

VOLUME

II

THE COMPANION TO LANGUAGE ASSESSMENT

APPROACHES AND
DEVELOPMENT

Edited by
Antony John Kunnan



WILEY Blackwell

The Companion to Language Assessment

Edited by
Antony John Kunnan

Volume II
Approaches and Development



WILEY Blackwell

This edition first published 2014
© 2014 John Wiley & Sons, Inc

Registered Office

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Offices

350 Main Street, Malden, MA 02148-5020, USA

9600 Garsington Road, Oxford, OX4 2DQ, UK

The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

The right of Antony John Kunnan to be identified as the author of the editorial material in this work has been asserted in accordance with the UK Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

The companion to language assessment / Edited by Antony John Kunnan.

pages cm

Includes bibliographical references and index.

ISBN 978-0-470-65533-7 (hardback : alk. paper) 1. Language and languages—Ability testing. 2. Language and languages—Examinations. I. Kunnan, Antony John.

P53.4.C58 2014

418.0076—dc23

2013017494

A catalogue record for this book is available from the British Library.

Cover image: Vol. 1 Abstract Painting 4, © Jumpingsack / Dreamstime.com; Vol. 2 Hermann Ayerbe, Inner Light 2004, © Hermann Ayerbe / Getty Images; Vol. 3 © Joe Cicak / istockphoto; Vol. 4 Yellow Land © Opas Chotiphantawanon / Shutterstock.
Cover design by Nicki Averill Design

Set in 10/12.5 pt Palatino by Toppan Best-set Premedia Limited
Printed and bound in Singapore by Markono Print Media Pte Ltd

About the Editor

Antony John Kunnan is Professor of English Language and Literature at the National Institute of Education, Nanyang Technological University, Singapore. He has authored and edited books and edited special issues of journals on test validation, test fairness, statistics for language assessment, differential item functioning, and structural equation modeling. His most recent book is *Workbook for Statistics in Language Assessment* (coauthored with Lyle Bachman, 2005). He was the founding editor of *Language Assessment Quarterly*, which he edited from 2003 to 2013, and is past president of the International Language Testing Association.

Introduction to Volume II

This volume presents chapters on overall approaches, assessment and learning, assessment development, and the use of technology. Specifically, the volume opens with chapters on large-scale assessment, norm- and criterion-referenced assessment, task-based assessment, and computer-assisted assessment. These chapters are critical as the ways in which assessments are designed, developed, scored, and interpreted depend largely on the overall approach to assessment. Chapters on different types of assessments are presented next. These include performance, portfolio, dynamic, and self- and peer assessment, monitoring progress and tracking achievement and growth, providing diagnostic feedback, and training test developers in assessment literacy. Chapters on the details of the development process follow. These include defining constructs and assessment design, writing assessment specifications, writing or selecting items and tasks, texts and response formats, and using test-taking strategies. Field testing, standards and guidelines, statistics and software, standard setting, administration, and detecting cheating complete the development process. The volume concludes with forward-looking chapters on the use of technology in language assessment, specifically the use of new media and corpora, eye-tracking, acoustic analysis, and computer-automated scoring of writing.

Brief Contents

Volume I

<i>About the Editor</i>	<i>xvii</i>
<i>Contributors</i>	<i>xviii</i>
<i>Introducing The Companion to Language Assessment</i>	<i>xli</i>
<i>Acknowledgments</i>	<i>xlvi</i>
<i>Introduction to Volume I</i>	<i>xlvi</i>
Abilities, Contexts, and Learners	1–530

Volume II

<i>About the Editor</i>	<i>xvii</i>
<i>Introduction to Volume II</i>	<i>xviii</i>
Approaches and Development	531–1078

Volume III

<i>About the Editor</i>	<i>xvii</i>
<i>Introduction to Volume III</i>	<i>xviii</i>
Evaluation, Methodology, and Interdisciplinary Themes	1079–604

Volume IV

<i>About the Editor</i>	<i>xvii</i>
<i>Introduction to Volume IV</i>	<i>xviii</i>
Assessment Around the World	1605–2086

<i>Index</i>	<i>2087</i>
--------------	-------------

Full Table of Contents

Volume I: Abilities, Contexts, and Learners

<i>About the Editor</i>	xvii
<i>Contributors</i>	xviii
<i>Introducing The Companion to Language Assessment</i>	xli
<i>Acknowledgments</i>	xlvi
<i>Introduction to Volume I</i>	xlvi

Part 1: Overview	1
-------------------------	----------

1 Fifty Years of Language Assessment, <i>Alan Davies</i>	3
--	---

Part 2: Assessing Abilities	23
------------------------------------	-----------

2 Assessing Aptitude, <i>Catherine J. Dougherty</i>	25
3 Assessing Listening, <i>Elvis Wagner</i>	47
4 Assessing Literacy, <i>Sara Cushing Weigle</i>	64
5 Assessing Responses to Literature, <i>Richard Beach</i>	83
6 Assessing Grammar, <i>James E. Purpura</i>	100
7 Assessing Pragmatics, <i>Carsten Roever</i>	125
8 Assessing Pronunciation, <i>Talia Isaacs</i>	140
9 Assessing Speaking, <i>Barry O'Sullivan</i>	156
10 Assessing Vocabulary, <i>David Beglar and Paul Nation</i>	172
11 Assessing Reading, <i>William Grabe and Xiangying Jiang</i>	185
12 Assessing Writing, <i>Deborah Crusan</i>	201
13 Assessing Integrated Skills, <i>Alister Cumming</i>	216
14 Assessing Language and Content, <i>Marguerite Ann Snow and Anne M. Katz</i>	230
15 Assessing Translation, <i>Juliane House</i>	248
16 Assessing Language Varieties, <i>Mingwei Pan and David D. Qian</i>	265

Part 3: Assessment Contexts	283
17 International Assessments, <i>Sauli Takala, Gudrun Erickson, and Neus Figueras</i>	285
18 English Language Proficiency Assessments as an Exit Criterion for English Learners, <i>Mikyung Kim Wolf and Timothy Farnsworth</i>	303
19 Tests of English for Academic Purposes in University Admissions, <i>Xiaoming Xi, Brent Bridgeman, and Cathy Wendler</i>	318
20 Government and Military Assessment, <i>Rachel L. Brooks and Mika Hoffman</i>	338
21 Language Assessment for Court Translators and Interpreters, <i>Piers Armstrong</i>	355
22 Language Testing for Immigration to Europe, <i>Piet Van Avermaet and Reinhilde Pulinx</i>	376
23 Language Testing for Immigration and Citizenship in the Netherlands, <i>Massimiliano Spotti</i>	390
24 Assessment in Asylum-Related Language Analysis, <i>Diana Eades</i>	404
Part 4: Assessing Learners	421
25 Developmental Considerations and Curricular Contexts in the Assessment of Young Language Learners, <i>Alison L. Bailey, Margaret Heritage, and Frances A. Butler</i>	423
26 Assessing Heritage Language Learners, <i>Lorena Llosa</i>	440
27 Assessing Teachers' Language Proficiency, <i>Cathie Elder and Sun Hee Ok Kim</i>	454
28 Assessing the Oral English Abilities of International Teaching Assistants in the USA, <i>Timothy Farnsworth</i>	471
29 Assessing the English Language Proficiency of International Aviation Staff, <i>Ruixia Yan</i>	484
30 Assessing Health Professionals, <i>Lynda Taylor and John Pill</i>	497
31 Assessing Test Takers With Communication Disorders, <i>John W. Oller, Jr.</i>	513

Volume II: Approaches and Development

<i>About the Editor</i>	<i>xvii</i>
<i>Introduction to Volume II</i>	<i>xviii</i>

Part 5: Assessment Approaches	531
32 Large-Scale Assessment, <i>Janna Fox</i>	533
33 Norm-Referenced Approach to Language Assessment, <i>Jungok Bae</i>	542

34	Criterion-Referenced Approach to Language Assessment, <i>Thom Hudson</i>	561
35	Task-Based Language Assessment, <i>Koen Van Gorp and Bart Deygers</i>	578
36	Computer-Assisted Language Testing, <i>Ruslan Suvorov and Volker Hegelheimer</i>	594
Part 6: Assessment and Learning		615
37	Performance Assessment in the Classroom, <i>Guoxing Yu</i>	617
38	Monitoring Progress in the Classroom, <i>Rama Mathew and Matthew E. Poehner</i>	631
39	Achievement and Growth in the Classroom, <i>Michael J. Kieffer</i>	646
40	Portfolio Assessment in the Classroom, <i>Muchun Yin</i>	659
41	Dynamic Assessment in the Classroom, <i>Matthew E. Poehner</i>	677
42	Diagnostic Feedback in the Classroom, <i>Eunice Eunhee Jang and Maryam Wagner</i>	693
43	Self-Assessment in the Classroom, <i>Mats Oscarson</i>	712
44	Peer Assessment in the Classroom, <i>Jette G. Hansen Edwards</i>	730
45	Test Development Literacy, <i>Kirby C. Grabowski and Jee Wha Dakin</i>	751
Part 7: Assessment Development		769
46	Defining Constructs and Assessment Design, <i>Joan Jamieson</i>	771
47	Effect-Driven Test Specifications, <i>Jiyoung Kim and Fred Davidson</i>	788
48	Writing Items and Tasks, <i>Mary Schedl and Jeanne Malloy</i>	796
49	Item Banking, <i>Susan Nissan and Robert French</i>	814
50	Adapting or Developing Source Material for Listening and Reading Tests, <i>Anthony Green</i>	830
51	Writing Scoring Criteria and Score Reports, <i>Megan Montee and Margaret E. Malone</i>	847
52	Response Formats, <i>David D. Qian and Mingwei Pan</i>	860
53	Field Testing of Test Items and Tasks, <i>Daniel J. Reed</i>	876
54	Using Test-Wiseness Strategy Research in Task Development, <i>Andrew D. Cohen</i>	893
55	Using Standards and Guidelines, <i>Nick Saville</i>	906
56	Statistics and Software for Test Revisions, <i>Yo In'nami and Rie Koizumi</i>	925
57	Standard Setting in Language Testing, <i>Dorry M. Kenyon and Anja Römhild</i>	944

58 Administration, Scoring, and Reporting Scores, <i>Ari Huhta</i>	962
59 Detecting Plagiarism and Cheating, <i>Ardeshir Geranpayeh</i>	980
Part 8: Technology and Assessment	995
60 New Media in Language Assessments, <i>Paul Gruba</i>	997
61 Using Corpora to Design Assessment, <i>Fiona Barker</i>	1013
62 Eye-Tracking Technology for Reading, <i>Paula M. Winke</i>	1029
63 Acoustic and Temporal Analysis for Assessing Speaking, <i>Okim Kang and Lucy Pickering</i>	1047
64 Computer-Automated Scoring of Written Responses, <i>Nathan T. Carr</i>	1063
 Volume III: Evaluation, Methodology, and Interdisciplinary Themes	
<i>About the Editor</i>	<i>xvii</i>
<i>Introduction to Volume III</i>	<i>xviii</i>
Part 9: Designing Evaluations	1079
65 Evaluation of Language Tests Through Validation Research, <i>Carol A. Chapelle and Erik Voss</i>	1081
66 Fairness and Justice in Language Assessment, <i>Antony John Kunnan</i>	1098
67 Accommodations in the Assessment of English Language Learners, <i>Jamal Abedi</i>	1115
68 Consequences, Impact, and Washback, <i>Liying Cheng</i>	1130
Part 10: Quantitative Analysis	1147
69 Classical Test Theory, <i>Yasuyo Sawaki</i>	1149
70 Classical Theory Reliability, <i>James Dean Brown</i>	1165
71 Score Dependability and Decision Consistency, <i>James Dean Brown</i>	1182
72 The Use of Generalizability Theory in Language Assessment, <i>George A. Marcoulides and Marsha Ing</i>	1207
73 Exploratory Factor Analysis and Structural Equation Modeling, <i>Gary J. Ockey</i>	1224
74 Questionnaire Development and Analysis, <i>Aek Phakiti</i>	1245
75 Item Response Theory in Language Testing, <i>David P. Ellis and Steven J. Ross</i>	1262
76 Differential Item and Testlet Functioning Analysis, <i>Hong Jiao and Ying-Fang Chen</i>	1282
77 Multifaceted Rasch Analysis for Test Evaluation, <i>Khaled Barkaoui</i>	1301

Part 11: Qualitative and Mixed Method Analysis	1323
78 Content Analysis, <i>Evelina D. Galaczi</i>	1325
79 Introspective Methods, <i>Miyuki Sasaki</i>	1340
80 Raters and Ratings, <i>Alistair Van Moere</i>	1358
81 Spoken Discourse, <i>Anne Lazaraton</i>	1375
82 Written Discourse, <i>Lia Plakans</i>	1390
83 Mixed Methods Research, <i>Carolyn E. Turner</i>	1403
84 Writing Research Reports, <i>Kyle McIntosh and April Ginther</i>	1418
Part 12: Interdisciplinary Themes	1431
85 Philosophy and Language Testing, <i>Glenn Fulcher</i>	1433
86 Cognition and Language Assessment, <i>James E. Purpura</i>	1452
87 Language Acquisition and Language Assessment, <i>Yasuhiro Shirai and Mary Lou Vercellotti</i>	1477
88 Bilingual Assessment, <i>Usha Lakshmanan</i>	1492
89 Classroom-Based Assessment Issues for Language Teacher Education, <i>Constant Leung</i>	1510
90 Program Evaluation and Language Assessment, <i>Brent A. Green</i>	1520
91 The Interface of Language Assessment and Forensic Contexts, <i>Margaret van Naerssen</i>	1534
92 Language Testing in the Dock, <i>Glenn Fulcher</i>	1553
93 The Influence of Ethics in Language Assessment, <i>Bernard Spolsky</i>	1571
94 Ongoing Challenges in Language Assessment, <i>Lyle F. Bachman</i>	1586
 Volume IV: Assessment Around the World	
<i>About the Editor</i>	<i>xvii</i>
<i>Introduction to Volume IV</i>	<i>xviii</i>
Part 13: Current Practices in Assessing English	1605
95 English as a Lingua Franca, <i>Jennifer Jenkins and Constant Leung</i>	1607
96 Assessing English in Australia and New Zealand, <i>Kathryn Hill and Rosemary Erlam</i>	1617
97 Assessing English in North America, <i>Samira ElAtia</i>	1629
98 Assessing English in Mexico and Central America, <i>Caroline Payant and Francisco Javier Barrón Serrano</i>	1638

99	Assessing English in the Middle East and North Africa, <i>Atta Gebril and Russanne Hozayin</i>	1649
100	Assessing English in South Asia, <i>Rama Mathew</i>	1658
101	Assessing English in East Asia, <i>Viphavee Vongpumivitch</i>	1671
102	Assessing English in Southeast Asia, <i>Richard Watson Todd and Chih-Min Shih</i>	1681
103	Assessing English in South America, <i>Matilde Scaramucci and Adriana Boffi</i>	1690
104	Assessing English in Europe, <i>Gad S. Lim</i>	1700
	Part 14: Current Practices in Assessing Languages in Africa	1709
105	Assessing Swahili, <i>Katrina Daly Thompson</i>	1711
106	Assessing Shona and Ndebele, <i>Galen Sibanda</i>	1722
	Part 15: Current Practices in Assessing Languages in North and South America	1731
107	Assessing American Sign Language, <i>Robert J. Hoffmeister, Marlon Kuntze, Sarah A. Fish</i>	1733
108	Assessing Hawaiian, <i>William H. Wilson</i>	1748
109	Assessing North American Indigenous Languages, <i>J. Dean Mellow and Kaitlyn Begg</i>	1759
110	Assessing North American Spanish, <i>M. Rafael Salaberry</i>	1769
	Part 16: Current Practices in Assessing Languages in the Middle East and South Asia	1779
111	Assessing Arabic, <i>Atta Gebril and Hanada Taha-Thomure</i>	1781
112	Assessing Farsi, <i>Hossein Farhady and Kobra Tavassoli</i>	1790
113	Assessing Hebrew, <i>Ofra Inbar-Lourie</i>	1799
114	Assessing Hindi, <i>Pritha Chandra</i>	1815
115	Assessing Malayalam, <i>Suchitra Sadanandan</i>	1823
116	Assessing Nepali, <i>Madhav P. Pokharel</i>	1832
117	Assessing Sinhala, <i>Priyanvada Abeywickrama and Dushyanthi Mendis</i>	1841
118	Assessing Tamil, <i>Vijayanthi Sankar</i>	1851
119	Assessing Telugu, <i>K. V. V. L. Narasimha Rao</i>	1863
	Part 17: Current Practices in Assessing Languages in Southeast and East Asia	1873
120	Assessing Bahasa Melayu and Bahasa Indonesia, <i>May Tan</i>	1875
121	Assessing Cantonese, <i>Wing Sat Chan</i>	1892

122	Assessing Japanese, <i>Yoshinori Watanabe and Kaoru Koyanagi</i>	1901
123	Assessing Korean, <i>Chungsook Kim and Junho Lee</i>	1909
124	Assessing Mandarin Chinese, <i>Luxia Qi and Kai Zhang</i>	1919
125	Assessing Hakka, Southern Min, and Taiwanese Indigenous Languages, <i>Jessica R. W. Wu</i>	1928
126	Assessing Thai, <i>Pranee Kullavanijaya, Viphavee Vongpumivitch, and Debi Jaratjarungkiat</i>	1938
Part 18: Current Practices in Assessing Languages in Australia and New Zealand		1949
127	Assessing Australian and New Zealand Indigenous Languages, <i>Gillian Wigglesworth and Peter Keegan</i>	1951
128	Assessing Māori Indigenous Language Learners, <i>Cath Rau</i>	1961
Part 19: Current Practices in Assessing Languages in Europe		1969
129	Assessing Armenian, <i>Rubina Gasparian and Mirtsa Halajyan</i>	1971
130	Assessing Finnish, <i>Mirja Tarnanen and Eija Aalto</i>	1980
131	Assessing French, <i>Eve Ryan</i>	1989
132	Assessing German, <i>Michaela Perlmann-Balme</i>	1999
133	Assessing Greek, <i>Spiros Papageorgiou</i>	2009
134	Assessing Italian, <i>Giuliana Grego Bolli</i>	2018
135	Assessing Norwegian, <i>Cecilie Carlsen and Eli Moe</i>	2029
136	Assessing Polish, <i>Jo Lewkowicz and Marcin Smolik</i>	2038
137	Assessing Portuguese, <i>Michele Back</i>	2047
138	Assessing Russian, <i>Irshat Madyarov</i>	2057
139	Assessing Spanish, <i>José Ramón Parrondo Rodríguez</i>	2068
140	Assessing Welsh, <i>Emyr Davies</i>	2077
<i>Index</i>		2087

Part 5

Assessment Approaches

Large-Scale Assessment

Janna Fox

Carleton University, Canada

Introduction

Kunnan (2008) defines large-scale language assessment as tests and testing practices that are designed and managed for “uniformity . . . across geographical regions, administration time, test raters and score interpretation” (p. 135). His definition is consistent with discussions of large-scale tests “as those that are administered to groups of examinees over multiple administrations . . . [and] require that the meaning of the test scores remains steady so that appropriate comparisons can be made and trends measured” (Wendler & Walker, 2006, p. 446). Although large-scale language assessment plays a critical role in contexts such as college admission, immigration, or licensing for occupations, Kunnan (2008) focuses on its role in general education and learning, noting that large-scale language assessment has “become increasingly important in the last 25 years in many parts of the world in school, college, and university contexts” (p. 135). He links the rise of large-scale language assessment to educational reform and accountability policies, but he also discusses its importance in considerations of test fairness—a key requirement for test quality and validity (see, for example, Messick, 1989; American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999).

Previous Views or Conceptualization

Fairness has long been a goal of large-scale language assessment. Davidson, Turner, and Huhta (1997) highlight its historical origins in the Chinese civil service examination system introduced in imperial China during the Han Dynasty, which

ruled from 202 BCE to 221 CE. The Chinese examination system was designed to select government bureaucrats on the basis of merit (rather than connections, or privilege) and test administration procedures were systematically defined to achieve fairness through uniformity. Durant gives a vivid description of one such procedure:

In the Examination Hall were ten thousand cells, in which the contestants, cribbed and confined, lived with their own food and bedding for three separate days, while they wrote essays or theses on subjects announced to them after their imprisonment. (Durant, 1935, p. 801)

As Durant's account suggests, large-scale language assessment has been associated with: (1) controls for uniformity across test versions, sites, administrations, raters, and interpretations; (2) large numbers of test takers; (3) high stakes; and (4) bureaucratic or accountability agendas. Further, as Davidson et al. (1997) point out, large-scale language assessment places "value [on] centralized control, because it is thought that such control helps guarantee test quality . . . [which] is often checked using statistics" (p. 305). Implicit in the views of these authors is the role that educational measurement has played in achieving (or attempting to achieve) *uniformity* in testing and testing practices. In the following section two key features of the measurement tradition in large-scale language assessment are briefly discussed: specifically, the psychometric properties of tests, development, and validation; and scale development, norm- and criterion-referenced.

Current Views or Conceptualization

Test Development in Large-Scale Language Testing: From Specifications to Validation

The development of a large-scale language test begins with a clear *mandate*, based on the determination that a need can best be addressed by testing. Next, the specific purpose of the test is defined, for example, to assess the proficiency of a second language (L2) applicant to an English-medium university, or to certify the interactional competence of a technician seeking a license to practice a trade. Then the *construct* and content domain are identified (for example the academic language required for undergraduate study, or the communicative skills required of an electrical technician). Given that uniformity across multiple versions and test sites is a goal of large-scale language assessment, the purpose, construct, and content of a test must be defined in sufficient detail regarding the knowledge, skills, processes, and so on, "so that it is clear whether or not any particular item, content, or skill falls within the scope of the test framework" (Linn, 2006, p. 28).

The next critical step is the development of *test specifications* that "delineate the format of items, tasks or questions; the response format or conditions of responding; and the type of scoring procedures" (AERA, APA, & NCME, 1999, p. 38). Test specifications provide test developers with the blueprint or recipe for the development of multiple versions of a test, which are as parallel or similar as possible. In order to ensure maximum similarity, large-scale testing requires ongoing research