessment in context than the language testing profession has been used to in recent decades. Thus Inbar-Lourie argued that the core competencies to be acquired through assessment training will "reflect current views about the social role of assessment in general and language assessment in particular, contemporary views about the nature of language knowledge, and give due emphases to both classroom and external assessment practices" (pp. 396–397). Her perspective chimes with other initiatives that seek to reconceptualize assessment paradigms and the competencies needed by those involved, including, among others, work on assessment for learning (Assessment Reform Group, 2002), classroom-based assessment (Rea-Dickins, 2008), alternative assessment (Fox, 2008), and dynamic assessment (Lantolf & Poehner, 2008).

This current trend in thinking seems to be that training for assessment literacy entails an appropriate balance of technical know-how, practical skills, theoretical knowledge, and understanding of principles, but all firmly contextualized within a sound understanding of the role and function of assessment within education and society. The exact nature of this balance and the level of literacy to be acquired may vary according to context, and according to whether test stakeholder constituencies are more or less directly involved in matters of test selection, test production, test score interpretation, or test evaluation.

Promoting Assessment Literacy Both Within and Beyond the Profession

As we have already seen, there is today no shortage of textbooks offering instruction in the principles and practice of assessment. Much of this material is specifically designed for those about to enter or already inducted into the language testing profession, that is, relative novices as well more established members committed to ongoing professional development. This constituency is relatively well-served in terms of access to assessment training, but if such materials are too specialized or technical for those outside the professional language testing community, the question remains how to encourage improved levels of assessment literacy more widely.

Malone (2008) reminded us that training in language assessment can be delivered not only through text-based materials, such as textbooks, but also via other

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

channels. Malone highlighted the contribution made by face-to-face workshops and self-instructional options, which often make use of new technologies and permit a wide range of materials to be made easily available (e.g., free downloading from the Internet). She cited the examples of the Council of Europe materials associated with the Common European Framework of Reference for Languages (CEFR) (http://www.coe.int/t/dg4/Linguistic), as well as those provided by the Centre for Canadian Language Benchmarks (CLB) (www.language.ca).

Reference frameworks such as the CEFR and CLB can themselves serve as communication tools and heuristic devices, aiding transparency and supporting the coherence of language learning, teaching, and assessment activities across communities and regions. Professor John Trim underlined this point when he described the original development of the CEFR:

> It was conceived . . . as a basis for reflection and communication. . . . What we were aiming at was something which will be a common reference point that people working in different fields and people using it for doing entirely different things and in very different ways could refer to in order to feel that they were part of a common universe. (Saville, 2005, pp. 280–281).

North (2008) described the three main aims of the CEFR: "to establish a metalanguage common across educational sectors, national and linguistic boundaries that could be used to talk about objectives and language levels; to encourage practitioners in the language field to reflect on their current practice . . .; [and] to agree common reference points." (p. 22). Not surprisingly, perhaps, views differ as to how far the aims of transparency and coherence have been achieved, and to what extent the CEFR risks being used instead as an instrument for harmonization, stifling diversity in relation to language teaching, learning, and assessment.

Other innovative approaches to help develop assessment literacy have included video projects such as *Mark My Words* (Melbourne LTC, 1997) and the ILTA *Video FAQs* (Fulcher & Thrasher, 1999/2000), though these types of resources need regular updating if they are to remain relevant and keep up with new developments in the field.

Professional language testing associations have taken on an increasingly active role in promoting assessment literacy more widely among language test developers, test users, and other stakeholders, especially teachers involved in classroom language assessment and other educators involved in developing language tests in national centers or institutional units. Training is now delivered, for example, through the preconference workshops of professional bodies such as the International Language Testing Association (ILTA) and the European Association of Language Testing and Assessment (EALTA), whose 2008 annual conference in Athens took assessment literacy as its theme. The Association of Language Testers in Europe

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

(ALTE), a consortium of European language testing agencies, provides a range of resources designed as communication tools for developing understanding and expertise, from publications such as item writing guidelines and test content analysis tools (freely downloadable from the ALTE Web site), to conference opportunities and a training seminar program for ALTE members and interested others, based around ALTE's Principles of Good Practice for ALTE Examinations, Minimum Quality Standards, and Quality Management System for test developers (www.alte.org). The publication in the late 1990s of the *Multilingual Glossary of Language Testing Terms* (ALTE Members, 1998) and the *Dictionary of Language Testing* (Davies et al., 1999) made available valuable practical resources for training specialists and nonspecialists alike in the terminology and discourse of language testing, a prerequisite for promoting assessment literacy. Language test providers such as examination boards and testing agencies (e.g., Cambridge ESOL) now provide a range of publications specially written to be accessible to a wide range of test stakeholder groups, from item writers and examiners, to learners and their parents, to teachers and publishers. In addition, they often offer conferences, seminar programs, and workshops for teachers, educational policymakers, and other test stakeholders.

An additional resource for developing assessment literacy is the series of narrative accounts that have recently emerged chronicling testing developments over the decades. The first of these by Spolsky (1995) provides an historical account of two influential and contrasting testing traditions on either side of the Atlantic. More recently published volumes in this historical narrative genre include the accounts of various Cambridge ESOL tests (Hawkey, 2009; Weir & Milanovic, 2003), the story of the communicative language testing movement (Hawkey, 2004), the development of testing English for business purposes (O'Sullivan, 2006), and the evolution of academic language proficiency assessment (Davies, 2008b; Taylor & Angelis, 2008). All these publications are helpful in describing and analyzing particular tests, as well as the more general business of testing and assessment, within the educational and sociopolitical contexts of their time. They contextualize the practice of language testing as a socially constructed and interpreted phenomenon, rather than treating it primarily as a pseudoscientific endeavor that is removed and isolated from human individuals and social values. It may well be that popular adaptations of this narrative, storytelling approach will prove a more effective means of developing assessment literacy among the wider stakeholder community in the future.

The overarching aim is not simply to train well-qualified engineers and technicians for quality test development, but to educate and build a much broader and inclusive community of good practice for the longer term, a community that is concerned with the impact of testing on society and with fairness issues in the broadest sense. Bachman commented as follows: "The advocacy for fairness by language testers and language test developers is thus, in my view, a critical element not only in the design and implementation of language assessment systems, but also, and perhaps more importantly, in the education and mentoring of those who will carry the field forward in the years to come" (Bachman, 2000, p. 40).

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

Conclusion: Reconceptualizing Test Wiseness

Traditionally, the term *test wiseness* has been used by the testing community to refer to the "amount and type of preparation or prior experience that a test taker has" (Davies et al., 1999, p. 210). Familiarity or lack of familiarity with the content and format of a test is acknowledged to impact on a test taker's performance, potentially leading to an over- or underestimation of the ability of interest. Perhaps we can borrow the term test wiseness and reconceptualize it to reflect an understanding of testing skills, knowledge and principles–a different sort of test wiseness on the part of test developers and test users, and one that helps encourage the positive impact of testing on people and society and enhances the value of test use.

Improving levels of assessment literacy, both within the language testing profession and in wider society, is about more than simply offering opportunities for individual professional development (Hamp-Lyons, 2008) or for further professionalization of the field. It may help to mitigate the potentially negative effects that risk resulting from the growing influence of testing and assessment in education and society worldwide.

Finally, in the specific context of this annual review, it is perhaps worth reminding ourselves that applied linguists are often at the same time academic staff members who find themselves using test scores for decisions about student admissions; or they may be tasked with creating assessment tools for end-of-year student evaluations. Thus, applied linguists too are among those who need to develop a good understanding of what test instruments can and cannot do, along with a correct understanding of how to interpret the scores generated. A sound understanding of measurement principles and practice is also essential for anyone who designs or uses tests for the purposes of applied linguistics research. Sadly, it is not uncommon to see the findings of otherwise interesting applied linguistics research studies undermined because of the poor quality of the research tools (i.e., testing instruments) that were used or the inappropriate interpretation of their outcomes (i.e., test scores). A better grasp of key measurement concepts such as validity and reliability would undoubtedly help to address this problem. This last point highlights the potential interface between language testing researchers and those from other fields of language study. As McNamara (2000) observed, language testing is not just a technical activity but "a site for research activity of a fundamental nature in applied linguistics" (p. 54). A stronger interface with other disciplines might also address the concerns raised by both Davies (2008a) and Spolsky (2008) over the apparent isolation of the language testing field, which in turn risks the marginalization of language testers themselves and any potential contribution for good they have to make to the wider academic and social community.

As more and more new test development projects are undertaken at national and regional level, so the need to understand the principles of sound assessment to implement these as good practice can only continue to grow among language teachers, applied linguists and a wide diversity of other test stakeholders. In light of this, there is an important and urgent role for the language testing and assessment

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

community to play in helping to ensure that the core knowledge, skills, and experience needed to underpin good quality and ethically sound assessment policy and practice are shared as widely as possible for the benefit of all. As Stansfield astutely observed: "This will allow us to reach our full potential, both as professionals and as contributors to an organized civil society" (Stansfield, 2008, p. 323).

ANNOTATED REFERENCES

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA, and Oxford, England: Blackwell.

This volume focuses on the social aspects of language testing, including the assessment of socially situated language use and the societal consequences of language tests. The authors argued that traditional approaches to ensuring social fairness in tests go some way to addressing social concerns, but a broader perspective is necessary to examine the functions of tests on a societal scale.

25th anniversary issue. (2008, July). *Language Testing, 25*(3). Sage.

This special issue of the journal *Language Testing* celebrates its 25th anniversary by bringing together various contributions (introduction, lecture, and articles) to summarise the recent history of the field, to discuss its current state and to speculate on its likely future. A persistent thread throughout the issue is that of assessment literacy and training in language assessment principles and practice, both within and outside the field.

Shohamy, E., & Hornberger, N. (2008). (Eds.). *Encyclopedia of language and education* (2nd ed.): *Vol. 7. Language testing and assessment*. New York: Springer Science+Business Media.

Volume 7–on language testing and assessment–forms part of the recently published, fully revised edition of the *Encyclopedia of Language and Education*. It contains several extremely useful chapters relevant to a discussion of how assessment literacy is developed, including contributions on language assessment culture, different methods and contexts of assessment, the role and impact of assessment in education and society, and related sociopolitical and ethical issues.

OTHER REFERENCES

AERA/APA/NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Alderson, J. C. (2000). *Assessing reading*. Cambridge, England: Cambridge University Press.

Alderson, J. C., & Wall, D. (Eds.). (1996). Special issue on washback. *Language Testing*, *13*(3).

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge, England: Cambridge University Press.

Allen, P., & Davies, A. (Eds.). (1977). *The Edinburgh course in applied linguistics: Vol. 4. Testing and experimental methods*. Oxford, England: Oxford University Press.

ALTE Members. (1998). *Multilingual glossary of language testing terms*. Cambridge, England: UCLES/Cambridge University Press.

Assessment Reform Group. (2002). Assessment for learning: 10 principles. Retrieved November 4, 2008, from http://k1.ioe.ac.uk/tlrp/arg/CIE3.PDF

Association of Language Testers in Europe (ALTE). (1994). Code of practice. Retrieved November 5, 2008, from http://www.alte.org/quality_assurance/index.php

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, England: Oxford University Press.

Bachman, L. F. (2000). What, if any, are the limits of our responsibility for fairness in testing? In A. J. Kunnan (Ed.), *Fairness and validation in language assessment*. Selected papers from the 19th LTRC, Orlando, Florida (pp. 39–41). Cambridge, England: UCLES/Cambridge University Press.

Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford, England: Oxford University Press.

Bailey, K. M., & Brown, J. D. (1995). Language testing courses. What are they? In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 236–256). Clevedon, UK: Multilingual Matters.

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

Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge, England: Cambridge University Press.

Buck, G. (2001). *Assessing listening*. Cambridge, England: Cambridge University Press.

Carroll, B. J. (1980). *Testing communicative performance*. Oxford, England: Pergamon.

Carroll, J. B. (1961). Fundamental considerations in testing for English language proficiency of foreign students. In *Testing the English proficiency of foreign students* (pp. 30–40). Washington, DC: Center for Applied Linguistics.

Carroll, B. J., & Hall, P. (1985). *Make your own language tests: A practical guide to writing language performance tests*. Oxford, England: Pergamon.

Chapelle, C., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge, England: Cambridge University Press.

Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge, England: UCLES/Cambridge University Press.

Cheng, L., Watanabe, Y., & Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. London: Erlbaum.

Code of Fair Testing Practices in Education. (2004). Washington, DC: Joint Committee on Testing Practices.

Cohen, A. (1994). *Assessing language ability in the classroom* (2nd ed.). Boston, MA: Heinle and Heinle.

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

Davidson, F., & Lynch, B. (2002) *Testcraft: A teacher's guide to writing and using language test specifications*. New Haven, CT, and London: Yale University Press.

Davies, A. (1990). *Principles of language testing*. Oxford, England: Blackwell.

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

Davies, A. (2008b). *Assessing academic English: Testing English proficiency 1950–1989–The IELTS solution*. Cambridge, England: UCLES/Cambridge University Press.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge, England: UCLES/Cambridge University Press.

Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge, England: Cambridge University Press.

Downing, S. M., & Haladyna, T. M. (Eds.). (2006). *Handbook of test development*. New York: Routledge.

European Association for Language Testing and Assessment (EALTA). (2006). Guidelines for good practice in language testing and assessment. Retrieved November 4, 2008, from http://www.ealta.eu.org/guidelines.htm

Fox, J. (2008). Alternative assessment. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.): *Vol. 7. Language testing and assessment* (pp. 97–109). New York: Springer Science+Business Media.

Fulcher, G. (2003). *Testing second language speaking*. London: Longman.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London: Routledge.

Fulcher, G., & Thrasher, R. (1999/2000). Video FAQs: Introducing topics in language testing. Retrieved November 4, 2008, from http://iltaonline.com/index.php?option=com_weblinks&catid=14&Itemid=23

Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge, England: UCLES/Cambridge University Press.

Hamp-Lyons, L. (2008). Best practice in writing assessment. In L. Taylor & C. J. Weir (Eds.), *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity. Proceedings of the ALTE Berlin Conference May 2005* (pp. 321–335). Cambridge, England: UCLES/Cambridge University Press.

Harris, D. (1969). *Testing English as a second language*. New York: McGraw-Hill.

Hawkey, R. (2004*). A modular approach to testing English language skills: The development of the certificates in English language skills (CELS) examinations.* Cambridge, England: UCLES/Cambridge University Press.

Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge, England: UCLES/Cambridge University Press.

Hawkey, R. (2009). *Examining FCE and CAE: Key issues and recurring themes in developing the First Certificate in English and Certificate in Advanced English Exams*. Cambridge, England: UCLES/Cambridge University Press.

Heaton, B. (1975). *Writing English language tests*. London: Longman.

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

Henning, G. (1987). *A guide to language testing*. Cambridge, MA: Newbury House.
Hughes, A. (1989). *Testing for language teachers*. Cambridge, England: Cambridge University Press.
Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge, England: Cambridge University Press.
Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385–402.
International Language Testing Association (ILTA). (2001). Code of ethics. Retrieved November 3, 2008, from http://iltaonline.com/index.php?option=com_content&task=view&id=57&Itemid=47
International Language Testing Association (ILTA). (2007). Guidelines for practice. Retrieved November 3, 2008, from http://iltaonline.com/index.php?option=com_content&task=view&id=122&Itemid=133
Kunnan, A. J. (2000). Fairness and justice for all. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment*. Selected papers from the 19th LTRC, Orlando, Florida (pp. 1–14). Cambridge, England: UCLES/Cambridge University Press.
Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. J. Weir (Eds.), *European language testing in a global context* (pp. 27–48). Cambridge, England: UCLES/Cambridge University Press.
Kunnan, A. J. (2008). Towards a model of test evaluation: Using the test fairness and test context frameworks. In L. Taylor & C. J. Weir (Eds.), *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity*. *Proceedings of the ALTE Berlin Conference May 2005* (pp. 229–251). Cambridge, England: UCLES/Cambridge University Press.
Lado, R. (1961). *Language testing*. London: Longman.
Lantolf, J., & Poehner, M. (2008). Dynamic assessment. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.): *Vol. 7. Language testing and assessment* (pp. 273–284). New York: Springer Science+Business Media.
Luoma, S. (2004). *Assessing speaking*. Cambridge, England: Cambridge University Press.
Madsen, H. (1983). *Techniques in testing*. New York and Oxford, England: Oxford University Press.
Malone, M. (2008). Training in language assessment. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.): *Vol. 7. Language testing and assessment* (pp. 225–239). New York: Springer Science+Business Media.
McKay, P. (2006). *Assessing young language learners*. Cambridge, England: Cambridge University Press.
McNamara, T. (1996). *Measuring second language performance*. London: Longman.
McNamara, T. (2000). *Language testing*. Oxford, England: Oxford University Press.
McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA, and Oxford, England: Blackwell.
Melbourne, LTC. (1997). *Mark my words* [Video series]. Melbourne: Melbourne Language Testing Centre.

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

Menken, K., & Shohamy, E. (Eds.) (2008). Special issue on No Child Left Behind and U.S. language education policy. *Language Policy, 7*(3).

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13–103). Washington, DC: American Council on Education and the National Council on Measurement in Education.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(4), 241–256.

North, B. (2008). The CEFR levels and descriptor scales. In L. Taylor & C. J. Weir (Eds.), *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity. Proceedings of the ALTE Berlin Conference May 2005* (pp. 21–66). Cambridge, England: UCLES/Cambridge University Press.

Oller, J. W. (1979). *Language tests at school: A pragmatic approach*. London: Longman.

O'Sullivan, B. (2006). *Issues in testing business English: The revision of the Cambridge Business English Certificates*. Cambridge, England: UCLES/Cambridge University Press.

Purpura, J. (2004). *Assessing grammar*. Cambridge, England: Cambridge University Press.

Rea-Dickins, P. (2008). Classroom-based language assessment. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.): *Vol. 7. Language testing and assessment* (pp. 257–271). New York: Springer Science+Business Media.

Read, J. (2000). *Assessing vocabulary*. Cambridge, England: Cambridge University Press.

Saville, N. (2005). An interview with John Trim at 80. *Language Assessment Quarterly, 2*(4), 263–288.

Shaw, S. D., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing*. Cambridge, England: UCLES/Cambridge University Press.

Shohamy, E. (2008). Introduction. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.): *Vol. 7. Language testing and assessment* (pp. xiii–xxii). New York: Springer Science+Business Media.

Spolsky, B. (1995). *Measured words*. Oxford, England: Oxford University Press.

Spolsky, B. (2008). Introduction–Language testing at 25: Maturity and responsibility? *Language Testing, 25*(3), 297–305.

Stansfield, C. W. (2008). Where we have been and where we should go. *Language Testing, 25*(3), 311–326.

Stiggins, G. (1997). *Student centered classroom assessment*. Upper Saddle River, NJ: Prentice Hall.

Stoynoff, S., & Chapelle, C. (2005). *ESOL tests and testing: A resource for teachers and program administrators*. Alexandria, VA: TESOL.

Taylor, C., & Angelis, P. (2008). The evolution of the TOEFL. In C. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language*. New York and London: Routledge.

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press

Vallette, R. (1967). *Modern language tests: A handbook*. New York: Harcourt Brace and World.

Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. Cambridge, England: UCLES/Cambridge University Press.

Weigle, S. C. (2002). *Assessing writing*. Cambridge, England: Cambridge University Press.

Weir, C. J. (1990). *Communicative language testing*. Upper Saddle River, NJ: Prentice Hall.

Weir, C. J. (1993). *Understanding and developing language tests*. Upper Saddle River, NJ: Prentice Hall.

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.

Weir, C. J., & Milanovic, M. (2003). *Continuity and innovation: Revising the Cambridge Proficiency in English Examination 1913–2002*. Cambridge, England: UCLES/Cambridge University Press.

https://doi.org/10.1017/S0267190509090035 Published online by Cambridge University Press