

Bibliography

- ACTFL. (2012). *ACTFL proficiency guidelines 2012*. Retrieved from <http://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012>
- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice Hall.
- ARG. (1999). *Assessment for learning: Beyond the black box*. Cambridge, England: University of Cambridge Faculty of Education.
- ARG. (2002a). *Assessment for learning: 10 principles*. Retrieved from <http://webarchive.nationalarchives.gov.uk/20101021152907/http://www.ttrb.ac.uk/ViewArticle2.aspx?ContentId=15313>
- ARG. (2002b). *Testing, motivation and learning*. Cambridge, England: University of Cambridge Faculty of Education.
- ARG. (2006). *The role of teachers in the assessment of learning*. London, England: University of London Institute of Education.
- Australian Council for Educational Research. (2002). *Report on the New Zealand national curriculum*. Melbourne, Australia: ACER.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, England: Oxford University Press.
- Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly*, 25, 671–704. <http://dx.doi.org/10.2307/3587082>
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1–42. <http://dx.doi.org/10.1177/026553220001700101>
- Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19(4), 453–476. <http://dx.doi.org/10.1191/0265532202lt240oa>
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, England: 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, England: Oxford University Press.
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university: What the student does* (4th ed.). Maidenhead, England: McGraw-Hill/Open University Press.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <http://dx.doi.org/10.1191/1478088706qp063oa>
- Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing*, 26(3), 341–366. <http://dx.doi.org/10.1177/0265532209104666>
- Brown, H. D. (2007). *Principles of language learning and teaching* (5th ed.). New York, NY: Pearson.

- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). New York, NY: Pearson.
- Bryman, A. (2004a). Member validation and check. In M. Lewis-Beck, A. Bryman, & T. Liao (Eds.), *Encyclopedia of social science research methods* (p. 634). Thousand Oaks, CA: Sage. <http://dx.doi.org/10.4135/9781412950589.n548>
- Bryman, A. (2004b). Triangulation. In M. B. Lewis-Beck, A. Bryman, & T. Liao (Eds.), *Encyclopedia of social science research methods* (pp. 1143–1144). Thousand Oaks, CA: Sage. <http://dx.doi.org/10.4135/9781412950589.n1031>
- Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, 10(1), 15–42.
- Buck, G. (1992). Translation as a language testing procedure: Does it work? *Language Testing*, 9(2), 123–148. <http://dx.doi.org/10.1177/026553229200900202>
- Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136–146). Oxford, England: Macmillan.
- Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks* (pp. 23–48). Harlow, England: Longman.
- Bygate, M. (2005). Oral second language abilities as expertise. In K. Johnson (Ed.), *Expertise in second language learning and teaching* (pp. 104–127). New York, NY: Palgrave Macmillan.
- Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Clevedon, England: Multilingual Matters.
- Byram, M. (2000). Assessing intercultural competence in language teaching. *Sprogforum*, 18(6), 8–13.
- Byram, M. (2008). *From foreign language education to education for intercultural citizenship: Essays and reflections*. Clevedon, England: Multilingual Matters.
- Byram, M. (2009). Intercultural competence in foreign languages: The intercultural speaker and the pedagogy of foreign language education. In D. K. Deardorff (Ed.), *The Sage handbook of intercultural competence* (pp. 321–332). Thousand Oaks, CA: Sage.
- Byram, M., Gribkova, B., & Starkey, H. (2002). *Developing the intercultural dimension in language teaching: A practical introduction for teachers*. Strasbourg, France: Council of Europe.
- Byram, M., Holmes, P., & Savvides, N. (2013). Intercultural communicative competence in foreign language education: Questions of theory, practice and research. *The Language Learning Journal*, 41(3), 251–253. <http://dx.doi.org/10.1080/09571736.2013.836343>
- Cambridge English language assessment. (2015). Retrieved from <http://www.cambridgeenglish.org/exams/>
- Canale, M. (1983). On some dimensions of language proficiency. In J. W. J. Oller (Ed.), *Issues in language testing research* (pp. 333–342). Rowley, MA: Newbury House.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47. <http://dx.doi.org/10.1093/applin/i.1.1>
- Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, 19, 254–272. <http://dx.doi.org/10.1017/s0267190599190135>
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38–54. <http://dx.doi.org/10.1080/09500789708666717>
- Clapham, C. (2000). Assessment and testing. *Annual Review of Applied Linguistics*, 20, 147–161. <http://dx.doi.org/10.1017/s0267190500200093>
- Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological testing and assessment: An introduction to tests and measurement* (6th ed.). New York, NY: McGraw Hill.
- Council of Europe. (1998). *Modern languages: Teaching, assessment. A common European framework of reference*. Strasbourg, France: Council of Europe.
- Council of Europe. (2001). *Common European framework of reference for languages*. Cambridge, England: Cambridge University Press.

- Crocker, L. (2002). Stakeholders in comprehensive validation of standards-based assessments: A commentary. *Educational Measurement: Issues and Practice*, 22, 5–6. <http://dx.doi.org/10.1111/j.1745-3992.2002.tb00079.x>
- Crooks, T. (2010). New Zealand: Empowering teachers and children. In I. C. Rotberg (Ed.), *Balancing change and tradition in global education reform* (2nd ed., pp. 281–310). Lanham, MD: Rowman and Littlefield Education.
- Csépes, I. (2002). Is testing speaking in pairs disadvantageous for students? Effects on oral test scores. *novELTj*, 9(1), 22–45.
- Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing*, 26(3), 367–396. <http://dx.doi.org/10.1177/0265532209104667>
- De Ridder, I., Vangehuchten, L., & Seseña Gómez, M. (2007). Enhancing automaticity through task-based language learning. *Applied Linguistics*, 28(2), 309–315. <http://dx.doi.org/10.1093/applin/aml057>
- DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). Cambridge, England: Cambridge University Press. <http://dx.doi.org/10.1017/cbo9781139524780.007>
- Denzin, N. K. (1970). *The research act in sociology*. Chicago, IL: Aldine.
- Dervin, F. (2010). Assessing intercultural competence in Language Learning and Teaching: A critical review of current efforts. In F. Dervin & E. Suomela-Salmi (Eds.), *New approaches to assessment in higher education* (pp. 157–173). Bern, Switzerland: Peter Lang.
- Dobric, K. (2006). Drawing on discourses: Policy actors in the debates over the National Certificate of Educational Achievement 1996–2000. *New Zealand Annual Review of Education*, 15, 85–109.
- Ducasse, A., & Brown, A. (2009). Assessing paired orals: Raters' orientation to interaction. *Language Testing*, 26(3), 423–443. <http://dx.doi.org/10.1177/0265532209104669>
- East, M. (2005). Using support resources in writing assessments: Test taker perceptions. *New Zealand Studies in Applied Linguistics*, 11(1), 21–36.
- East, M. (2007). Bilingual dictionaries in tests of L2 writing proficiency: Do they make a difference? *Language Testing*, 24(3), 331–353. <http://dx.doi.org/10.1177/0265532207077203>
- East, M. (2008a). *Dictionary use in foreign language writing exams: Impact and implications*. Amsterdam, Netherlands/Philadelphia, PA: John Benjamins. <http://dx.doi.org/10.1075/llt.22>
- East, M. (2008b). Language evaluation policies and the use of support resources in assessments of language proficiency. *Current Issues in Language Planning*, 9(3), 249–261. <http://dx.doi.org/10.1080/14664200802139539>
- East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing*, 14(2), 88–115. <http://dx.doi.org/10.1016/j.asw.2009.04.001>
- East, M. (2012). *Task-based language teaching from the teachers' perspective: Insights from New Zealand*. Amsterdam, Netherlands/Philadelphia, PA: John Benjamins. <http://dx.doi.org/10.1075/tblt.3>
- East, M. (2013, August 24). *The new NCEA 'interact' standard: Teachers' thinking about assessment reform*. Paper presented at the New Zealand Association of Language Teachers (NZALT) Auckland/Northland Region language seminar, Auckland.
- East, M. (2014a, July 6–9). *To interact or not to interact? That is the question*. Keynote address at the New Zealand Association of Language Teachers (NZALT) Biennial National Conference, *Languages Give You Wings*, Palmerston North, NZ.
- East, M. (2014b). Working for positive outcomes? The standards-curriculum alignment for *Learning Languages* and its reception by teachers. *Assessment Matters*, 6, 65–85.
- East, M. (2015a). Coming to terms with innovative high-stakes assessment practice: Teachers' viewpoints on assessment reform. *Language Testing*, 32(1), 101–120. <http://dx.doi.org/10.1177/0265532214544393>
- East, M. (2015b). Taking communication to task – again: What difference does a decade make? *The Language Learning Journal*, 43(1), 6–19. <http://dx.doi.org/10.1080/09571736.2012.723729>

- East, M., & Scott, A. (2011a). Assessing the foreign language proficiency of high school students in New Zealand: From the traditional to the innovative. *Language Assessment Quarterly*, 8(2), 179–189. <http://dx.doi.org/10.1080/15434303.2010.538779>
- East, M., & Scott, A. (2011b). Working for positive washback: The standards-curriculum alignment project for *Learning Languages*. *Assessment Matters*, 3, 93–115.
- Edge, J., & Richards, K. (1998). May I see your warrant please?: Justifying outcomes in qualitative research. *Applied Linguistics*, 19, 334–356. <http://dx.doi.org/10.1093/applin/19.3.334>
- Education Counts. (2012). *Subject enrolment*. Retrieved from http://www.educationcounts.govt.nz/statistics/schooling/july_school_roll_returns/6052
- Együd, G., & Glover, P. (2001). Readers respond. Oral testing in pairs – Secondary school perspective. *ELT Journal*, 55(1), 70–76. <http://dx.doi.org/10.1093/elt/55.1.70>
- Elder, C. (1997). What does test bias have to do with fairness? *Language Testing*, 14(3), 261–277. <http://dx.doi.org/10.1177/026553229701400304>
- Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: What does the test-taker have to offer? *Language Testing*, 19(4), 347–368. <http://dx.doi.org/10.1191/0265532202lt2350a>
- Ellis, R. (Ed.). (2005). *Planning and task performance in a second language*. Amsterdam, Netherlands/Philadelphia, PA: John Benjamins. <http://dx.doi.org/10.1075/llt.11>
- Foot, M. C. (1999). Relaxing in pairs. *ELT Journal*, 53(1), 36–41. <http://dx.doi.org/10.1093/elt/53.1.36>
- Fulcher, G. (1996). Testing tasks: Issues in task design and the group oral. *Language Testing*, 13(1), 23–51. <http://dx.doi.org/10.1177/026553229601300103>
- Fulcher, G. (2003). *Testing second language speaking*. Harlow, England: Pearson. <http://dx.doi.org/10.4324/9781315837376>
- Galaczi, E. D. (2010). Paired speaking tests: An approach grounded in theory and practice. In J. Mader & Z. Ürkün (Eds.), *Recent approaches to teaching and assessing speaking*. IATEFL TEA SIG conference proceedings. Canterbury, England: IATEFL Publications.
- Galaczi, E. D., & ffrench, A. (2011). Context validity. In L. Taylor (Ed.), *Examining speaking: Research and practice in assessing second language speaking* (pp. 112–170). Cambridge, England: Cambridge University Press.
- Gardner, J., Harlen, W., Hayward, L., & Stobart, G. (2008). *Changing assessment practice: Process, principles and standards*. Belfast, Northern Ireland: Assessment Reform Group.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London, England: The Falmer Press. <http://dx.doi.org/10.4324/9780203486009>
- Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Buckingham, England: Open University Press.
- Gov.UK. (2015). *Get the facts: GCSE reform*. Retrieved from <https://www.gov.uk/government/publications/get-the-facts-gcse-and-a-level-reform/get-the-facts-gcse-reform>
- Graham, J. W. (2012). *Missing data: Analysis and design*. New York, NY: Springer.
- Haertel, E. H. (2002). Standard setting as a participatory process: Implications for validation of standards-based accountability programs. *Educational Measurement: Issues and Practice*, 22, 16–22. <http://dx.doi.org/10.1111/j.1745-3992.2002.tb00081.x>
- Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. In L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P. M. Johnston, & M. Rees (Eds.), *Tertiary assessment and higher education student outcomes: Policy, practice and research* (pp. 259–275). Wellington, NZ: Ako Aotearoa.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <http://dx.doi.org/10.3102/003465430298487>
- Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford, England: Oxford University Press.
- Higgs, T. V. (Ed.). (1984). *Teaching for proficiency: The organizing principle*. Lincolnwood, IL: National Textbook Company.

- Hinkel, E. (2010). Integrating the four skills: Current and historical perspectives. In R. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (2nd ed., pp. 110–123). Oxford, England: Oxford University Press. <http://dx.doi.org/10.1093/oxfordhb/9780195384253.013.0008>
- Hipkins, R. (2013). *NCEA one decade on: Views and experiences from the 2012 NZCER National Survey of Secondary Schools*. Wellington, NZ: New Zealand Council for Educational Research.
- Hu, G. (2013). Assessing English as an international language. In L. Alsagoff, S. L. McKay, G. Hu, & W. A. Renandya (Eds.), *Principles and practices for teaching English as an international language* (pp. 123–143). New York, NY: Routledge.
- Hunter, D. (2009). *Communicative language teaching and the ELT Journal: A corpus-based approach to the history of a discourse*. Unpublished doctoral thesis. University of Warwick, Warwick, England.
- Iwashita, N. (1996). The validity of the paired interview in oral performance assessment. *Melbourne Papers in Language Testing*, 5(2), 51–65.
- Jacoby, S., & Ochs, E. (1995). Co-construction: An introduction. *Research on Language and Social Interaction*, 28(3), 171–183.
- Kane, M. J. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, 21(1), 31–42. <http://dx.doi.org/10.1111/j.1745-3992.2002.tb00083.x>
- Kaplan, R. M., & Saccuzzo, D. P. (2012). *Psychological testing: Principles, applications, and issues* (8th ed.). Belmont, CA: Wadsworth, Cengage Learning.
- Klapper, J. (2003). Taking communication to task? A critical review of recent trends in language teaching. *Language Learning Journal*, 27, 33–42. <http://dx.doi.org/10.1080/09571730385200061>
- Kline, P. (2000). *Handbook of psychological testing* (2nd ed.). London, England: Routledge. <http://dx.doi.org/10.4324/9781315812274>
- Koefoed, G. (2012). Policy perspectives from New Zealand. In M. Byram & L. Parmenter (Eds.), *The common European framework of reference: The globalisation of language education policy* (pp. 233–247). Clevedon, England: Multilingual Matters.
- Kramsch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal*, 70(4), 366–372. <http://dx.doi.org/10.1111/j.1540-4781.1986.tb05291.x>
- Kramsch, C. (1987). The proficiency movement: Second language acquisition perspectives. *Studies in Second Language Acquisition*, 9(3), 355–362. <http://dx.doi.org/10.1017/s0272263100006732>
- Kramsch, C. (2005). Post 9/11: Foreign languages between knowledge and power. *Applied Linguistics*, 26(4), 545–567. <http://dx.doi.org/10.1093/applin/ami026>
- Kunnan, A. J. (2000). Fairness and justice for all. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 1–14). Cambridge, England: Cambridge University Press.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <http://dx.doi.org/10.2307/2529310>
- Language Testing International. (2014). *ACTFL Oral Proficiency Interview by Computer (OPIC)*. Retrieved from <http://www.languagetesting.com/oral-proficiency-interview-by-computer-opic>
- Lazaraton, A. (1995). Qualitative research in applied linguistics: A progress report. *TESOL Quarterly*, 29(3), 455–472. <http://dx.doi.org/10.2307/3588071>
- Lazaraton, A. (2002). *A qualitative approach to the validation of oral language tests*. Cambridge, England: Cambridge University Press.
- Leaper, D. A., & Riazi, M. (2014). The influence of prompt on group oral tests. *Language Testing*, 31(2), 177–204. <http://dx.doi.org/10.1177/0265532213498237>
- Leung, C. (2005). Convivial communication: Recontextualizing communicative competence. *International Journal of Applied Linguistics*, 15(2), 119–144. <http://dx.doi.org/10.1111/j.1473-4192.2005.00084.x>
- Leung, C. (2007). Dynamic assessment: Assessment for and as teaching? *Language Assessment Quarterly*, 4(3), 257–278. <http://dx.doi.org/10.1080/15434300701481127>
- Leung, C., & Lewkowicz, J. (2006). Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly*, 40(1), 211–234. <http://dx.doi.org/10.2307/40264517>

- Lewkowicz, J. (2000). Authenticity in language testing: Some outstanding questions. *Language Testing*, 17(1), 43–64. <http://dx.doi.org/10.1177/026553220001700102>
- Liddicoat, A. (2005). Teaching languages for intercultural communication. In D. Cunningham & A. Hatoss (Eds.), *An international perspective on language policies, practices and proficiencies* (pp. 201–214). Belgrave, Australia: Fédération Internationale des Professeurs de Langues Vivantes (FIPLV).
- Liddicoat, A. (2008). Pedagogical practice for integrating the intercultural in language teaching and learning. *Japanese Studies*, 28(3), 277–290. <http://dx.doi.org/10.1080/10371390802446844>
- Liddicoat, A., & Crozet, C. (Eds.). (2000). *Teaching languages, teaching cultures*. Melbourne, Australia: Language Australia.
- Lo Bianco, J., Liddicoat, A., & Crozet, C. (Eds.). (1999). *Striving for the third place: Intercultural competence through language education*. Melbourne, Australia: Language Australia.
- Long, M. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, 4(2), 126–141. <http://dx.doi.org/10.1093/applin/4.2.126>
- Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York, NY: Academic.
- Long, M., & Norris, J. (2000). Task-based teaching and assessment. In M. Byram (Ed.), *Routledge encyclopedia of language teaching and learning* (pp. 597–603). London, England: Routledge.
- Luoma, S. (2004). *Assessing speaking*. Cambridge, England: Cambridge University Press. <http://dx.doi.org/10.1017/cbo9780511733017>
- Madaus, G. F., & Kellaghan, T. (1992). Curriculum evaluation and assessment. In P. W. Jackson (Ed.), *Handbook on research on curriculum* (pp. 119–154). New York, NY: Macmillan.
- Mangubhai, F., Marland, P., Dashwood, A., & Son, J. B. (2004). Teaching a foreign language: One teacher's practical theory. *Teaching and Teacher Education*, 20, 291–311. <http://dx.doi.org/10.1016/j.tate.2004.02.001>
- Martinez-Flor, A., Usó-Juan, E., & Alcón Soler, E. (2006). Towards acquiring communicative competence through speaking. In E. Usó-Juan & A. Martínez-Flor (Eds.), *Studies on language acquisition: Current trends in the development and teaching of the four language skills* (pp. 139–157). Berlin, Germany/New York, NY: Walter de Gruyter. <http://dx.doi.org/10.1515/9783110197778.3.139>
- May, L. (2009). Co-constructed interaction in a paired speaking test: The rater's perspective. *Language Testing*, 26(3), 397–422. <http://dx.doi.org/10.1177/0265532209104668>
- May, L. (2011). Interactional competence in a paired speaking test: Features salient to raters. *Language Assessment Quarterly*, 8(2), 127–145. <http://dx.doi.org/10.1080/15434303.2011.565845>
- McNamara, T. (1996). *Measuring second language performance*. London, England: Longman.
- McNamara, T. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics*, 18(4), 446–466. <http://dx.doi.org/10.1093/applin/18.4.446>
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell.
- Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749. <http://dx.doi.org/10.1037//0003-066x.50.9.741>
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. <http://dx.doi.org/10.1177/026553229601300302>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

- Ministry of Education. (1993). *The New Zealand curriculum framework*. Wellington, NZ: Learning Media.
- Ministry of Education. (1995a). *Chinese in the New Zealand curriculum*. Wellington, NZ: Learning Media.
- Ministry of Education. (1995b). *Spanish in the New Zealand curriculum*. Wellington, NZ: Learning Media.
- Ministry of Education. (1998). *Japanese in the New Zealand curriculum*. Wellington, NZ: Learning Media.
- Ministry of Education. (2002a). *French in the New Zealand curriculum*. Wellington, NZ: Learning Media.
- Ministry of Education. (2002b). *German in the New Zealand curriculum*. Wellington, NZ: Learning Media.
- Ministry of Education. (2007). *The New Zealand curriculum*. Wellington, NZ: Learning Media.
- Ministry of Education. (2010). *Learning languages – Curriculum guides*. Retrieved from <http://learning-languages-guides.tki.org.nz/>
- Ministry of Education. (2011a). *Ministry of Education position paper: Assessment (schooling sector)*. Wellington, NZ: Ministry of Education.
- Ministry of Education. (2011b). *New Zealand curriculum guides senior secondary: Learning languages*. Retrieved from <http://seniorsecondary.tki.org.nz/Learning-languages>
- Ministry of Education. (2012a). *Secondary student achievement*. Retrieved from <http://nzcurriculum.tki.org.nz/System-of-support-incl.-PLD/School-initiated-supports/Professional-learning-and-development/Secondary-achievement>
- Ministry of Education. (2012b). *What's new or different?* Retrieved from <http://seniorsecondary.tki.org.nz/Learning-languages/What-s-new-or-different>
- Ministry of Education. (2014a). *Learning languages – Achievement objectives*. Retrieved from <http://nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum/Learning-areas/Learning-languages/Achievement-objectives>
- Ministry of Education. (2014b). *Resources for internally assessed achievement standards*. Retrieved from <http://ncea.tki.org.nz/Resources-for-Internally-Assessed-Achievement-Standards>
- Mislevy, R., Wilson, M. R., Ercikan, K., & Chudowsky, N. (2003). Psychometric principles in student assessment. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (Vol. 9, pp. 489–531). Dordrecht, Netherlands: Kluwer Academic Publishers. http://dx.doi.org/10.1007/978-94-010-0309-4_31
- Mitchell, R., & Martin, C. (1997). Rote learning, creativity and 'understanding' in classroom foreign language teaching. *Language Teaching Research*, 1(1), 1–27. <http://dx.doi.org/10.1177/136216889700100102>
- Mochizuki, N., & Ortega, L. (2008). Balancing communication and grammar in beginning-level foreign language classrooms: A study of guided planning and relativization. *Language Teaching Research*, 12(1), 11–37. <http://dx.doi.org/10.1177/1362168807084492>
- Morrow, K. (1991). Evaluating communicative tests. In S. Anivan (Ed.), *Current developments in language testing* (pp. 111–118). Singapore, Singapore: SEAMEO Regional Language Centre.
- Nakatsuhara, F. (2009). *Conversational styles in group oral tests: How is the conversation constructed?* Unpublished doctoral thesis. University of Essex, Essex, England.
- National Foundation for Educational Research. (2002). *New Zealand stocktake: An international critique*. Retrieved from <http://www.educationcounts.govt.nz/publications/curriculum/9137>
- Newton, P., & Shaw, S. (2014). *Validity in educational and psychological assessment*. London, England: Sage.
- Nitta, R., & Nakatsuhara, F. (2014). A multifaceted approach to investigating pre-task planning effects on paired oral test performance. *Language Testing*, 31(2), 147–175. <http://dx.doi.org/10.1177/0265532213514401>
- Norris, J. (2002). Interpretations, intended uses and designs in task-based language assessment. *Language Testing*, 19(4), 337–346. <http://dx.doi.org/10.1191/0265532202lt234ed>

- Norris, J. (2008). *Validity evaluation in language assessment*. Frankfurt am Main, Germany: Peter Lang.
- Norris, J., Bygate, M., & Van den Branden, K. (2009). Introducing task-based language teaching. In K. Van den Branden, M. Bygate, & J. Norris (Eds.), *Task-based language teaching: A reader* (pp. 15–19). Amsterdam, Netherlands/Philadelphia, PA: John Benjamins.
- Norton, J. (2005). The paired format in the Cambridge Speaking Tests. *ELT Journal*, 59(4), 287–297. <http://dx.doi.org/10.1093/elt/cci057>
- Nunan, D. (2004). *Task-based language teaching*. Cambridge, England: Cambridge University Press. <http://dx.doi.org/10.1017/cbo9780511667336>
- NZQA. (2014a). *Assessment and moderation best practice workshops*. Retrieved from <http://www.nzqa.govt.nz/about-us/events/best-practice-workshops/>
- NZQA. (2014b). *External examinations*. Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/ncea-exams-and-portfolios/external/>
- NZQA. (2014c). *External moderation*. Retrieved from <http://www.nzqa.govt.nz/providers-partners/assessment-and-moderation/managing-national-assessment-in-schools/secondary-moderation/external-moderation/>
- NZQA. (2014d). *History of NCEA*. Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/history-of-ncea/>
- NZQA. (2014e). *Internal moderation*. Retrieved from <http://www.nzqa.govt.nz/providers-partners/assessment-and-moderation/managing-national-assessment-in-schools/secondary-moderation/external-moderation/internal-moderation/>
- NZQA. (2014f). *Languages – Clarifications*. Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/subjects/languages/clarifications/>
- NZQA. (2014g). *Languages – Moderator’s newsletter*. Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/subjects/languages/moderator-newsletters/october-2014/>
- NZQA. (2014h). *NCEA subject resources*. Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/subjects>
- NZQA. (2014i). *Search framework*. Retrieved from <http://www.nzqa.govt.nz/framework/search/index.do>
- NZQA. (2014j). *Secondary school qualifications prior to 2002*. Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/results-2/secondary-school-qualifications-prior-to-2002/>
- NZQA. (2014k). *Standards*. Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/how-ncea-works/standards/>
- O’Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing*, 19(3), 277–295. <http://dx.doi.org/10.1191/0265532202lt205oa>
- Ockey, G. J. (2001). Is the oral interview superior to the group oral? *Working Papers, International University of Japan*, 11, 22–40.
- Pardo-Ballester, C. (2010). The validity argument of a web-based Spanish listening exam: Test usefulness evaluation. *Language Assessment Quarterly*, 7(2), 137–159. <http://dx.doi.org/10.1080/15434301003664188>
- Philp, J., Adams, R., & Iwashita, N. (2014). *Peer interaction and second language learning*. New York, NY: Routledge. <http://dx.doi.org/10.4324/9780203551349>
- Pinter, A. (2005). Task repetition with 10-year old children. In C. Edwards & J. Willis (Eds.), *Teachers exploring tasks in English language teaching* (pp. 113–126). New York, NY: Palgrave Macmillan.
- Pinter, A. (2007). What children say: Benefits of task repetition. In K. Van den Branden, K. Van Gorp, & M. Verhelst (Eds.), *Tasks in action: Task-based language education from a classroom-based perspective* (pp. 131–158). Newcastle, England: Cambridge Scholars Publishing.
- Poehner, M. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. New York, NY: Springer.
- Poehner, M., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, 9(3), 233–265. <http://dx.doi.org/10.1191/1362168805lr166oa>

- Popham, W. J. (2006). *Assessment for educational leaders*. Boston, MA: Pearson.
- Rea-Dickins, P. (1997). So, why do we need relationships with stakeholders in language testing? A view from the UK. *Language Testing*, 14(3), 304–314. <http://dx.doi.org/10.1177/026553229701400307>
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249–258. <http://dx.doi.org/10.1191/0265532204lt283ed>
- Resnick, R. (2012). *Comparison of postal and online surveys: Cost, speed, response rates and reliability*. Sweet Springs, MO: Education Market Research/MCH Strategic Data.
- Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge, England: Cambridge University Press. <http://dx.doi.org/10.1017/cbo9780511667220>
- Richards, J. C. (2006). *Communicative language teaching today*. Cambridge, England: Cambridge University Press.
- Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching* (3rd ed.). Cambridge, England: Cambridge University Press.
- Roeber, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, 28(4), 463–481. <http://dx.doi.org/10.1177/0265532210394633>
- Ryan, K. (2002). Assessment validation in the context of high-stakes assessment. *Educational Measurement: Issues and Practice*, 22, 7–15. <http://dx.doi.org/10.1111/j.1745-3992.2002.tb00080.x>
- Sakuragi, T. (2006). The relationship between attitudes toward language study and cross-cultural attitudes. *International Journal of Intercultural Relations*, 30, 19–31. <http://dx.doi.org/10.1016/j.ijintrel.2005.05.017>
- Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Basingstoke, England: Palgrave Macmillan. <http://dx.doi.org/10.1057/9780230596429>
- Savignon, S. (2005). Communicative language teaching: Strategies and goals. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 635–651). Mahwah, NJ: Lawrence Erlbaum.
- Saville, N., & Hargreaves, P. (1999). Assessing speaking in the revised FCE. *ELT Journal*, 53(1), 42–51. <http://dx.doi.org/10.1093/elt/53.1.42>
- Scott, A., & East, M. (2009). The standards review for learning languages: How come and where to? *The New Zealand Language Teacher*, 39, 28–33.
- Scott, A., & East, M. (2012). Academic perspectives from New Zealand. In M. Byram & L. Parmenter (Eds.), *The common European framework of reference: The globalisation of language education policy* (pp. 248–257). Clevedon, England: Multilingual Matters.
- Segalowitz, N. (2005). Automaticity and second languages. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 381–408). Oxford, England: Blackwell. <http://dx.doi.org/10.1002/9780470756492.ch13>
- Sercu, L. (2010). Assessing intercultural competence: More questions than answers. In A. Paran & L. Sercu (Eds.), *Testing the untestable in language education* (pp. 17–34). Clevedon, England: Multilingual Matters.
- Shearer, R. (n.d.). The New Zealand curriculum framework: A new paradigm in curriculum policy development. *ACE Papers, Issue 7* (Politics of curriculum, pp. 10–25). Retrieved from <https://researchspace.auckland.ac.nz/handle/2292/25073>
- Shohamy, E. (2000). Fairness in language testing. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 15–19). Cambridge, England: Cambridge University Press.
- Shohamy, E. (2001a). *The power of tests: A critical perspective on the uses of language tests*. Harlow, England: Longman/Pearson. <http://dx.doi.org/10.4324/9781315837970>
- Shohamy, E. (2001b). The social responsibility of the language testers. In R. L. Cooper (Ed.), *New perspectives and issues in educational language policy* (pp. 113–130). Amsterdam, Netherlands/Philadelphia, PA: John Benjamins Publishing Company. <http://dx.doi.org/10.1075/z.104.09sho>
- Shohamy, E. (2006). *Language policy: Hidden agendas and new approaches*. New York, NY: Routledge. <http://dx.doi.org/10.4324/9780203387962>

- Shohamy, E. (2007). Tests as power tools: Looking back, looking forward. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. E. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 141–152). Ottawa, Canada: University of Ottawa Press.
- Skehan, P. (2001). Tasks and language performance assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 167–185). London, England: Longman.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510–532. <http://dx.doi.org/10.1093/applin/amp047>
- Smallwood, I. M. (1994). Oral assessment: A case for continuous assessment at HKCEE level. *New Horizons: Journal of Education, Hong Kong Teachers' Association*, 35, 68–73.
- Spada, N. (2007). Communicative language teaching: Current status and future prospects. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 271–288). New York, NY: Springer. http://dx.doi.org/10.1007/978-0-387-46301-8_20
- Spolsky, B. (1985). The limits of authenticity in language testing. *Language Testing*, 2(1), 31–40. <http://dx.doi.org/10.1177/026553228500200104>
- Spolsky, B. (1995). *Measured words*. Oxford, England: Oxford University Press.
- Sunstein, B. S., & Lovell, J. H. (Eds.). (2000). *The portfolio standard: How students can show us what they know and are able to do*. Portsmouth, NH: Heinemann.
- Swain, M. (1984). Large-scale communicative language testing: A case study. In S. Savignon & M. Burns (Eds.), *Initiatives in communicative language teaching: A book of readings* (pp. 185–201). Reading, MA: Addison-Wesley.
- Swain, M. (2001). Examining dialogue: Another approach to content specification and to validating inferences drawn from test scores. *Language Testing*, 18(3), 275–302. <http://dx.doi.org/10.1177/026553220101800302>
- Taylor, L. (2001). The paired speaking test format: Recent studies. *Research Notes*, 6, 15–17.
- Taylor, L., & Wigglesworth, G. (2009). Are two heads better than one? Pair work in L2 assessment contexts. *Language Testing*, 26(3), 325–339. <http://dx.doi.org/10.1177/0265532209104665>
- The University of Queensland. (2012). *About flipped classrooms*. Retrieved from <http://www.uq.edu.au/tediteach/flipped-classroom/what-is-fc.html>
- Tomlinson, B. (Ed.). (2011). *Materials development in language teaching* (2nd ed.). Cambridge, England: Cambridge University Press.
- Torrance, H. (Ed.). (2013a). *Educational assessment and evaluation: Major themes in education* (Purposes, functions and technical issues, Vol. 1). London, England/New York, NY: Routledge.
- Torrance, H. (Ed.). (2013b). *Educational assessment and evaluation: Major themes in education* (Current issues in formative assessment, teaching and learning, Vol. 4). London, England/New York, NY: Routledge.
- Turner, J. (1998). Assessing speaking. *Annual Review of Applied Linguistics*, 18, 192–207. <http://dx.doi.org/10.1017/s0267190500003548>
- University of Cambridge. (2014). *IGCSE syllabus for Dutch, French, German and Spanish*. Cambridge, England: University of Cambridge International Examinations.
- Van den Branden, K. (2006). Introduction: Task-based language teaching in a nutshell. In K. Van den Branden (Ed.), *Task-based language education: From theory to practice* (pp. 1–16). Cambridge, England: Cambridge University Press. <http://dx.doi.org/10.1017/cbo9780511667282.002>
- van Lier, L. (1989). Reeling, writhing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly*, 23, 489–508. <http://dx.doi.org/10.2307/3586922>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wajda, E. (2011). New perspectives in language assessment: The interpretivist revolution. In M. Pawlak (Ed.), *Extending the boundaries of research on second language learning and teaching* (pp. 275–285). Berlin: Springer. http://dx.doi.org/10.1007/978-3-642-20141-7_21

- Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*. San Francisco, CA: Jossey-Bass.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke, England: Palgrave Macmillan.
- Willis, D., & Willis, J. (2007). *Doing task-based teaching*. Oxford, England: Oxford University Press.
- Winke, P. (2011). Evaluating the validity of a high-stakes ESL test: Why teachers' perceptions matter. *TESOL Quarterly*, 45(4), 628–660. <http://onlinelibrary.wiley.com/doi/10.5054/tq.2011.268063/abstract>
- Wood, R. (1993). *Assessment and testing*. Cambridge, England: Cambridge University Press.
- Yoffe, L. (1997). An overview of the ACTFL proficiency interview: A test of speaking ability. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 1(2), 2–13.

Index

A

Accountability, 9, 12, 30–32, 35, 53, 81, 87, 197, 199, 202, 204
Accuracy, 5, 25, 86, 113, 114, 119, 148, 160–162, 167, 176, 180, 186, 193, 207
Alternative assessment, 36, 52
American Council on the Teaching of Foreign Languages (ACTFL), 4–5, 27, 30, 41
Anxiety, 33, 43, 81, 113, 115, 122, 129, 176, 180, 181, 193, 200, 201
Assessment blueprints, 1, 7, 56, 69, 72
Assessment Reform Group, 33, 52
Audio-lingualism, 3, 5
Authenticity, 6, 38, 65, 79, 81–83, 86–88, 101, 106, 109, 110, 112, 113, 118, 119, 123, 147, 150, 166, 173, 180–182, 192, 193, 198, 205
Automaticity, 4, 5, 27, 28, 191, 199, 200, 204, 205

B

Best Practice Workshops, 71, 171

C

Common European Framework of Reference for languages (CEFR), 5, 27, 58, 85, 129, 148, 205
Communicative competence, 3, 4, 25–30, 34, 61
Communicative interaction, 28, 82
Communicative Language Teaching (CLT), 3–6, 25, 43, 58, 82

Communicative proficiency, 5, 28, 40, 44, 45, 58, 59, 175, 194
Complexity, 162, 199, 207
Consequences, 7, 9, 12–15, 17, 20, 33, 45, 81, 189, 198, 201, 207, 208
Consequential validity, 207
Construct, 110
 irrelevant variance, 13, 14, 44, 197
 under-representation, 13, 14, 42, 194
 validity, 10, 87, 101, 192
Criterion-referencing, 56
Curriculum, 127, 178

D

Discourse competence, 26, 28
Dynamic, 30, 33, 35–37, 40, 45, 51, 78, 84, 187, 189, 190, 196, 199–201, 205, 207
Dynamic assessment, 31

F

Fairness, 11, 14, 15, 32, 44, 130, 195
Feedback, 33–35, 37, 52, 55, 72, 84, 86–88, 114, 115, 132, 141, 151, 190, 200, 204
Feedforward, 33, 52, 55, 72, 86, 88, 114, 168, 190
Fit for purpose, 2, 8, 11, 32, 58, 78, 83, 111, 118, 130, 144, 161, 187, 190, 202, 207
Fluency, 5, 25, 85, 87, 113, 114, 117, 148, 159, 162, 163, 170, 176, 183, 185, 186, 191, 193, 207
Formative, 31, 33, 34, 40, 54, 86

G

Grammar-translation, 3, 5, 58
 Grammatical competence, 26

H

High-stakes assessment, 1, 2, 6, 20, 42, 51,
 53–56, 77, 83, 86, 129, 168, 186, 195,
 198, 201, 202

I

IGCSE, 132
 Impact, 13, 17, 33, 44, 45, 69, 79, 81, 84, 87,
 88, 101, 106, 110, 111, 114, 115, 120,
 121, 123, 126, 127, 129, 130, 135, 137,
 138, 156, 173, 178, 179, 181, 186, 189,
 192–194, 196, 201, 205
 Impracticality, 125–127, 131, 137, 144, 202,
 203, 208
 Interaction hypothesis, 4, 191
 Interactional authenticity, 82, 83, 152
 Interactional competence, 27–29, 42, 43, 45,
 81, 87, 148, 170
 Interactional proficiency, 42, 86, 147–149,
 157, 169, 176, 182, 199
 Interactiveness, 79, 81, 83, 88, 101, 106, 109,
 110, 112, 113, 118, 121, 123, 173, 182,
 192, 193
 Intercultural communicative competence, 29
 Intercultural competence, 28
 Interlocutor effects, 44
 Interlocutor variables, 44, 45, 129, 130, 138,
 143, 178, 184, 186, 195, 205
 Internal assessments, 6, 53, 59, 87, 101, 128
 Interview test, 1, 41–43, 45, 58, 77, 84, 87,
 111, 193, 197, 199

K

Key competencies, 61, 64

L

Learning languages, 7, 62, 64, 67, 68

M

Ministry of Education, 52, 63, 64, 66, 202

N

National Certificate of Educational
 Achievement (NCEA), 6, 51, 56, 66,
 147, 197

New Zealand Curriculum Framework (NZCF),
 53, 55

New Zealand Qualifications Authority
 (NZQA), 2, 55, 202, 203

Norm-referenced, 54, 197

Norm-referencing, 56

O

Oral Proficiency Interview test, 41

P

Paired/group, 30, 44, 45, 51, 187
 Peer-to-peer interactions, 1, 4, 77, 115, 130,
 183, 187, 193, 201
 Performance outcomes, 9–11, 13, 14, 32,
 36, 81
 Performance scores, 2, 11, 12, 207
 Performance-based assessment,
 36, 37, 204
 Portfolio, 34–36, 116, 122, 127, 128, 132,
 135–137, 144, 194, 195
 Practicality, 79, 81, 82, 87, 88, 101, 106–111,
 131, 137, 192, 193, 202, 203
 Pragmatic competence, 28
 Proficiency, 32
 Proficiency movement, 3, 4, 36
 Psychometric model, 10, 31, 33
 Psychometrics, 12

Q

Qualitative research, 16

R

Rebuttals, 88
 Reliability, 10–12, 16, 32, 33, 35–37, 53,
 78–80, 83, 87, 101, 109, 110, 116, 130,
 173, 191–193, 201, 202
 Role-play, 38, 113, 134, 135, 140, 149

S

Scaffolding, 62, 93, 131–134, 141, 142, 144,
 150, 156, 183, 200
 Situational authenticity, 82, 148, 205
 Sociolinguistic competence, 26, 28, 29, 86
 Spoken communicative proficiency, 1, 6,
 11, 14, 17, 25–30, 36, 37, 39–41,
 43, 45, 51, 53, 61, 77, 79, 81, 83,
 84, 87, 96, 97, 101, 104, 106, 111,
 147, 173, 187, 189–193, 202, 204,
 206, 208

Spontaneity, 42, 86, 87, 113, 122, 133,
139–143, 147, 148, 150, 156–159, 161,
162, 165, 167, 170, 179–181, 183, 193,
201

Standards, 56, 66

Standards-Curriculum Alignment Languages
Experts (SCALEs) project, 64

Static, 30, 32, 33, 35–37, 40, 45, 51, 78, 187,
189, 190, 197, 199–202, 205, 207

Strategic competence, 26, 27, 113, 114, 120,
161, 184, 200

Stress, 81, 106, 115, 129, 158, 173, 175, 177,
192, 193, 195, 196, 201

Summative, 6, 31–33, 35, 40, 53, 54, 61, 63,
65, 77, 86, 116, 193, 197, 199, 201,
202, 204

T

Task-based language assessment (TBLA),
37–39, 155

Task-based language teaching (TBLT), 5–7,
25, 37–39, 62

TLU domain, 30, 32, 38, 39, 82, 83

Triangulation, 95, 97, 207

U

Usefulness, 8, 13, 17, 45, 78, 79, 81, 84, 87,
88, 91, 101, 103, 107, 109, 111, 112,
126, 173, 175, 185, 187, 190, 192, 193,
200, 202, 206, 207

V

Validation, 3, 12, 14, 16, 17, 21, 80, 207

Validity, 2, 8, 10–18, 20, 32, 33, 35–37, 41,
44, 53, 78–80, 87, 109, 110, 116, 123,
130, 173, 191, 193, 195, 202, 203,
206–208

W

Warrants, 88

Washback, 33, 41, 43, 59, 69, 72, 116,
122–124, 168–170, 178, 191, 194,
195, 203

Z

Zone of proximal development (ZPD),
34, 191