Editorial

The Evolving Role of Language Assessment in Skills Development in the Modern World: Introduction to the Special Issue

Michael Milanovic
LanguageCert, UK

Wen Zhao
Jinan University, China

Peter Falvey
The Education University of Hong Kong, China

David Coniam*
LanguageCert, UK

In order to make it easier for readers to access this special issue that focuses on *The Evolving Role of Language Assessment in Skills Development* in the Modern World, the thirteen articles are divided into four groups: background research, test analysis, assessment, and test development. Within each group, a variety of topics are discussed.

The first group comprises six papers dealing with background research in: use of connectors in writing in China; a survey of online invigilation in China; interactional competence in speaking; language standards; cognitive processes in reading; and eye-tracking when reading. Readers will note that the CEFR (Common European Framework of Reference) for languages is quite commonly referred to as its levels have become universally recognized as a popular international standard for describing language ability. In this special issue, other tests are sometimes compared with CEFR levels. In the second group on test analysis, consisting of three articles, the first two focus on vocational English in China. The third paper in this group analyses speaking skills in a test for a large cohort of young learners in Uruguay. The third group deals with assessment and comprises three papers, focusing on: self-assessment; equating standards; and examiner quality in writing tests. The single paper in the area of test development lies within the context of vocational education in China. A fuller description of the papers in the special issue is provided below.

Group 1: Background Research

This, the largest group, comprises six diverse, background research papers ranging across topics such as: the use of connectors in the writing of Chinese college students; a study of the perceptions of China candidates to online invigilation; interactional competence in LanguageCert examinations; a description of WIDA's

*Corresponding author. Email: david.coniam@PeopleCert.org

Language Standards Frameworks and Language Assessment; an investigation of Austrian students' cognitive processes and FL reading test taking behaviour; and the opportunities and obstacles of eye-tracking in determining the cognitive processes of L2 students taking online multiple-choice reading tests. The purpose of including these papers is to present as wide a range as possible of on-going research projects that provide readers with up-to-date descriptions and findings of a range of research studies from different countries.

In the first paper in this group, Nairui CHEN and Wen ZHAO describe how Chinese college students use cohesive devices in their writing by analysing a learner corpus of 20,000 English texts of university students from Years 1-4, taken from *pigai wang*, an online automatic scoring system designed to help campus students with their English writing. Comparisons were made in terms of connector use patterns from the corpus with reference to CEFR and CSE descriptors (the China Standards of English Language Ability) to provide empirical connector use patterns that students demonstrate in their writing and how the descriptors used in the CSE can better reflect students' actual language use across levels.

The second paper, written by David CONIAM, investigates the attitudes and perceptions of Chinese participants to the online invigilation of English language examinations. Since the advent of the Covid pandemic, online assessment and invigilation has grown exponentially and such studies help to assess the impact that online invigilation has had on test participants.

The third paper, by Leda LAMPROPOULOU, investigates the operationalizing of interactional competence (IC) through role play, sourcing its data from the suite of LanguageCert International ESOL Speaking exams at CEFR levels B1 and C1. Issues such as scalability, discriminability, score separability, and the interlocutor effect are discussed. The paper confirms previous studies that the measurement of interactional competence with a specific task has found role play activity to be a realistic and authentic task type, able to tap into most of the IC characteristics testers would wish to elicit. In addition, the role play task has been found to afford a less unequal interaction than other types of tasks, an important consideration when assessing interactional competence in speaking examinations.

Paper four in this group by Margo GOTTLIEB and Mark CHAPMAN, provides a detailed description of work done by WIDA (World-class Instructional Design and Assessment) that is an alternative to the CEFR in the context of Language Standards Frameworks and Language Assessment. In this paper, the authors highlight the role of the WIDA consortium in the development of language standards in assessment practices within the context of the United States. The paper provides a useful and comprehensive overview of a system different from that presented by the CEFR.

The penultimate paper in this group comes from Austria where Klaus SILLER, Ulrike KIPMAN, and Andrea KULMHOFER-BOMMER investigate the cognitive processes and FL reading test taking behaviour of Austrian school learners of English. The paper emphasises the importance of explicit, guided foreign language reading instruction that focuses not only on the product of reading (comprehension) but also on the processes involved.

The final paper in this group, by Nicola LATIMER and Sathena CHAN investigates eye-tracking in the cognitive processes used by L2 students engaged in reading test tasks because it offers an opportunity to gather evidence to support the validity argument of the tests as well as to inform learning-oriented assessment. Using Weir and Khalifa's model of reading (2009), the paper studies the mental processes test-takers at B1, B2 and C1 CEFR levels actually use in comprehending texts when engaging in different types of real-life reading.

Group 2: Test Analysis

This group comprises three papers. The first two focus on vocational education in China. First, Haixia LIU focuses on the washback effect of the Beijing English Academic Proficiency Test on English Teaching in Vocational Schools (BEAPT). The paper's contribution to the field of testing lies in the

investigation of the washback effect of the test in particular by comparing and contrasting the expected effect of the test with the real feelings and behaviours of teachers, thus providing implications for English teaching and learning.

Second, Duqin WANG, continuing the theme of vocational education in China, discusses Vocational English Assessment in Senior Middle Schools and investigates, through a review of the research literature and a quantitative online survey whether and how a more integrative approach to assessment can develop through more formative assessment methods.

Third, Nahal KHABBAZBASHI, Fumiyo NAKATSUHARA, Chihiro INOUE, Anthony GREEN, and Gabriela KAPLAN, in a paper that moves away from vocational education, describe the design and validation of an online speaking test for young learners in Uruguay, a joint effort between the University of Bedfordshire's CRELLA (Centre for Research in English Language and Assessment) and Plan Ceibal (ceibal.edu.uy) in Uruguay that 'provides programs, educational resources and teacher training courses that transform the ways of teaching and learning'. The paper provides extensive descriptions of the initiation, development, construction, administration, and evaluation of a test for a whole cohort of young learners and discusses the challenges and innovations that occurred.

Group 3: Assessment

Group 3, consisting of three papers, focuses on the topic of assessment. In the first paper, Wen ZHAO and David CONIAM discuss how using self-Assessments to establish a dataset can shed light on the comparability of the CEFR and CSE (the China Standards of English). The more often different examinations can be equated with international standards, the better. This paper describes how this can be done by linking both tests and sets of self-assessments to a single frame of reference via a third test whose robustness and values had been previously established.

In the second paper, Tony LEE reveals how using a mixture of Rasch Analysis and expert judgement in English language tests can attempt to solve the vexed question of equating tests with no common anchoring structures by using externally-referenced anchoring methods. This innovative method shows how such equating can be established and sheds light on a difficult issue.

The third paper in this group, by Yiannis PAPARGYRIS and Zi YAN, focuses on the assessment of examiner quality and consistency in LanguageCert writing tests. The paper reports on a study of the training and standardisation of examiners who mark *LanguageCert*'s International ESOL (IESOL) suite of English language tests linked to the Common European Framework of Reference (CEFR). Examiner quality and consistency is fundamental as more and more tests move away from routine multiple-choice formats to the more valid use of criteria-referenced performative assessment.

Group 4: Test Development

In this single entry, Yongjie CHEN describes, within the context of vocational education in China, the development of PRETCO: An English Test for Vocational College Students. PRETCO (the Practical English Test for Colleges) is a nation-wide standardized English test specially designed for non-English majors of vocational and technical colleges, with a candidature of 3 million annually. Chen describes the test's purpose, administration, development, test design, and provides an overall evaluation including a description of the considerable amount of research carried out on various aspects of PRETCO including validity studies and studies on the test's washback and impact. Chen concludes by stating that with the publication of the CSE (the China Standards of English), the issue facing PRETCO test developers

involves how to relate the test to the CSE while implementing the requirements set out in the English Curriculum of Higher Vocational Education.

Michael Milanovic is Chairman of LanguageCert and a member of its Advisory Council. Previously CEO of Cambridge Assessment English, he has been working extensively with PeopleCert since 2015. He worked closely with the Council of Europe on its Common European Framework of Reference, has held, and still holds a number of key external roles. ORCID: 0000-0002-5722-1811

Wen Zhao is Dean of the School of Foreign Studies at Jinan University, Guangzhou. Her main publication and research interests are in corpus linguistics, English curriculum and instruction, and EFL writing. She has been working and researching in English language teaching and learning, and has been involved in national English curriculum development for senior secondary vocational education and College English education. ORCID: 0000-0003-4965-0146

Peter Falvey is an Adjunct Associate Professor at The Education University of Hong Kong. He is a teacher educator, and a former Head of Department in the Faculty of Education. His main publication and research interests are in language assessment, second language writing methodology, and text linguistics. ORCID: 0000-0002-2017-8450

David Coniam is Head of Research at LanguageCert. He has been working and researching in English language teaching, education and assessment for almost 50 years. His main publication and research interests are in language assessment, language teaching methodology and academic writing and publishing. ORCID: 0000-0003-4480-1742