Annual Review of Applied Linguistics (2009) **29**, 90–100. Printed in the USA.
Copyright © 2009 Cambridge University Press 0267-1905/09 $16.00
doi:10.1017/S0267190509090084

7. LANGUAGE ASSESSMENT IN EDUCATION: TESTS, CURRICULA, AND TEACHING

Alister Cumming

Over the past decade, many concerted policy efforts have aimed to change the status and functions of language assessment in school systems or higher education, redefining relationships among language tests, curriculum policies, and classroom teaching practices. Conventionally, formal language tests describe individual proficiency levels in reference to normative standards for purposes of certifying abilities; screening applicants for higher education, employment, or immigration decisions; or monitoring the results of educational systems. Recently, many curriculum policies have been reconceptualized in reference to attainment or benchmark standards that specify (usually functional, communicative) goals for language education, learners' achievements, and program accountability. These innovations have adopted principles of criterion-referenced rather than norm-referenced assessment, creating new relations (as well as dilemmas) between language assessment and new curriculum policies, highlighting the nature of language assessment practices in programs, classrooms, or other learning contexts, particularly the foundation bases for defining language proficiency, alignment between assessments and curricula, the formative purposes of assessment in pedagogy, and the situations and interests of particular learner populations.

Introduction

What should the relationships be among formal language tests, curricula for language learning, and pedagogical functions of formative assessment? Many differing realizations are possible. Some may be preferable. Most are contestable, contingent, and susceptible to situational variability. Recent debates in language education have centered on two fundamental issues amid efforts to change, redefine, or align these three elements. One issue concerns the basis for defining standards of language proficiency and learning opportunities: Should these come primarily from tests (as normative standards relative to the performance of learner populations on particular assessment instruments)? From policies (such as curriculum frameworks or attainment standards officially prescribed)? From teaching (by standards of

https://doi.org/10.1017/S0267190509090084 Published online by Cambridge University Press

pedagogical practice or as negotiated in unique circumstances of teaching and learning)? Or from some combination of these sources, as may often be the case, because each element inevitably informs the other?

The second issue concerns how integrally language tests, curricula, and pedagogical practices should be aligned and what benefits or consequences may arise when they are. Most recent publications have approached this issue from one of three perspectives. The first approach is to establish policies to align tests, curricula, and pedagogy through criterion-referenced assessment and standards-based curricula. The second approach has been to inquire into ongoing practices of teaching and learning to describe their qualities and determine their mutual influences, as in studies of teachers' assessment practices or of test washback. The third approach has been to adapt assessment policies and practices for particular populations, such as test accommodations for certain learner groups or setting performance standards for occupational purposes.

Bases for Defining Language Proficiency

The conventional rationale for curriculum planning (e.g., outlined in Tyler, 1949) is that educators determine the nature of educational programs first by establishing relevant objectives, then by creating and sequencing appropriate teaching and learning activities, and then later evaluating the extent to which students have achieved the objectives intended as a result of participating in the activities prescribed. Education in second and foreign languages, however, has often stood apart from this paradigm because schooling is not the sole (or necessarily the most effective) basis by which people develop language proficiency. People may develop language proficiency independent of schooling through interactions at home or in local subcommunities or while traveling. Moreover, many language programs are for mature learners, either in higher education or for settlement or work-related purposes, who tend to differ greatly in their abilities, backgrounds, and levels of language proficiency.

In these latter circumstances, evaluation assumes roles beyond the conventional function of determining students' achievements directly from an educational program: Language assessment commonly serves purposes of screening applicants for admission to programs of higher education or suitability for employment or immigration, monitoring outputs or variability within educational systems, supporting independent learning, or accommodating specific populations. These latter functions of language assessment have become prominent over the past two decades for a variety of reasons: increased international migration and mobility, the globalization of economic and educational opportunities in respect to a few lingua franca languages (such as English, French, Arabic, Chinese, and Spanish), the emergence of local varieties of these languages, concerns for the survival of other minority languages, and calls for accountability in education (Canagarajah, 2005, 2006; Elder & Davies, 2006; Shohamy, 2001, 2007). In short, language assessment itself has become a major policy issue in educational programs.

https://doi.org/10.1017/S0267190509090084 Published online by Cambridge University Press

Aligning Tests, Curricula, and Pedagogy

Most educational systems have, over the past two decades, established new curriculum policies that define standards for students' intended achievements in languages in terms of benchmarks, competencies, or attainment levels. A guiding rationale for these policies is to make explicit the results expected for students' achievements at specific points in educational programs so that these can be taught, studied, assessed, reported, as well as monitored and accounted for, it is hoped, with increased clarity, consistency, and precision (Brindley, 1998, 2000; Council of Europe, 2001; McKay, 2007; McKay et al., 2001; McNamara & Roever, 2006; Nunan, 2007; Trim, 1998). These policies can permit the alignment of curricula and assessment through criterion-referenced forms of assessment. That is, curriculum standards set "criteria" for teaching and for student evaluation (for placement, diagnosis, and achievement), rather than the normative-based approaches that have dominated the design and development of most major tests of language abilities, founded on piloting and analyses with populations outside of educational programs, which inevitably present disjunctures with educational programs (Brown & Hudson, 2002; Lynch & Davidson, 1994; Gipps & Cumming, 2005). In turn, new curriculum standards can promote functionally oriented conceptualizations of language abilities for teaching and assessments, focused on communicative competence in languages or arising from analyses of optimal teaching and learning practices (Brown & Hudson; McKay, 2007; Johnstone, 2000; North, 2000; Nunan, 2007; Purpura, 2008; Trim, 1998).

Research into the implementation of these curriculum policies, however, has exposed a variety of dilemmas. One dilemma is the relative imprecision of descriptive criteria, rating scales, and even sample benchmarks in view of the complexity of languages and human behavior. Moreover, most curriculum standards have been established through professional judgment or consensus rather than systematic empirical or theoretical inquiry. This dilemma challenges the transposition of curriculum standards into assessment instruments, as reported in many projects to implement the Common European Framework of Reference for Languages (CEFR) (Alderson, 2005; Byrnes, 2007). Indeed, variability tends to characterize teachers' interpretations and uses of language standards, rather than the expected consistency (Alabau et al., 2002; Brindley, 1998, 2000; Brindley & Burrows, 2001; Cumming, 2001a, 2001b; Davison, 2004; North, 2000). Moreover, fears exist that assessments fixed to language standards narrow the scope of teaching and learning, as in education at large (Cummins, 2000; Leung & Lewkowicz, 2006; Linn, 2000; Hamp-Lyons, 2007; Hillocks, 2002). In the United States, policies and legislation for No Child Left Behind have required simply that school systems demonstrate "improvement," resulting in great discrepancies in the ways and extent to which states and school boards have specified and acted on curriculum standards and relevant assessments to meet this policy (Harper, Platt, Naranjo, & Boynton, 2007; Rivera & Collum, 2006). This raises a second dilemma concerning the appropriate scope and legitimate political jurisdiction of curriculum standards, particularly when standards extend across vast geopolitical domains (such as all of Europe) or there is competition between educational authorities, teachers' professional associations, and local school or community concerns (Alabau et al., 2002; Byrnes, 2007; Canagarajah, 2005, 2006;

https://doi.org/10.1017/S0267190509090084 Published online by Cambridge University Press

Darling-Hammond, Ancess, & Falk, 1995; McKay, 2007; Moore, 2005; Shohamy, 2007).

A third dilemma is that successful adoption of new curriculum standards requires enormous investments in resources for teachers' professional development (Brindley & Burrows, 2001; Murray, 2008; North, 2000; Nunan, 2007; Slayter, 2003; Short et al., 2000; TESOL, 1998, 2001; Yoshida, 2003). A final dilemma arises implicitly from conceptualizations of "dynamic assessment," which argue, from principles of sociocultural theory, that language teaching and learning need to be conceived as integrally interactive, jointly constructed, and negotiated processes between teachers and learners, which cannot be prescribed or predicted by general curriculum policies (Lantolf & Poehner, 2008; Leung, 2007; Leung & Mohan, 2004; Poehner & Lantolf, 2005).

Pedagogical Practices

These curriculum innovations and dilemmas have prompted recent research to begin to describe classroom practices for language assessment. Descriptive case studies and surveys have started to document the functions and formats of teachers' assessment practices in language courses and in view of curriculum standards (Breen, Hird, Milton, Oliver, & Thwaite, 2001; Brindley, 2000; Colby-Kelly & Turner, 2007; Cumming, 2001b; Davison, 2004; Edelenbos & Kubanek-German, 2004; Grierson, 1995; Leung & Mohan, 2004; Rea-Dickins, 2001). A rallying slogan has been that assessment should be *for* learning (i.e., formative in nature) rather than *of* learning (i.e., not summative, as in conventional tests, even though few tests of language achievement actually exist—in the sense of being linked directly to curricula taught) (Colby-Kelly & Turner, 2007; Gipps & Cumming, 2005; Rea-Dickins, 2007). Despite these exploratory accounts, teachers' practices for responding to second-language students' writing remains one of the few areas that has been studied with sufficient scope to foster general principles, and even these are tentative and controversial (Ferris, 2003; Goldstein, 2005; Hyland & Hyland, 2006). Another development has been renewed interest in promoting learners' self-assessment, facilitated by the explicit descriptive criteria in curriculum standards (Alderson, 2005; Ekbatani & Pierson, 2000; Little, 2005; Ross, 1998). Studies investigating the potential washback of language tests on teaching and learning have proliferated, but evidence for direct causal impacts from assessments remains diffuse given the complex range of interrelated variables associated with assessment in language education (Bailey, 1996; Cheng, 2005; Cheng, Watanabe, & Curtis, 2004; Shi, 2007; Wall, 2000; Wall & Horak, 2006). Perhaps most revealing in this regard are studies of how language assessments or curriculum standards are interpreted and acted upon either by students (Huhta, Kalaja, & Pitkanen-Huhta, 2006; Johns, 1991; Shi, 2007) or by teachers (Brindley, 2000; Coniam & Falvey, 2007; Gardner, 2008; Wall & Horak, 2006).

Accommodations for Particular Populations or Purposes

Recent emphases on language assessment and curriculum standards have also given rise to two types of concerns for particular populations. For minority-language

https://doi.org/10.1017/S0267190509090084 Published online by Cambridge University Press

children learning English in mainstream education, numerous analyses have documented the unfairness inherent in applying curriculum standards and assessments designed for majority populations, and various adaptations and accommodations have been recommended for minority students (Abedi, Hofstetter, & Lord, 2004; Artiles & Ortiz, 2002; Bailey & Butler, 2004; Butler & Stevens, 2001; Cheng, Klinger, & Zheng, 2007; Cummins, 2000; McKay, 2005; Garcia, McKoon & August, 2006; McKay, 2005; Riveria & Collum, 2006; Solano-Flores, 2008; Solano-Flores & Trumbull, 2003; Stansfield, 2003). A wholly different concern arises in uses of language assessments to set minimum performance standards for certification for employment, professional practice, or curricula. In this vein, researchers have followed principles established in Cizek (2001) to formally set performance standards as minimum benchmark levels on major language tests, for example, among teachers or health professionals (Coniam & Falvey, 2007; Epp, Stawychny, Bonham, & Cumming, 2002; O'Neil, Buckendahl, Plake, & Taylor, 2007; Xi, 2007).

Summary

Language assessment has, in recent decades, become a major policy issue in education around the world, raising numerous fundamental issues about the functions, conceptualizations, and applications of assessment in respect to curricula, teaching, and learning. As educational programs adopt curriculum standards that define explicit criteria for students' expected language achievements, various challenges emerge: in aligning curricula and tests or the compatibility of different sets of language standards, in describing or promoting optimal pedagogical practices and conditions for learning, and in accommodating unique student populations. More fundamentally, questions arise about how constructs of language proficiency should be defined, by whom, in what contexts, and on the basis of what evidence. Although innovative, criterion-referenced assessments serve pedagogical purposes within educational systems, the authority of norm-referenced language tests persists—particularly where benchmark norms are required apart from those defined by or within educational programs—for such consequential policy decisions as selection into programs, professional licensure, or monitoring educational systems themselves.

ANNOTATED REFERENCES

Alderson, J. C. (2005). *Diagnosing foreign language proficiency*. London: Continuum.

Alderson outlined the value, uses, and limitations of assessments of language abilities for diagnostic purposes, describing research and problematic issues related to the DIALANG project, a Web-based self-assessment instrument in 14 European languages related to the CEFR.

https://doi.org/10.1017/S0267190509090084 Published online by Cambridge University Press

Brindley, G. (Ed.). (2000). *Studies in immigrant English language assessment* (Vol. 1). Sydney, Australia: National Centre for English Language Teaching and Research, Macquarie University.

Brindley, G., & Burrows, C. (Eds.). (2001). *Studies in immigrant English language assessment* (Vol. *2*). Sydney, Australia: National Centre for English Language Teaching and Research, Macquarie University.

These companion volumes describe a range of empirical studies as well as the rationales and contextual factors related to assessment policies and practices in the Adult Migrant Education Program in Australia, particularly issues related to implementation of the Certificates in Spoken and Written English.

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

Brown and Hudson described the rationale, various methods of test development and analysis, and educational functions of criterion-referenced assessment in language programs. In addition to recounting some of their own studies in Hawaii, the volume is directed at readers wanting to learn about criterion-referenced language assessments to use them in language programs.

North, B. (2000). *The development of a common framework scale of language proficiency*. New York: Peter Lang.

North described a variety of empirical studies as well as theoretical and policy rationales that underpin the initial development and subsequent refinements of descriptions of language proficiency and rating scales in the CEFR.

Rivera, C., & Collum, E. (Eds.). (2006). *State assessment policy and practice for English language learners: A national perspective*. Mahwah, NJ: Erlbaum.

Contributors to this book report on surveys that describe policies and practices to account for English language learners in the various states of the United States, providing definitive and comprehensive details about these assessments but revealing how they demonstrate considerable variation at the levels of state, district, and school.

OTHER REFERENCES

Abedi, J., Hofstetter, C., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*(1), 1–28.

https://doi.org/10.1017/S0267190509090084 Published online by Cambridge University Press

Alabau, I., Bonnet, G., de Bot, K., Bramsbye, J., Dauphin, L, Erickson, G., et al. (2002). *The assessment of pupils' skills in English in eight European countries*. Paris: European Network of Policy Makers for the Evaluation of Education Systems.

Artiles, A. J., & Ortiz, A. A. (Eds.). (2002). *English language learners with special education needs: Identification, assessment, and instruction*. Washington, DC: Center for Applied Linguistics and Delta Systems.

Bailey, A., & Butler, F. (2004). Ethical considerations in the assessment of the language and content knowledge of U.S. school-age learners. *Language Assessment Quarterly, 1*(2&3), 177–193.

Bailey, K. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing, 13*(2), 257–279.

Breen, M., Hird, B., Milton, M., Oliver, R., & Thwaite, A. (2001). Making sense of language teaching: Teachers' principles and classroom practices. *Applied Linguistics, 22*(4), 470–501.

Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programmes: A review of the issues. *Language Testing, 15*(1), 45–85.

Byrnes, H. (Ed.). (2007). Perspectives. *Modern Language Journal, 91*(4), 641–685.

Butler, F., & Stevens, R. (2001). Standardized assessment of the content knowledge of English language learners K–12: Current trends and old dilemmas. *Language Testing, 18*, 409–427.

Canagarajah, S. (Ed.). (2005). *Reclaiming the local in language policy and practice*. Mahwah, NJ: Erlbaum.

Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly, 3*(3), 229–242.

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

Cheng, L., Klinger, D., & Zheng, Y. (2007). The challenges of the Ontario Secondary School Literacy Test for second language students. *Language Testing, 24*(2), 185–208.

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

Cizek, B. (Ed.). (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Erlbaum.

Colby-Kelly, C., & Turner, C. (2007). AFL research in the L2 classroom and evidence of usefulness: Taking formative assessment to the next level. In A. Cumming & M. Laurier (Eds.), *Language assessment*, special issue of *Canadian Modern Language Review, 64*(1), 7–36.

Coniam, D., & Falvey, P. (2007). High-stakes testing and assessment: English language teacher benchmarking. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching: Part one* (pp. 456–471). New York: Springer.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge, England: Cambridge University Press.

https://doi.org/10.1017/S0267190509090084 Published online by Cambridge University Press

Cumming, A. (2001a). The difficulty of standards, for example in second language writing. In T. Silva & P. Matsuda (Eds.), *On second language writing* (pp. 209–229). Mahwah, NJ: Erlbaum.

Cumming, A. (2001b). ESL/EFL instructors' practices for writing assessment: Specific or general purposes? *Language Testing, 18*,(2), 207–224.

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Clevedon, England: Multilingual Matters.

Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work*. New York: Teachers College Press.

Davison, C. (2004). The contradictory culture of classroom-based assessment: Teacher assessment practices in senior secondary English. *Language Testing, 21*(3), 304–334.

Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: The concept of "diagnostic competence." *Language Testing, 21*(2), 259–283.

Ekbatani, G., & Pierson, H. (Eds.) (2000). *Learner-directed assessment in ESL.* Mahwah, NJ: Erlbaum.

Elder, C., & Davies, A. (2006). Assessing English as a lingua franca. *Annual Review of Applied Linguistics, 26*, 282–301.

Epp, L., Stawychny, M., Bonham, A., & Cumming, A. (2002). Benchmarking the English language demands of the nursing profession across Canada. Winnipeg: Red River Community College. Report submitted to the Centre for Canadian Language Benchmarks. Retrieved February 18, 2009, from http://www.language.ca.

Ferris, D. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Erlbaum.

Gardner, S. (2008). Transforming talk and phonics practice: Or, how *do* crabs clap? *TESOL Quarterly, 42*(2), 261–284.

Garcia, G., McKoon, G., & August, D. (2006). Language and literacy assessment of language-minority students. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-minority Children and Youth* (pp. 597–624). Mahwah, NJ: Erlbaum.

Gipps, C., & Cumming, J. J. (2005). Assessing literacies. In N. Bascia, A. Cumming, A. Datnow, K. Leithwood, & D. Livingstone (Eds.), *International handbook of educational policy* (Vol. 2, pp. 695–713). Dordrecht, Netherlands. Springer.

Goldstein, L. (2005). *Teacher written commentary in second language writing classrooms*. Ann Arbor: University of Michigan Press.

Grierson, J. (1995). Classroom-based assessment in Intensive English Centres. In G. Brindley (Ed.), *Language assessment in action* (pp. 195–236). Macquarie University: National Centre for English Language Teaching and Research, Australia.

Hamp-Lyons, L. (2007). The impact of testing practices on teaching: Ideologies and alternatives. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching: Part one* (pp. 487–504). New York: Springer.

Harper, C., Platt, E., Naranjo, C., & Boynton, S. (2007). Marching in unison: Florida ESL teachers and No Child Left Behind. *TESOL Quarterly, 41*(3), 642–651.

https://doi.org/10.1017/S0267190509090084 Published online by Cambridge University Press

Hillocks, G. (2002). *The testing trap: How state writing tests control learning*. New York: Teachers College Press.

Huhta, A., Kalaja, P., & Pitkanen-Huhta, A. (2006). The discursive construction of a high-stakes test: The many faces of a test-taker. *Language Testing, 23*(3), 326–350.

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching, 39*(2), 83–101.

Johns, A. (1991). Interpreting an English competency exam: The frustrations of an ESL science student. *Written Communication, 8*(3), 379–401.

Johnstone, R. (2000). Context-sensitive assessment of modern languages in primary (elementary) and early secondary education: Scotland and the European experience. *Language Testing, 17*(2), 123–143.

Lantolf, J., & Poehner, M. (2008). Dynamic assessment. In E. Shohamy (Vol. Ed.) & N. Hornberger (Series Ed.), *Language testing and assessment: Vol. 7. Encyclopedia of language and education* (2nd ed., pp. 273–284). New York: Springer.

Leung, C. (2007). Dynamic assessment: Assessment *for* and *as* teaching. *Language Assessment Quarterly, 4*(3), 257–278.

Leung, C., & Lewkowicz, J. (2006). Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly, 40*(1), 211–234.

Leung, C., & Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts: Assessment *as* discourse and assessment *of* discourse. *Language Testing, 21*(3), 335–359.

Linn, R. L. (2000). Assessment and accountability. *Educational Researcher, 29*(2), 4–16.

Little, D. (2005). The Common European Framework and the European Language Portfolio: Involving learners and their judgements in the assessment process. *Language Testing, 22*(3), 321–336.

Lynch, B., & Davidson, F. (1994). Criterion-reference test development: Linking curricula, teachers, and tests. *TESOL Quarterly, 28*(4), 727–743.

McKay, P. (2005). *Assessing young language learners*. New York: Cambridge University Press.

McKay, P. (2007). The standards movement and ELT for school-aged learners: Cross national perspectives. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching: Part one* (pp. 439–456). New York: Springer.

McKay, P., Coppari, P., Cumming, A., Graves, K., Lopriore, L., & Short, D. (2001). Language standards: An international perspective. Parts 1 & 2. *TESOL Matters, 11*(2), *1–4*(3), 11, 15.

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

Moore, H. (2005). Telling what is real: Competing views in assessing ESL development. *Linguistics and Education, 8*(2), 189–228.

Murray, D. (Ed.). (2008). *Planning change, changing plans: Innovations in second language teaching*. Ann Arbor: University of Michigan Press.

https://doi.org/10.1017/S0267190509090084 Published online by Cambridge University Press

Nunan, D. (2007). Standards-based approaches to the evaluation of ESL instruction. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching: Part one* (pp. 421–438). New York: Springer.

O'Neil, T., Buckendahl, C., Plake, B., & Taylor, L. (2007). Recommending a nursing-specific passing standard for the IELTS examination. *Language Assessment Quarterly, 4*(4), 295–317.

Poehner, M., & Lantolf, J. (2005). Dynamic assessment in the language classroom. *Language Teaching Research, 9*(3), 233–265.

Purpura, J. (2008). Assessing communicative language ability: Models and their components. In E. Shohamy (Vol. Ed.) & N. Hornberger (Series Ed.), *Language testing and assessment: Vol. 7. Encyclopedia of language and education* (2nd ed., pp. 53–68). New York: Springer.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18*(4), 429–462.

Rea-Dickins, P. (2007). Classroom-based assessment: Possibilities and pitfalls. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching: Part one* (pp. 505–520). New York: Springer.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing, 15*(1), 1–20.

Shi, C. (2007). A new washback model of students' learning. In A. Cumming & M. Laurier (Eds.), *Language assessment*, special issue of *Canadian Modern Language Review, 64*(1), 133–160.

Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. New York: Pearson Education.

Shohamy, E. (2007). The power of language tests, the power of the English language and the role of ELT. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching: Part one* (pp. 521–531). New York: Springer.

Short, D., Gomez, E., Cloud, N., Katz, A., Gottlieb, M., & Malone, M. (2000). *Training others to use the ESL standards: A professional development manual*. Alexandria, VA: TESOL.

Slayter, H. (2003). Responding to change in English language assessment. *Prospect, 18*(1), 42–52.

Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilities views of language in the testing of English language learners. *Educational Researcher, 37*(4), 189–199.

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher, 32*(2), 3–13.

Stansfield, C. (2003). Test translation and adaptation in public education in the USA. *Language Testing, 20*(2), 188–206.

Teachers of English to Speakers of Other Languages (TESOL). (1998). *Managing the assessment process: A framework for measuring student attainment of the ESL standards*. Alexandria, VA: TESOL.

Teachers of English to Speakers of Other Languages (TESOL). (2001). *Scenarios for ESL standards-based assessment*. Alexandria, VA: TESOL.

https://doi.org/10.1017/S0267190509090084 Published online by Cambridge University Press

Trim, J. (1998). European perspectives on modern language learning: Contributions to the Modern Languages Project of the Council of Europe. *Language Teaching, 31*(2), 206–217.

Tyler, R. (1949). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.

Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System, 28*(4), 499–509.

Wall, D., & Horak, T. (2006). *The impact of changes in the TOEFL examination on teaching and learning in central and eastern Europe: Phase 1, the baseline study*. TOEFL Monograph MS 34. Princeton, NJ: Educational Testing Service. Retrieved February 18, 2009, from http://www.ets.org/portal/site/ets

Yoshida, K. (2003). Language education policy in Japan—The problem of espoused objectives versus practice. *Modern Language Journal, 87*(2), 290–292.

Xi, X. (2007). Validating TOEFL iBT speaking and setting score requirements for ITA screening. *Language Assessment Quarterly, 4*(4), 318–351.

https://doi.org/10.1017/S0267190509090084 Published online by Cambridge University Press