ILTA Language Testing Bibliography 1990-1999

First Edition

Edited by:  Jay Banerjee
Caroline Clapham
Phoebe Clapham
Dianne Wall

Language Testing Update
INTRODUCTION TO THE FIRST EDITION

This bibliography grew from a set of references produced by Kathleen and Kenji Kitao for Language Testing Update 20. Later issues of Language Testing Update (LTU) included further titles, kindly sent in by language testers from around the world, and it soon became evident that LTU was building up an invaluable resource. We therefore decided that all these references should be conflated into one large bibliography, and, with the help of the International Language Testing Association (ILTA), we have now produced this large collection of references from 1990 to 1999. This is the first edition of the bibliography; we know that not only are there omissions and miscategorisations in this edition but also that since we started on this collection new works have been published. We therefore propose to issue a second edition in the near future in which as many errors and omissions as possible will be rectified. We therefore ask readers to send their corrections and additions to the address overleaf.

In the long term, we hope to install this bibliography on the World Wide Web so that language testers can access it, and so that it can be readily updated. However, we know that for the moment not everyone has access to the Internet, so for the foreseeable future the bibliography will also be available in book form.

Since we want to make sure that the references are as widely accessible as possible, we have focused on published works, and have omitted papers presented at conferences. Some of these papers have since been published, and we should be grateful if readers could tell us in which books or journals they now appear.

Because we want the works to be easily accessible, we have also, after much discussion, restricted the bibliography to publications written in English. We know that there are many interesting language testing articles published in different languages, but unfortunately our knowledge about such works is restricted, and we feel that it would be unhelpful and unfair to include works in certain languages but not in others. However, we are most happy to include bibliographies of publications in languages other than English in LTU. So please do send the LTU editors your lists of references of works in languages other than English.

Organisation of the volume

In the first half of this volume the references are presented under headings such as ‘Testing Listening’, and ‘Test Reliability and Validity’. Some references appear under more than one heading. Although these categories are very general, some of the works proved difficult to categorise. In the second half of the volume, therefore, the references are listed in alphabetical order by author. We hope that in this way all the references will be easily identified.
If readers are uncertain about which sub-categories there are in each main category, there is a Classification Index on pages vii and viii which shows which specific areas of language testing fall under which main headings.

The editors

Please send amendments and additions to:

caroline.clapham@lancaster.ac.uk

Caroline Clapham
Department of Linguistics and Modern English Language
Lancaster University
Lancaster
LA1 4YT
United Kingdom
# INDEX OF CONTENTS

1. General ............................................................................................................................................. 1
2. Testing Listening ............................................................................................................................. 4
3. Testing Reading ............................................................................................................................... 6
4. Testing Speaking ........................................................................................................................... 10
5. Testing Writing .............................................................................................................................. 15
6. Testing Language Elements ........................................................................................................... 18
7. Purposes for Testing ...................................................................................................................... 20
8. Specific Purposes Testing ............................................................................................................ 24
9. Alternative Approaches to Assessment ......................................................................................... 30
10. Principles of Test Design ............................................................................................................ 35
11. Factors Affecting Test Performance ............................................................................................ 41
12. Rating and Rating Scales .............................................................................................................. 43
13. Test Reliability and Validity ........................................................................................................ 48
14. Washback .................................................................................................................................. 57
15. Ethics and Standards .................................................................................................................. 60
CLASSIFICATION INDEX

Achievement tests
see – purposes for testing

Aptitude testing
see – purposes for testing

Authenticity
see – principles of test design

Bias
see – ethics and standards
factors affecting test performance

Classroom assessment
see – alternative approaches to assessment

Cloze tests
see – principles of test design
testing reading

Collections of Papers
see – general

Communicative competence, assessment of
see – principles of test design
purposes for testing
specific purposes testing
test reliability and validity

Computer assisted testing
see – alternative approaches to assessment

Concurrent validity
see – test reliability and validity

Construct validity
see – test reliability and validity

Content, assessment of
see – specific purposes testing

Content validity
see – test reliability and validity

Criterion-referenced testing
see – principles of test design
test reliability and validity

C-test
see – principles of test design
testing reading

Culture
see – principles of test design
factors affecting test performance
specific purposes testing

Dictation
see – testing listening
testing writing

English for academic purposes, Assessment of
see – specific purposes testing

Ethics
see – ethics and standards

Gender
see – factors affecting test performance

General Language Testing Books
see – general

Generalisability theory
see – test reliability and validity

Grammar
see – testing language elements

Impact
see – washback

Integrated testing
see – principles of test design

Item analysis
see – test reliability and validity

Item construction
see – principles of test design

Item Response theory
see – test reliability and validity

International teaching assistants
see – specific purposes testing

Item types
see – principles of test design

Language acquisition theory
see – principles of test design

Literacy, assessment of
see – specific purposes testing

Overview of the field
see – general

Performance testing
see – principles of test design
specific purposes testing
test reliability and validity

Placement testing
see – purposes for testing

Portfolio assessment
see – alternative approaches to assessment
testing writing

Proficiency testing
see – purposes for testing

Programme evaluation
see – ethics and standards

Qualitative validation procedures
see – test reliability and validity

Questionnaire validation
see – test reliability and validity

Rasch analysis
see – test reliability and validity

Rater behaviour
see – rating and rating scales
Rating scales
  see – rating and rating scales
test reliability and validity
Research instrument design
  see – test reliability and validity
Research methodology
  see – test reliability and validity
Rule-Space methodology
  see – test reliability and validity
Self assessment
  see – alternative approaches to assessment
Special Issues of Language Testing
  see – general
Speech and communication disorders
  see – factors affecting test performance
Standards
  see – ethics and standards
Structural Equation Modelling
  see – test reliability and validity
Subject specific testing
  see – specific purposes testing
Summary tasks
  see – testing writing
Task types
  see – principles of test design
Task variability
  see – principles of test design
test reliability and validity
Teachers’ abilities, assessment of
  see – specific purposes testing
Test developers’ research reports
  see – principles of test design
Test methods
  see – principles of test design
Test takers
  see – factors affecting test performance
test reliability and validity
Test taking strategies
  see – factors affecting test performance
Test tasks
  see – principles of test design
testing listening
testing reading
testing speaking
testing writing
Translation
  see – testing listening
testing writing
Vocabulary
  see – testing language elements
Young learners, assessment of
  see – specific purposes testing
Part I: Bibliography by Topic

1. General

Airasian, P.W. (1991) Classroom Assessment, McGraw-Hill, New York.

Alderson, J.C. (1991) Language Testing in the 1990s: How far have we come? How much further have we to go? in Anivan, S. (ed.), Current Developments in Language Testing, RELC Anthology Series 23, RELC.

Alderson, J.C. (1993) The State of Language Testing in the 1990s, in Huhta, A., Sajavaara, K. and Takala, S. (eds.), Language Testing: New Openings, Institute for Educational Research, University of Jyväskylä, Finland.

Alderson, J.C., Clapham, C.M. and Wall, D. (1995) Language Test Construction and Evaluation, Cambridge University Press, Cambridge.

Alderson, J.C. and North, B. (eds.) (1991) Language Testing in the 1990s: The Communicative Legacy, Macmillan, London.

Alderson, J.C. and Wall, D. (guest eds.) (1996) Washback, Language Testing: Special Issue, 13,3.

Anivan, S. (ed.) (1991) Current Developments in Language Testing, RELC Anthology Series 23, RELC.

Bachman, L. (1990) Fundamental Considerations in Language Testing, Oxford University Press, Oxford.

Bachman, L.F. (1990) Assessment and Evaluation, Annual Review of Applied Linguistics, 10: 210-226.

Bachman, L. (1991) What Does Language Testing Have to Offer? TESOL Quarterly, 25,4: 671-704.

Bachman, L.F. and Cohen, A.D. (eds.) (1998) Interfaces between Second Language Acquisition and Language Testing Research, Cambridge University Press, Cambridge.

Bachman, L.F. and Cohen, A.D. (1998) Language testing –CLA Interfaces: An update, in Bachman, L.F. and Cohen, A.D. (eds.) Interfaces between SLA and Language Testing Research, Cambridge University Press, Cambridge.

Bachman, L. and Palmer, A. (1996) Language Testing in Practice, Oxford University Press, Oxford.

Boyle, J. and Falvey, F. (eds.) (1994) English Language Testing in Hong Kong, The Chinese University Press, Hong Kong.

Broadfoot, P. (1996) Performance Assessment in Perspective, in Craft, A. (ed.), Primary Education: Assessing and Planning Learning, Routledge in association with the Open University, London.

Brown, J.D. (1996) Testing in Language Programs, Prentice Hall, Upper Saddle River, NJ.

Brown, J.D. (1997) Computers in Language Testing: Present Research and Some Future Directions, Language Learning and Technology, 1,1: 44-59.

Brown, J.D. and Hudson, T. (1998) Alternatives in Language Assessment, TESOL Quarterly, 32,4: 653-675.

Brown, J.D. and Hudson, T. (1998) The Alternatives in Language Testing: Advantages and Disadvantages, University of Hawaii Working Papers in ESL, 16,2: 79-103.

Brown, J.D. and Yamashita, S.O. (eds.) (1995) Language Testing in Japan, Japan Association for Language Teaching, Tokyo.

Cascallar, E.C. (guest ed.) (1995) Language Aptitude, Language Testing: Special Issue, 12,3.
Chalhoub-Deville, M. (1997) Recent Trends in Assessment, in Coombe, C. and Troudi, S. (eds.) *Tradition and Innovation*, TESOL ARABIA, Al-Ain, United Arab Emirates, 20-29.

Chalhoub-Deville, M. and Deville C. (1999) Computer Adaptive Testing and its Application in the field of second language testing, *Annual Review of Applied Linguistics*, 19: 273-299.

Chapelle, C.A. and Douglas, D. (1993) Foundations and Directions for a New Decade of Language Testing, in Douglas, D. and Chapelle, C.A. (eds.), *A New Decade of Language Testing Research*, TESOL Publications, Alexandria, VA.

Clapham, C.M. and Corson, D. (eds.) (1997) *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Cohen, A. (1994) *Assessing Language Ability in the Classroom*, Heinle and Heinle, Boston.

Cohen, A. (guest ed.) (1994) *Focusing on Language and Language Processing: Beyond Test Scores*, Language Testing: Special Issue, 11,2.

Coombe, C. (ed.) (1998) *Current Trends in English Language Testing*, Proceedings of the CTELT Conference 1997 and 1998.

Cumming, A. and Berwick, R. (eds.) (1995) *Validation in Language Testing*, Multilingual Matters, Clevedon, England.

Davies, A. (1990) *Principles of Language Testing*, Blackwell, Oxford.

Davies, A. (guest ed.) (1997) *Ethics*, Language Testing: Special Issue, 14,3.

Douglas, D. (ed.) (1990) *English Language Testing in U.S. Colleges and Universities*, NAFSA, Washington, DC.

Douglas, D. (1995) Developments in Language Testing, *Annual Review of Applied Linguistics*, 15: 167-187.

Douglas, D. and Chapelle, C. (eds.) (1993) *A New Decade of Language Testing Research*, TESOL Publications, Alexandria, VA.

Haertel, E. (1992) Performance Measurement, in Alkin, M.C. (ed.), *Encyclopedia of Educational Research* (6th edition): 984-9.

Hamayan, E.V. (1995) Approaches to Alternative Assessment, *Annual Review of Applied Linguistics*, 15: 212-226.

Harrison, A. (1991) Language Assessment as Theatre: Ten Years of Communicative Testing, in Alderson, J.C. and North, B. (eds.), *Language Testing in the 1990s: The Communicative Legacy*, Macmillan, London.

Heaton, J.B. (1990) *Classroom Testing*, Longman, London.

Heaton, J.B. (1990) *Writing English Language Tests* (second edition), Longman, London.

Hill, C. and Parry, K. (eds.) (1994) *From Testing to Assessment: English as an International Language*, Longman, London.

Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.) (1997) *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Huhta, A., Sajavaara, K. and Takala, S. (eds.) (1993) *Language Testing: New Openings*, Institute for Educational Research, University of Jyväskylä, Finland.

Kunnan, A.J. (guest ed.) (1998) *Structural Equation Modelling: Theory and Practice*, Language Testing: Special Issue, 15,3.
Kunnan, A.J. (ed.) (1998) *Validation in Language Assessment*, Lawrence Erlbaum, Mahwah, NJ.

Laveault, D., Zumbo, B.D., Gessaroli, M.E. and Boss, M.W. (eds.) (1994) *Modern Theories of Measurement: Problems and Issues*, University of Ottawa Press, Ottawa.

McNamara, T.F. (1996) *Measuring Second Language Performance*, Longman, London.

McNamara, T.F. (1997) Performance Testing, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Milanovic, M. and Saville, N. (eds.) (1996) *Performance Testing, Cognition and Assessment*, Cambridge University Press, Cambridge.

Millman, J. (1994) Criterion-Referenced Testing 30 Years Later: Promise Broken, Promise Kept, *Educational Measurement: Issues and Practice*, 13,19-20, 39.

Oller, J.W. (1991) Foreign Language Testing, Part 1: Its Breadth, *ADFL Bulletin*, 22: 33-38.

Quinn, T.J. (1993) The Competency Movement, Applied Linguistics and Language Testing: some Reflections and Suggestions for a Possible Research Agenda, *Melbourne Papers in Language Testing*, 2,2.

Rea-Dickins, P. and Germane, K. (1992) Evaluation, Oxford University Press, Oxford.

Shohamy, E. (1995) Performance Assessment in Language Testing, *Annual Review of Applied Linguistics*, 15: 188-211.

Shohamy, E. (1997) Second Language Assessment, in Tucker, G.R. and Corson, D. *Second Language Education*, Encyclopaedia of Language and Education, Vol. 4, Kluwer, Dordrecht.

Skehan, P. (1991) Progress in Language Testing: the 1990s, in Alderson, J.C. and North, B. (eds.), *Language Testing in the 1990s: The Communicative Legacy*, Macmillan, London.

Spolsky, B. (1995) *Measured Words: The Development of Objective Language Testing*, Oxford University Press, Oxford.

Weir, C. (1990) *Communicative Language Testing*, Prentice Hall, Englewood Cliffs, NJ.

Weir, C.J. (1992) Language Testing: the Decade Past the Decade to Come, in Hill, D.A. (ed.) *The State of the Art*, Modern English Publications, London, 107-115.

Weir, C. (1993) *Understanding and Developing Language Tests*, Prentice Hall, London.

Weir, C.J. and Roberts, J.R. (1994) *Evaluation in ELT*, Blackwell, Oxford.

Wigglesworth, G. and Elder, C. (eds.) (1996) *The Language Testing Cycle: from Inception to Washback*, Australian Review of Applied Linguistics, Australian National University, Canberra.
2. Testing Listening

Bacon, S.M. (1991) Assessing Foreign Language Listening: Processes, Strategies and Comprehension, in Teschner, R.V. (ed.), Assessing Foreign Language Proficiency of Undergraduates, Heinle & Heinle, Boston.

Boyle, J. and Suen, D.L.K. (1994) Communicative Considerations in a Large-Scale Listening Test, in Boyle, J. and Falvey, P. (eds.), English Language Testing in Hong Kong, The Chinese University Press, Hong Kong.

Brindley, G. (1998) Assessing Listening Abilities, Annual Review of Applied Linguistics, 18: 171-91.

Brindley, G. and Nunan, D. (1992) Draft Bandscales for Listening, IELTS Research Projects: Project I, National Centre for English Language Teaching and Research, Macquarie University, Sydney, Australia.

Buck, G. (1990) The Testing of Second Language Listening Comprehension, unpublished Ph.D. thesis, University of Lancaster.

Buck, G. (1991) The Testing of Listening Comprehension: an Introspective Study, Language Testing, 8,1: 67-91.

Buck, G. (1992) Listening Comprehension: Construct Validity and Trait Characteristics, Language Learning, 42,3: 313-357.

Buck, G. (1994) The Appropriacy of Psychometric Measurement Models for Testing Second Language Listening Comprehension, Language Testing, 11,2: 145-170.

Buck, G. (1997) The Testing of Listening in a Second Language, in Clapham, C.M. and Corson, D. (eds.), Language Testing and Assessment, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Buck, G. and Tatsuoka, L. (1998) Application of the Rule-space Procedure to Language Testing: Examining Attributes of a Free Response Listening Test, Language Testing, 15,2: 119-157.

Buck, G., Tatsuoka, K., Kostin, I. and Phelps, M. (1997) The Sub-Skills of Listening: Rule-Space Analysis of a Multiple-Choice Test of Second Language Listening Comprehension, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Dunkel, P. (1991) Computerized Testing of Non-Participatory L2 Listening Comprehension Proficiency: an ESL Prototype Development Effort, Modern Language Journal, 75,1: 64-73.

Dunkel, P. (1991) Listening in the Native and Second/Foreign Language: Towards an Integration of Research and Practice, TESOL Quarterly, 25,3: 431-457.

Dunkel, P. (1992) The Use of PC-Generated Speech Technology in the Development of an L2 Listening Comprehension Proficiency Test: A Prototype Design Effort, in Pennington, M.C. and Stevens, V. (eds.), Computers in Applied Linguistics: An International Perspective, Multilingual Matters, Clevedon, England.

Dunkel, P. (1996) Checking the Utility and Appropriacy of the Content and Measurement Models Used to Develop L2 Listening Comprehension CATS: Implications for Further Development of Comprehensive CATs, in Chalhoub-Deville, M. (ed.), Issues in Computer Adaptive Testing of Second Language Reading Proficiency Conference Seminar Papers, University of Minnesota, Bloomington, MN.

Dunkel, P., Henning, G. and Chaudron, C. (1993) The Assessment of an L2 Listening Comprehension Construct: Tentative Model for Test Specification and Development, The Modern Language Journal, 77,2: 180-191.

Flowerdew, J. (ed.) (1994) Academic Listening: Research Perspectives, Cambridge University Press, Cambridge.

Gruba, P. (1997) The Role of Video Media in Listening Assessment, System, 25,3: 335-345 [An earlier version appeared in 1995 under the title ‘The Role of Visual Media in Listening Assessment’, Melbourne Papers in Language Testing, 4,1].
Hale, G.A. and Courtney, R (1994) The Effects of Note-Taking on Listening Comprehension in the Test of English as a Foreign Language, *Language Testing*, 11,1: 29-48.

Hanson, C. and Jensen, C. (1994) Evaluating Lecture Comprehension, in Flowerdew, J. (ed.), *Academic Listening: Research Perspectives*, Cambridge University Press, Cambridge.

Jensen, C. and Hanson C. (1995) The Effect of Prior Knowledge on EAP Listening Performance, *Language Testing*, 12,1: 99-119.

Jensen, C., Hansen, C., Green, S.B. and Akey, T. (1997) An Investigation of Item Difficulty Incorporating the Structure of Listening Tests: A Hierarchical Linear Modeling Analysis, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Kellerman, S. (1990) Lip Service: the Contribution of the Visual Modality to Speech Perception and its Relevance to the Teaching and Testing of Foreign Language Listening Comprehension, *Applied Linguistics*, 11: 272-80.

Kitao, S.K. and Kitao, K. (1996) Testing Listening, *The Internet TESL Journal*, 2,7.

Leung, C. and Teasdale, A. (1997) What Do Teachers Mean by Speaking and Listening? A Contextualised Study of Assessment in Multilingual Classrooms in the English National Curriculum, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Lewkowicz, J. (1997) *Investigating Authenticity in Language Testing*, unpublished Ph.D. thesis, University of Lancaster.

Lewkowicz, J. (1997) Authentic for whom? Does Authenticity Really Matter?, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Lewkowicz, J. (1997) The Integrated Testing of a Second Language, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Lund, R.J. (1991) A Comparison of Second Language Listening and Reading Comprehension, *The Modern Language Journal*, 75,2: 196-204.

McNamara, T.F. (1991) Test Dimensionality: IRT Analysis of an ESP Listening Test, *Language Testing*, 8,2: 139-159.

Mets, J., van den Bosch, L. and de Jong, J.H.A.L. (1993) The Listening and Speaking Tests in the New Examinations for Dutch as a Second Language, *Toegepaste Taalwetenschap in Artikelen*, 46/47: 207-221.

Scott, M.L., Stansfield, C.W. and Kenyon, D. (1996) Examining Validity in a Performance Test: The Listening Summary Translation Exam (LSTE) - Spanish Version, *Language Testing*, 13,1: 83-109.

Sherman, J. (1997) The Effect of Question Preview in Listening Comprehension Tests, *Language Testing*, 14,2: 185-213.

Shohamy, E. and Inbar, O. (1991) Validation of Listening Comprehension Tests: The Effect of Text and Question Types, *Language Testing*, 8,1: 23-40.

Thompson, I. (1995) Assessment of Second/Foreign Language Listening Comprehension, in Mendelsohn, D.J. and Rubin, J. (eds.), *A Guide for the Teaching of Second Language Listening*, Dominie Press, San Diego.

Vogely, A. (1995) Perceived Strategy Use during Performance on Three Authentic Listening Comprehension Tasks, *The Modern Language Journal*, 79,1: 41-56.

Yi’an, W. (1998) What do Tests of Listening Comprehension Reveal? - A Retrospection Study of EFL Test-takers Performing a Multiple-choice Task, *Language Testing*, 15,1: 21-44.
3. **Testing Reading**

Alderson, J.C. (1990) Testing Reading Comprehension Skills (Part One), *Reading in a Foreign Language*, 6,2: 425-38.

Alderson, J.C. (1991) Letter, *Reading in a Foreign Language*, 7,2: 599-603.

Alderson, J.C. (1991) Testing Reading Comprehension Skills (Part Two), *Reading in a Foreign Language*, 7,1: 465-503.

Alderson, J.C. (1993) Judgments in Language Testing, in Douglas, D. and Chapelle, C. (eds.), *A New Decade of Language Testing*, TESOL Publications, Alexandria, VA.

Alderson, J.C. (1993) The Relationship Between Grammar and Reading in an English for Academic Purposes Test Battery, in Douglas, D. and Chapelle, C. (eds.), *A New Decade of Language Testing Research*, TESOL, Alexandria, VA.

Alderson, J.C. (1996) The testing of reading, in Nuttall, C. (ed.), *Teaching Reading Skills in a Foreign Language*, Heinemann, London.

Alderson, J.C. (in press) *Assessing Reading*, Cambridge University Press, Cambridge.

Anderson, N.J. (1991) Individual Differences in Strategy Use in Second Language Reading and Testing, *Modern Language Journal*, 75,4: 460-472.

Anderson, N.J., Bachman, L.F., Perkins, K. and Cohen, A. (1991) An Exploratory Study into the Construct Validity of a Reading Comprehension Test: Triangulation of the Data Sources, *Language Testing*, 8,1: 41-66.

Bernhardt, E.B. (1991) *Reading Development in a Second Language: Theoretical, Empirical and Classroom Perspectives*, Ablex, Norwood, NJ.

Carver, R.P. (1992) What Do Standardised Tests of Reading Comprehension Measure in Terms of Efficiency, Accuracy and Rate? *Reading Research Quarterly*, 27,4: 347-59.

Chalhoub-Deville, M. (1996) Constructing an Operational Framework for a CAT of L2 Reading Proficiency, in Chalhoub-Deville, M. (ed.), *Issues in Computer Adaptive Testing of Second Language Reading Proficiency Conference Seminar Papers*, University of Minnesota, Bloomington, MN.
Chalhoub-Deville, M., Alcaya, C. and Lozier, V. (1996) An Operational Framework for Constructing a Computer-Adaptive Test of L2 Reading Ability: Theoretical and Practical Issues, CARLA working paper series #1, Center for Advanced Research on Language Acquisition, University of Minnesota, Minneapolis, Minnesota.

Chalhoub-Deville, M., Alcaya, C. and Lozier, V.M. (1997) Language and Measurement Issues in Developing Computer-Adaptive Tests of Reading Ability: The University of Minnesota Model, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Choi, I.C. and Bachman, L.F. (1992) An Investigation into the Adequacy of Three IRT Models for Data from Two EFL Reading Tests, Language Testing, 9:1: 51-78.

Clapham, C.M. (1991) The Effect of Academic Discipline on Reading Test Performance, in Selected Papers from the 13th Annual Language Testing Research Colloquium, Educational Testing Service, Princeton, NJ.

Clapham, C.M. (1996) The Development of IELTS: A Study of the Effect of Background Knowledge on Reading Comprehension, Cambridge University Press, Cambridge.

Clapham, C.M. (1996) What Makes an ESP Reading Test Appropriate for its Candidates? in Cumming, A. and Berwick, R. (eds.), Validation in Language Testing, Multilingual Matters, Clevedon, England.

Deville, C. and Chalhoub-Deville, M. (1993) Modified Scoring, Traditional Item Analysis and Sato’s Caution Index Used to Investigate the Reading Recall Protocol, Language Testing, 10:2: 117-132.

Dollerup, C., Glahn, E. and Rosenberg Hansen, C. (1994) ‘Sprogtest’: a Smart Test (or How to Develop a Reliable and Anonymous EFL Reading Test), Language Testing, 11:1: 65-81.

Eignor, D.R. (1996) Adaptive Assessment of Reading Comprehension for TOEFL, in Chalhoub-Deville, M. (ed.), Issues in Computer Adaptive Testing of Second Language Reading Proficiency Conference Seminar Papers, University of Minnesota, Bloomington, MN.

Freedle, R. and Kostin, I. (1993) The Prediction of TOEFL Reading Item Difficulty: Implications for Construct Validity, Language Testing, 10:2: 133-170.

Fulcher, G. (1998) Widdowson’s model of communicative competence and the testing of reading: an exploratory study, System, 26: 281 – 302.

Gordon, C.M. and Hanauer, D. (1995) The Interaction between Task and Meaning Construction in EFL Reading Comprehension Tests, TESOL Quarterly, 29:2: 299-324.

Grabe, W. (1991) Current Developments in Second Language Reading Research, TESOL Quarterly, 25:3: 375-406.

Henning, G., Anbar, M., Helm, C.E. and D’Arcy, S.J. (1993) Computer-Assisted Testing of Reading Comprehension: Comparisons among Multiple-Choice and Open-Ended Scoring Methods, in Douglas, D. and Chapelle, C.A. (eds.), A New Decade of Language Testing Research, TESOL Publications, Alexandria, VA.

Hill, C. and Parry, K. (1992) The Test at the Gate: Models of Literacy in Reading Assessment, TESOL Quarterly, 26:3: 433-461.

Hill, C. and Parry, K. (1994) Models of Literacy: The Nature of Reading Tests, in Hill, C. and Parry, K. (eds.), From Testing to Assessment: English as an International Language, Longman, London.

Hudson, T. (1993) Testing the Specificity of ESP Reading Skills, in Douglas, D. and Chapelle, C.A. (eds.), A New Decade of Language Testing Research, TESOL Publications, Alexandria, VA.

Hughes, A. (1993) Testing the Ability to Infer when Reading in a Second or Foreign Language, Journal of English and Foreign Languages, 10&11: 13-20.

Huhta, A. and Randell, E. (1996) Multiple-Choice Summary: a Measure of Text Comprehension, in Cumming, A. and Berwick, R, (eds.), Validation in Language Testing, Multilingual Matters, Clevedon, England.
Jones, S. (1990) Evaluation and Assessment for ESL Literacy, *TESL Talk*, 20: 294-304.

Kaya-Carton, E., Carton, A.S. and Dandonoli, P. (1991) Developing a Computer-Adaptive Test of French Reading Proficiency, in Dunkel, P. (ed.), *Computer-Assisted Language Learning And Testing: Research Issues and Practice*, Newbury House, New York.

Lumley, T. (1993) The Notion of Subskills in Reading Comprehension Tests: an EAP Example, *Language Testing*, 10,3: 211-234.

Lumley, T. (1993) Reading Comprehension Sub-Skills: Teachers’ Perceptions of Content in an EAP Test, *Melbourne Papers in Language Testing*, 2,1.

Lund, R.J. (1991) A Comparison of Second Language Listening and Reading Comprehension, *The Modern Language Journal*, 75,2: 196-204.

Madsen, H.S. (1991) Computer-Adaptive Testing of Listening and Reading Comprehension, in Dunkel, P. (ed.), *Computer-Assisted Language Learning And Testing: Research Issues and Practice*, Newbury House, New York.

Matthews, M. (1991) Skills Taxonomies and Problems for the Testing of Reading, *Reading in a Foreign Language*, 7,1: 511-517.

Mommers, M. and Oud, J. (1992) Modeling and Predicting Reading Difficulties, in Verhoeven, W. and de Jong, J.H.A.L. (eds.), *The Construct of Language Proficiency: Application of Psychological Models to Language Assessment*, John Benjamins, Amsterdam.

Oh, J. (1992) The Effects of L2 Reading Assessment Methods on Anxiety Level, *TESOL Quarterly*, 26,1: 172-76.

Paribakht, T. and Wesche, M. (1993) The Relationship between Reading Comprehension and Second Language Development in a Comprehension-Based ESL Program, *TESL Canada Journal*, 11,1: 9-29.

Paribakht, T. and Wesche, M. (1997) Vocabulary Enhancement Activities and Reading for Meaning in Second Language Acquisition, in Coady, J. and Huckin, T. (eds.), *Second Language Vocabulary Acquisition: a Rationale for Pedagogy*, Cambridge University Press, Cambridge.

Peretz, A.S. and Shoham, M. (1990) Testing Reading Comprehension in LSP: Does Topic Familiarity Affect Assessed Difficulty and Actual Performance? *Reading in a Foreign Language*, 7,1: 447-455.

Perkins, K. (1992) The Effect of Passage and Topical Structure Types on ESL Reading Comprehension Difficulty, *Language Testing*, 9,2: 163-172.

Perkins, K., Gupta, L. and Tammana, R. (1995) Predicting Item Difficulty in a Reading Comprehension Test with an Artificial Neural Network, *Language Testing*, 12,1: 34-53.

Peirce, B.N. (1992) Demystifying the TOEFL Reading Test, *TESOL Quarterly*, 26,4: 665-689.

Peirce, B.N. (1994) The Test of English As a Foreign Language: Developing Items for Reading Comprehension, in Hill, C. and Parry, K. (eds.), *From Testing to Assessment: English as an International Language*, Longman, London.

Pollitt, A. and Taylor, L. (1993) Question Level Bias in Cloze Questions: An L1 Transfer Effect, in Sajavaara, K., Takala, S. and Huhta, A. (eds.), *Language Testing: New Openings*, Institute for Educational Research, University of Jyväskylä, Finland.
Pollitt, A.B. and Taylor, L.B. (1997) Good Reading Tests aren’t as Good as They Seem to Be, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.) Current Developments and Alternatives in Language Assessment, University of Jyväskylä.

Riley, G.L. and Lee, J.F. (1996) A Comparison of Recall and Summary Protocols as Measures of Second Language Reading Comprehension, Language Testing, 13,2: 173-190.

Rost, D.H. (1993) Assessing the Different Components of Reading Comprehension: Fact or Fiction, Language Testing, 10,1: 79-92.

Schedl, M., Gordon, A., Carey, P.A. and Tang, K.L. (1996) An Analysis of the Dimensionality of TOEFL Reading Comprehension Items, Research Report 95-27, Educational Testing Service, Princeton, NJ.

Storey, P. (1995) Investigating Construct Validity through Test Taker Introspection, unpublished Ph.D. thesis, University of Reading.

Storey, P. (1997) Examining the Test-Taking Process: a Cognitive Perspective on the Discourse Cloze Test, Language Testing, 14,2: 214-231.

Tan, S.H. (1990) The Role of Prior Knowledge and Language Proficiency as Predictors of Reading Comprehension among Undergraduates, in de Jong, J.H.A.L. and Stevenson, D.K. (eds.), Individualizing the Assessment of Language Abilities, Multilingual Matters, Clevedon, England.

Urquhart, A. and Weir, C.J. (1998) Reading in a Second Language: Process, Product and Practice, Longman.

Weir, C.J. (1997) The Testing of Reading in a Second Language, in Clapham, C.M. and Corson, D. (eds.), Language Testing and Assessment, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Weir, C.J. (1998) The Selection of Texts and Tasks for Testing and Teaching Academic Reading Ability in English, in Quality in Learning in English Medium Higher Education, Bilkent University Press, 224-233.

Weir, C.J., Hughes, A. and Porter, D. (1990) Reading Skills: Hierarchies, Implicational Relationships and Identifiability, Reading in a Foreign Language, 7,1: 505-510.

Weir, C.J. and Porter, D. (1994) The Multi-Divisible or Unitary Nature of Reading: The Language Tester between Scylla and Charybdis, Reading in a Foreign Language, 10,2: 1-20.

West, R. (1991) Developments in the Testing of Reading, English as a World Language, 1,1: 60-70.

Wijgh, I.F. (1996) A Communicative Test in Analysis: Strategies in Reading Authentic Texts, in Cumming, A. and Berwick, R. (eds.), Validation in Language Testing, Multilingual Matters, Clevedon, England.

Wolf, D. F. (1993) Issues in Reading Comprehension Assessment: Implications for the Development of Research Instruments and Classroom Tests, Foreign Language Annals, 26,3: 322-31.

Wolf, D.F. (1993) A Comparison of Assessment Tasks Used to Measure FL Reading Comprehension, Modern Language Journal, 77,4: 473-489.
4. Testing Speaking

Bachman, L.F., Lynch, B. and Mason, M. (1995) Investigating Variability in Tasks and Rater Judgements in a Performance Test of Foreign Language Speaking. Language Testing, 12,2: 238-257.

Berry, V. (1997) Ethical Considerations when Assessing Oral Proficiency in Pairs, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Berwick, R. and Ross, S. (1996) Cross-Cultural Pragmatics in Oral Proficiency Interview Strategies, in Milanovic, M. and Saville, N. (eds.), Performance Testing, Cognition and Assessment, Cambridge University Press, Cambridge.

Brown, A. (1993) The Role of Test-Taker Feedback in the Test Development Process: Test-Takers’ Reactions to a Tape-Mediated Test of Proficiency in Spoken Japanese, Language Testing, 10,3: 277-303.

Brown, A. and Lumley, T. (1997) Interviewer Variability in Specific-Purpose Language Performance Tests, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Casselar, M.I. (1997) Modified Oral Proficiency Interview: Its Purpose, Development and Description, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Chalhoub-Deville, M. (1993) Performance Assessment and the Components of the Oral Construct Across Different Tasks and Rater Groups, ERIC Document Reproduction Service No. 360 830.

Chalhoub-Deville, M. (1995) A Contextualized Approach to Describing Oral Language Proficiency, Language Learning, 45: 251-281.

Chalhoub-Deville, M. (1995) Deriving Oral Assessment Scales across Different Tests and Rater Groups, Language Testing, 12,1: 16-33.

Chalhoub-Deville, M. (1996) Performance Assessment and the Components of the Oral Construct across Different Tests and Rater Groups, in Milanovic, M. and Saville, N. (eds.), Performance Testing, Cognition and Assessment, Cambridge University Press, Cambridge.

Chambers, F. and Richards, B. (1993) Oral Assessment: the Views of Language Teachers, Language Learning Journal, 7: 22-26.

Clark, J.L.D. (1992) ‘Screen-to-screen’ testing: an exploration of speaking proficiency measures in the TOELF context, TOEFL Research Report 4, Educational Testing Service, Princeton, NJ.

Cohen, A.D. and Olshtain, E. (1993) The Production of Speech Acts by EFL Learners, TESOL Quarterly, 27,1: 33-56.

Dandonoli, P. and Henning, G. (1990) An Investigation of the Construct Validity of the ACTFL Proficiency Guidelines and Oral Interview Procedure, Foreign Language Annals, 23,1: 11-22.

de Jong, J.H.A.L. and van Ginkel, L.W. (1992) Dimensions in Oral Foreign Language Proficiency, in Verhoeven, L. and de Jong, J.H.A.L. (eds.), The Construct of Language Proficiency: Applications of Psychological Models to Language Assessment, John Benjamins Publishing Co., Amsterdam.

Dick, R.C. and Robinson, B.M. (1994) Oral Proficiency Requirements for ITAs in U.S. Colleges and Universities: An Issue in Speech Communication, JACA (Journal for the Association for Communication Administration), 2: 77-86.

Douglas, D. (1994) Quantity and Quality in Speaking Test Performance, Language Testing, 11,2: 125-144.
Douglas, D. (1997) *Testing Speaking Ability in Academic Contexts: Theoretical Considerations*, TOEFL Monograph Series, Number 8, Educational Testing Service, Princeton, NJ.

Douglas, D. and Selinker, L. (1992) Analysing Oral Proficiency Test Performance in General and Specific Purpose Contexts, *System*, 20: 317-328.

Douglas, D. and Selinker, L. (1993) Performance on General versus Field-Specific Tests of Speaking Proficiency, in Douglas, D. and Chapelle, C.A. (eds.), *A New Decade of Language Testing Research*, TESOL Publications, Alexandria, VA.

Douglas, D. and Selinker, L. (1994) Native and Non-native Teaching Assistants: A Case Study of Discourse Domains and Genres, in Madden, C. and Myers, C. (eds.) *Discourse and Performance of International Teaching Assistants*, TESOL Publications, Alexandria, VA, 221-230.

Douglas, D. and Smith, J. (1997) *Theoretical Underpinnings of the Test of Spoken English Revision Project*, TOEFL Monograph Series, Number 9, Educational Testing Service, Princeton, NJ.

Ellerton, A.W. (1997) *Considerations in the Validation of Semi-Direct Oral Testing*, unpublished Ph.D. thesis, University of Reading.

Fulcher, G. (1993) *The Construct Validation of Rating Scales for Oral Tests in English as a Foreign Language*, unpublished Ph.D. thesis, University of Lancaster.

Fulcher, G. (1994) Some Priority Areas for Oral Language Testing, *Language Testing Update*, 15: 39-47.

Fulcher, G. (1996) Invalidation Validity Claims for the ACTFL Oral Rating Scale, *System*, 24,2: 163-172.

Fulcher, G. (1996) Testing Tasks: Issues in Task Design and the Group Oral, *Language Testing*, 13,1: 23-51.

Fulcher, G. (1996) Does Thick Description lead to Smart Tests? A Data-Based Approach to Rating Scale Construction, *Language Testing*, 13,2: 208-238.

Fulcher, G. (1997) The Testing of Speaking in a Second Language, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Gallego, J.C., Goodwin, J. and Turner, J. (1991) ITA Oral Assessment: The Examinee As Perspective In Preparing the Professoriate of Tomorrow to Teach, in Nyquist J.D., Abbott, R.D., Wulff, D.H. and Sprague, J. (eds.) *Selected Readings in TA Training* (From the Second National Conference on the Training and Employment of Graduate Teaching Assistants), Kendall/Hunt Publishing Company, Dubuque, Iowa, 404-412.

Halleck, G.B. (1995) Assessing Oral Proficiency: a Comparison of Holistic and Objective Measures, *Modern Language Journal*, 79,2: 223-234.

Halleck, G.B. and Moder, C.L. (1995) Testing Language and Teaching Skills of International Teaching Assistants: the Limits of Compensatory Strategies, *TESOL Quarterly*, 29,4: 733-758.

Hasselgren, A. (1997) Oral Test Subskill Scores: What They Tell Us About Raters and Pupils, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Hilsdon, J. (1991) The Group Oral Exam: Advantages and Limitations, in Alderson, J.C. and North, B. (eds.), *Language Testing in the 1990s: the Communicative Legacy*, Macmillan, London.

Hoecherl-Alden, G. and Boyce, J. (1994) What’s the Story? A Suggestion for an Oral Exam and its Assessment in Beginning and Intermediate German Classes, *Unterrichtspraxis*, 27: 62-68.

Hoekeje, B. and Linnell, K. (1994) ‘Authenticity’ in Language Testing: Evaluating Spoken Language Tests for International Teaching Assistants, *TESOL Quarterly*, 28,1: 103-126.
Ingram, D. and Wylie, E. (1993) Assessing Speaking Proficiency in the International English Language Testing System, in Douglas, D. and Chapelle, C.A. (eds.), A New Decade of Language Testing Research, TESOL Publications, Alexandria, VA.

Iwashita, N. (1996) The Validity of the Paired Interview Format in Oral Performance Assessment, Melbourne Papers in Language Testing, 5,2.

Iwashita, N. (1997) Assessment of Oral Communication Skills in LOTE Settings in Australia, Melbourne Papers in Language Testing, 6,2: 37-44.

Johnson, K.E. (1991) Modifying the SPEAK Test for International Teaching Assistants, TESOL Matters, 5,1: 8.

Knight, B. (1992) Assessing Speaking Skills: a Workshop for Teacher Development, English Language Teaching Journal, 46,3: 294-302.

Lazaraton, A. (1991) A Conversation Analysis of Structure and Interaction in the Language Interview, unpublished Ph.D. dissertation, University of California, Los Angeles.

Lazaraton, A. (1992) The Structural Organization of a Language Interview: a Conversation Analytic Perspective, System, 20,3: 373-386.

Lazaraton, A. (1996) A Qualitative Approach to Monitoring Examiner Conduct in CASE, in Milanovic, M. and Saville, N. (eds.), Performance Testing, Cognition and Assessment, Cambridge University Press, Cambridge.

Lazaraton, A. (1996) Interlocutor Support in Oral Proficiency Interviews: the Case of CASE, Language Testing, 13,2: 151-172.

Lunt, H. (1994) The Investigation of Oral Proficiency and Language Learning Strategies in a Migrant ESL Context, Melbourne Papers in Language Testing, 3,2.

McNamara, T.F. (1997) Interaction in Second Language Performance Assessment: Whose Performance? Applied Linguistics.

McNamara, T.F. and Lumley, T. (1997) The Effect of Interlocutor and Assessment Mode Variables in Overseas Assessments of Speaking Skills in Occupational Settings, Language Testing, 14,2: 140-156.

Meldman, M.A. (1991) The Validation of Oral Performance Tests for Second Language Learners, in McGroarty, M.E. and Faltis, C.J. (eds.), Languages in School and Society: Policy and Pedagogy, Mouton de Gruyter, Berlin.

Meredith, R.A. (1990) The Oral Proficiency Interview in Real Life: Sharpening the Scale, Modern Language Review, 74,3: 288-296.

Mets, J., van den Bosch, L. and de Jong, J.H.A.L. (1993) The Listening and Speaking Tests in the New Examinations for Dutch as a Second Language, Toegepaste Taalwetenschap in Artikelen, 46/47: 207-221.

Milleret, M., Stansfield, C.W. and Kenyon, D.M. (1991) The Validity of the Portuguese Speaking Test for Use in a Summer Study Abroad Program, Hispania, 74: 778-87.

Munro, M.J. and Derwing, T.M. (1994) Evaluation of Foreign Accent in Extemporaneous and Read Material, Language Testing, 11,3: 253-266.

North, B. (1994) Scales of Language Proficiency: A Survey of Some Existing Systems, Strasbourg: Council of Europe CC-LANG (94) 24.
North, B. (1995) The Development of a Common Framework Scale of Descriptors of Language Proficiency based on a Theory of Measurement, System 23, 4: 445-465.

North, B. (1996) The Development of a Common Framework Scale of Descriptors of Language Proficiency Based on a Theory of Measurement, unpublished Ph.D. thesis, Thames Valley University.

North, B. (1995) Scales of Language Proficiency, Melbourne Papers in Language Testing, 4,2.

North, B. (1997) The Development of a Common Framework Scale of Descriptors of Language Proficiency Based on a Theory of Measurement, in Huita, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

North, B. and Schneider, G. (1998) Scaling Descriptors for Language Proficiency Scales, Language Testing, 15,2: 217-262.

O'Loughlin, K. (1994) Lexical Density in Candidate Output On Two Versions of an Oral Proficiency Test, Melbourne Papers in Language Testing, 3,1.

O'Loughlin, K. (1995) Lexical Density in Candidate Output on Direct and Semi-Direct Versions of an Oral Proficiency Test, Language Testing, 12,2: 217-237.

O'Loughlin, K. (1997) The Equivalence of Two Versions of an Oral Proficiency Test, unpublished Ph.D. thesis, University of Melbourne.

Pavlou, P. and Rasi, S. (1996) Considerations in Developing Writing Prompts for Simulated Oral Proficiency Interviews, Melbourne Papers in Language Testing, 5,2.

Pavlou, P. (1997) Do Different Speech Interactions in an Oral Proficiency Test Yield Different Kinds of Language? in Huhta, A., Kohonen, V., Kurki-Suonio, K. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Perrett, G. (1990) The Language Testing Interview: a Reappraisal, in de Jong, J.H.A.L. and Stevenson, D. K. (eds.), Individualizing the Assessment of Language Abilities, Multilingual Matters, Philadelphia.

Porter, D. (1991) Affective Factors in Language Testing, in Alderson, J.C. and North, B. (eds.), Language Testing in the 1990s: The Communicative Legacy, Macmillan, London.

Reed, D.J. and Halleck, G.B. (1997) Probing Above the Ceiling in Oral Interviews: What’s Up There?, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Rekart, D. and Dunkel, P. (1992) The Utility of Objective (Computer) Measures of the Fluency of Speakers of English as a Second Language, Applied Language Learning, 3: 65-85.

Reves, T. (1991) From Testing Research to Educational Policy: a Comprehensive Test of Oral Proficiency, in Alderson, J.C. and North, B. (eds.), Language Testing in the 1990s: The Communicative Legacy, Macmillan, London.

Ross, S. (1992) Accommodative Questions in Oral Proficiency Interviews, Language Testing, 9,2: 173-186.

Ross, S. and Berwick, R. (1992) The Discourse of Accommodation in Oral Proficiency Interviews, Studies in Second Language Acquisition, 14: 159-176.

Saleva, M. (1997) Now They’re Talking: Testing Oral Proficiency in a Language Laboratory, University of Jyväskylä.

Sarwark, S.M., Smith, J., MacCallum, R. and Cascallar, E.C. (1995) A Study of Characteristics of the SPEAK Test, Research Reports (Report no. 49), Educational Testing Service, Princeton, NJ.

Shohamy, E. (1994) The Validity of Direct Versus Semi-Direct Oral Tests, Language Testing, 11,2: 99-123.
Shohamy, E. and Stansfield, C.W. (1990) The Hebrew Speaking Test: An Example of International Cooperation in Test Development and Validation, *AILA Review*, 7: 79-90.

Spolsky, B. (1990) Oral Examinations: an Historical Note, *Language Testing*, 7,2: 158-173.

Stansfield, C.W. (1991) A Comparative Analysis of Simulated and Direct Oral Proficiency Interviews, in Anivan, S. (ed.), *Current Developments in Language Testing*, RELC Anthology Series 23, RELC.

Stansfield, C.W. (1993) An Approach to Performance Testing of Oral Language Proficiency for Bilingual Education Teacher Certification, *Proceedings of the Third National Research Symposium on Limited English Proficient Student Issues: Focus on Middle and High School Issues*, United States Department of Education, Office of Bilingual Education and Minority Language Affairs, Washington, DC.

Stansfield, C.W. and Kenyon, D.M. (1992) Research on the Comparability of the Oral Proficiency Interview and the Simulated Oral Proficiency Interview, *System*, 20,3: 347-364.

Stansfield, C.W. and Kenyon, D.M. (1996) Comparing the Scaling of Speaking Tasks by Language Teachers and by the ACTFL Guidelines, in Cumming, A. and Berwick, R. (eds.), *Validation in Language Testing*, Multilingual Matters, Clevedon, England.

Thompson, I. (1995) A Study of Interrater Reliability of the ACTFL Oral Proficiency Interview in Five European Languages: Data from ESL, French, German, Russian and Spanish, *Foreign Language Annals*, 28,3: 407-422.

Tonkyn, A. (1996) The Oral Language Development of Instructed Second Language Learners: The Quest for a Progress-Sensitive Proficiency Measure, in Coleman, H. and Cameron, L. (eds.), *Change and Language*, British Studies in Applied Linguistics 10, Multilingual Matters, Clevedon, England.

van Gelderen, A. (1994) Prediction of Global Ratings of Fluency and Delivery in Narrative Discourse by Linguistic and Phonetic Measures - Oral Performances of Students Aged 11-12 years, *Language Testing*, 11,3: 291-319.

Wall, D. (1994) The Testing of Speaking, *GRETA: Revista para Profesores do Ingles*, 1,2, Granada.

Weir, C.J. and Bygate, M. (1993) Meeting the Criteria of Communicativeness in a Spoken Language Test, *Journal of English and Foreign Languages*, Special Issue on Evaluation in the Language Curriculum Hyderabad, 27-43.

Wigglesworth, G. (1993) Exploring Bias Analysis as a Tool for Improving Rater Consistency in Assessing Oral Interaction, *Language Testing*, 10,3: 305-336.

Wigglesworth, G. (1997) An Investigation of Planning Time and Proficiency Level on Oral Test Discourse, *Language Testing*, 14,1: 85-106.

Wigglesworth, G. and O’Loughlin, K. (1993) An Investigation into the Comparability of Direct and Semi-Direct Versions of an Oral Interaction Test in English, *Melbourne Papers in Language Testing*, 2,1.

Young, R. and Milanovic, M. (1992) Discourse Variation in Oral Proficiency Interviews, *Second Language Acquisition*, 14: 403-424.

Young, R.E (1994) Conversational Styles in Language Proficiency Interviews, *Language Learning*, 45,1: 3-42.

Young, R. and He, A.W. (eds.) (1998) *Talking and Testing: Discourse Approaches to the Assessment of Oral Proficiency*, John Benjamins Publishing Company, Amsterdam.
5. **Testing Writing**

Allaei, S.K., and Connor, U. (1991) Using Performative Assessment Instruments with ESL Student Writers, in Hamp-Lyons, L. (ed.), *Assessing Second Language Writing in Academic Contexts*, Ablex Norwood, NJ.

Brown, J.D., Hilgers, T. and Marsella, J. (1991) Essay Prompts and Topics: Minimizing the Effect of Differences. *Written Communication*, 8,4: 532-555.

Chastain, K. (1990) Characteristics of Graded and Ungraded Compositions, *Modern Language Journal*, 74,1: 10-14.

Connor-Linton, J. (1995) Looking Behind the Curtain: What Do L2 Composition Ratings Really Mean? *TESOL Quarterly*, 29,4: 762-5.

Cumming, A. (1990) Expertise in Evaluating Second Language Compositions, *Language Testing*, 7,1: 31-51.

Cumming, A. (1997) The Testing of Writing in a Second Language, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Cumming, A. and Mellow, D. (1996) An Investigation into the Validity of Written Indicators of Second Language Proficiency, in Cumming, A. and Berwick, R. (eds.), *Validation in Language Testing*, Multilingual Matters, Clevendon, England.

Cushing Weigle, S. (1994) *Effects of Training on Raters of ESL Compositions: an Investigation of Construct Validity*, unpublished Ph.D. thesis, University of California, Los Angeles.

Cushing Weigle, S. (1994) Effects of Training on Raters of ESL Compositions, *Language Testing*, 11,2: 197-223.

Educational Testing Service (1992) *TOEFL Test of Written English* (third edition), Educational Testing Service, Princeton, NJ.

Freedman, S. (1991) *Evaluating Writing: Linking Large-scale Testing and Classroom Assessment*, Occasional Paper No. 27, Center for the Study of Writing, Berkeley, CA.

Fulcher, G. (1997) Assessing Writing, in Fulcher, G. (ed.), *Writing in the English Language Classroom*, Prentice Hall, Europe in association with the British Council.

Giffert, S. and Harada, K. (1992) Two Composition Scoring Methods: the Analytic versus the Holistic Method, *The Bulletin of Faculty in Foreign Languages*, 1: 17-22.

Ginther, A. and Grant, L. (1997) The Influences of Proficiency, Language Background, and Topic on the Production of Grammatical Form and Error on the Test of Written English, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Grabe, W. and Kaplan, R.B. (1996) *Theory and Practice of Writing*, Longman, London.

Hamp-Lyons, L. (1990) Second Language Writing: Assessment Issues, in Kroll, B. (ed.), *Second Language Writing: Research Insights for the Classroom*, Cambridge University Press, Cambridge.

Hamp-Lyons, L. (ed.) (1991) *Assessing Second Language Writing in Academic Contexts*, Ablex, Norwood, NJ.

Hamp-Lyons, L. (1991) Scoring Procedures for ESL Contexts, in Hamp-Lyons, L. (ed.), *Assessing Second Language Writing in Academic Contexts*, Ablex, Norwood, NJ.

Hamp-Lyons, L. (1994) Exploring Bias in Essay Tests Through Student Interviews, in Guerra, J., Severino, C. and Butler, J. (eds.), *Writing in Multicultural Settings*, Modern Language Association, New York.

Hamp-Lyons, L. (1994) The Challenges of Second Language Writing Assessment, in Lutz, W., Kamusikiri, S. and White, E. (eds.), *The Practices and Politics of Writing Assessment*, Modern Language Association, New York.
Hamp-Lyons, L. (1995) Rating Nonnative Writing: The Trouble with Holistic Scoring, *TESOL Quarterly*, 29,4: 759-62.

Hamp-Lyons, L. (1996) Applying Ethical Standards to Portfolio Assessment of Writing in English as a Second Language, in Milanovic, M. and Saville, N. (eds.), *Performance Testing, Cognition and Assessment*, Cambridge University Press, Cambridge.

Hamp-Lyons, L. (1996) The Challenges of Second Language Writing Assessment, in White, E.M., Lutz, W.D., and Kamusikiri, S. (eds.) *Assessment of Writing: Politics, Policies, Practices*, Modern Language Association, New York, 226-240.

Hamp-Lyons, L. (1997) Exploring Bias in Essay Tests Through Student Interviews, in Severino, C., Guerra, J., and Butler, J. (eds.) *Writing in Multicultural Settings*, Modern Language Association, New York, 51-66.

Hamp-Lyons, L. and Condon, W. (1993) Questioning Assumptions about Portfolio-Based Assessment, *College Composition and Communication*, 44,2: 176-90.

Hamp-Lyons, L. and Condon, W. (1999) *Assessing College Writing Portfolios: Principles for Practice, Theory, Research*, Hampton Press, Cresskill, NJ.

Hamp-Lyons, L. and Henning, G. (1991) Communicative Writing Profiles: An Investigation into the Transferability of a Multiple-Trait Scoring Instrument across ESL Writing Contexts, *Language Learning*, 41,3: 337-373.

Hamp-Lyons, L. and Prochnow, S. (1991) Combining Holistic Scoring with Diagnostic Feedback in a Large-Scale Writing Test, *Language Testing Update*, 9: 12-15.

Hamp-Lyons, L. and Prochnow, S. (1991) The Difficulties of Difficulty: Prompts in Writing Assessment, in Anivan, S. (ed.), *Current Developments in Language Testing*, RELC Anthology Series 23, RELC/SEAMEO, 58-76.

Henning, G. (1991) Issues in Evaluating and Maintaining an ESL Writing Assessment Program, in Hamp-Lyons, L. (ed.), *Assessing Second Language Writing in Academic Contexts*, Ablex, Norwood, NJ.

Herter, R.J. (1991) Writing Portfolios: Alternatives to Testing, *English Journal*, 80,1: 90-91.

Ishikawa, S. (1995) Objective Measurement of Low-Proficiency EFL Narrative Writing, *Journal of Second Language Writing*, 4: 51-69.

Johns, A. (1991) Faculty Assessment of ESL Literacy Skills: Implications for Writing Assessment, in Hamp-Lyons, L. (ed.), *Assessing Second Language Writing in Academic Contexts*, Ablex, Norwood, NJ.

Kobayashi, H. and Rinnert, C. (1996) Factors Affecting Composition Evaluation in an EFL Context: Cultural Rhetorical Pattern and Readers' Background, *Language Learning*, 46: 397-437.

Kroll, B. and Reid, J. (1994) Guidelines for Writing Prompts: Clarifications, Caveats, and Cautions, *Journal of Second Language Writing*, 3: 231-55.

Lewkowicz, J. (1995) Authenticity of Task vs. Quality of Written Output, in Tickoo, M. (ed.), *Reading and Writing: Theory into Practice*, SEAMEO, Singapore.

Milanovic, M., Saville, N. and Shuhong, S. (1996) A Study of the Decision-Making Behaviour of Composition Markers, in Milanovic, M. and Saville, N. (eds.), *Performance Testing, Cognition and Assessment*, Cambridge University Press, Cambridge.

Polio, C. (1997) Measures of Linguistic Accuracy in Research on Second Language Writing, *Language Learning*, 47: 101-143.

Raimes, A. (1990) The TOEFL Test of Written English: Causes for Concern, *TESOL Quarterly*, 24,3: 427-442.
Read, J. (1990) Providing Relevant Content in an EAP Writing Test, *English for Specific Purposes*, 9: 109-121.

Read, J. (1991) The Validity of Writing Tasks, in Anivan, S. (ed.), *Current Developments in Language Testing*, RELC Anthology Series 23, RELC.

Salvi, R. (1991) A Communicative Approach to Testing Written English in Nonnative Speakers, *Rassegna Italiana di Linguistica Applicata*, 23: 67-91.

Schoonen, R., Vergeer, M. and Eiting, M. (1997) The Assessment of Writing Ability: Expert Readers versus Lay Readers, *Language Testing*, 14,2: 157-184.

Shohamy, E., Gordon, C. and Kraemer, R. (1992) The Effect of Raters' Background and Training on the Reliability of Direct Writing Tests, *Modern Language Journal*, 76,1: 27-33.

Simmons, J. (1990) Portfolios as Large-Scale Assessment, *Language Arts*, 67,3: 262-68.

Storch, N. (1993) The Development and Validation of a Writing Test and an Analytic Scoring Scheme used in the ESL Program, the University of Melbourne, *Melbourne Papers in Language Testing*, 2,2.

Song, B. and Caruso, I. (1996) Do English and ESL Faculty Differ in Evaluating the Essays of Native English-Speaking and ESL Students? *Journal of Second Language Writing*, 5: 163-82.

Spaan, M. (1993) The Effect of the Prompt in Essay Examinations, in Douglas, D. and Chapelle, C.A. (eds.), *A New Decade of Language Testing Research*, TESOL, Alexandria, Virginia.

Taylor, C.A. (1996) A Report of Writing Tasks Assigned in Academic Degree Programs: a Report on Stage 1 of the Project, in Milanovic, M. and Saville, N. (eds.), *Performance Testing, Cognition and Assessment*, Cambridge University Press, Cambridge.

Tedick, D. (1990), ESL Writing Assessment: Subject-Matter Knowledge and its Impact on Performance, *English for Specific Purposes*, 9: 123-144.

Tedick, D. and Mathison, M. (1995) Holistic Scoring in ESL Writing Assessment: What Does an Analysis of Rhetorical Features Reveal? in Belcher, D. and Braine, G. (eds.), *Academic Writing in a Second Language: Essays on Research and Pedagogy*, Ablex, Norwood, NJ.

Tyndall, B. and Kenyon, D.M. (1996) Validation of a New Holistic Rating Scale Using Rasch Multi-faceted Analysis, in Cumming, A. and Berwick, R. (eds.), *Validation in Language Testing*, Multilingual Matters, Clevedon, England.

Vaughan, C. (1991) Holistic Assessment: What Goes on in the Raters Minds? in Hamp-Lyons, L. (ed.), *Assessing Second Language Writing in Academic Contexts*, Ablex, Norwood, NJ.
6. Testing Language Elements

Alderson, J.C. (1993) The Relationship Between Grammar and Reading in an English for Academic Purposes Test Battery, in Douglas, D. and Chapelle, C. (eds.), A New Decade of Language Testing Research, TESOL, Alexandria, VA.

Chapelle, C.A. (1994) Are C-tests Valid Measures for L2 Vocabulary Research? Second Language Research, 10,2: 157-187.

Chapelle, C.A. (1998) Construct Definition and Validity Inquiry in SLA Research, in Bachman, L.F. and Cohen, A.D. (eds.), Interfaces between Second Language Acquisition and Language Testing Research, Cambridge University Press, Cambridge.

Cushing Weigle, S. and Lynch, B. (1995) Hypothesis Testing in Construct Validation, in Cumming, A. and Berwick, R. (eds.), Validation in Language Testing, Multilingual Matters, Clevedon, England.

Harrison, A. (1991) Language Assessment as Theatre: Ten Years of Communicative Testing, in Alderson, J.C. and North, B. (eds.), Language Testing in the 1990s: The Communicative Legacy, Macmillan, London.

Harrison, A. (1991) Response to Rea-Dickins, in Alderson, J.C. and North, B. (eds.), Language Testing in the 1990s: The Communicative Legacy, Macmillan, London.

Henning, G.H. (1991) A Study of the Effects of Contextualization and Familiarization on Responses to the TOEFL Vocabulary Test Items, TOEFL Research Reports, No. 35, Educational Testing Service, Princeton, NJ.

Kitao, S.K. and Kitao, K. (1996) Testing Grammar, The Internet TESL Journal, 2,6.

Meara, P. (1990) Some Notes on the Eurocentres Vocabulary Tests, in Tommola, J. (ed.), Foreign Language Comprehension and Production, Turku, Finland.

Meara, P. (1992) EFL Vocabulary Tests, University College Swansea, Centre for Applied Language Studies.

Meara, P. (1994) LLEX: Threshold Vocabulary Tests, Swansea, Centre for Applied Language Studies.

Meara, P. and Jones, G. (1990) Eurocentres Vocabulary Size Test 10Ka (version E1.1/K10, MSDOS), Zurich, Eurocentres.

Meara, P., Lightbown, P.M. and Halter, R.H. (1994) The Effect of Cognates on the Applicability of YES/NO Vocabulary Tests, Canadian Modern Language Review, 50,2: 296-311.

Nation, P. (1993) Using Dictionaries to Estimate Vocabulary Size: Essential, but Rarely Followed, Procedures, Language Testing, 10,1: 27-40.

Paribakht, T. and Wesche, M. (1997) Vocabulary Enhancement Activities and Reading for Meaning in Second Language Acquisition, in Coady, J. and Huckin, T. (eds.), Second Language Vocabulary Acquisition: a Rationale for Pedagogy, Cambridge University Press, Cambridge.

Read, J. (1993) The Development of a New Measure of L2 Vocabulary Knowledge, Language Testing, 10,3: 355-371.

Read, J. (1995) Refining the Word Associates Format as a Measure of Depth of Vocabulary Knowledge, New Zealand Studies in Applied Linguistics, 1: 1-17.

Read, J. (1997) Assessing Vocabulary in a Second Language, in Clapham, C.M. and Corson, D. (eds.), Language Testing and Assessment, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Read, J. (in press) Assessing Vocabulary, Cambridge University Press, Cambridge.

Rea-Dickins, P. (1991) What makes a Grammar Test Communicative? in Alderson, J.C. and North, B. (eds.), Language Testing in the 1990s: The Communicative Legacy, Macmillan, London.
Rea-Dickins, P. (1997) The Testing of Grammar in a Second Language, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Singleton, D. and Little, D. (1991) The Second Language Lexicon: some Evidence from University Level Learners of French and German, *Second Language Research*, 7: 62-81.

Stansfield, C.W., Karl, J. and Kenyon, D.M. (1990) *The Guam Educators’ Test of English Proficiency (GETEP)*, Final Project Report, Revised, Center for Applied Linguistics, Washington, DC.

Thomas, M. (1994) Assessment of L2 Proficiency in Second Language Acquisition Research, *Language Learning*, 44,2: 307-336.

Turner, C. E. and Upshur, J. A. (1995) Some Effects of Task Type on the Development of Communicative Effectiveness and Grammatical Accuracy in Intensive ESL Classes, *TESOL Canada Journal*, 12,2: 18-31.

Verhoeven, L. and de Jong, J. H. A. L. (eds.) (1992) *The Construct of Language Proficiency: Applications of Psychological Models to Language Assessment*, John Benjamins Publishing Company, Amsterdam.

Wesche, M. and Paribakht, T.S. (1996) Assessing Second Language Vocabulary Knowledge: Depth versus Breadth, *Canadian Modern Language Review*, 53,1: 13-40.
7. Purposes for Testing

Alderson, J.C. and Clapham, C.M. (1992) Applied Linguistics and Language Testing: a Case Study of the ELTS Test, *Applied Linguistics*, 13,2: 149-167.

Alderson, J.C. and Clapham, C.M. (1992) *Examining the ELTS Test: An Account of the First Stage of the ELTS Revision Project*, IELTS Research Report 2, Cambridge: The British Council, University of Cambridge Local Examinations Syndicate and the International Development Program of Australian Universities and Colleges.

Bachman, L.F., Davidson, F. and Milanovic, M. (1996) The Use of Test Method Characteristics in the Content Analysis and Design of EFL Proficiency Tests, *Language Testing*, 13,2: 125-150.

Blais, J.G. and Laurier, M.D. (1995) The Dimensionality of a Placement Test from Several Analytical Perspectives, *Language Testing*, 12, 1: 72-98.

Bradshaw, J. (1990) Test-Takers’ Reactions to a Placement Test, *Language Testing*, 7,1: 13-30.

Briggs, S.L. (1994) Using Performance Assessment Methods to screen ITAs, in Madden, C.G. and Myers, C.L. (eds.), *Discourse and Performance of International Teaching Assistants*, TESOL Publications, Alexandria, VA. 63-80.

Brindley, G. (1991) Assessing Achievement in a Learner-Centred Curriculum, in Alderson, J.C. and North, B. (eds.), *Language Testing in the 1990s: The Communicative Legacy*, Macmillan, London.

Broadfoot, P. (1996) Performance Assessment in Perspective, in Craft, A. (ed.), *Primary Education: Assessing and Planning Learning*, Routledge in association with the Open University, London.

Brown, A. and Lumley, T. (1994) How Can an English Proficiency Test Be Made More Culturally Appropriate? a Case Study, in Kaur Gill, S. (ed.), *National and International Challenges and Responses. Proceedings of the International English Language Education Conference*, Language Centre Universiti Kebangsaan, Kuala Lumpur.

Brown, J.D. (1990) The Placement of ESL Students in a Writing across the Curriculum Program, *University of Hawaii Working Papers in ESL*, 9,1: 35-58.

Cascallar, E.C. (1995) Introduction, in Cascallar, E.C. (guest ed.), *Special Issue on Language Aptitude, Language Testing*, 12,3: 259-63.

Cascallar, E.C. (1997) The Assessment of Language Learning Aptitude, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Chalhoub-Deville, M. (1997) The Minnesota Articulation Project and its Proficiency-Based Assessment Instruments, *Foreign Language Annals*, 30: 492-502.

Chalhoub-Deville, M. (1997) Theoretical Models, Assessment Frameworks and Test Construction, *Language Testing*, 14,1: 3-22.

Clapham, C.M. (1996) *The Development of IELTS: A Study of the Effect of Background Knowledge on Reading Comprehension*, Cambridge University Press, Cambridge.

Coleman, J., Grotjahn, R., Klein-Braley, C. and Raatz, U. (1994) The European Language Proficiency Survey: a Comparative Investigation of Foreign Language Learners in Schools and Universities in Several European Countries, *Language Testing Update*, 16: 49-55.

Cumming, A. (1995) Changing Definitions of Language Proficiency: Functions of Language Assessment in Educational Programs for Recent Immigrant Learners of English in Canada, *Journal of the Canadian Association of Applied Linguistics*, 17: 35-48.

D’Anglejan, A., Harley, B. and Shapson, S. (1990) Student Evaluation in a Multidimensional Core French Curriculum, *Canadian Modern Language Review*, 47,1: 106-124.
David, G. and Hudak, I. (1993) The Test of Language Competence for Teacher Trainees at Eötvös Lorand University, *Language Testing Update*, 13: 27-29.

Davidson, F. (1994) Norms Appropriacy of Achievement Tests: Spanish-Speaking Children and English Children’s Norms, *Language Testing*, 11,1: 83-95.

de Jong, J.H.A.L. (1992) Assessment of Language Proficiency in the Perspective of the 21st Century, *AILA Review*, 9: 39-45.

de Jong, J.H.A.L. (1992) Certificates in Dutch as a Second Language, *Language Testing Update*, 11: 25-26.

de Jong, J.H.A.L. and Verhoeven, L.W. (1992) Modeling and Assessing Language Proficiency, in Verhoeven, L.W. and de Jong, J.H.A.L. (eds.), *The Construct of Language Proficiency*, John Benjamins, Amsterdam.

DeMauro, G. (1992) Examination of the Relationships among TSE, TWE and TOEFL scores, *Language Testing*, 9,2: 149-162.

DeVincenzi, F. (1995) Examining Standardised Test Content: some Advice for Teachers, in Peirce, B.N. (ed.), *Language Tests and ESL Teaching*, TESOL Quarterly, 29,1: 180-184.

Douglas, D. and Selinker, L. (1992) Analysing Oral Proficiency Test Performance in General and Specific Purpose Contexts, *System*, 20: 317-328.

Douglas, D. and Selinker, L. (1993) Performance on General versus Field-Specific Tests of Speaking Proficiency, in Douglas, D. and Chapelle, C.A. (eds.), *A New Decade of Language Testing Research*, TESOL Publications, Alexandria, VA.

Educational Testing Service (1991) *Computerised Placement Tests*, Educational Testing Service, Princeton, NJ.

Elder, C. (1991) Second Language Proficiency Testing in Relation to Classroom Performance: Two NLIA Language Centre Projects, *Language Testing Update*, 10: 34-38.

Elder, C. (1993) Language Proficiency as a Predictor of Performance in Teacher Education, *Melbourne Papers in Language Testing*, 2,1.

Elder, C. (1993) How do Subject Specialists Construe Classroom Language Proficiency? *Language Testing*, 10,3: 235-254 [An earlier version appeared in 1992 under the title ‘How do Subject Specialists Construe Language Proficiency?’ *Melbourne Papers in Language Testing*, 1,1].

Elder, C. (1995) The Effect of Language Background on ‘Foreign’ Language Test Performance: Problems of Classification and Measurement, *Language Testing Update*, 17: 36-38.

English Proficiency Tests in Canada (1992) *Language Testing Update*, 11: 6-21.

Fox, J. (1992) English Proficiency Tests in Canada, *Language Testing Update*, 11: 6-9.

Fulcher, G. (1997) An English Language Placement Test: Issues in Reliability and Validity, *Language Testing*, 14,2: 113-139.

Gardner, R.C. (1991) Second Language Learning in Adults: Correlates of Proficiency, *Applied Language Learning*, 2: 1-28.

Geranpayeh, A. (1994) Are Score Comparisons across Language Proficiency Test Batteries Justified? an IELTS-TOEFL Comparability Study, Edinburgh Working Papers in Applied Linguistics, 5: 50-65.

Gronlund, N.E. (1993) *How to Make Achievement Tests and Assessments*, Allyn and Bacon, Boston.

Hamers, J.H.M., Sijtsma, K. and Ruijssenaars, A.J.J.M. (eds.) (1993) *Learning Potential Assessment: Theoretical, Methodological and Practical Issues*, Swets and Zeitlinger B.V., Amsterdam.
Harrison, A. (1991) Language Assessment as Theatre: Ten Years of Communicative Testing, in Alderson, J.C. and North, B. (eds.), *Language Testing in the 1990s: The Communicative Legacy*, Macmillan, London.

Harrison, A. (1991) Response to Rea-Dickins, in Alderson, J.C. and North, B. (eds.), *Language Testing in the 1990s: The Communicative Legacy*, Macmillan, London.

Kitao, S.K. and Kitao, K. (1996) Testing Communicative Competence, *The Internet TESL Journal*, 2,5.

Kunnan, A.J. (1990) DIF in Native Language and Gender Groups in an ESL Placement Test, *TESOL Quarterly*, 24,4: 740-746.

Kyllonen, P.C. (1993) Aptitude Testing Inspired by Information Processing: a Test of the Four-Sources Model, *The Journal of General Psychology*, 120: 375-405.

Lazaraton, A. (1996) A Qualitative Approach to Monitoring Examiner Conduct in CASE, in Milanovic, M. and Saville, N. (eds.), *Performance Testing, Cognition and Assessment*, Cambridge University Press, Cambridge.

Lissik-Gasparro, J.E. (1995) Practical Approaches to Outcomes Assessment: the Undergraduate Major in Foreign Languages and Literatures, *ADFL Bulletin*, 26,2: 21-27.

Lussier, D., Turner, C.E. and Desharnais, S. (1993) Measuring Second Language Proficiency in High School Level Exchange Students, *Canadian Modern Language Review*, 49,3: 526-549.

MacWhinney, B. (1996) Language-Specific Prediction in Foreign Language Learning, in Cascallar, E.C. (guest ed.), *Special Issue on Language Aptitude, Language Testing*, 12,3: 292-320.

McLaughlin, B. (1995) Aptitude from an Information-Processing Perspective, in Cascallar, E.C. (guest ed.), *Special Issue on Language Aptitude, Language Testing*, 12,3: 370-87.

North, B. and Schneider, G. (1998) Scaling Descriptors for Language Proficiency Scales, *Language Testing*, 15,2: 217-262.

O’Mara, F.E., Lett, J.A. and Alexander, E.E. (1994) The Prediction of Language Learning Success at DLIFLC, in U.S. Department of Defense Report, *The Language Skill Change Project, LSCP Report II*, Defense Language Institute Foreign Language Center, Monterey CA.

Paltridge, B. (1992) EAP Placement Testing: An Integrated Approach, *English for Specific Purposes*, 11: 243-268.

Parry, T.S. and Stansfield, C.W. (eds.) (1990) *Language Aptitude Reconsidered*, Prentice Hall Regents, Englewood Cliffs, NJ.

Sasaki, M. (1993) Relationships among Second Language Proficiency, Foreign Language Aptitude, and Intelligence: A Structural Equation Modeling Approach, *Language Learning*, 43: 313-44.

Sasaki, M. (1996) *Second Language Proficiency, Foreign Language Aptitude and Intelligence: Quantitative and Qualitative Analyses*, Theoretical Studies in Second Language Acquisition Vol. 6, Peter Lang, New York.

Shohamy, E. (1992) Beyond Proficiency Testing: A Diagnostic Feedback Testing Model for Assessing Foreign Language Learning, *The Modern Language Journal*, 76,4: 513-521.

Shohamy, E. (1995) Language Testing: Matching Assessment Procedures with Language Knowledge, in Birenbaum, M. and Dochy, F.R.J. (eds.), *Alternatives in Assessment of Achievement, Prior Knowledge and Learning Procedures*, Kluwer, Boston.

Spaan, M. (1994) The Examination for the Certificate of Competence in English (ECCE), *Language Testing Update*, 15: 25-28.
Sparks, R.L., Artzer, M., Ganschow, L., Siebenhar, D., Plageman, M. and Patton, J. (1998) Differences in Native-language Skills, Foreign-language Aptitude, and Foreign-language Grades among High-, Average-, and Low-proficiency Foreign-language Learners: Two Studies, *Language Testing*, 15,2: 181-216.

Spolsky, B. (1990) The Prehistory of TOEFL, *Language Testing*, 7,1: 98-118.

Spolsky, B. (1995) Prognostication and Language Aptitude Testing, 1925-62, in Cascallar, E.C. (guest ed.), *Special Issue on Language Aptitude, Language Testing*, 12,3: 321-40.

St. John, M.J. (1990) UET(O) to UETESOL: the Revised JMB Examination in English for Academic Purposes, *Language Testing Update*, 8: 10-14.

Stansfield, C.W., Karl, J. and Kenyon, D.M. (1990) *The Guam Educators’ Test of English Proficiency (GETEP)*, Final Project Report, Revised, Center for Applied Linguistics, Washington, DC.

Sternberg, R.J. (1995) Styles of Thinking and Learning, in Cascallar, E.C. (guest ed.), *Special Issue on Language Aptitude, Language Testing*, 12,3: 265-91.

Tabori, J.R. and Cascallar, M.I. (1996) *Factors that Participate in the Prediction of Language Learning Success: Aptitude, Motivation, and Background*, Technical Report, Federal Bureau of Investigation, Washington DC.

Tharu, J. (1993) Tests of English Proficiency: the Problem of Standards, *Journal of English as a Foreign Language*, 3/4: 59-78.

Thomas, M. (1994) Assessment of L2 Proficiency in Second Language Acquisition Research, *Language Learning*, 44,2: 307-336.

Van Weeren, J. (1995) European Language Qualifications: Information on Standards of Proficiency, *System*, 23,4: 481-490.

Verhoeven, L. and de Jong, J. H. A. L. (eds.) (1992) *The Construct of Language Proficiency: Applications of Psychological Models to Language Assessment*, John Benjamins Publishing Company, Amsterdam.

Wall, D., Clapham, C.M., and Alderson, J.C. (1994) Evaluating a Placement Test, *Language Testing*, 11,3: 321-344.

Wesche, M., Paribakht, T.S. and Ready, D. (1996) A Comparative Study of four ESL Placement Instruments, in Milanovic, M. and Saville, N. (eds.), *Performance Testing, Cognition and Assessment*, Cambridge University Press, Cambridge.

Wette, R. and Daley, T. (1994) Achievement and Proficiency Assessment: Adult New Migrants, *Language Testing Update*, 15: 19-20.

Wolf, R. (1992) Performance Assessment in IEA Studies, *IEA Bulletin*, 1,2:1-2.
8. Specific Purposes Testing

Absalom, D. and Brice, A. (1997) Sharpening our Pragmatic Instruments: Additive Assessment in ESL, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Ainscough, V. (1992) A New English Language Test for Upper Secondary Schools in the Congo: the Overriding Issue of Practicality, Language Testing Update, 12: 45-51.

Bachman, L. (1991) What Does Language Testing Have to Offer? TESOL Quarterly, 25,4: 671-704.

Barrs, M. (1990) Words not Numbers: Assessment in English, National Association of Advisers in English.

Briggs, S.L. (1994) Using Performance Assessment Methods to screen ITAs, in Madden, C.G. and Myers, C.L. (eds.), Discourse and Performance of International Teaching Assistants, TESOL Publications, Alexandria, VA. 63-80.

Brindley, G. (ed.) (1995) Language Assessment in Action, National Centre for Language Teaching and Research, Macquarie University, Sydney, Australia.

Brown, J.D. (1991) Content-Based Language Testing, Cross Currents, 18,2: 174-178.

Brown, J.D. (1991) Performance of ESL Students on a State Minimal Competency Test, Issues in Applied Linguistics, 2,1: 31-47.

Brown, J.D. (1992) Identifying ESL Students at Linguistic Risk on a State Minimal Competency Test, TESOL Quarterly, 26,1: 167-172.

Brown, J.D. and Wolfe-Quintero, K. (1997) Teacher Portfolios for Evaluation: A Great Idea? Or a Waste of Time? The Language Teacher, 21: 28-30.

Brunfitt, C.J. (ed.) (1991) Assessment in Literature Teaching, Modern English Publications, Macmillan, London.

Bruton, A. (1991) Continuous assessment in Spanish State Schools, Language Testing Update, 10: 14-20.

Bunton, D. (1995) Teaching for the 'Practical Skills' Exam, Hong Kong Examinations Authority, Hong Kong.

Carpenter, K., Fujii, N., and Kataoka, H. (1995) An Oral Interview Procedure for Assessing Second Language Abilities in Children, Language Testing, 12,2: 157-175.

Cawood, G. and Moore, J. (1994) Diplomatic Service Language Allowance (DSLA) Examinations, Language Testing Update, 15: 21-24.

Clapham, C.M. (1993) Can ESP Testing be Justified? in Douglas, D. and Chapelle, C.A. (eds.), A New Decade of Language Testing Research, TESOL Publications, Alexandria, VA.

Clapham, C.M. (1996) What Makes an ESP Reading Test Appropriate for its Candidates? in Cumming, A. and Berwick, R. (eds.), Validation in Language Testing, Multilingual Matters, Clevedon, England.

Cohen, A.D. (1994) English for Academic Purposes in Brazil: the Use of Summary Tasks, in Hill, C. and Parry, K. (eds.), From Testing to Assessment: English as an International Language, Longman, London.

D’Anglejan, A., Harley, B. and Shapson, S. (1990) Student Evaluation in a Multidimensional Core French Curriculum, Canadian Modern Language Review, 47,1: 106-124.

Daugherty, R. (1995) National Curriculum Assessment: A Review of Policy 1987 - 1994, The Falmer Press, London.

David, G. and Hudak, I. (1993) The Test of Language Competence for Teacher Trainees at Eötvös Lorand University, Language Testing Update, 13: 27-29.
Davis, W.E. (1991) Comparing Language Proficiency Scores and Student Evaluations to Determine Policy for International Teaching Assistants, *College Student Journal*, 25,4: 489-495.

Dick, R.C. and Robinson, B.M. (1994) Oral Proficiency Requirements for ITAs in U.S. Colleges and Universities: An Issue in Speech Communication, *JACA* (Journal for the Association for Communication Administration), 2: 77-86.

Douglas, D. (1997) Language for Specific Purposes Testing, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Douglas, D. (1997) *Testing Speaking Ability in Academic Contexts: Theoretical Considerations*, TOEFL Monograph Series, Number 8, Educational Testing Service, Princeton, NJ.

Douglas, D. (1998) Testing Methods in Context-Based Second Language Research, in Bachman, L.F. and Cohen, A.D. (eds.), *Interfaces between Second Language Acquisition and Language Testing Research*, Cambridge University Press, Cambridge.

Douglas, D. (in press) *Testing Language for Specific Purposes: Theory and Practice*, Cambridge University Press, Cambridge.

Douglas, D. and Selinker, L. (1992) Analysing Oral Proficiency Test Performance in General and Specific Purpose Contexts, *System*, 20: 317-328.

Douglas, D. and Selinker, L. (1993) Performance on General versus Field-Specific Tests of Speaking Proficiency, in Douglas, D. and Chapelle, C.A. (eds.), *A New Decade of Language Testing Research*, TESOL Publications, Alexandria, VA.

Douglas, D. and Selinker, L. (1994) Native and Non-native Teaching Assistants: A Case Study of Discourse Domains and Genres, in Madden, C. and Myers, C. (eds.) *Discourse and Performance of International Teaching Assistants*, TESOL Publications, Alexandria, VA, 221-230.

Duncan, K. (1995) The Role of Testing in the Evaluation of a Primary Education Project: The Case of Molteno, in Rea-Dickens, P. and Lwaitama, A.F.L. (eds.), *Evaluation for Development in English Language Teaching*, ELT documents, Macmillan, London.

Edelenbosch, P. and de Jong, J.H.A.L. (eds.) (1995) Special Issue: Foreign Languages Examinations at the End of Secondary Education, *Tijdschrift voor Onderwijswetenschappen*, 24, 5/6.

Elder, C. (1994) Are Raters’ Judgements of Language Teacher Effectiveness Wholly Language-Based? *Melbourne Papers in Language Testing*, 3,2.

Elder, C. (1994) Performance Testing as a Benchmark for LOTE Teacher Education, *Melbourne Papers in Language Testing*, 3,1.

Elder, C. and Brown, A. (1997) Performance Testing for the Professions: Language Proficiency or Strategic Competence? *Melbourne Papers in Language Testing*, 6,1.

Elder, C., Davies, A., Hajek, J., Manwaring, D. and Warren, J. (1996) Is Grammar Good For You? The Relationship Between Metalinguistic Knowledge and Success in Studying a Language at University, *Melbourne Papers in Language Testing*, 5,1.

Fradd, S.H., McGee, P.L. and Wilen, D.K. (1994) *Instructional Assessment: an Integrative Approach to Evaluating Student Performance*, Addison-Wesley, Reading, Mass.

Fulcher, G. (1999) Assessment in English for Academic Purposes: Putting content validity in its place, *Applied Linguistics*, 20,2: 221-236.

Gallego, J.C., Goodwin, J. and Turner, J. (1991) ITA Oral Assessment: The Examinee As Perspective In Preparing the Professoriate of Tomorrow to Teach, in Nyquist J.D., Abbott, R.D., Wulff, D.H. and Sprague, J. (eds.) *Selected
Readings in TA Training (From the Second National Conference on the Training and Employment of Graduate Teaching Assistants), Kendall/Hunt Publishing Company, Dubuque, Iowa, 404-412.

Genese, F. and Hamayan, E.V. (1994) Classroom-Based Assessment, in Genese, F. (ed.), Educating Second Language Children: The Whole Child, The Whole Curriculum, The Whole Community, Cambridge University Press, Cambridge.

Genese, F. and Upshur, J.A. (1996) Classroom-Based Evaluation in Second Language Education: Three Approaches to Validity, Cambridge University Press, Cambridge.

Gipps, C.V. and Stobart, G. (1993) Assessment, A Teacher’s Guide to the Issues, Hodder and Stoughton, London.

Gottlieb, M. and Ewy, C. (1996) Multi-Level and Multi-Faceted: State and Local Assessment of Linguistically and Culturally Diverse Students, Linguathon, 11,1: 3-4.

Grant, L. (1997) Testing the Language Proficiency of Bilingual Teachers: Arizona’s Spanish Proficiency Test, Language Testing, 14,1: 23-46.

Griffin, P., Smith, P.G. and Burrill, L.E. (1995) The American Literacy Profile Scales, Heinemann, Portsmouth, NH.

Hall, D. (1995) Assessing the Needs of Bilingual Pupils. Living in Two Languages, David Fulton Publishers, London.

Halleck, G.B. and Moder, C.L. (1995) Testing Language and Teaching Skills of International Teaching Assistants: the Limits of Compensatory Strategies, TESOL Quarterly, 29,4: 733-758.

Hamilton, J., Lopes, M., McNamara, T.F. and Sheridan, E. (1993) Rating Scales and Native Speaker Performance on a Communicatively Oriented EAP Test, Language Testing, 10,3: 337-353.

Hoekje, B. and Linnell, K. (1994) ‘Authenticity’ in Language Testing: Evaluating Spoken Language Tests for International Teaching Assistants, TESOL Quarterly, 28,1: 103-126.

Holland, K., Bloome, D. and Solsken, J. (1994) Alternative Perspectives in Assessing Children's Language and Literacy, Ablex, Norwood, NJ.

Holland, D. and Street, B. (1994) Assessing Adult Literacy in the United Kingdom: The Progress Profile, in Hill, C. and Parry, K. (eds.), From Testing to Assessment: English as an International Language, Longman, London.

Hudson, T., Detmer, E. and Brown, J.D. (1992) A Framework for Testing Cross-Cultural Pragmatics, Technical Report #2, University of Hawai‘i at Manoa, Second Language Teaching and Curriculum Center, Honolulu.

Hudson, T., Detmer, E. and Brown, J.D. (1995) Developing Prototypic Measures of Cross-Cultural Pragmatics, Technical Report #7, University of Hawai‘i at Manoa, Second Language Teaching and Curriculum Center, Honolulu.

Jacoby, S. and McNamara, T.F. (1999) Locating competence. English for Specific Purposes, 18,2.

Johnson, K.E. (1991) Modifying the SPEAK Test for International Teaching Assistants, TESOL Matters, 5,1: 8.

Jones, S. (1990) Evaluation and Assessment for ESL Literacy, TESL Talk, 20: 294-304.

Kenyon, D.M. and Stansfield, C.W. (1993) A Method for Improving Tasks on Performance-Based Assessments through Field-Testing, in Huhta, A., Sajavaara, K. and Takala, S. (eds.), Language Testing: New Openings, Institute for Educational Research, University of Jyväskylä, Finland.

Law, B. and Eckes, M. (1995) Assessment and ESL, Peguis, Winnipeg.

Leung, C. and Teasdale, A. (1996) English as an Additional Language within the National Curriculum: A Study of Assessment Practices, Prospect, 12,2: 58-68.
Leung, C. and Teasdale, A. (1997) What Do Teachers Mean by Speaking and Listening? A Contextualised Study of Assessment in Multilingual Classrooms in the English National Curriculum, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Liskin-Gasparro, J.E. (1995) Practical Approaches to Outcomes Assessment: the Undergraduate Major in Foreign Languages and Literatures, ADFL Bulletin, 26,2: 21-27.

Low, L., Duffield, J., Brown, S. and Johnstone, R. (1993) Evaluating Foreign Languages in Primary Schools, Scottish CILT, University of Stirling, Stirling.

Lukmani, Y. (1996) Linguistic Accuracy versus Coherence in Assessing Examination Answers in Content Subjects, in Milanovic, M. and Saville, N. (eds.), Performance Testing, Cognition and Assessment, Cambridge University Press, Cambridge.

Lumley, T. (1993) Reading Comprehension Sub-Skills: Teachers’ Perceptions of Content in an EAP Test, Melbourne Papers in Language Testing, 2,1.

Lumley, T. (1995) The Judgements of Language-trained Raters and Doctors in a Test of English for Health Professionals, Melbourne Papers in Language Testing, 4,1.

Lumley, T. and Brown, A. (1998) An Investigation of the Authenticity of Discourse in a Specific-Purpose Language Performance Test and its Relevance to Test Validity, in Li, E.S.L. and James, G. (eds.) Testing and Evaluation in Second Language Education, Hong Kong: Language Centre, Hong Kong University of Science and Technology, 22-33.

Lumley, T., Raso, E. and Mincham, L. (1993) Exemplar Assessment Activities, in NLLIA (National Languages and Literacy Institute of Australia): NLLIA ESL Development: Language and Literacy in Schools, National Languages and Literacy Institute of Australia, Canberra.

Lussier, D., Turner, C.E. and Desharnais, S. (1993) Measuring Second Language Proficiency in High School Level Exchange Students, Canadian Modern Language Review, 49,3: 526-549.

Madden, C.G. and Myers, C.L. (eds.) (1994) Discourse and Performance of International Teaching Assistants, TESOL Publications, Alexandria, VA.

McCann, P. and Teasdale, A. (1992) Eurocontrol Standard Testing Project: Exit Test in English for Air Traffic Control, Language Testing Update, 12: 52.

McKay, P. (1995) Developing ESL Proficiency Descriptions for the School Context: The NLLIA ESL Bandscales, in Brindley, G. (ed.), Language Assessment in Action, National Centre for Language Teaching and Research, Macquarie University, Sydney.

McKay, P. (1997) The Second Language Assessment of Minority Children, in Clapham, C.M. and Corson, D. (eds.), Language Testing and Assessment, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

McNamara, T.F. (1990) Assessing the Second Language Proficiency of Health Professionals, unpublished Ph.D. thesis, University of Melbourne.

McNamara, T.F. (1990) Item Response Theory and the Validation of an ESP Test for Health Professionals, Language Testing, 7,1: 52-76.

McNamara, T.F. (1991) Test Dimensionality: IRT Analysis of an ESP Listening Test, Language Testing, 8,2: 139-159.

McNamara, T. (1997) Problematising Content Validity: the Occupational English Test (OET) as a Measure of Medical Communication, Melbourne Papers in Language Testing, 6,1.
Mincham, L. (1995) ESL Student Needs Procedures: An Approach to Language Assessment in Primary and Secondary School Contexts, in Brindley, G. (ed.), Language Assessment in Action, National Centre for Language Teaching and Research, Macquarie University, Sydney.

Moore, H. (1996) Telling What is Real: Competing Views in Assessing ESL Development, Linguistics and Education, 8,2: 189-228.

Paltridge, B. (1992) EAP Placement Testing: An Integrated Approach, English for Specific Purposes, 11: 243-268.

Peretz, A.S. and Shoham, M. (1990) Testing Reading Comprehension in LSP: Does Topic Familiarity Affect Assessed Difficulty and Actual Performance? Reading in a Foreign Language, 7,1: 447-455.

Plakans, B. and Abraham, R. (1990) The Testing and Evaluation of International Teaching Assistants, in Douglas, D. (ed.), English Language Testing in U.S. Colleges and Universities, NAFSA, Washington, DC.

Read, J. (1990) Providing Relevant Content in an EAP Writing Test, English for Specific Purposes, 9: 109-121.

Rea-Dickins, P. and Rixon, S. (1997) The Assessment of Young Learners of English as a Foreign Language, in Clapham, C.M. and Corson, D. (eds.), Language Testing and Assessment, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Robson, A. (1995) The Assessment of Bilingual Children, in Verma, M.K., Corrigan, K.P. and Firth, S. (eds.), Working with Bilingual Children, Multilingual Matters, Clevedon, England.

Sajavaara, K. (1992) Designing Tests to Match the Needs of the Workplace, in Shohamy, E. and Walton, A.R. (eds.), Language Assessment for Feedback: Testing and Other Strategies, Kendall/Hunt, Dubuque, IA.

Shohamy, E. (1993) The Power of Tests: The Impact of Language Tests on Teaching and Learning, NFLC Occasional Papers, The National Foreign Language Center, Washington DC.

Short, D. (1993) Assessing Integrated Language and Content Instruction, TESOL Quarterly, 27,3: 627-656.

Smith, K. (1996) Action Research on the Use of Portfolio for Assessing Foreign Language Learners, Testing SIG Newsletter, April, IATEFL, 17-24.

Smith, K. (1996) Assessing and Testing Young Learners: Can We? Should We? in Allan, D. (ed.), Entry Points, Papers from a Symposium of the Research, Testing and Young Learners Special Interest Groups, Cambridge, IATEFL, Whitstable, Kent.

Spielman, S.J. (1993) Adapting Benchmarks to French Immersion Education, Orbit, 24,2: 25-28.

Stansfield, C.W. (1993) An Approach to Performance Testing of Oral Language Proficiency for Bilingual Education Teacher Certification, Proceedings of the Third National Research Symposium on Limited English Proficient Student Issues: Focus on Middle and High School Issues, United States Department of Education, Office of Bilingual Education and Minority Language Affairs, Washington, DC.

Sutton, R. (1995) Assessment for Learning, RS Publications, Salford.

Tan, S.H. (1990) The Role of Prior Knowledge and Language Proficiency as Predictors of Reading Comprehension among Undergraduates, in de Jong, J.H.A.L. and Stevenson, D.K. (eds.), Individualizing the Assessment of Language Abilities, Multilingual Matters, Clevedon, England.

Teasdale, A. (1996) Content Validity in Tests for Well-Defined LSP Domains: an Approach to Defining What is to be Tested, in Milanovic, M. and Saville, N. (eds.), Performance Testing, Cognition and Assessment, Cambridge University Press, Cambridge, 211-230.

Tedick, D. (1990), ESL Writing Assessment: Subject-Matter Knowledge and its Impact on Performance, English for Specific Purposes, 9: 123-144.
TESOL (1996) Promising Futures. ESL Standards for Pre-K-12 students, TESOL Professional Papers #1.

Towler, L. and Broadfoot, P. (1992) Self-Assessment in the Primary School, *Educational Review*, 44,2: 137-151.

van Gelderen, A. (1994) Prediction of Global Ratings of Fluency and Delivery in Narrative Discourse by Linguistic and Phonetic Measures - Oral Performances of Students Aged 11-12 years, *Language Testing*, 11,3: 291-319.

Valdés, G. and Figueroa, R.A. (1994) *Bilingualism and Testing. A Special Case of Bias*, Ablex, Norwood, NJ.

Weir, C.J. (1998) The Selection of Texts and Tasks for Testing and Teaching Academic Reading Ability in English, in *Quality in Learning in English Medium Higher Education*, Bilkent University Press, 224-233.

Woods, C. and Neather, T. (1994) Target Language Testing in KS4 Examinations, *Language Learning Journal*, 10: 19-21.
9. Alternative Approaches to Assessment

Alderson, J.C. and Clapham, C.M. (1995) Assessing Student Performance in the ESL Classroom, *TESOL Quarterly*, 29,1: 184-187.

Alderson, J.C. and Windeatt, S. (1991) ‘Learner-Adaptive’ Computer Based Language Tests, *Language Testing Update*, 9: 18-23.

Alderson, J.C. and Windeatt, S. (1991) Computers and Innovation in Language Testing in Alderson, J.C. and North, B. (eds.), *Language Testing in the 1990s: The Communicative Legacy*, Macmillan, London.

Alderson, J.C. and Windeatt, S. (1995) Is There an Interest in Innovative Items? *Language Testing Update*, 17: 58-59.

Berberich, F. (1995) Computer Adaptive Testing and Its Extension to a Teaching Model in CALL, *CAELL Journal*, 6: 11-18.

Blanche, P. (1990) Using Standardized Achievement and Oral Proficiency Tests for Self-Assessment Purposes: the DLIFLC study, *Language Testing*, 7,2: 202-229.

Blue, G.M. (1994) Self-Assessment of Foreign Language Skills: Does it work? CLE Working Papers, no. 3, University of Southampton, UK.

Brown, A. and Iwashita, N. (1996) The Role of Language Background in the Validation of a Computer-Adaptive Test, *Melbourne Papers in Language Testing*, 5,1.

Brown, A. and Iwashita, N. (1996) Language Background and Item Difficulty: The Development of a Computer-adaptive Test of Japanese, *System*, 24: 199-206.

Brown, J.D. (1992) Classroom-Centered Language Testing, *TESOL Journal*, 1,4: 12-15.

Brown, J.D. (1992) Using Computers in Language Testing, *Cross Currents*, 19,1: 92-99.

Brown, J.D. (1997) Computers in Language Testing: Present Research and Some Future Directions, *Language Learning and Technology*, 1,1: 44-59.

Brown, J.D. (ed.) (1998) *New Ways of Classroom Assessment*, TESOL, Alexandria, VA.

Brown, J.D. and Hudson, T. (1998) Alternatives in Language Assessment, *TESOL Quarterly*, 32,4: 653-675.

Brown, J.D. and Hudson, T. (1998) The Alternatives in Language Testing: Advantages and Disadvantages, *University of Hawaii Working Papers in ESL*, 16,2: 79-103.

Brown, J.D. and Wolfe-Quintero, K. (1997) Teacher Portfolios for Evaluation: A Great Idea? Or a Waste of Time? *The Language Teacher*, 21: 28-30.

Burstein, J., Frase, L., Ginther, A. and Grant, L. (1996) Technologies for Language Assessment, *Annual Review of Applied Linguistics*, 16: 240-260.

Chalhoub-Deville, M. (1996) Constructing an Operational Framework for a CAT of L2 Reading Proficiency, in Chalhoub-Deville, M. (ed.), *Issues in Computer Adaptive Testing of Second Language Reading Proficiency Conference Seminar Papers*, University of Minnesota, Bloomington, MN.

Chalhoub-Deville, M. (1997) Important Considerations in Constructing Second Language Computer Adaptive Tests, *Shiken*, The Japan Association of Language Teachers, 17: 11-15.

Chalhoub-Deville, M., (ed.) (in press) Issues in Computer Adaptive Testing of Reading Proficiency, Cambridge University Press, NY.
Chalhoub-Deville, M., Alcaya, C. and Lozier, V. (1996) An Operational Framework for Constructing a Computer-Adaptive Test of L2 Reading Ability; Theoretical and Practical Issues, CARLA working paper series #1, Center for Advanced Research on Language Acquisition, University of Minnesota, Minneapolis, Minnesota.

Chalhoub-Deville, M., Alcaya, C. and Lozier, V.M. (1997) Language and Measurement Issues in Developing Computer-Adaptive Tests of Reading Ability: The University of Minnesota Model, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Chalhoub-Deville, M. and Deville C. (1999) Computer Adaptive Testing and its Application in the field of second language testing, Annual Review of Applied Linguistics, 19: 273-299.

Coombe, C. (1992) The Relationship between Self-Assessment Ratings of Functional Skills and Basic English Skills Test Results in Adult Refugee ESL Learners, unpublished Ph.D. thesis, Graduate School, Ohio State University.

Corbel, C. (1993) Computer-Enhanced Language Assessment, National Centre for English Language Testing and Research, Sydney.

Cram, B. (1992) Training Learners for Self-Assessment, TESOL in Context, 2,2: 30-33.

Davies, A. (1995) Computers in Language Assessment, Language Testing Update, 18: 51-58.

Dunkel, P. (1992) The Use of PC-Generated Speech Technology in the Development of an L2 Listening Comprehension Proficiency Test: A Prototype Design Effort, in Pennington, M.C. and Stevens, V. (eds.), Computers in Applied Linguistics: An International Perspective, Multilingual Matters, Clevedon, England.

Dunkel, P. (1996) Checking the Utility and Appropriacy of the Content and Measurement Models Used to Develop L2 Listening Comprehension CATS: Implications for Further Development of Comprehensive CATs, in Chalhoub-Deville, M. (ed.), Issues in Computer Adaptive Testing of Second Language Reading Proficiency Conference Seminar Papers, University of Minnesota, Bloomington, MN.

Educational Testing Service (1991) Computerised Placement Tests, Educational Testing Service, Princeton, NJ.

Eignor, D.R. (1996) Adaptive Assessment of Reading Comprehension for TOEFL, in Chalhoub-Deville, M. (ed.), Issues in Computer Adaptive Testing of Second Language Reading Proficiency Conference Seminar Papers, University of Minnesota, Bloomington, MN.

Eignor, D.R., Stocking, M.L., Way, W.D. and Steffen, M. (1993) Case Studies in Computer Adaptive Test Design Through Simulation, Research Report 93-56, Educational Testing Service, Princeton.

Frase, L.T., Chapelle, C.A., Ginther, A., Chawla, P., Burstein, J., Klavans, J.L. and Bernstein, J. (1997) Technology for Language Assessment and Learning, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Fulcher, G. (1991) The Role of Assessment by Teachers in Schools, in Caudery, T. (ed.), New Thinking in TEFL, The Dolphin Series No. 21, Aarhus University Press, Denmark.

Gearhart, M. and Herman, J. (1995) Portfolio Assessment: Whose Work Is It? Issues in the Use of Classroom Assignments for Accountability, UCLA Center for the Study of Evaluation, Los Angeles.

Genesee, F. and Hamayan, E.V. (1994) Classroom-Based Assessment, in Genesee, F. (ed.), Educating Second Language Children: The Whole Child, The Whole Curriculum, The Whole Community, Cambridge University Press, Cambridge.

Genesee, F. and Upshur, J.A. (1996) Classroom-Based Evaluation in Second Language Education: Three Approaches to Validity, Cambridge University Press, Cambridge.

Gipps, C. (1995) Beyond Testing: Towards a Theory of Educational Assessment, The Falmer Press, London.
Gottlieb, M. (1997) A Peek into Portfolio Practice, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Grierson, J. (1995) Classroom-based Assessment in Intensive English Centres, in Brindley, G. (ed.), *Language Assessment in Action*, National Centre for English Language Teaching and Research, Macquarie University.

Griffee, D. (1995) Classroom Testing for Teachers who Hate Testing: Criterion-Referenced Test Construction and Evaluation (ED 385140, in the ERIC microfiche files).

Gruba, P. and Corbel, C. (1997) Computer-Based Testing, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Hamayan, E.V. (1995) Approaches to Alternative Assessment, *Annual Review of Applied Linguistics*, 15: 212-226.

Hamp-Lyons, L. (1996) Applying Ethical Standards to Portfolio Assessment of Writing in English as a Second Language, in Milanovic, M. and Saville, N. (eds.), *Performance Testing, Cognition and Assessment*, Cambridge University Press, Cambridge.

Hamp-Lyons, L. and Condon, W. (1993) Questioning Assumptions about Portfolio-Based Assessment, *College Composition and Communication*, 44,2: 176-90.

Hamp-Lyons, L. and Condon, W. (1999) *Assessing College Writing Portfolios: Principles for Practice, Theory, Research*, Hampton Press, Cresskill, NJ.

Harris, M. (1997) Self-Assessment of Language Learning in Formal Settings, *English Language Teaching Journal*, 51,1: 12-20.

Heilenman, L.K. (1990) Self-Assessment of Second Language Ability: The Role of Response Effects, *Language Testing*, 7,2: 174-201.

Heilenman, L.K. (1991) Self-Assessment and Placement: A Review of the Issues, in Teschner, R.V. (ed.), *Assessing Foreign Language Proficiency of Undergraduates*, Heinele and Heinle, Boston.

Henning, G. (1991) Validating an Item Bank in a Computer-Assisted or Computer-Adaptive Test, in Dunkel, P. (ed.), *Computer-Assisted Language Learning And Testing: Research Issues and Practice*, Newbury House, New York.

Henning, G., Anbar, M., Helm, C.E. and D’Arcy, S.J. (1993) Computer-Assisted Testing of Reading Comprehension: Comparisons among Multiple-Choice and Open-Ended Scoring Methods, in Douglas, D. and Chapelle, C.A. (eds.), *A New Decade of Language Testing Research*, TESOL Publications, Alexandria, VA.

Henning, G., Johnson, P.J., Boutin, A.J. and Rice, H.R. (1994) Automated Assembly of Pre-equated Language Proficiency Tests, *Language Testing*, 11,1: 15-28.

Herter, R.J. (1991) Writing Portfolios: Alternatives to Testing, *English Journal*, 80,1: 90-91.

Hill, R.A. (1991) ToPE: Test of Proficiency in English: The Development of an Adaptive Test, in Alderson, J.C. and North, B. (eds.), *Language Testing in the 1990s: the Communicative Legacy*, Macmillan, London.

Hudson, T. and Brown, J.D. (1990) *The Application of Computers in Second Language Reading*, Report on the Educational Improvement Fund 1989/1990, University of Hawai’i at Manoa.

Johnston, C. (1996) Computerized Testing on a Large Network: Issues for Today and Tomorrow, in Chalhoub-Deville, M. (ed.), *Issues in Computer Adaptive Testing of Second Language Reading Proficiency Conference Seminar Papers*, University of Minnesota, Bloomington, MN.
Kaya-Carton, E., Carton, A.S. and Dandonoli, P. (1991) Developing a Computer-Adaptive Test of French Reading Proficiency, in Dunkel, P. (ed.), Computer-Assisted Language Learning And Testing: Research Issues and Practice, Newbury House, New York.

Krausert, S.R. (1991) Determining the Usefulness of Self-Assessment of Foreign Language Skills: Post-Secondary ESL Students’ Placement Contribution, unpublished Ph.D. thesis, University of Southern California.

Kubiszyn, T. and Borich, G. (1990) Educational Testing and Measurement: Classroom Application and Practice (third edition), Harper Collins, USA.

Latomaa, S. (1996) Self-Reported Second Language Proficiency among Four Immigrant Groups in the Nordic Countries, Jyväskylä Cross Language Studies, 17: 169-78.

Laurier, M. (1991) What We Can Do with Computerized Adaptive Testing . . . And What We Cannot Do! in Anivan, S. (ed.), Current Developments in Language Testing, SEAMEO Regional Language Centre, Singapore.

MacIntyre, P.D., Noels, K.A., and Clément, R. (1997) Biases in Self-Ratings of Second Language Proficiency: the Role of Language Anxiety. Language Learning, 47,2:

Madsen, H.S. (1991) Computer-Adaptive Testing of Listening and Reading Comprehension, in Dunkel, P. (ed.), Computer-Assisted Language Learning And Testing: Research Issues and Practice, Newbury House, New York.

Moeller, A.J. and Reschke, C. (1993) A Second Look at Grading and Classroom Performance: Report of a Research Study, Modern Language Journal, 77,2: 163-169.

Moritz, C. (1995) Self-Assessment of Foreign Language Proficiency: A Critical Analysis of Issues and a Study of Cognitive Orientations of French Learners, unpublished Ph.D. thesis, Cornell University.

North, B. (1994) Itembanker: a Testing Tool for Language Teachers, Language Testing Update, 16: 85-97.

Oscarson, M. (1997) Self-Assessment of Foreign and Second Language Proficiency, in Clapham, C.M. and Corson, D. (eds.), Language Testing and Assessment, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Oscarson, M. (1999) A Review of R&D Work in the Area of Self-Assessment of Foreign Language Ability, Working Paper, The Dialang Project, Centre for Applied Language Studies, University of Jyväskylä and Department of Education, Göteborg University.

Peirce, B.N., Swain, M., and Hart, D. (1993) Self-Assessment in Two French Immersion Programmes, Applied Linguistics, 14,1: 25-42.

Pollari, P. (1997) Could Portfolio Assessment Empower EFL Learners? Portfolios in the Teaching of English as a Foreign Language in Finnish Upper Secondary School, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Rekart, D. and Dunkel, P. (1992) The Utility of Objective (Computer) Measures of the Fluency of Speakers of English as a Second Language, Applied Language Learning, 3: 65-85.

Rolfé, T. (1990) Self- and Peer-Assessment in the ESL Curriculum, in Brindley, G. (ed.), The Second Language Curriculum in Action, National Centre for English Language Teaching and Research, Macquarie University, Sydney.

Ross, S. (1998) Self-assessment in Second Language Testing: a Meta-analysis and Analysis of Experiential Factors, Language Testing, 15,1: 1-20.

Shameem, N. (1998) Validating Self-reported Language Proficiency by Testing Performance in an Immigrant Community: the Wellington Indo-Fijians, Language Testing, 15,1: 86-108.

Simmons, J. (1990) Portfolios as Large-Scale Assessment, Language Arts, 67,3: 262-68.
Spath Hirschman, S.A. and Traversa, A. (1997) When a Portfolio Programme Adopts the Portfolio, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Stiggins, R.J. (1997) Student-Centred Classroom Assessment (second edition), Merrill, Upper Saddle River, NJ.

Stocking, M.L. and Swanson, L. (1993) A Method for Severely Constrained Item Selection in Adaptive Testing’, Applied Psychological Measurement, 17: 277-292.

Strong-Krause, D. (1995) Computer-Adaptive Placement Exams, Language Testing Update, 18: 63.

Strong-Krause, D. (1999) Exploring the Effectiveness of Self-Assessment Strategies in ESL Placement, in Ekbatani, G. and Pierson, H. (eds.) Learner-Directed Assessment in ESL, Lawrence Erlbaum, Mawah, NJ.

Towler, L. and Broadfoot, P. (1992) Self-Assessment in the Primary School, Educational Review, 44,2: 137-151.

Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R.J., Steinberg, L. and Thissen, D. (eds.) (1990) Computerized Adaptive Testing: A Primer, Lawrence Erlbaum, Hillsdale, NJ.

Wainer, H., Dorans, N.J., Green, B.F., Mislevy, R.J. and Steinberg, L. (1990) Future Challenges, in Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R.J., Steinberg, L. and Thissen, D. (eds.), Computerized Adaptive Testing: A Primer, Lawrence Erlbaum, Hillsdale, NJ.

Wilson, K.M. (1996) Validity of Global Self-Ratings of ESL Speaking Proficiency Based on an FSI/ILR-Referenced Scale: an Empirical Assessment. In collaboration with R. Lindsay. Educational Testing Service, Princeton, NJ.

Wylie, E. (1997) Test and Non-Test Assessment for the Professional Second Language User, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Yoes, M. (1996) Exploring New Item-Types for Computerized Testing: New Possibilities and Challenges, in Chalhoub-Deville, M. (ed.), Issues in Computer Adaptive Testing of Second Language Reading Proficiency Conference Seminar Papers, University of Minnesota, Bloomington, MN.
10. Principles of Test Design

Abraham, R.G. and Chapelle, C.A. (1992) The Meaning of Cloze Test Scores: and Item Difficulty Perspective, *Modern Language Journal*, 76,4: 468-479.

Alderson, J.C. (1996) Do corpora have a role in language assessment? in Thomas, J. and Short, M. (eds.) *Using Corpora for Language Research*. Longman, Harlow, 248-259

Alderson, J.C. and Clapham, C.M. (1992) Applied Linguistics and Language Testing: a Case Study of the ELTS Test, *Applied Linguistics*, 13,2: 149-167.

Benander, R. (1990) Cultural Expression in Speech Behavior: Methods of Inquiry, *Penn Working Papers in Educational Linguistics*, 6,2.

Bensoussan, M. (1993) A Guided Summary Completion Test for Long Academic Texts, *Language Testing Update*, 14: 59-61.

Bonheim, H. (1995) Ambiguity Items, *Language Testing Update*, 18, 27-32.

Brindley, G. (1994) Task-Centred Assessment in Language Learning: The Promise and the Challenge, in Bird, N., Tsui, A.B.M., Allison, D. and McNeill, A. (eds.) *Language and Learning*, Hong Kong: Institute of Language in Education, Hong Kong Education Department, 73-94.

Brindley, G. (ed.) (1997) *Access: Issues in Test Design and Delivery*, Macquarie University, National Centre for English Language Teaching and Research, Sydney, Australia.

Brindley, G. (1998) Assessment in the AMEP: Current Trends and Future Directions, *Prospect*, 13,3: 59-72.

Brown, A. and Lumley, T. (1994) How Can an English Proficiency Test Be Made More Culturally Appropriate? a Case Study, in Kaur Gill, S. (ed.), *National and International Challenges and Responses. Proceedings of the International English Language Education Conference*, Language Centre Universiti Kebangsaan, Kuala Lumpur.

Brown, A. and McNamara, T. (1992) The Role of Test-Taker Feedback in the Test Development Process, *Melbourne Papers in Language Testing*, 1,1.

Brown, J.D. (1990) Where do Tests Fit into Language Programs? *JALT Journal*, 12,1: 121-140.

Brown, J.D. (1992) What Text Characteristics Predict Human Performance on Cloze Test Items, in the *Proceedings of the Third Conference on Language Research in Japan*, Japan: International University Japan, Urasa, 1-26.

Brown, J.D. (1993) A Comprehensive Criterion-Referenced Language Testing Project, in Douglas, D. and Chapelle, C.A. (eds.), *A New Decade of Language Testing Research*, TESOL Publications, Alexandria, VA., 163-184.

Brown, J.D. (1993) What are the Characteristics of Natural Cloze Tests? *Language Testing*, 10,2: 93-116.

Brown, J.D. (1994) A Closer Look at Cloze: Validity and reliability, in Oller, J. W. and Jonz, J. (eds.) (1994) *Cloze and Coherence*, Bucknell University Press, Lewisburg, Reprinted by permission from the original article: Brown, J.D. (1983) A closer look at cloze: Validity and reliability, in Oller, J.W. Jr. (ed.) *Issues in Language Testing Research*, Newbury House, Rowley, MA, 237-250.

Brown, J.D. (1995) Developing Norm-Referenced Language Tests for Program-Level Decision Making, in Brown, J.D., and Yamashita, S.O. (eds.) *Language Testing in Japan*, Japan Association for Language Teaching, Tokyo, 40-47.

Brown, J.D. (1995) Differences Between Norm-Referenced and Criterion-Referenced tests? in Brown, J.D. and Yamashita, S.O. (eds.) *Language Testing in Japan*, Japan Association for Language Teaching, Tokyo, 12-19.
Brown, J.D. (1995) English Language Entrance Examinations in Japan: Myths and Facts, *The Language Teacher*, 19,10: 21-26.

Brown, J.D. (1996) English Language Entrance Examinations in Japan: Problems and Solutions, in van Troyer, G. (ed.) *JALT 95: Curriculum and evaluation*, Japan Association for Language Teaching, Tokyo, 272-283.

Brown, J.D. (1996) Japanese Entrance Exams: A Measurement Problem? *The Daily Yomiuri* (Educational Supplement, February 5), 15.

Brown, J.D. (1998) An EFL Readability Index. *JALT Journal*, 20,2: 7-36.

Brown, J.D., Cook, H.G., Lockhart, C. and Ramos, T. (1991) Southeast Asian Languages Proficiency Examinations, in Anivan, S. (ed.), *Current Developments in Language Testing*, RELC Anthology Series 23, RELC/SEAMEO, 210-226.

Brown, J.D. and Hudson, T. (1998) Deciding which Language Test to Use, *Kanda Institute of Foreign Languages Academic Journal*, 7: 1-12.

Brown, J.D. and Yamashita, S.O. (1995) English Language Entrance Examinations at Japanese Universities: 1993 and 1994, in Brown, J.D. and Yamashita, S.O. (eds.) (1995) *Language Testing in Japan*, Japan Association for Language Teaching, Tokyo, 86-100.

Brown, J.D. and Yamashita, S.O. (1995) Introduction to Language Testing in Japan, in Brown, J.D. and Yamashita, S.O. (eds.) (1995) *Language Testing in Japan*, Japan Association for Language Teaching, Tokyo, 5-10.

Brown, J.D. and Yamashita, S.O. (1995) English Language Entrance Examinations at Japanese Universities: What do we Know About Them? *JALT Journal*, 17,1: 7-30.

Brown, J.D. and Yamashita, S.O. (1995) The authors respond to O'Sullivan's letter to JALT Journal: Out of Criticism Comes Knowledge. *JALT Journal*, 17,2: 257-260.

Buck, G. (1992) Translation as a Language Testing Procedure: does it work? *Language Testing*, 9,2: 123-148.

Chalhoub-Deville, M. (1997) The Minnesota Articulation Project and its Proficiency-Based Assessment Instruments, *Foreign Language Annals*, 30: 492-502.

Chalhoub-Deville, M. (1997) Theoretical Models, Assessment Frameworks and Test Construction, *Language Testing*, 14,1: 3-22.

Chalhoub-Deville, M. (in press) Task-Based Assessment: A Link To Second Language Instruction, in Skehan, P., Swain, M. and Bygate, M. (eds.) *Applied Language Studies: Task Based Research*, Longman, NY.

Chalhoub-Deville, M. and Tarone, E. (1997) Assessment Measures for Specific Contexts of Language Use, ERIC Document Reproduction Service No. 401 721.

Chapelle, C.A. (1994) Are C-tests Valid Measures for L2 Vocabulary Research? *Second Language Research*, 10,2: 157-187.

Chapelle, C.A. and Abraham, R.G. (1990) Cloze Method: What Difference Does it Make? *Language Testing*, 7,2: 121-146.

Chitravelu, N., Khong C.P. and Tan, S.H. (1992) Issues in the Design of a Large-Scale Test of English, *Melbourne Papers in Language Testing*, 1,2.

Clapham, C.M. and Alderson, J.C. (1997) IELTS Research Report 3, The British Council, the University of Cambridge Local Examinations Syndicate and the International Development Project for Australian Universities and Colleges, Cambridge.
Cohen, A. (1993) The Role of Instructions in Testing Summarizing Ability, in Douglas, D. and Chapelle, C. (eds.), A New Decade of Language Testing Research, TESOL Publications, Alexandria, VA.

Cohen, A.D. (1994) English for Academic Purposes in Brazil: the Use of Summary Tasks, in Hill, C. and Parry, K. (eds.), From Testing to Assessment: English as an International Language, Longman, London.

Davidson, F. and Lynch, B.K. (1993) Criterion-Referenced Language Test Development: A Prolegomenon, in Huhta, A., Sajavaara, K. and Takala, S. (eds.), Language Testing: New Openings, University of Jyväskylä, Institute for Educational Research, Jyväskylä, Finland.

Davies, A. (1995) Testing Communicative Language or Testing Language Communicatively: What? How? Melbourne Papers in Language Testing, 4,1.

Davies, A. (1997) The Competing Claims of Accuracy and Fluency in the Constructions of Performance Tests of Language Proficiency: Two Cheers for Robert Lado! Melbourne Papers in Language Testing, 6,2: 1-20.

Dörnyei, Z. and Katona, L. (1992) Validation of the C-test amongst Hungarian ELL Learners, Language Testing, 9,2: 187-206.

Douglas, D. (1992) Review of the ‘Bilingual Home Inventory’, Eleventh Annual Mental Measurements Yearbook, Buros Institute, 108-109.

Douglas, D. (1992) Review of the ‘Test of English for International Communication’ (TOEIC), Eleventh Annual Mental Measurements Yearbook, Buros Institute, 950-951.

Ellis, R. (1990) Individual Learning Styles in Classroom Second Language Development, in de Jong, J.H.A.L. and Stevenson, D.K. (eds.), Individualizing the Assessment of Language Abilities, Multilingual Matters, Clevedon, England.

Farhady, H. and Keramati, M.N. (1996) A Text-Driven Method for the Deletion Procedure in Cloze Passages, Language Testing, 13,2: 191-207.

Finnish National Board of Education (1995) The Framework of the Finnish National Foreign Language Certificate, Syrinx Oy/Libris Oy, Helsinki, Finland.

Fulcher, G. (1995) Variable Competence and Second Language Acquisition: a Problem for Research Methodology, System, 25,1: 25-33.

Fulcher, G. (1996) Testing Tasks: Issues in Task Design and the Group Oral, Language Testing, 13,1: 23-51.

Glaser, R. (1994) Criterion-Referenced Tests: Part I, Origins; Part II, Unfinished Business, Educational Measurement: Issues and Practice, 13: 9-11, 27-30.

Gonzalez, V., Brusca-Vega, R. and Yawkey, T. (1997) Assessment and Instruction of Culturally and Linguistically Diverse Students with or at Risk of Learning Problems: from Research to Practice, Allyn and Bacon, Needham Heights, MA.

Grove, E. (1996) Creativity and Accountability in Competency-based Language Programmes: Issues of Curriculum and Assessment, Melbourne Papers in Language Testing, 5,1.

Gunn, M. (1995) Criterion-Based Assessment: A Classroom Teacher's Perspective, in Brindley, G. (ed.), Language Assessment in Action, Research Series 8, National Centre for English Language Teaching and Research, Sydney.

Hamp-Lyons, L. and Prochnow, S. (1991) The Difficulties of Difficulty: Prompts in Writing Assessment, in Anivan, S. (ed.), Current Developments in Language Testing, RELC Anthology Series 23, RELC/SEAMEO, 58-76.

Hawthorne, L. (1997) Defining the Target Domain: What Language Skills are Required of Engineers and Nurses? Melbourne Papers in Language Testing, 6,1.
Hill, K. (1997) From Job Analysis to Task Design: Different Approaches to Simulating Teacher Language Behaviour, *Melbourne Papers in Language Testing*, 6,1: 44-52.

Holmes, J.L. (1996) Studying in Two Languages: Study Summaries in the Classroom, unpublished Ph.D. thesis, Lancaster University.

Huhta, A. and Randell, E. (1996) Multiple-Choice Summary: a Measure of Text Comprehension, in Cumming, A. and Berwick, R. (eds.), *Validation in Language Testing*, Multilingual Matters, Clevedon, England.

Hudson, T. (1993) Nothing does not Equal Zero: Problems with Applying Developmental Sequence Findings to Assessment and Pedagogy, *Studies in Second Language Acquisition*, 15,4: 461-493.

Jafarpu, A. (1995) Is C-testing Superior to Cloze? *Language Testing*, 12,2: 194-216.

Jonz, J. (1990) Another Turn in the Conversation: What Does Cloze Measure? *TESOL Quarterly*, 24,1: 61-83.

Jonz, J. (1991) Cloze Item Types and Second Language Comprehension, *Language Testing*, 8,1: 1-22.

Kaga, M. (1991) Dictation as a Measure of Japanese Proficiency, *Language Testing*, 8,2: 112-124.

Kenyon, D.M. and Stansfield, C.W. (1993) A Method for Improving Tasks on Performance-Based Assessments through Field-Testing, in Huhta, A., Sajavaara, K. and Takala, S. (eds.), *Language Testing: New Openings*, Institute for Educational Research, University of Jyväskylä, Finland.

Kirsch, I.S., Jenkins, L., Jungeblut, A. and Kolstad, A. (1993) *Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey*, National Center for Educational Statistics, Washington, DC.

Kirschner, M., Wexler, C. and Spector-Cohen, E. (1992) Avoiding Obstacles to Student Comprehension of Test Questions, *TESOL Quarterly*, 26,3: 537-56.

Klein-Braley, C. (1997) C-Tests in the Context of Reduced Redundancy Testing: an Appraisal, *Language Testing*, 14,1: 47-84.

Köberl, J. and Sigott, G. (1994) Word Frequency, Transitional Probability, and Item Facility in C-tests, *Language Testing Update*, 16: 56-62.

Kohonen, V. (1997) Authentic Assessment as an Integration of Language Learning, Teaching, Evaluation and the Teacher’s Professional Growth, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Kunnan, A.J. (1993) The Development of a Criterion-Referenced Language Test for Non-Native English Speaking Graduate Students, *Journal of English as a Foreign Language*, 10&11:118-135.

Milanovic, M., Saville, N. and Taylor, L.B. (1997) The Alte User’s Guide for Examiners, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

National Assessment of Educational Progress (1993) *NAEP 1992 Technical Report*, Educational Testing Service, Princeton.

Norris, J.M., Brown, J.D., Hudson, T. and Yoshioka, J. (1998) *Designing Second Language Performance Assessments*, University of Hawai’i, Honolulu.

Norris, J.M., Brown, J.D. and Hudson, T. (in press) *Constructing and Validating Second Language Performance Assessments: Tasks, Tests, and Rating Criteria*, University of Hawai’i Press, Honolulu, HI.

O’Malley, J.M. and Valdez Pierce, L. (1996) *Authentic Assessment for English Language Learners: Practical Approaches for Teachers*, Addison-Wesley, New York.
Oller, J.W. and Jonz, J. (eds.) (1994) Cloze and Coherence, Bucknell University Press, Lewisburg.

Perkins, K., Bruten, S.R. and Gass, S.M. (1996) An Investigation of Patterns of Discontinuous Learning: Implications for ESL Measurements, Language Testing, 13,1: 63-82.

Pienemann, M. (1992) Assessing Second Language Acquisition through 'Rapid Profile', LARC Occasional Papers no. 3, Language Acquisition Research Centre, University of Sidney, Sidney.

Lessard-Clouston, M. (1992) Assessing Culture Learning: Issues and Suggestions, Canadian Modern Language Review, 48,2: 326-341.

Lewkowicz, J. (1995) Authenticity of Task vs. Quality of Written Output, in Tickoo, M. (ed.), Reading and Writing: Theory into Practice, SEAMEO, Singapore.

Lewkowicz, J. (1997) Investigating Authenticity in Language Testing, unpublished Ph.D. thesis, University of Lancaster.

Lewkowicz, J. (1997) Authentic for whom? Does Authenticity Really Matter?, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Lewkowicz, J. (1997) The Integrated Testing of a Second Language, in Clapham, C.M. and Corson, D. (eds.), Language Testing and Assessment, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Lumley, T. and Brown, A. (1998) An Investigation of the Authenticity of Discourse in a Specific-Purpose Language Performance Test and its Relevance to Test Validity, in Li, E.S.L. and James, G. (eds.) Testing and Evaluation in Second Language Education, Hong Kong: Language Centre, Hong Kong University of Science and Technology, 22-33.

Lynch, B.K. and Davidson, F. (1994) Criterion-Referenced Language Test Development: Linking Curricula, Teachers, and Tests, TESOL Quarterly, 28,4: 727-743.

Lynch, B.K. and Davidson, F. (1997) Criterion Referenced Testing, in Clapham, C.M. and Corson, D. (eds.), Language Testing and Assessment, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

McKenna, M.C. and Layton, K. (1990) Concurrent Validity of Cloze as a Measure of Intersentential Comprehension, Journal of Educational Psychology, 82,2: 372-377.

Pavlou, P. and Rasi, S. (1996) Considerations in Developing Writing Prompts for Simulated Oral Proficiency Interviews, Melbourne Papers in Language Testing, 5,2.

Pollitt, A. and Taylor, L. (1993) Question Level Bias in Cloze Questions: An L1 Transfer Effect, in Sajavaara, K., Takala, S. and Huhta, A. (eds.), Language Testing: New Openings, Institute for Educational Research, University of Jyväskylä, Finland.

Sarwark, S.M., Smith, J., MacCallum, R. and Cascellar, E.C. (1995) A Study of Characteristics of the SPEAK Test, Research Reports (Report no. 49), Educational Testing Service, Princeton, NJ.

Shohamy, E. (1998) Language Testing and SLA Interfaces: the Case of Discourse, in Bachman, L.F. and Cohen, A.D. (eds.), Interfaces between Second Language Acquisition and Language Testing Research, Cambridge University Press, Cambridge.

Sigott, G. and Köberl, J. (1993) Validating the X-Test, Language Testing Update, 14: 54-58.

Skehan, P. (1998) Task-Based Approaches to Language Testing, in Li, E.S.L. and James, G. (eds.) Testing and Evaluation in Second Language Education, Language Centre, Hong Kong University of Science and Technology. Hong Kong.

Stansfield, C.W. (1992) The Measurement of Translation Ability, Modern Language Journal, 76,4: 455-467.
Stemmer, B. (1991) *What's on a C-Test Taker's Mind? Mental Processes in C-Test Taking*, Bochum, Germany: Universitätsverlag Dr. N. Brockmeyer.

Storey, P. (1997) Examining the Test-Taking Process: a Cognitive Perspective on the Discourse Cloze Test, *Language Testing*, 14,2: 214-231.

Swain, M. (1993) Second Language Testing and Second Language Acquisition: is there a Conflict with Traditional Psychometrics? *Language Testing*, 10,2: 193-207.

Wall, D., Kalmberzina, V., Mazuoliene, Z. and Truus, K. (1996) The Baltic States Year 12 Examination Project, *Language Testing Update*, 19: 15-27.

Weir, C.J. and Bygate, M. (1993) Meeting the Criteria of Communicativeness in a Spoken Language Test, *Journal of English and Foreign Languages*, Special Issue on Evaluation in the Language Curriculum Hyderabad, 27-43.

Wood, R. (1991) *Assessment and Testing: A Survey of Research Commissioned by the University of Cambridge Local Examinations Syndicate*, Cambridge University Press, Cambridge.

Xiao, S. and Oller, J.W. (1994) Can Relatively Perfect Translation between English and Chinese be Achieved? *Language Testing*, 11,3: 267-290.
11. Factors Affecting Test Performance

Allan, A.I.C.G. (1992) Development and Validation of a Scale to Measure Test-Wiseness in EFL/ESL Reading Test Takers, *Language Testing*, 9,2: 101-122.

Amer, A.A. (1993) Teaching EFL Students to Use a Test-Taking Strategy, *Language Testing*, 10,1: 71-78.

Anderson, N.J. (1991) Individual Differences in Strategy Use in Second Language Reading and Testing, *Modern Language Journal*, 75,4: 460-472.

Baker, R. (1996) Language Testing and the Assessment of Dementia in Second Language Settings: a Case Study, *Language Testing*, 13,1: 3-22.

Brown, A. and Iwashita, N. (1996) The Role of Language Background in the Validation of a Computer-Adaptive Test, *Melbourne Papers in Language Testing*, 5,1.

Brown, J.D. (1994) Differential Subtest Performance and Test-Taker Characteristics? *University of Hawaii Working Papers in ESL*, 12,1: 139-169.

Brown, J.D. (in press) Relative Importance of Persons, Items, Subtests and Language to TOEFL Test Variance. *Language Testing*, 16,2.

Brown, J.D., Robson, G. and Rosenkjar P. (1996) Personality, Motivation, Anxiety, Strategies, and Language Proficiency of Japanese Students, *University of Hawaii Working Papers in ESL*, 15,1: 33-72.

Cohen, A.D. (1994) The Language Used to Perform Cognitive Operations during Full Immersion Maths Tasks, *Language Testing*, 11,2: 171-196.

Cohen, A. (1997) Towards Enhancing Verbal Reports as a Source of Insights on Test-Taking Strategies, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Cohen, A.D. (1998) Strategies and Processes in Test-Taking and SLA, in Bachman, L.F. and Cohen, A.D. (eds.), *Interfaces between Second Language Acquisition and Language Testing Research*, Cambridge University Press, Cambridge.

Edwards, S. and Knott, R. (1994) Assessing Spontaneous Language Abilities of Aphasic Speakers, *Language Testing*, 11,1: 49-64.

Elder, C. (1995) Do Bilinguals Perform Better than Monolinguals on School Foreign Language Examinations? *Melbourne Papers in Language Testing*, 4,1.

Elder, C. (1995) The Effect of Language Background on ‘Foreign’ Language Test Performance: Problems of Classification and Measurement, *Language Testing Update*, 17: 36-38.

Fox, J., Pychyl, T. and Zumbo, B. (1997) An Investigation of Background Knowledge in the Assessment of Language Proficiency, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Gintner, A. and Grant, L. (1997) The Influences of Proficiency, Language Background, and Topic on the Production of Grammatical Form and Error on the Test of Written English, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Henning, G. (1990) National Issues in Individual Assessment: the Consideration of Specialisation Bias in University Language Screening Tests, in de Jong, J.H.A.L. and Stevenson, D.K. (eds.), *Individualising the Assessment of Language Abilities*, Multilingual Matters, Clevedon, England.

Kunnan, A.J. (1990) DIF in Native Language and Gender Groups in an ESL Placement Test, *TESOL Quarterly*, 24,4: 740-746.
Kunnan, A.J. (1994) Modelling Relationships among some Test-Taker Characteristics and Performance on EFL Tests: an Approach to Construct Validation, *Language Testing*, 11,3: 225-252.

Kunnan, A.J. (1995) *Test Taker Characteristics and Test Performance: A Structural Modeling Approach*, Studies in *Language Testing* 2, Cambridge University Press, Cambridge.

Lumley, T. and McNamara, T.F. (1995) Rater Characteristics and Rater Bias: Implications for Training, *Language Testing*, 12,1: 54-71.

Pollitt, A. and Taylor, L. (1993) Question Level Bias in Cloze Questions: An L1 Transfer Effect, in Sajavaara, K., Takala, S. and Huhta, A. (eds.), *Language Testing: New Openings*, Institute for Educational Research, University of Jyväskylä, Finland.

Purpura, J.E. (1996) *Modeling the Relationships between Test Takers' Reported Cognitive and Metacognitive Strategy Use and Performance on Language Tests*, unpublished Ph.D. thesis, University of California, Los Angeles.

Purpura, J. (in press) *Learner Strategies and Performance on Language Tests: A Structural Equation Modeling Approach*, Cambridge University Press, Cambridge.

Stemmer, B. (1991) *What's on a C-Test Taker's Mind? Mental Processes in C-Test Taking*, Bochum, Germany: Universtitätsverlag Dr. N. Brockmeyer.

Storey, P. (1997) Examining the Test-Taking Process: a Cognitive Perspective on the Discourse Cloze Test, *Language Testing*, 14,2: 214-231.

Sunderland, J. (1994) Gender and Language Testing: Any Connection? Preliminary explorations, *Language Testing Update*, 16: 46-56.

Sunderland, J. (1995) Gender and Language Testing, *Language Testing Update*, 17: 24-35.

Tedick, D. (1990), ESL Writing Assessment: Subject-Matter Knowledge and its Impact on Performance, *English for Specific Purposes*, 9: 123-144.

Vogely, A. (1995) Perceived Strategy Use during Performance on Three Authentic Listening Comprehension Tasks, *The Modern Language Journal*, 79,1: 41-56.

Wennerstrom, A.K. and Heiser, P. (1992) ESL Student Bias in Instructional Evaluation, *TESOL Quarterly*, 26,2: 271-288.

Wigglesworth, G. (1997) An Investigation of Planning Time and Proficiency Level on Oral Test Discourse, *Language Testing*, 14,1: 85-106.

Wijgh, I.F. (1996) A Communicative Test in Analysis: Strategies in Reading Authentic Texts, in Cumming, A. and Berwick, R. (eds.), *Validation in Language Testing*, Multilingual Matters, Clevedon, England.

Wolf, D.F. (1993) A Comparison of Assessment Tasks Used to Measure FL Reading Comprehension, *Modern Language Journal*, 77,4: 473-489.
12. Rating and Rating Scales

Alderson, J.C. (1991) Bands and Scores, in Alderson, J.C. and North, B. (eds.), Language Testing in the 1990s: The Communicative Legacy, Macmillan, London.

Alderson, J.C. (1991) Dis-sporting Life. Response to Pollitt, in Alderson, J.C. and North, B. (eds.), Language Testing in the 1990s: The Communicative Legacy, Macmillan, London.

Alderson, J.C. (1993) Judgments in Language Testing, in Douglas, D. and Chapelle, C. (eds.), A New Decade of Language Testing, TESOL Publications, Alexandria, VA.

Allaei, S.K., and Connor, U. (1991) Using Performative Assessment Instruments with ESL Student Writers, in Hamp-Lyons, L. (ed.), Assessing Second Language Writing in Academic Contexts, Ablex Norwood, NJ.

Australian Education Council (1994) ESL Scales, Curriculum Corporation, Melbourne.

Bachman, L.F., Lynch, B. and Mason, M. (1995) Investigating Variability in Tasks and Rater Judgements in a Performance Test of Foreign Language Speaking, Language Testing, 12,2: 238-257.

Beaton, A.E. and Allen, N.L. (1992) Interpreting Scales Through Scale Anchoring, Journal of Educational Statistics, 17: 191-204.

Briggs, S.L. (1994) Using Performance Assessment Methods to screen ITAs, in Madden, C.G. and Myers, C.L. (eds.), Discourse and Performance of International Teaching Assistants, TESOL Publications, Alexandria, VA. 63-80.

Brindley, G. (1998) Assessment and Reporting in Language Learning Programs: Purposes, Problems and Pitfalls, in Li, E.S.L. and James, G. (eds.) Testing and Evaluation in Second Language Education, Language Centre, Hong Kong University of Science and Technology, Hong Kong, 133-162.

Brindley, G. (1998) Describing Language Development? Rating Scales and SLA, in Bachman, L.F. and Cohen, A.D. (eds.), Interfaces between Second Language Acquisition and Language Testing Research, Cambridge University Press, Cambridge.

Brindley, G. (1998) Outcomes-based Assessment and Reporting in Language Learning Programmes: a Review of the Issues, Language Testing, 15,1: 45-85.

Brown, A. (1993) LSP Testing: the Role of Linguistic and Real-World Criteria, Melbourne Papers in Language Testing, 2,2.

Brown, A. (1995) The Effect of Rater Variables in the Development of an Occupation-Specific Language Performance Test, Language Testing, 12,1: 1-15.

Brown, J.D. (1991) Do English and ESL Instructors Rate Writing Samples Differently? TESOL Quarterly, 25,4: 587-603. [An earlier version appeared in 1989 under the title "Rating writing samples: An interdepartmental perspective" in University of Hawaii Working Papers in ESL, 8,2: 119-141].

Citizenship and Immigration Canada (1996) Canadian Language Benchmarks: English as a Second Language for Adults/English as a Second Language for Literacy Learners, Working Document, Minister of Supply and Services, Ottawa, Canada.

The College Board (1993) Levels of English Proficiency (LOEP), The College Entrance Examination Board, New York.

Coniam, D. (1995) Towards a Common Ability Scale for Hong Kong English Secondary-School Forms, Language Testing, 12,2: 182-193.

Corbel, C. (1995) Exrater: A knowledge-based system for language assessors, in Brindley, G. (ed.) Assessment in Action, National Centre for English Language Testing and Research, Sydney.
Elder, C. (1993) How do Subject Specialists Construe Classroom Language Proficiency? Language Testing, 10,3: 235-254 [An earlier version appeared in 1992 under the title ‘How do Subject Specialists Construe Language Proficiency?’ Melbourne Papers in Language Testing, 1,1].

Elder, C. (1994) Are Raters’ Judgements of Language Teacher Effectiveness Wholly Language-Based? Melbourne Papers in Language Testing, 3,2.

Forsyth, R.A. (1991) Do NAEP Scales Yield Valid Criterion-Referenced Interpretations? Educational Measurement: Issues and Practice, 10, 3-9, 16.

Fulcher, G. (1993) The Construct Validation of Rating Scales for Oral Tests in English as a Foreign Language, unpublished Ph.D. thesis, University of Lancaster.

Fulcher, G. (1996) Does Thick Description lead to Smart Tests? A Data-Based Approach to Rating Scale Construction, Language Testing, 13,2: 208-238.

Fulcher, G. (1996) Invalidating Validity Claims for the ACTFL Oral Rating Scale, System, 24,2: 163-172.

Giffert, S. and Harada, K. (1992) Two Composition Scoring Methods: the Analytic versus the Holistic Method, The Bulletin of Faculty in Foreign Languages, 1: 17-22.

Griffin, P., Smith, P.G. and Burrill, L.E. (1995) The American Literacy Profile Scales, Heinemann, Portsmouth, NH.

Halleck, G.B. (1995) Assessing Oral Proficiency: a Comparison of Holistic and Objective Measures, Modern Language Journal, 79,2: 223-234.

Hamilton, J., Lopes, M., McNamara, T.F. and Sheridan, E. (1993) Rating Scales and Native Speaker Performance on a Communicatively Oriented EAP Test, Language Testing, 10,3: 337-353.

Hamp-Lyons, L. (1991) Scoring Procedures for ESL Contexts, in Hamp-Lyons, L. (ed.), Assessing Second Language Writing in Academic Contexts, Ablex, Norwood, NJ.

Hamp-Lyons, L. (1995) Rating Nonnative Writing: The Trouble with Holistic Scoring, TESOL Quarterly, 29,4: 759-62.

Hamp-Lyons, L. and Henning, G. (1991) Communicative Writing Profiles: An Investigation into the Transferability of a Multiple-Trait Scoring Instrument across ESL Writing Contexts, Language Learning, 41,3: 337-373.

Hamp-Lyons, L. and Prochnow, S. (1991) Combining Holistic Scoring with Diagnostic Feedback in a Large-Scale Writing Test, Language Testing Update, 9: 12-15.

Hasselgren, A. (1997) Oral Test Subskill Scores: What They Tell Us About Raters and Pupils, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Henning, G. (1996) Accounting for Nonsystematic Error in Performance Ratings, Language Testing, 13,1: 53-62.

Hill, K. (1995) Scales and Tests: a Case Study, Melbourne Papers in Language Testing, 4,2.

Hill, K. (1997) Who Should be the Judge? The Use of Non-Native Speakers as Raters on a Test of English as an International Language, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere [An earlier version appeared in 1996 under the title Who Should be the Judge? The Use of Non-Native Speakers as Raters on a test of English as an International Language, Melbourne Papers in Language Testing, 5,2].

Hill, K. and Storch, N. (1994) Analytic Rating Scales: How Diagnostic are they? Melbourne Papers in Language Testing, 3,1.
Huot, B. (1990) Reliability, Validity, and Holistic Scoring: What We Know and What We Need to Know, *College Composition and Communication*, 41,2: 201-213.

Ingram, D. (1990) The Australian Second Language Proficiency Ratings, in de Jong, J.H.A.L. (ed.), *Standardization in Language Testing*, AILA Review 7, Free University Press, Amsterdam.

Ingram, D. (1995) Scales, *Melbourne Papers in Language Testing*, 4,2.

Ishikawa, S. (1995) Objective Measurement of Low-Proficiency EFL Narrative Writing, *Journal of Second Language Writing*, 4: 51-69.

Jacoby, S. and McNamara, T.F. (1999) Locating competence. *English for Specific Purposes*, 18,2.

Kenyon, D. (1997) Further Research on the Efficacy of Rater Self-Training, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Kirsch, I.S., Jungeblut, A. and Mosenthal, P.B. (1998) The Measurement of Adult Literacy, in *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey*, US Department of Education.

Kobayashi, H. and Rinnert, C. (1996) Factors Affecting Composition Evaluation in an EFL Context: Cultural Rhetorical Pattern and Readers’ Background, *Language Learning*, 46: 397-437.

Lazaraton, A. (1996) A Qualitative Approach to Monitoring Examiner Conduct in CASE, in Milanovic, M. and Saville, N. (eds.), *Performance Testing, Cognition and Assessment*, Cambridge University Press, Cambridge.

Lazaraton, A. (1996) Interlocutor Support in Oral Proficiency Interviews: the Case of CASE, *Language Testing*, 13,2: 151-172.

Leung, C. and Teasdale, A. (1997) Raters’ Understanding of Rating Scales as Abstracted Concept and as Instruments for Decision Making, *Melbourne Papers in Language Testing*, Melbourne University, 6,2: 45-72.

Lukmani, Y. (1996) Linguistic Accuracy versus Coherence in Assessing Examination Answers in Content Subjects, in Milanovic, M. and Saville, N. (eds.), *Performance Testing, Cognition and Assessment*, Cambridge University Press, Cambridge.

Lumley, T. (1995) The Judgements of Language-trained Raters and Doctors in a Test of English for Health Professionals, *Melbourne Papers in Language Testing*, 4,1.

Lumley, T., Lynch, B. and McNamara, T. (1994) A New Approach To Standard-setting In Language Assessment, *Melbourne Papers in Language Testing*, 3,2.

Lumley, T. and McNamara, T.F. (1995) Rater Characteristics and Rater Bias: Implications for Training, *Language Testing*, 12,1: 54-71.

McKay, P. (1995) Developing ESL Proficiency Descriptions for the School Context: The NLLIA ESL Bandscales, in Brindley, G. (ed.), *Language Assessment in Action*, National Centre for Language Teaching and Research, Macquarie University, Sydney.

McKay, P. and Scarino, A. (1991) *The ESL Framework of Stages*, Curriculum Corporation, Melbourne.

McNamara, T.F. and Adams, R.J. (1997) New Approaches to the Analysis of Task- and Rater-Related Dependencies in Performance Assessments, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Milanovic, M., Saville, N., Pollitt, A. and Cook, A. (1996) Developing Rating Scales for CASE: Theoretical Concerns and Analyses, in Cumming, A. and Berwick, R. (eds.), *Validation in Language Testing*, Multilingual Matters, Clevedon, England.
Milanovic, M., Saville, N. and Shen, S. (1996) A Study of the Decision-Making Behaviour of Composition Markers, in Milanovic, M. and Saville, N. (eds.), Performance Testing, Cognition and Assessment, Cambridge University Press, Cambridge.

North, B. (1991) Standardisation of Continuous Assessment Grades, in Alderson, J.C. and North, B. (eds.), Language Testing in the 1990s: The Communicative Legacy, Macmillan, London.

North, B. (1993) The Development of Descriptor Scales of Language Proficiency, The National Foreign Language Center, Washington.

North, B. (1994) Empirical Development of a Common Framework Scale, Language Testing Update, 16: 36-48.

North, B. (1994) Scales of Language Proficiency: A Survey of Some Existing Systems, Strasbourg: Council of Europe CC-LANG (94) 24.

North, B. (1995) The Development of a Common Framework Scale of Descriptors of Language Proficiency based on a Theory of Measurement, System 23, 4: 445-465.

North, B. (1996) The Development of a Common Framework Scale of Descriptors of Language Proficiency Based on a Theory of Measurement, unpublished Ph.D. thesis, Thames Valley University.

North, B. (1995) Scales of Language Proficiency, Melbourne Papers in Language Testing, 4,2.

North, B. (1997) The Development of a Common Framework Scale of Descriptors of Language Proficiency Based on a Theory of Measurement, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

North, B. and Schneider, G. (1998) Scaling Descriptors for Language Proficiency Scales, Language Testing, 15,2: 217-262.

Norton Peirce, B.N. and Stewart, G. (1997) The Development of the Canadian Language Benchmarks Assessment, TESL Canada Journal, 14,1: 17-31.

Oltman, P.K. and Stricker, L.J. (1990) Developing Homogenous TOEFL Scales by Multidimensional Scaling, Language Testing, 7,1: 1-12.

O'Loughlin, K. (1992) Do English and ESL Teachers Rate Essays Differently? Melbourne Papers in Language Testing, 1,2.

Pollitt, A. (1991) Giving Students a Sporting Chance: Assessment by Counting and by Judging, in Alderson, J.C. and North, B. (eds.), Language Testing in the 1990s: The Communicative Legacy, Macmillan, London.

Pollitt, A. (1991) Response to Alderson, in Alderson, J.C. and North, B. (eds.), Language Testing in the 1990s: The Communicative Legacy, Macmillan, London.

Pollitt, A.B. and Murray, N.L. (1996) What Raters Really Pay Attention to, in Milanovic, M. and Saville, N. (eds.), Performance Testing, Cognition and Assessment, Cambridge University Press, Cambridge.

Scarino, A. (1995) Language Scales and Language Tests: Development in LOTE, Melbourne Papers in Language Testing, 4,2.

Schoonen, R., Vergeer, M. and Eiting, M. (1997) The Assessment of Writing Ability: Expert Readers versus Lay Readers, Language Testing, 14,2: 157-184.

Sheehan, K.M. (1996) A Tree-Based Approach to Proficiency Scaling, Educational Testing Service, Princeton, NJ.

Shohamy, E., Gordon, C. and Kraemer, R. (1992) The Effect of Raters' Background and Training on the Reliability of Direct Writing Tests, Modern Language Journal, 76,1: 27-33.
Song, B. and Caruso, I. (1996) Do English and ESL Faculty Differ in Evaluating the Essays of Native English-Speaking and ESL Students? *Journal of Second Language Writing*, 5: 163-82.

Stansfield, C.W. and Kenyon, D.M. (1996) Comparing the Scaling of Speaking Tasks by Language Teachers and by the ACTFL Guidelines, in Cumming, A. and Berwick, R. (eds.), *Validation in Language Testing*, Multilingual Matters, Clevedon, England.

Storch, N. (1993) The Development and Validation of a Writing Test and an Analytic Scoring Scheme used in the ESL Program, the University of Melbourne, *Melbourne Papers in Language Testing*, 2,2.

Tatsuoka, K.K., Birenbaum, M., Lewis, C. and Sheehan, K.M. (1993) *Proficiency Scaling Based on Conditional Probability Functions for Attributes*, Research Report 93-50-ONR, Educational Testing Service, Princeton, NJ.

Trim, J.L.M. (1997) The Proposed Common European Framework for the Description of Language Learning, Teaching and Assessment, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Turner, C.E. and Upshur, J.A. (1996) Developing Rating Scales for the Assessment of Second Language Performance, in Wigglesworth, G. and Elder, C. (eds.), *The Language Testing Cycle: from Inception to Washback*, Australian Review of Applied Linguistics (ARAL) Series, S, 13: 55-79.

Upshur, J.A. and Turner, C.E. (1995) Constructing Rating Scales for Second Language Tests, *English Language Teaching Journal*, 49,1: 3-12.

Upshur, J.A. and Turner, C.E. (1999) Systematic Effects and Bias in the Rating of Second Language Speaking Ability: Test Method and Learner Discourse, *Language Testing*, 16,1: 82-111

Ward, S. (1997) Inter-Rater Reliability in an ESP Context, *Melbourne Papers in Language Testing*, 6,2: 73-91.

Wigglesworth, G. (1993) Exploring Bias Analysis as a Tool for Improving Rater Consistency in Assessing Oral Interaction, *Language Testing*, 10,3: 305-336.

Yule, G. and Hoffman, P. (1993) Enlisting the Help of U.S. Undergraduates in Evaluating International Teaching Assistants, *TESOL Quarterly*, 27,2: 323-327.
13. Test Reliability and Validity

Abraham, R.G. and Chapelle, C.A. (1992) The Meaning of Cloze Test Scores: and Item Difficulty Perspective, *Modern Language Journal*, 76,4: 468-479.

Aida, Y. (1994) Examination of Horwitz, Horwitz and Cope’s Construct of Foreign Language Anxiety: the Case of Students of Japanese, *Modern Language Journal*, 78,2: 155-168.

Allan, A.I.C.G. (1995) Begging the Questionnaire: Instrument Effect on Readers’ Responses to a Self-Report Checklist, *Language Testing*, 12,2: 133-156.

Bachman, L.F. (1997) Generalizability Theory, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Bachman, L.F., Davidson, F. and Milanovic, M. (1996) The Use of Test Method Characteristics in the Content Analysis and Design of EFL Proficiency Tests, *Language Testing*, 13,2: 125-150.

Bachman, L.F., Davidson, F., Ryan, K. and Choi, I.C. (1994) *An Investigation into the Comparability of Two Tests of English as a Foreign Language: The Cambridge-TOEFL Comparability Study*, Cambridge University Press, Cambridge.

Bachman, L.F. and Eignor, D.R. (1997) Recent Advances in Quantitative Test Analysis, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Bachman, L.F., Lynch, B. and Mason, M. (1995) Investigating Variability in Tasks and Rater Judgements in a Performance Test of Foreign Language Speaking, *Language Testing*, 12,2: 238-257.

Baker, R. (1997) *Classical Test Theory and Item Response Theory in Test Analysis*, Special Report 2, Language Testing Update, Department of Linguistics, Lancaster University, Lancaster, England.

Banerjee, J. and Luoma, S. (1997) Qualitative Approaches to Test Validation, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Beland, A. and Mislevy, R. J. (1996) Probability-Based Inference in a Domain of Proportional Reasoning Tasks, *Journal of Educational Measurement*, 33: 3-27.

Berger, M.P.F. and Knol, D.L. (1990) On the Assessment of Dimensionality in Multidimensional Item Response Theory Models. Research Report, No. 90-8. Enschede: University of Twente, Department of Education.

Boldt, R.F. (1992) Cross-Validation of Item Response Curve Models using TOEFL Data, *Language Testing*, 9,1: 79-96.

Bollen, K.A. and Long, J.S. (1993) *Testing Structural Equation Models*, Sage Publications, Newbury Park.

Brown, A., Elder, C., Lumley, T., McNamara, T. and McQueen, J. (1992) Mapping Abilities and Skill Levels using Rasch Techniques, *Melbourne Papers in Language Testing*, 1,1: 37-69.

Brown, A. and Iwashita, N. (1996) The Role of Language Background in the Validation of a Computer-Adaptive Test, *Melbourne Papers in Language Testing*, 5,1.

Brown, J.D. (1990) Short-cut Estimates of Criterion-Referenced Test Consistency, *Language Testing*, 7,1: 77-97.

Brown, J.D. (1990) The Use of Multiple T-Tests in Language Research, *TESOL Quarterly*, 24,4: 770-773.

Brown, J.D. (1992) What Text Characteristics Predict Human Performance on Cloze Test Items, in the *Proceedings of the Third Conference on Language Research in Japan*, Japan: International University Japan, Urasa, 1-26.
Brown, J.D. (1994) A Closer Look at Cloze: Validity and reliability, in Oller, J. W. and Jonz, J. (eds.) (1994) Cloze and Coherence, Bucknell University Press, Lewisburg. Reprinted by permission from the original article: Brown, J.D. (1983) A closer look at cloze: Validity and reliability, in Oller, J.W, Jr. (ed.) Issues in Language Testing Research, Newbury House, Rowley, MA, 237-250.

Brown, J.D. (1997) Statistics Corner: Questions and Answers about Language Testing Statistics (Skewness and kurtosis), Shiken, 1,1: 16-18.

Brown, J.D. (1997) Statistics Corner: Questions and Answers about Language Testing Statistics (Reliability of surveys), Shiken, 1,2: 17-19.

Brown, J.D. (1998) Statistics Corner: Questions and Answers about Language Testing Statistics (Reliability and Cloze Test Length), Shiken, 2,1: 19-22.

Brown, J.D. (in press) Relative Importance of Persons, Items, Subtests and Language to TOEFL Test Variance. Language Testing, 16,2.

Brown, J.D. (in press). Statistics Corner. Questions and Answers about Language Testing Statistics: Standard Error of What? Shiken, 2,2.

Brown, J.D. and Bley-Vroman, R. (1995) Predicting Achievement in the ESL MA Program at UHM, University of Hawaii Working Papers in ESL, 14,1.

Brown, J.D., Hilgers, T. and Marsella, J. (1991) Essay Prompts and Topics: Minimizing the Effect of Differences. Written Communication, 8,4: 532-555.

Brown, J.D., Robson, G. and Rosenkjar P. (1996) Personality, Motivation, Anxiety, Strategies, and Language Proficiency of Japanese Students, University of Hawaii Working Papers in ESL, 15,1: 33-72.

Brown, J.D. and Ross, J.A. (1996) Decision Dependability of Item Types, Sections, Tests, and the Overall TOEFL Test Battery, in Milanovic, M. and Saville, N. (eds.) Performance Testing, Cognition and Assessment, Cambridge University Press, Cambridge, 231-265.

Buck, G. and Tatsuoka, L. (1998) Application of the Rule-space Procedure to Language Testing: Examining Attributes of a Free Response Listening Test, Language Testing, 15,2: 119-157.

Buck, G., Tatsuoka, K., Kostin, I. and Phelps, M. (1997) The Sub-Skills of Listening: Rule-Space Analysis of a Multiple-Choice Test of Second Language Listening Comprehension, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Chalhoub-Deville, M. (1993) Performance Assessment and the Components of the Oral Construct Across Different Tasks and Rater Groups, ERIC Document Reproduction Service No. 360 830.

Chalhoub-Deville, M. (1995) A Contextualized Approach to Describing Oral Language Proficiency, Language Learning, 45: 251-281.

Chalhoub-Deville, M. (1996) Performance Assessment and the Components of the Oral Construct across Different Tests and Rater Groups, in Milanovic, M. and Saville, N. (eds.), Performance Testing, Cognition and Assessment, Cambridge University Press, Cambridge.

Chalhoub-Deville, M. (1999) Investigating the properties of Assessment Instruments and the Setting of Proficiency Standards for Admission into University Second Language Courses, in Heilenman, K. (ed.) Research Issues in Language Program Direction, AAUSC.

Chapelle, C.A. (1998) Construct Definition and Validity Inquiry in SLA Research, in Bachman, L.F. and Cohen, A.D. (eds.), Interfaces between Second Language Acquisition and Language Testing Research, Cambridge University Press, Cambridge.
Choi, I.C. (1992) *An Application of Item Response Theory to Language Testing*, Theoretical Studies in Second Language Acquisition Vol. 2, Peter Lang, New York.

Clapham, C.M. (1996) What Makes an ESP Reading Test Appropriate for its Candidates? in Cumming, A. and Berwick, R. (eds.), *Validation in Language Testing*, Multilingual Matters, Clevedon, England.

Cohen, A. (1997) Towards Enhancing Verbal Reports as a Source of Insights on Test-Taking Strategies, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Cohen, A.D. (1998) Strategies and Processes in Test-Taking and SLA, in Bachman, L.F. and Cohen, A.D. (eds.), *Interfaces between Second Language Acquisition and Language Testing Research*, Cambridge University Press, Cambridge.

Cushing Weigle, S. (1998) Using FACETS to Model Rater Training Effects, *Language Testing*, 15,2: 263-287.

Cushing Weigle, S. and Lynch, B. (1995) Hypothesis Testing in Construct Validation, in Cumming, A. and Berwick, R. (eds.), *Validation in Language Testing*, Multilingual Matters, Clevedon, England.

Dandonoli, P. and Henning, G. (1990) An Investigation of the Construct Validity of the ACTFL Proficiency Guidelines and Oral Interview Procedure, *Foreign Language Annals*, 23,1: 11-22.

Davies, A. (1992) Is Language Proficiency Always Achievement? *Melbourne Papers in Language Testing*, 1,1.

Davies, A. (1992) Speculation and Empiricism in Applied Linguistics, *Melbourne Papers in Language Testing*, 1,2.

Davies, A. (1996) What Second Language Learners Can Tell us about the Native Speakers: Identifying and Describing Exceptions, *Melbourne Papers in Language Testing*, 5,2.

Davies, A. (1997) The Competing Claims of Accuracy and Fluency in the Constructions of Performance Tests of Language Proficiency: Two Cheers for Robert Lado! *Melbourne Papers in Language Testing*, 6,2: 1-20.

Davis, W.E. (1991) Comparing Language Proficiency Scores and Student Evaluations to Determine Policy for International Teaching Assistants, *College Student Journal*, 25,4: 489-495.

de Gruijter, D.N.M. and de Jong, J.H.A.L (1991) Item Test Regressions, Item Response Curves and the Relation between Test Forms, *Applied Psychological Measurement*, 15: 25-34.

de Jong, J.H.A.L. (1990) Linguistic Theory and Psychometric Models, in de Jong, J.H.A.L and Stevenson, D.K. (eds.), *Individualizing the Assessment of Language Abilities*, Multilingual Matters, Clevedon, England.

de Jong, J.H.A.L. (1991) *Defining a Variable of Foreign Language Ability: An Application of Item Response Theory*, Ph.D. thesis, Twente University, published by J.H.A.L. de Jong.

de Jong, J.H.A.L and Stevenson, D.K. (eds.) (1990) *Individualizing the Assessment of Language Abilities*, Multilingual Matters, Clevedon, England.

DeKeyser, R. (1990) Towards a Valid Measurement of Monitored Knowledge, *Language Testing*, 7,2: 147-157.

DeMauro, G. (1992) Examination of the Relationships among TSE, TWE and TOEFL scores, *Language Testing*, 9,2: 149-162.

Douglas, D. and Smith, J. (1997) *Theoretical Underpinnings of the Test of Spoken English Revision Project*, TOEFL Monograph Series, Number 9, Educational Testing Service, Princeton, NJ.

Duncan, K. (1994) Both sides now: a Qualitative and Quantitative Evaluation Compared, *Language Testing Update*, 16: 29-35.
Elder, C. (1993) Language Proficiency as a Predictor of Performance in Teacher Education, *Melbourne Papers in Language Testing*, 2,1.

Ellerton, A.W. (1997) *Considerations in the Validation of Semi-Direct Oral Testing*, unpublished Ph.D. thesis, University of Reading.

Fox, J., Pychyl, T. and Zumbo, B. (1997) An Investigation of Background Knowledge in the Assessment of Language Proficiency, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Frederiksen, N., Mislevy, R.J. and Bejar, I.I. (1993) *Test Theory for a New Generation of Tests*, Lawrence Erlbaum, Hillsdale, NJ.

Freedle, R. and Kostin, I. (1993) The Prediction of TOEFL Reading Item Difficulty: Implications for Construct Validity, *Language Testing*, 10,2: 133-170.

Fulcher, G. (1996) Invalidating Validity Claims for the ACTFL Oral Rating Scale, *System*, 24,2: 163-172.

Fulcher, G. (1997) An English Language Placement Test: Issues in Reliability and Validity, *Language Testing*, 14,2: 113-139.

Fulcher, G. (1999) Assessment in English for Academic Purposes: Putting content validity in its place, *Applied Linguistics*, 20,2: 221-236.

Ginther, A. and Grant, L. (1997) The Influences of Proficiency, Language Background, and Topic on the Production of Grammatical Form and Error on the Test of Written English, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Genesee, F. and Upshur, J.A. (1996) *Classroom-Based Evaluation in Second Language Education: Three Approaches to Validity*, Cambridge University Press, Cambridge.

Geranpayeh, A. (1994) Are Score Comparisons across Language Proficiency Test Batteries Justified? an IELTS-TOEFL Comparability Study, Edinburgh Working Papers in Applied Linguistics, 5: 50-65.

Glas, C.A.W. and ver Helst, N.D. (1991) *Using the Rasch Model for Dichotomous Data for Analysing Polytomous Responses*, Measurement and Research Department Report 91-3, CITO, Arnhem.

Glas, R. (1994) Criterion-Referenced Tests: Part I, Origins; Part II, Unfinished Business, *Educational Measurement: Issues and Practice*, 13: 9-11, 27-30.

Green, A. (1998) *Verbal Protocol Analysis in Language Testing Research*, Cambridge University Press, Cambridge.

Halvari, A. and Tarananen, M. (1997) Some Aspects on using Qualitative Procedures to Ensure Comparability Across Languages Within a Testing System, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Hambleton, R.K., Swaminathan, H. and Rogers, H.J. (1991) *Fundamentals of Item Response Theory*, Sage Publications, Newbury Park.

Hamp-Lyons, L. and Lynch, B.K. (1998) Perspectives on validity: A historical analysis of language testing conference abstracts, in Kunnan, A.J. (ed.), *Issues in Language Testing Research: Conventional Validity and Beyond*, Lawrence Erlbaum Inc., Mahwah, NJ, 253-276.

Henning, G. (1990) National Issues in Individual Assessment: the Consideration of Specialisation Bias in University Language Screening Tests, in de Jong, J.H.A.L. and Stevenson, D.K. (eds.), *Individualising the Assessment of Language Abilities*, Multilingual Matters, Clevedon, England.
Henning, G. (1992) Dimensionality and Construct Validity of Language Tests, *Language Testing*, 9,1: 1-11.

Henning, G. and Cascallar, E.C. (1992) *A Preliminary Study of the Nature of Communicative Competence*, TOEFL Research Reports No. 36, Educational Testing Service, Princeton, NJ.

Hudson, T.D. (1991) Relationships among IRT Item Discrimination and Item Fit Indices in Criterion-referenced Language Testing, *Language Testing*, 8,2: 160-181.

Hudson, T.D. (1993) Surrogate Indices for Item Information Functions in Criterion-Referenced Language Testing, *Language Testing*, 10,2: 171-192.

Iwashita, N. (1996) The Validity of the Paired Interview Format in Oral Performance Assessment, *Melbourne Papers in Language Testing*, 5,2.

Iwashita, N. and Elder, C. (1997) Expert Feedback? Assessing the Role of Test-Taker Reactions to a Proficiency Test for Teachers of Japanese, *Melbourne Papers in Language Testing*, 6,1.

Jensen, C., Hansen, C., Green, S.B. and Akey, T. (1997) An Investigation of Item Difficulty Incorporating the Structure of Listening Tests: A Hierarchical Linear Modeling Analysis, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Jones, N. (1994) *An Item Bank for Testing English Language Proficiency*, unpublished Ph.D. thesis, University of Edinburgh.

Jonz, J. (1990) Another Turn in the Conversation: What Does Cloze Measure? *TESOL Quarterly*, 24,1: 61-83.

Kenyon, D.M. and Stansfield, C.W. (1993) A Method for Improving Tasks on Performance-Based Assessments through Field-Testing, in Huhta, A., Sajavaara, K. and Takala, S. (eds.), *Language Testing: New Openings*, Institute for Educational Research, University of Jyväskylä, Finland.

Kirsch, I.S., Jenkins, L., Jungeblut, A. and Kolstad, A. (1993) *Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey*, National Center for Educational Statistics, Washington, DC.

Kunnan, A.J. (1992) An Investigation of a Criterion-Referenced Test using G-theory, and Factor and Cluster Analysis, *Language Testing*, 9,1: 30-50.

Kunnan, A.J. (1994) Modelling Relationships among some Test-Taker Characteristics and Performance on EFL Tests: an Approach to Construct Validation, *Language Testing*, 11,3: 225-252.

Kunnan, A.J. (1995) *Test Taker Characteristics and Test Performance: A Structural Modeling Approach*, *Studies in Language Testing* 2, Cambridge University Press, Cambridge.

Lewkowicz, J. (1995) Authenticity of Task vs. Quality of Written Output, in Tickoo, M. (ed.), *Reading and Writing: Theory into Practice*, SEAMEO, Singapore.

Linacre, J.M. (1995) Paired Comparisons with Ties: Bradley-Terry and Rasch, *Rasch Measurement Transactions*, 9, 425.

Linacre, J.M. and Wright, B.D. (1992) *A User's Guide to Facets: Rasch Measurement Computer Program*, MESA, Chicago.

Linn, R.L. (1994) Criterion-Referenced Measurement: A Valuable Perspective Clouded by Surplus Meaning, *Educational Measurement: Issues and Practice*, 13: 12-14.

Linn, R.L., Baker, E.L. and Dunbar, S.B. (1991) Complex Performance-Based Assessment: Expectations and Validation Criteria, *Educational Researcher*, 20,8: 15-21.

Lowenberg, P. (1993) Issues of Validity in Tests of English, *World Englishes*, 12,1: 95-106.
Lozier, V. and Chalhoub-Deville, M. (1996) Preliminary Item Response Theory Analysis of the University of Minnesota CLA Language Proficiency Tests in French, German and Spanish, CARLA working paper series #5, Center for Advanced Research on Language Acquisition, University of Minnesota, Minneapolis.

Lumley, T. and Brown, A. (1998) An Investigation of the Authenticity of Discourse in a Specific-Purpose Language Performance Test and its Relevance to Test Validity, in Li, E.S.L. and James, G. (eds.) Testing and Evaluation in Second Language Education, Hong Kong: Language Centre, Hong Kong University of Science and Technology, 22-33.

Lynch, B.K. and McNamara, T.F. (1998) Using G-theory and Many-facet Rasch Measurement in the Development of Performance Assessments of the ESL Speaking Skills of Immigrants, Language Testing, 15,2: 158-180.

Madden, C.G. and Myers, C.L. (eds.) (1994) Discourse and Performance of International Teaching Assistants, TESOL Publications, Alexandria, VA.

Marcoulides, G.A. (1993) Maximizing Power in Generalizability Studies under Budget Constraints, Journal of Educational Statistics, 18: 102-09.

Marcoulides, G.A. and Drezner, Z. (1996) A Method for Analyzing Performance Assessments, in Wilson, M. Draney, K. and Engelhard, G. (eds.), Objective Measurement: Theory into Practice, Ablex, Norwood, NJ.

Masters, G. (1990) Psychometric Aspects of Individual Assessment, in de Jong, J.H.A.L. and Stevenson, D.K. (eds.), Individualising the Assessment of Language Abilities, Multilingual Matters, Cledendon, England.

McKenna, M.C. and Layton, K. (1990) Concurrent Validity of Cloze as a Measure of Intersentential Comprehension, Journal of Educational Psychology, 82,2: 372-377.

McNamara, T.F. (1990) Item Response Theory and the Validation of an ESP Test for Health Professionals, Language Testing, 7,1: 52-76.

McNamara, T.F. (1995) Modelling Performance: Opening Pandora's Box, Applied Linguistics, 16,2: 159-179.

McNamara, T.F. (1997) Interaction in Second Language Performance Assessment: Whose Performance? Applied Linguistics.

McNamara, T. (1997) Problematising Content Validity: the Occupational English Test (OET) as a Measure of Medical Communication, Melbourne Papers in Language Testing, 6,1.

McNamara, T.F. and Adams, R.J. (1997) New Approaches to the Analysis of Task- and Rater-Related Dependencies in Performance Assessments, in Hultta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Messick, S. (1992) Validity of Test Interpretation and Use, in Alkin, M.C. (ed.), Encyclopaedia of Educational Research (sixth edition), Macmillan, NewYork.

Messick, S. (1994) The Interplay of Evidence and Consequences in the Validation of Performance Assessments, Educational Researcher, 23,2: 13-23.

Messick, S. (1996) Validity and Washback in Language Testing, Language Testing, 13,3: 241-256.

Milanovic, M., Savile, N. and Shen, S. (1996) A Study of the Decision-Making Behaviour of Composition Markers, in Milanovic, M. and Savile, N. (eds.), Performance Testing, Cognition and Assessment, Cambridge University Press, Cambridge.

Mislevy, R.J. (1995) Test Theory and Language-Learning Assessment, Language Testing, 12,3: 341-369.

Mislevy, R.J. and Bock, R.D. (1990) BILOG 3: Item Analysis and Test Scoring with Binary Logistic Models, Scientific Software, Mooresville, IN.
Moss, P. (1994) Can there be Validity without Reliability? *Educational Researcher*, 23,2: 5-12.

Moss, P. (1995) Themes and Variations in Validity Theory, *Educational Measurement: Issues and Practice*, 14,2: 5-13.

Morrow, K. (1991) Evaluating Communicative Tests, in Anivan, S. (ed.), *Current Developments in Language Testing*, RELC Anthology Series 23, RELC.

Muraki, E. and Bock, R.D. (1993) *PARSCALE: IRT Based Test Scoring and Item Analysis*, Scientific Software, Chicago.

O'Loughlin, K. (1997) *The Equivalence of Two Versions of an Oral Proficiency Test*, unpublished Ph.D. thesis, University of Melbourne.

Perkins, K. and Brutten, S. (1993) A Comparison of Indices for the Identification of Misfitting Items, in Douglas, D. and Chapelle, C.A. (eds.), *A New Decade of Language Testing Research*, TESOL Publications, Alexandria, VA.

Pollitt, A.B. (1997) Rasch Measurement in Latent Trait Models, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Purpura, J.E. (1996) *Modeling the Relationships between Test Takers' Reported Cognitive and Metacognitive Strategy Use and Performance on Language Tests*, unpublished Ph.D. thesis, University of California, Los Angeles.

Purpura, J. (in press) *Learner Strategies and Performance on Language Tests: A Structural Equation Modeling Approach*, Cambridge University Press, Cambridge.

Reed, D.J. and Halleck, G.B. (1997) Probing Above the Ceiling in Oral Interviews: What’s Up There?, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Reynolds, T., Perkins, K. and Brutten, S. (1994) A Comparative Item Analysis of a Language Testing Instrument, *Language Testing*, 11,1: 1-14.

Ryan, K.E. and Bachman, L.F. (1992) Differential Item Functioning on two tests of EFL Proficiency, *Language Testing*, 9,1: 12-29.

Sasaki, M. (1991) A Comparison of Two Methods for Detecting Differential Item Functioning in an ESL Placement Test, *Language Testing*, 8,2: 95-111.

Shohamy, E. (1994) The Validity of Direct Versus Semi-Direct Oral Tests, *Language Testing*, 11,2: 99-123.

Swanson, L. and Stocking, M.L. (1993) A Model and Heuristic for Solving Very Large Item Selection Problems, *Applied Psychological Measurement*, 17: 151-166.

Schiis, E.D.J., van der Poel, M.G.M. and Weltens, B. (1991) The Reliability Ritual, *Language Testing*, 8,2: 125-138.

Scott, M.L., Stansfield, C.W. and Kenyon, D. (1996) Examining Validity in a Performance Test: The Listening Summary Translation Exam (LSTE) - Spanish Version, *Language Testing*, 13,1: 83-109.

Shameem, N. (1998) Validating Self-reported Language Proficiency by Testing Performance in an Immigrant Community: the Wellington Indo-Fijians, *Language Testing*, 15,1: 86-108.

Shavelson, R.J. and Webb, N.M. (1991) *Generalizability Theory: A Primer*, Sage Publications, Newbury Park, CA.

Shen, Z., Weir, C.J. and Green, R. (1997) *The Test for English Majors Validation Project*, Shanghai Foreign Language Education Press.
Shepard, L. (1993) Evaluating Test Validity, *Review of Research in Education*, 19: 405-450.

Storey, P. (1995) Investigating Construct Validity through Test Taker Introspection, unpublished Ph.D. thesis, University of Reading.

Tan, S.H. (1990) The Role of Prior Knowledge and Language Proficiency as Predictors of Reading Comprehension among Undergraduates, in de Jong, J.H.A.L. and Stevenson, D.K. (eds.), *Individualizing the Assessment of Language Abilities*, Multilingual Matters, Cleveldon, England.

Tang, K.L. (1995) *Multidimensional IRT Models and Their Potential Application to TOEFL 2000*, Report Prepared for TOEFL 2000 Project, Educational Testing Service, Princeton, NJ.

Tang, K.L. and Eignor, D.R. (1996) Concurrent Calibration of Dichotomously and Polytomously Scored TOEFL Items Using IRT Models, Educational Testing Service, Princeton, NJ.

Tarone, E. (1998) Research on Interlanguage Variation: Implications for Language Testing, in Bachman, L.F. and Cohen, A.D. (eds.) *Interfaces between SLA and Language Testing Research*, Cambridge University Press, Cambridge.

Timminga, E. and Adema, J.J. (1994) Test Construction from Item Banks, in Fischer, G.H. and Molenaar, I.W. (eds.), *Rasch Models: Foundations, Recent Developments, and Applications*, Springer-Verlag, New York.

Turner, C. (1992) Rasch model programs and scalar analysis, *Language Testing Update*, 12: 62-64.

van der Linden, W.J. (1994) Fundamental Measurement and the Fundamentals of Rasch Measurement, in Wilson, M. (ed.), *Objective Measurement: Theory into Practice: Volume 2*, Ablex, Norwood, NJ.

van der Linden, W.J. and Hambleton, R.K. (1996) *Handbook of Modern Item Response Theory*, Springer-Verlag, New York.

van Hest, E. (1997) Self-Repair as a Measure of L2 Proficiency, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

Verhelst, N.D., Glas, C.A.W. and Verstralen, H.H.F.M. (1994) *OPLM: Computer Program and Manual*, CITO, Arnhem.

Verhoeven, L. and de Jong, J. H. A. L. (eds.) (1992) *The Construct of Language Proficiency: Applications of Psychological Models to Language Assessment*, John Benjamins Publishing Company, Amsterdam.

Wainer, H. and Lewis, C. (1990) Toward a Psychometrics for Testlets, *Journal of Educational Measurement*, 27: 1-14.

Wall, D., Clapham, C.M. and Alderson, J.C. (1991) Validating Tests in Difficult Circumstances, in Alderson, J.C. and North, B. (eds.), *Language Testing in the 1990s: The Communicative Legacy*, Macmillan, London.

Wall, D., Clapham, C.M., and Alderson, J.C. (1994) Evaluating a Placement Test, *Language Testing*, 11,3: 321-344.

Ward, S. (1997) Inter-Rater Reliability in an ESP Context, *Melbourne Papers in Language Testing*, 6,2: 73-91.

Weir, C.J. and Yan, J. (in press) An Empirical Investigation of the Componentiality of L2 Reading in English For Academic Purposes, Cambridge University Press, Cambridge.

Wennerstrom, A.K. and Heiser, P. (1992) ESL Student Bias in Instructional Evaluation, *TESOL Quarterly*, 26,2: 271-288.

Wigglesworth, G. and O’Loughlin, K. (1993) An Investigation into the Comparability of Direct and Semi-Direct Versions of an Oral Interaction Test in English, *Melbourne Papers in Language Testing*, 2,1.
Wilson, K.M. (1996) Validity of Global Self-Ratings of ESL Speaking Proficiency Based on an FSI/ILR-Referenced Scale: an Empirical Assessment. In collaboration with R. Lindsay. Educational Testing Service, Princeton, NJ.

Wright, B.D. and Linacre, J.M. (1991) *BIGSTEPS Rasch Analysis computer program*, MESA Press, Chicago.

Yang, H. and Weir, C.J. (1998) *The Validity of the College English Test*, Shanghai Foreign Language Education Press.

Yen, W.M. (1993) Scaling Performance Assessments: Strategies for Managing Local Item Dependence, *Journal of Educational Measurement*, 30: 187-213.
14. Washback

Akoha, J. (1991) Curriculum Innovation and Examination Reform in Benin, West Africa, in Alderson, J.C. and North, B. (eds.), Language Testing in the 1990s: the Communicative Legacy, Macmillan, London.

Alderson, J.C. and Hamp-Lyons, L. (1996) TOEFL Preparation Courses: A Study of Washback, Language Testing 13,3: 280-297.

Alderson, J.C. and Wall, D. (1993) Does Washback Exist? Applied Linguistics 14,2: 115-129.

Allen, K. (1994) English Education in Zimbabwe: Testing Communicative Competence, in Hill, C. and Parry, K. (eds.), From Testing to Assessment: English as an International Language, Longman, London.

Andrews, S. and Fullilove, J. (1994) Assessing Spoken English in Public Examinations - Why and How? in Boyle, J. and Falvey, P. (eds.), English Language Testing in Hong Kong, The Chinese University Press, Hong Kong.

Bailey, K.M. (1996) Working for Washback: a Review of the Washback Concept in Language Testing, Language Testing 13,3: 257-279.

Boyle, J. and Suen, D.L.K. (1994) Communicative Considerations in a Large-Scale Listening Test, in Boyle, J. and Falvey, P. (eds.), English Language Testing in Hong Kong, The Chinese University Press, Hong Kong.

Brown, J.D. (1993) Language Test Hysteria in Japan. The Language Teacher, 17,12: 41-43.

Brown, J.D. (1995) A Gaijin Teacher's Guide to the Vocabulary of Entrance Examinations, The Language Teacher, 19,12: 25.

Brown, J.D. (1997) Do Tests Washback on the Language Classroom? TESOLANZ Journal, 5: 63-80. [An earlier version appeared in 1997 under the title "The washback effect of language tests" in University of Hawaii Working Papers in ESL, 16,1: 27-45.]

Brown, J.D. (1997) The Washback Effect of Language Tests, University of Hawaii Working Papers in ESL, 16,1: 27-45.

Brown, J.D. (1997) Testing Washback in Language Education. PASAA Journal, 27: 64-79.

Brown, J.D. (1998) Language Testing: Purposes, Effects, Options, and Constraints, TESOLANZ Journal, 6: 13-30.

Brown, J.D. (1998) University Entrance Examinations and their Effect on English Language Teaching in Japan. in Kahny, J. and James, M. (eds.) Perspectives on Secondary School EFL: A Publication in Commemoration of the 30th Anniversary of the Language Institute of Japan, Language Institute of Japan, Odawara, Japan, 20-27.

Brown, J.D. (in press) Using Tests to Unify Language Curriculum. Proceedings of the Fourth International ELT Conference, Bilkent University, Ankara, Turkey.

Bunton, D. (1995) Teaching for the ‘Practical Skills’ Exam, Hong Kong Examinations Authority, Hong Kong.

Chalhoub-Deville, M. (1996) Test Interpretation, Test Use, and Pedagogical Implications, in Wigglesworth, G. and Elder, C. (eds.), The Language Testing Cycle: from Inception to Washback, Special Issue of Applied Linguistics Association of Australia, Series no. 13.

Cheng, L. (1997) How Does Washback Influence Teaching? Implications for Hong Kong, Language and Education 235 11,1: 38-54.

Cheng, L. (1998) The Washback Effect of Public Examination Change on Classroom Teaching, unpublished Ph.D. thesis, University of Hong Kong.

DeVincenzi, F. (1995) Examining Standardised Test Content: some Advice for Teachers, in Peirce, B.N. (ed.), Language Tests and ESL Teaching, TESOL Quarterly, 29,1: 180-184.
Eckstein, M.A. and Noah, H.J. (1993) *Secondary School Examinations: International Perspectives on Policies and Practice*, Yale University Press, New Haven and London.

Eisemon, T.O. (1990) Examinations Policies to Strengthen Primary Schooling in African Countries, *International Journal of Educational Development*, 10,1: 69-82.

Haladyna, T.M., Nolan, S.B. and Haas, N.S. (1991) Raising Standardized Achievement Test Scores and the Origins of Test Score Pollution, *Educational Researcher*, 20,5: 2-7.

Hamp-Lyons, L. (1997) Washback, Impact and Validity: Ethical Concerns, *Language Testing*, 14,3: 295-303.

Kellaghan, T. and Greaney, V. (1992) *Using Examinations to Improve Education: A Study of Fourteen African Countries*, World Bank, Washington, DC.

Li, X. (1990) How Powerful Can a Language Test Be? The MET in China, *Journal of Multilingual and Multicultural Development*, 11,5: 393-404.

Messick, S. (1996) Validity and Washback in Language Testing, *Language Testing*, 13,3: 241-256.

Peirce, B.N. (1992) Demystifying the TOEFL Reading Test, *TESOL Quarterly*, 26,4: 665-689.

Popham, W.J. (1994) The Instructional Consequences of Criterion-Referenced Clarity, *Educational Measurement: Issues and Practice*, 13, 15-18, 30.

Raimies, A. (1990) The TOEFL Test of Written English: Causes for Concern, *TESOL Quarterly*, 24,3: 427-442.

Shohamy, E. (1992) Beyond Proficiency Testing: A Diagnostic Feedback Testing Model for Assessing Foreign Language Learning, *The Modern Language Journal*, 76,4: 513-521.

Shohamy, E., Donitsa-Schmidt, S. and Ferman, I. (1996) Test Impact Revisited: Washback Effect over Time, *Language Testing*, 13,3: 298-317.

Smith, M.L. (1991) Put to the Test: The Effects of External Testing on Teachers, *Educational Researcher*, 20,5: 8-11.

Wall, D. (1994) Measuring examination ‘washback’: The Sri Lankan Evaluation Project’, in Flavell, R.H. (ed.) *ELT Policy and its Impact: A Case Study*, Modern English Publications in association with The British Council, London.

Wall, D. (1996) Introducing New Tests into Traditional Systems: Insights from General Education and from Innovation Theory, *Language Testing*, 13,3: 334-354.

Wall, D. (1997) Impact and Washback in Language Testing, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Wall, D. (1998) The Impact of High-Stakes Testing on Teaching and Learning, in Coombe, C. (ed.) *Current Trends in English Language Testing*, Proceedings of the CTELT Conference 1997 and 1998, 98-115.

Wall, D. and Alderson, J.C. (1993) Examining Washback: The Sri Lankan Impact Study, *Language Testing*, 10,1: 41-69.

Wall, D., Kalnberzina, V., Mazuoliene, Z. and Truus, K. (1996) The Baltic States Year 12 Examination Project, *Language Testing Update*, 19: 15-27.

Watanabe, Y. (1996) Does Grammar-Translation Come from the Entrance Examination? Preliminary Findings from Classroom-based Research, *Language Testing*, 13,3: 319-333.

Watanabe, Y. (1997) The Washback Effects of the Japanese University Entrance Examinations of English - Classroom-Based Research, unpublished Ph.D. thesis, University of Lancaster.
Xiaoju, L. (1990) How Powerful can a Language Test be? The MET in China, *Journal of Multilingual and Multicultural Development*, 11,5: 393-404.
15. Ethics and Standards

Alderson, J.C. and Beretta, A. (eds.) (1992) Evaluating Second Language Education, Cambridge University Press, Cambridge.

Alderson, J.C. and Buck, G. (1993) Standards in Testing: A Study of the Practice of UK Examination Boards in EFL/ESL Testing, Language Testing, 10,1: 1-26.

Association of Language Testers in Europe (ALTE) (1995) The ALTE Code of Practice, ALTE, Cambridge.

Allen, K. (1994) English Education in Zimbabwe: Testing Communicative Competence, in Hill, C. and Parry, K. (eds.), From Testing to Assessment: English as an International Language, Longman, London.

Bailey, K. and Brown, J.D. (1995) Language testing courses: What are they? in Cumming, A. and Berwick, R. (eds.), Validation in Language Testing, Multilingual Matters, Clevedon, England.

Berry, V. (1997) Ethical Considerations when Assessing Oral Proficiency in Pairs, in Huita, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Brindley, G. (1991) Defining Language Ability: The Criteria for Criteria, in Anivan, S. (ed.) Current Developments in Language Testing, Anthology Series 25, SEAMEO/RELC, Singapore, 139-164.

Broadfoot, P. (1992) A Question of Quality: The Changing Role of Assessment in Education, ACSA Workshop Report No. 4, Australian Curriculum Studies Association, Canberra.

Brown, J.D. (1994) Problems in Language Program Evaluation, University of Hawaii Working Papers in ESL, 13,1: 1-22.

Brown, J.D. (1995). Language Program Evaluation: Problems and Solutions, in Grabe, W. (ed.) Annual Review of Applied Linguistics, 15: 227-248) [An earlier version appeared in 1994 under the title "Problems in language program evaluation" in University of Hawaii Working Papers in ESL, 13,1: 1-22].

Brown, J.D. (1997) Designing Surveys for Language Programs, in Nunan, D. and Griffie, D. (eds.) Classroom Teachers and Classroom Research, Japan Association for Language Teaching, Tokyo, 55-70.

Brown, J.D. (in press) The Roles and Responsibilities of Assessment in Foreign Language Education, JALT Journal, 2.

Brown, J.D. and Pennington, M.C. (1991) Developing Effective Evaluation Systems for Language Programs, in Pennington, M.C. (ed.) Building Better English Language Programs: Perspectives on Evaluation in ESL, NAFSA, Washington, DC, 3-18.

Buros Institute (1992) Mental Measurement Yearbook, Buros Institute of Mental Measurements, Lincoln, Nebraska.

Challhoub-Deville, M. (1999) Investigating the properties of Assessment Instruments and the Setting of Proficiency Standards for Admission into University Second Language Courses, in Heilenman, K. (ed.) Research Issues in Language Program Direction, AAUSC.

Cumming, A. (1994) Does Language Assessment Facilitate Recent Immigrants' Participation in Canadian Society? TESL Canada Journal, 2,2: 117-133.

Darling-Hammond, L. (1994) Performance-Based Assessment and Educational Equity, Harvard Educational Review, 64,1: 5-30.

Davidson, F. (1994) Task Force on Test Standards (TFTS), Language Testing Update, 16: 8-13.

Davidson, F., Alderson, J.C., Douglas, D., Huhta, A., Turner, C. E. and Wylie, E. (1995) Report of the Task Force on Testing Standards to the International Language Testing Association, ILTA, Melbourne.
Davidson, F., Alderson, J.C., Douglas, D., Huhta, A., Turner, C. E. and Wylie, E. (1995) An International Survey of Language Assessment Standards, *TESOL Journal*, 5,1: 6-7.

Davidson, F. and Bachman, L. (1990) The Cambridge-TOEFL Comparability Study: An Example of the Cross-National Comparison of Language Tests, in de Jong, J.H.A.L. (ed.), *Standardization in Language Testing*, AILA Review, 7: 24-45.

Davidson, F., Turner, C.E. and Huhta, A. (1997) Language Testing Standards, in Clapham, C.M. and Corson, D. (eds.), *Language Testing and Assessment*, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Davies, A. (1997) The Education (and Training) of Language Testers, *Melbourne Papers in Language Testing*, 6,1.

Davies, A. (1997) Demands of being Professional in Language Testing, *Language Testing*, 14,3: 328-339.

Davies, A. (1997) The Limits of Ethics in Language Testing, *Language Testing*, 14,3: 235-241.

Davies, A. (1997) Australian Immigrant Gatekeeping Through English Language Tests: How Important is Proficiency? in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), *Current Developments and Alternatives in Language Assessment*, Universities of Tampere and Jyväskylä, Tampere.

de Jong, J.H.A.L. (ed.) (1990) *Standardization in Language Testing*, AILA Review, No.7.

de Jong, J.H.A.L. (1991) Standards for Language Testing, *Language Testing Update*, 9: 7-11.

de Jong, J.H.A.L. (1995) Foreign Language Exams and the Need for Standards in the Dutch Educational System, *Tijdschrift voor Onderwijswetenschappen*, 24, 5/6: 196-202.

de Jong, J.H.A.L. (1995) The Need for Standards in Language Education, *System*, 23,4: 441-444.

de Jong, J.H.A.L and Oscaron, M. (1990) Cross-National Standards: a Dutch-Swedish Collaborative Effort in National Standardized Testing, *AILA Review*, 7: 62-78.

Douglas, D. (1993) An International Association of Language Testing Specialists, the Formalization of the Language Testing Research Colloquium, and the Establishment of Standards for Language Testing: A Summary of the Debate so far, Appendix 3 to Davidson, F. (ed.) Testing English Across Cultures: Summary and Comments, *World Englishes*, 12: 113-125.

Earl, L.M. (1995) Assessment and Accountability in Education in Ontario, *Canadian Journal of Education*, 20,1: 45-55.

Elder, C. (1997) What does Test Bias have to do with Fairness? *Language Testing*, 14,3: 261-277.

Elson, N. (1992) The Failure of Tests: Language Tests and Post-Secondary Admissions of ESL Students, in Burnaby, B. and Cumming, A. (eds.), *Socio-Political Aspects of ESL in Canada*, OISE Press, Toronto.

Fulcher, G. and Bamford, R. (1997) I Didn't Get the Grade I Need. Where's My Solicitor? *System*, 24,4: 437-448

Gearhart, M. and Herman, J. (1995) *Portfolio Assessment: Whose Work Is It? Issues in the Use of Classroom Assignments for Accountability*, UCLA Center for the Study of Evaluation, Los Angeles.

Gipps, C. (1995) Beyond Testing: Towards a Theory of Educational Assessment, The Falmer Press, London.

Guiton, G. and Oakes, J. (1995) Opportunity to Learn and Conceptions of Educational Equality, *Educational Evaluation and Policy Analysis*, 17,3: 323-336.

Hamp-Lyons, L. (1996) Applying Ethical Standards to Portfolio Assessment of Writing in English as a Second Language, in Milanovic, M. and Saville, N. (eds.), *Performance Testing, Cognition and Assessment*, Cambridge University Press, Cambridge.
Hamp-Lyons, L. (1997) Ethics in Language Testing, in Clapham, C.M. and Corson, D. (eds.), Language Testing and Assessment, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.

Hamp-Lyons, L. (1997) Washback, Impact and Validity: Ethical Concerns, Language Testing, 14,3: 295-303.

Hawthorne, L. (1997) The Political Dimension of English Language Testing in Australia, Language Testing, 14,3: 248-260.

Hill, K., Davies, A., Oldfield, J. and Watson, N. (1997) Questioning an Early Start: the Transition from Primary to Secondary Foreign Language Learning, Melbourne Papers in Language Testing, 6,2: 21-36.

Holland, D. and Street, B. (1994) Assessing Adult Literacy in the United Kingdom: The Progress Profile, in Hill, C. and Parry, K. (eds.), From Testing to Assessment: English as an International Language, Longman, London.

International Association for the Evaluation of Educational Achievement (IEA) (1994) Standards for the Design and Operations in IEA Studies, IEA, The Hague.

Joint Advisory Committee (1993) Principles for Fair Student Assessment Practices for Education in Canada, Joint Advisory Committee, Edmonton.

Kachru, B. (1993) Ethical Issues for Applying Linguistics Afterword, Issues in Applied Linguistics, 4,2: 283-294.

Katz, A. and Short, D. (1996) ESL Standards and the TESOL Assessment Guidelines Project, TESOL Matters, 6, 2, 1 and 14.

Kunnan, A.J. (1997) Connecting Fairness With Validation in Language Assessment, in Huhta, A., Kohonen, V., Kurki-Suonio, K. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Lacelle-Peterson, M.W. and Rivera, C. (1994) Is it Real for All Kids? A Framework for Equitable Assessment Policies for English Language Learners, Harvard Educational Review, 64,1: 55-75.

Larter, S. and Donnelly, J. (1993) Demystifying the Goals of Education: Toronto's Benchmark Program, Orbit, 24,2: 22 - 28.

Lynch, B.K. (1996) Language Program Evaluation: Theory and Practice, Cambridge University Press, Cambridge.

Lynch, B.K. (1997) In Search of the Ethical Test, Language Testing, 14,3: 315-527.

Madaus, G. (1995) A Technological and Historical Consideration of Equity Issues Associated with Proposals to Change our Nation's Testing Policy, in Nettles, M.T. and Nettles, A.L. (eds.), Equity and Excellence in Educational Testing and Assessment, Kluwer, Norwell, MA.

Matthews, M. (1990) The Measurement of Productive Skills: Doubts Concerning the Assessment Criteria of Certain Public Examinations, ELT Journal, 44,2: 117 - 121.

Moore, H. (1996) Telling What is Real: Competing Views in Assessing ESL Development, Linguistics and Education, 8,2: 189-228.

McNamara, T.F. (1998) Policy and Social Consideration in Language Assessment, Annual Review of Applied Linguistics, 18: 304-319.

Milanovic, M., Saville, N. and Taylor, L.B. (1997) The Alte User’s Guide for Examiners, in Huhta, A., Kohonen, V., Kurki-Suonio, L. and Luoma, S. (eds.), Current Developments and Alternatives in Language Assessment, Universities of Tampere and Jyväskylä, Tampere.

Norton, B. (1997) Accountability in Language Assessment, in Clapham, C.M. and Corson, D. (eds.), Language Testing and Assessment, Encyclopaedia of Language and Education, Vol. 7, Kluwer, Dordrecht.
Norton, B. and Starfield, S. (1997) Covert Language Assessment in Academic Writing, *Language Testing*, 14,3: 278-294.

Peirce, B.N. and Stein, P. (1995) Why the Monkeys Passage Bombed: Tests, Genres and Teaching, *Harvard Educational Review*, 65,1: 50-65.

Pennington, M. (ed.) (1991) Building Better English Language Programs: Perspectives on Evaluation in ESL, NAFSA, Washington, D.C.

Pennington, M.C. and Brown, J.D. (1991) Excellence in Language Education: A Function of Curriculum Process, in Pennington, M. (ed.) (1991) Building Better English Language Programs: Perspectives on Evaluation in ESL, NAFSA, Washington, D.C., 57-74 [An earlier version appeared in 1987 under the title "Excellence in language programs: A function of curriculum process" in *University of Hawaii Working Papers in ESL*, 6,1: 81-105].

Rea-Dickins, P. (1997) So, why do we need Relationships with Stakeholders in Language Testing? A View from the UK, *Language Testing*, 14,3: 301-314.

Reardon, S., Scott, K. and Verre, J. (1994) Symposium: Equity in Educational Assessment, *Harvard Educational Review*, 64,1: 1-4.

Schmeiser, C.B. (1992) Ethical Codes in the Professions, *Educational Measurement: Issues and Practice*, 11,3: 5-11.

Schools Examination and Assessment Commission (1993) Mandatory Code of Practice for the GCSE, SEAC, London.

Shohamy, E. (1990) Language Testing Priorities: a Different Perspective, *Foreign Language Annals*, 23: 385-394.

Shohamy, E. (1993) The Exercise of Power and Control in the Rhetorics of Testing, in Huhta, A., Sajavaara, K. and Takala, S. (eds.), *Language Testing: New Openings*, Institute for Educational Research, University of Jyväskylä, Finland.

Shohamy, E. (1997) Testing Methods, Testing Consequences: are they Ethical? are they Fair? *Language Testing*, 14,3: 340-349.

Spolsky, B. (1990) Social Aspects of Individual Assessment, in de Jong, J.H.A.L. and Stevenson, D.K. (eds.), *Individualizing the Assessment of Language Abilities*, Multilingual Matters, Clevedon, England.

Spolsky, B. (1997) The ethics of gatekeeping tests: What have we learned in a hundred years? *Language Testing*, 14,3: 242-247.

Stansfield, C.W. (1993) Ethics, standards, and professionalism in language testing, *Issues in Applied Linguistics*, 4,2: 189-206.

Task Force on Testing Standards (TFTS) (1995) Report of the Task Force on Testing Standards (TFTS) to the International Language Testing Association (ILTA), ILTA, Melbourne.