This bibliography has been compiled for ILTA members only and is not to be distributed or posted elsewhere without the permission of ILTA.

New entries and corrections to errata in the bibliography may be submitted at https://bit.ly/3r9AMQB.
### Editorial Board

<table>
<thead>
<tr>
<th>Years</th>
<th>Editor(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999–2012</td>
<td>Annie Brown</td>
</tr>
<tr>
<td>2013–2018</td>
<td>Tineke Brunfaut</td>
</tr>
<tr>
<td>2019</td>
<td>Olena Rossi and Wenjun (Elyse) Ding</td>
</tr>
<tr>
<td>2020</td>
<td>ILTA GSA Bibliography Committee</td>
</tr>
<tr>
<td></td>
<td>Dylan Burton (chief editor) Minh Ngo</td>
</tr>
<tr>
<td></td>
<td>Jorge Beltrán</td>
</tr>
<tr>
<td></td>
<td>Hyunah Kim</td>
</tr>
<tr>
<td></td>
<td>Wenjun (Elyse) Ding</td>
</tr>
<tr>
<td></td>
<td>Haoshan (Sally) Ren</td>
</tr>
<tr>
<td></td>
<td>Olena Rossi</td>
</tr>
<tr>
<td></td>
<td>Yan (Niles) Zhao</td>
</tr>
</tbody>
</table>

Questions about the bibliography may be directed to Dylan Burton: burtonjd@msu.edu
Contents

Editorial Board......................................................................................................................... 1
Classification Index .................................................................................................................... 2
1. General................................................................................................................................. 5
2. Testing Listening / Sign Comprehension ............................................................................. 14
3. Testing Reading .................................................................................................................... 19
4. Testing Speaking / Signing ................................................................................................. 26
5. Testing Writing .................................................................................................................... 40
6. Testing Language Elements ............................................................................................... 57
7. Purposes for Testing........................................................................................................... 64
8. Specific Purpose Testing ..................................................................................................... 69
9. Standards and Frameworks................................................................................................. 79
10. Principles of Test Design ................................................................................................... 85
11. Factors Affecting Test Performance .................................................................................. 92
12. Rating and Rating Scales ................................................................................................. 100
13. Test Reliability and Validity ............................................................................................. 111
14. Washback ........................................................................................................................ 124
15. Ethics ................................................................................................................................ 128
16. Classroom Based Assessment ......................................................................................... 132
17. Technology and Language Testing .................................................................................. 140
18. Testing Young Learners .................................................................................................... 147
19. Language Testing Policy ................................................................................................... 151
20. Research Methodology ..................................................................................................... 154

Classification Index

Accommodations see – factors affecting test performance
Accountability see – purposes for testing, ethics
Achievement tests see – purposes for testing, principles of test design
Admission test see – purposes for testing
Alternative assessment see – principles of test design, test reliability and validity, ethics, classroom based assessment
Aptitude testing see – purposes for testing, ethics
Assessment literacy see – general, classroom based assessment
Authenticity see – specific purpose testing, principles of test design
Automated scoring see – technology and language testing, testing speaking, testing writing
Bias see – factors affecting test performance, rating and rating scales, test reliability and validity, washback, ethics
Can do statements see – standards and frameworks
Cheating see – ethics
Cloze tests see – testing reading
Conference proceedings see – general
Computer adaptive testing, Computer based testing

see – principles of test design, rating and rating scales, test reliability and validity, technology and language testing

Conference proceedings

see – general

Construct

see – principles of test design, rating and rating scales, test reliability and validity, classroom based assessment, technology and language testing

Construct validity

see – test reliability and validity

Content, assessment of

see – specific purpose testing, rating and rating scales, ethics, classroom based assessment

Content validity

see – test reliability and validity

Corpus, Corpora

see – specific purpose testing, test reliability and validity, technology and language testing

Criterion-referenced testing

see – principles of test design, test reliability and validity

C-test

see – testing reading

Curriculum

see – washback, classroom based assessment

Dependability

see – test reliability and validity

Diagnosis, Diagnostic

see – purposes for testing

DIF

see - factors affecting test performance, test reliability and validity

Disabilities

see – principles of test design, factors affecting test performance, ethics

Discourse, discourse analysis

see – testing speaking, testing writing, rating and rating scales, test reliability and validity, classroom based assessment

English for academic purposes, EAP

see – specific purpose testing

Entrance test

see – purposes for testing

Fairness

see – test reliability and validity, ethics

Formative assessment

see – classroom based assessment

Gender

see – factors affecting test performance

Generalisability theory, G theory, GENOVA

see – test reliability and validity

Grammar

see – testing language elements

Immigration

see – specific purpose testing, ethics, language testing policy

Impact

see - washback

Integrated skills, integrated testing

see – principles of test design, test reliability and validity, testing reading, testing writing

Item Response Theory, IRT

see – test reliability and validity

Item types

see – principles of test design, test reliability and validity

Language assessment literacy

see - general

Languages for specific purposes, LSP

see – specific purpose testing

Literacy, assessment of

see – testing reading, testing writing

Outcomes, Outcomes-based assessment

see – standards and frameworks, washback, classroom based assessment
<table>
<thead>
<tr>
<th>Term</th>
<th>See</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer assessment</td>
<td>see – rating and rating scales, classroom based assessment</td>
</tr>
<tr>
<td>Performance testing</td>
<td>see – principles of test design, test reliability and validity</td>
</tr>
<tr>
<td>Placement testing</td>
<td>see – purposes for testing</td>
</tr>
<tr>
<td>Policy</td>
<td>see – language testing policy</td>
</tr>
<tr>
<td>Portfolio assessment</td>
<td>see – testing writing, rating and rating scales, classroom based assessment</td>
</tr>
<tr>
<td>Pragmatics</td>
<td>see – testing speaking</td>
</tr>
<tr>
<td>Proficiency testing</td>
<td>see – purposes for testing</td>
</tr>
<tr>
<td>Program evaluation</td>
<td>see – specific purpose testing</td>
</tr>
<tr>
<td>Qualitative methods, Qualitative validation procedures</td>
<td>see – test reliability and validity</td>
</tr>
<tr>
<td>Rasch analysis</td>
<td>see – factors affecting test performance, rating and rating scales, test reliability and validity</td>
</tr>
<tr>
<td>Rater, Rater training</td>
<td>see – rating and rating scales, test reliability and validity</td>
</tr>
<tr>
<td>Rating scales</td>
<td>see – rating and rating scales, test reliability and validity</td>
</tr>
<tr>
<td>Reform, educational / language assessment</td>
<td>see – language testing policy</td>
</tr>
<tr>
<td>Self assessment</td>
<td>see – rating and rating scales, classroom based assessment</td>
</tr>
<tr>
<td>Speech and communication disorders</td>
<td>see – specific purpose testing</td>
</tr>
<tr>
<td>Standards, Standards frameworks, Standard setting</td>
<td>see – standards and frameworks</td>
</tr>
<tr>
<td>Statistics</td>
<td>see – principles of test design, test reliability and validity</td>
</tr>
<tr>
<td>Strategies</td>
<td>see – factors affecting test performance</td>
</tr>
<tr>
<td>Structural Equation Modelling</td>
<td>see – test reliability and validity</td>
</tr>
<tr>
<td>Task types, Task based assessment</td>
<td>see – principles of test design</td>
</tr>
<tr>
<td>Task difficulty</td>
<td>see – principles of test design</td>
</tr>
<tr>
<td>Test taking strategies</td>
<td>see – factors affecting test performance</td>
</tr>
<tr>
<td>Test translation</td>
<td>see – principles of test design</td>
</tr>
<tr>
<td>Translation tests</td>
<td>see – specific purpose testing, test reliability and validity</td>
</tr>
<tr>
<td>Video</td>
<td>see – testing listening</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>see – testing language elements</td>
</tr>
</tbody>
</table>
1. General


Brunfaut, T., & Harding, L. (2018). Teachers setting the assessment (literacy) agenda: a case study of a teacher-led national test development project in Luxembourg. In D. Xerri, & P. Vella Briffa (Eds.), Teacher involvement in high stakes language testing (pp. 155-172). Springer.


2. Testing Listening / Sign Comprehension


Qi, L. (1999). Experimenting on the listening section of the National Matriculation English Test. *English for Primary and Middle Schools, 1*, 3-4.


3. Testing Reading


https://doi.org/10.1080/15434303.2020.1731753


4. Testing Speaking / Signing


Liu, W., & Qian, D. D. (2009). The influence of raters’ levels of occupational knowledge on the process of assessing ESP oral proficiency. In M. Chan, G. Humphreys, T. Lo, J. Lung, C. Wong, & X. Xu (Eds.), Proceedings of the 3rd International Symposium on Teaching English at Tertiary Level (pp. 41-47). Hong Kong: Hong Kong Polytechnic University.


https://doi.org/10.1177/0265532218775764


5. Testing Writing


Bae, J. (2006). The validation of an alternative writing task for beginning level students: Write as many words as you can. *New Studies of English Language and Literature*, 33, 229-256.


Lockwood, J. (2013). The Diagnostic English Language Tracking Assessment (DElTA) writing project: A case for post-entry policies and practices in Hong Kong Universities. Papers in Language Testing and Assessment, 2(1) 30-49.


© ILTA, 2021


© ILTA, 2021


6. Testing Language Elements


7. Purposes for Testing


© ILTA, 2021

64


8. Specific Purpose Testing


Barber, I., & Bernstein, J. (2000). Design and development parameters for a rapid automatic screening test for prospective simultaneous interpreters. *Interpreting, 5*(2), 221-238.


© ILTA, 2021

69


9. Standards and Frameworks


Bejar, K., Selvachandran, J., Clea...


Fulcher, G. (2010). The reification of the Common European Framework of Reference (CEFR) and effect-driven testing. In A. Psylou-Joycey & M. Matthiasoudakis (Eds.), Advances in research on language acquisition and teaching (pp. 15–26). Greece, Thessaloniki: GALA.


10. Principles of Test Design


Bailey, A., Butler, F., & Sato, E. (2007). St...


Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. Assessing Writing, 10(1), 5-43.


11. Factors Affecting Test Performance


Brown, J. D. (1999). The relative importance of persons, items, subtests and languages to TOEFL test variance. Language Testing, 16(2), 217-238.


Zumbo, B. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly, 4*(2), 223-233.
12. Rating and Rating Scales


language education: Meeting the challenges of research and practice. Hong Kong: Hong Kong University of Science and Technology.


Liu, W., & Qian, D. D. (2009). The influence of raters’ levels of occupational knowledge on the process of assessing ESP oral proficiency. In M. Chan, G. Humphreys, T. Lo, J. Lung, C. Wong, & X. Xu (Eds.), *Proceedings of the 3rd International Symposium on Teaching English at Tertiary Level* (pp. 41-47). Hong Kong: Hong Kong Polytechnic University.


13. Test Reliability and Validity


Bae, J. (2006). The validation of an alternative writing task for beginning level students: Write as many words as you can. *New Studies of English Language and Literature, 33*, 229-256.


Brown, J. D. (1999). The relative importance of persons, items, subtests and languages to TOEFL test variance. Language Testing, 16(2), 217-238.


Cheng, L., & Ma, J. (2015). "It puts us in our students' shoes": Listening to voices from teacher candidates on their test-taking experience. *International Journal of Humanities and Social Science Research, 1*, 46-56.


D’Este, C., & Ludbrook, G. (2013). Fairness and validity in testing students with SpLDs: a case study from Italy. In D. Tsagari & G. Spanoudis (Eds.), *Assessing L2 students with learning and other disabilities* (pp. 169-188). Newcastle upon Tyne: Cambridge Scholars Publishing.


14. Washback


Merrifield, G. (2011). An impact study into the use of IELTS by professional associations and registration entites: Canada, the United Kingdom and Ireland. IELTS Research Reports, 11, 21-72.


15. Ethics


16. Classroom Based Assessment


17. Technology and Language Testing


© ILTA, 2021


© ILTA, 2021

144


Henshey, PA; IGI Global.


18. Testing Young Learners


https://doi.org/10.15434/2019.1619180


19. Language Testing Policy


20. Research Methodology

© ILTA, 2021