ILTA Bibliography of PhDs in Language Testing

This bibliography is provided by ILTA as a service to students of language testing. In order to ensure that it is as comprehensive as possible, when you complete your PhD, or if you know of any others that have not been included, please send the bibliographic details to o.rossi@lancaster.ac.uk or wenjun.ding@bristol.ac.uk. The bibliography is updated annually.


Assiri, Mohammed (2011) Test-taking strategy use on the reading section of the TOEFL-iBT: A study of Arab ESL learners. Oklahoma State University.


Batty, Aaron Olaf (2017) The impact of visual cues on item response in video-mediated tests of foreign language listening comprehension. Lancaster University.


Coombe, Christine (1992) The relationship between self-assessment ratings of functional literacy skills and basic English skills test results in adult refugee ESL learners. The Ohio State University.


Elham Fruzande (2008) TEFL program evaluation at Master’s level in Iran. Shiraz University, Iran.


Fox, Janna (2001) It’s all about meaning: L2 test validation in and through the landscape of an evolving construct. McGill University.

Froetscher, Doris (2017) An investigation into the washback of a standardized national exam on the classroom testing of reading. Lancaster University.

Frost, Kellie (2017) Test impact as dynamic process: individual experiences of the English test requirements for permanent skilled migration in Australia. The University of Melbourne.


Grabowski, Kirby (2009) Investigating the construct validity of a test designed to measure grammatical and pragmatic knowledge in the context of speaking. Teachers College, Columbia University.


Gui, Min (2011) Investigating the construct validity of the reading comprehension section of the College English Test in China: A structural equation modeling approach. The University of Texas at Austin.


Huang, Heng-Tsung Danny (2010) Modeling the relationships among topical knowledge, anxiety, and integrated speaking test performance: A structural equation modeling approach. The University of Texas at Austin.


Hyland, Theresa A. (2005) A study of the practices of EL1 and EL2 students in reporting information from sources in a timed writing proficiency assessment. OISE at University of Toronto.

Im, Gwan Hyeok (2019) Stakeholder voices: Validity argument for score meaning and use of the Test of English for International Communication (TOEIC) in international business workplaces. Queen’s University.


In’mani, Yo (2007). The effects of task types on listening test performance: a quantitative and qualitative study. University of Tsukuba, Japan.


Jang, Eunice (2005) A validity narrative: effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL. University of Illinois at Urbana-Champaign.


Ji, Yong-II (2012). Comparing strategic processes in the iBT speaking test and in the academic classroom. University of Leicester.


Joo, M.-J. (2007) Korean university students’ attitudes to and performances on a Face-to-Face Interview (FTFI) and a Computer Administered Oral Test (CAOT). Institute of Education, University of London.


Kangzhou, Peng (2014) A validation study of listening comprehension based on criterion-referenced language assessment which has been done in 2010. Shanghai International Studies University.


Kasai, Masahiro (1997) Application of the rule space model to the reading comprehension section of The Test of English as a Foreign Language (TOEFL). University of Illinois at Urbana-Champaign.


Kattan, Jeanne (1990) The construction and validation of an EAP test for second year english and nursing majors at Bethlehem University. Lancaster University.


Khalifa, Hanan (1997) A study in the construct validation of the reading module of an EAP proficiency test battery: Validation from a variety of perspectives. Reading University.


Kinzley, Simon (2011) The impact of a university pre-sessional course on the academic writing behaviours of a group of Chinese undergraduate students studying for a degree in media and cultural studies. Lancaster University.


Kollias, Charalambos (2017) Virtual standard setting in language testing: Exploring the feasibility and appropriateness of setting cut scores in synchronous audio and audio-visual environments. Lancaster University.


Updated December 2019


Leta, Dejenie (1990) Achievement and proficiency in a school leaving examination: A case of the nature of innovation in an Ethiopian setting. Lancaster University.


Liao, Yen-Fen (2009) A construct validation study of the GEPT reading and listening sections: re-examining the models of L2 reading and listening abilities and their relations to lexico-grammatical knowledge. Teachers College, Columbia University.


Lin, Chih-Kai (2014) Issues and challenges in current generalizability theory applications in rated measurement. University of Illinois at Urbana-Champaign.


Lin, Yu-Chen Tina (2016) Decoding complex constructs of the integrated reading-writing task: Modeling relationships among second language reading/writing proficiency, familiarity
with text features, perceived cognitive operation difficulty, and summary writing through the structural equation modeling approach. Indiana University Bloomington.


Lu, Yang (2006) A discourse analytic study of EFL test-takers’ spoken discourse competence and its impact on their oral proficiency and spoken grammatical competence. The University of Reading.


Nakatsuha, Fumiyo (2009) Conversational styles in group oral tests: How is the conversation co-constructed? The University of Essex.


Park, Sang Bok (2011) Using motivation theories to analyse students’ perceptions of an examination and their inclination to study for it. Lancaster University.


Schoonen, Rob (1991) De evaluatie van schrijfvaardigheidsmetingen. Een empirische studie naar betrouwbaarheid, validiteit en bruikbaarheid van schrijfvaardigheidsmetingen in de achtste groep van het basisonderwijs [The evaluation of writing assessments. An empirical study into the reliability, validity and utility of writing assessments in final grade of elementary school.] University of Amsterdam. [Dutch, summary in English]


Shahmirzadi, Niloufar (2019). A simulation and real data study on reading comprehension of National PhD Admission Test under Cognitive Diagnostic Assessment (CDA) through Differential Attribute Functioning (DAF) and Differential Item Functioning (DIF). Islamic Azad University at Central Tehran Branch.


So, Young Jang (2010) The development and evaluation of a systematic training program for increasing both rater reliability and rating accuracy. University of Illinois at Urbana-Champaign.

Song, Xiaomei (2014) Test fairness in a large-scale high-stakes languages. Queen's University.

Souza, Paula Ribeiro (2018) O efeito retroativo do EPLIS nas percepções, atitudes e ações de professores e alunos de um curso de formação em controle de tráfego aéreo. [The washback of EPLIS on teachers’ and learners’ perceptions, attitudes and actions in an air traffic control initial training course]. Universidade Estadual de Campinas.


Souza, Paula Ribeiro (2018) O efeito retroativo do EPLIS nas percepções, atitudes e ações de professores e alunos de um curso de formação em controle de tráfego aéreo. [The washback of EPLIS on teachers’ and learners’ perceptions, attitudes and actions in an air traffic control initial training course]. Universidade Estadual de Campinas.


Tsagari, Dina (2007) Investigating the washback effect of a high-stakes EFL exam in the Greek context: participants' perceptions, material design and classroom applications. Lancaster University.

Tsai, Constance (2004) Investigating the relationships between ESL writers' strategy use and their second language writing ability. Teachers College, Columbia University.


Van Mael, Jan (1999) Understanding the assessment of oral foreign language proficiency. an investigation into the rater’s judgmental contents structure. Université catholique de Louvain.


Voss, Erik (2012) A validity argument for score meaning of a computer-based ESL collocational ability test based on a corpus-driven approach to test design. Iowa State University.

Updated December 2019


Waller, Daniel (2015) Investigation into the features of written discourse at levels B2 and C1 of the CEFR. University of Bedfordshire.


Weiler, Theresa (2018) Investigating the construct tested through four item types used to assess lexicogrammatical competence in English as a foreign language. Lancaster University.


Updated December 2019

Xu, Jing (2015) Predicting ESL learners’ oral proficiency by measuring the collocations in their spontaneous speech. Iowa State University.


Yang, Rui (2009) A many-facet Rasch analysis of rater effects on an oral English proficiency test. Purdue University.


