Introduction

Coreen Docherty
Fiona Barker
Cambridge English Language Assessment

Multilingualism and language assessment was the theme of the 5th International Conference of the Association of Language Testers in Europe (ALTE) held in Paris in April 2014. The concept of and discussion around multilingualism is ubiquitous, particularly in education, the workplace and society more generally. This is because the notion of multilingualism is multifaceted and linked to issues such as international mobility, migration and social integration, all of which have clear implications for assessment. As increasing numbers of people move from one country to another to work, study or live, there are pressures to manage this flow of people and to develop strategies to address the needs of new migrants in a host country. Assessment often has a significant and complex role in a multilingual world, particularly related to issues of fairness.

ALTE promotes multilingualism in a number of ways: by setting quality standards for language assessment; by supporting its members to deliver language tests; by providing training and expertise; and, most pertinent to this volume, by providing opportunities for discussion and debate of all aspects of creating multilingual societies, including holding an international conference every three years, which this volume is a product of. The 5th ALTE Conference provided an opportunity to discuss issues around multilingualism and assessment, many of which continue the themes of the previous ALTE conference in Krakow on language frameworks. Over 100 papers and plenaries were presented at the Paris Conference and this volume represents a subset of these, selected on the basis of the four core strands addressed during the conference.

This volume supports ALTE's work, its membership and the wider community of language teachers, assessment experts and policy makers by contributing to the ongoing debate on the theory of multilingualism and its impact on the linked areas of language education and migration policy, which are areas of growing concern for Europe and beyond. The ALTE Paris Conference organisers wanted to foreground contemporary research on language policy and practice and through this volume we can highlight the relevance and benefits that research projects undertaken in local contexts have more widely, also making research accessible to a wider audience than

Research & Validation products

attended the Paris event. It is important to note that the majority of ALTE
members are multilingual and this diversity is reflected in the contributors
to this volume, with most chapters not written in the authors' first language
and many of the papers focusing on the teaching, learning and assessment of
languages besides English. In total, this volume covers 22 languages, written
by contributors based in 16 countries. The aim of this volume is therefore
to explore the role of multiculturalism in language learning, teaching and
assessment through a collection of 25 edited papers based on presentations
given at this event. Although the editors have grouped the papers by theme,
the themes and paper topics are not mutually exclusive and alternative
groupings would have been possible.

Section 1 addresses the issue of intercultural competence and how pluri-
lingualism and pluriculturalism are developed and supported in society. **Neil
Jones**' paper begins the section by focusing on the role of intercultural com-
petence in language education. He argues that it should be treated as an inte-
gral part of language education because learning a language is not simply
about focusing on form and the transmission of knowledge but on helping
learners develop new ways of understanding the world and others. Jones
points out that the language classroom is an ideal place to discuss both lan-
guage and culture with the aim being to encourage the development of inter-
cultural understanding. As part of his argument, Jones reminds us of the core
values contained within the Common European Framework of Reference
for Languages (CEFR) which emphasise plurilingualism and pluricultural-
ism and the acceptance of 'otherness'.

Elaine Boyd and **David Donnarumma** build on Jones' overview of inter-
cultural communication by providing a practical example of the challenges
faced by employees communicating in an international work setting. The
authors describe a study which took place in an international hotel group
in India. Although employees were highly proficient in English, which was
used as a *lingua franca*, there were issues of miscommunication between hotel
guests and staff. The paper highlights not only the challenges of successfully
communicating in an international setting but the complexity involved in
identifying and overcoming the underlying causes of this miscommunication.
Using a mixed method research design, the authors explore the interactions
of the hotel staff with guests and identify the types of (mis)communication
that take place, and then through drama activities explore communication
strategies used by staff, encouraging them to reflect on their own interac-
tions and strategy use. The authors found that the loci of miscommunication
related to culture and power relationships typical in this workplace context.
The expectation of guests, who were also often using English as a *lingua
franca*, and their power position in the transactional interactions meant that
it became the staff's responsibility to repair miscommunication. Boyd and
Donnarumma highlight the difficulties of trying to address intercultural

2

miscommunication within an international workplace but also how important it is to understand these types of interactions as the intercultural workplace is increasingly becoming the norm.

The next two papers move the discussion on to how culture and multilingualism can be conceptualised for assessment purposes. **Maria Stathopoulou** deals with the issue of mediation as a learning outcome that can be assessed. The author describes the Greek national foreign exams which incorporate cross-language written mediation tasks where candidates are given, for example, a text in one language and asked to do something with the content and present it in another language. The author investigates what characterises successful mediation by focusing on strategy use. Stathopoulou's paper not only highlights the importance of mediation as a test feature but also the implications of mediation for teaching. In recent years within language education increasing prominence has been given to making full use of learners' language background, recognising that learners do not use language in a vacuum and that more often than not learners find themselves mediating language for others or for themselves.

Trisevgeni Liontou's paper, which is also related to the Greek context, looks more specifically at the role of culturally familiar reading texts on reading comprehension. Again, focused on testing, Liontou investigates whether texts that are culturally familiar have an effect on test scores and whether they are linguistically comparable to more general texts. The author's premise is that culturally familiar texts may reduce candidate anxiety and provide a fairer test-taking experience for this particular context. Liontou touches on an important issue of fitness for purpose. If the test is being undertaken for local purposes – employment or education in Greece – then would it not be more authentic and appropriate to provide candidates with input texts that are relevant for this context rather than more general educational texts, provided there is no bias? Although this paper taps into the issue of fairness, it highlights the need to consider the purpose of tests in particular contexts and what is fair in an international standardised test may not be fair in a specific context where tests are being used for a particular purpose.

This section ends with a paper which looks more generally at the issue of multilingualism in Europe and whether it is in crisis. **Anne Gallagher** argues that the emergence of English as a *lingua franca* seems to have had a negative effect for the learning of other languages in some European countries, and argues it is in Europe's best interest to ensure the promotion of multilingualism, both at the individual and societal level. Besides merely economic reasons, the author highlights the importance of sustaining the rich cultural and linguistic diversity that has always characterised Europe and that is essential for the maintenance of democracy. In order to ensure a multilingual Europe in the future, the author stresses the importance of adapting language

education to the communicative needs of the learners and of implementing valid, reliable and transparent assessment systems. Finally, she emphasises the need for enhanced support from policy makers and governments to encourage stronger collaboration between language testers and teachers, who will be the key contributors towards a Europe more multilingual than ever.

Section 2 builds on the papers from Section 1 but shifts the focus to the implications of international mobility or migration on language learning and how to address the various needs of children and migrants who have very different backgrounds and purposes for emigrating.

Masashi Negishi and **Yukio Tono** begin the section. Although their paper is not directly related to mobility, its focus on the development of the CEFR-J, the adaptation of the CEFR in English language teaching/learning in Japanese contexts, has implications for researchers who are considering modifying the CEFR for migrants and for those who are interested in language learning for school-aged pupils. This project resulted in a range of lower-level descriptors below A1 and the rest of the levels being divided into multiple branching levels. In their paper, Negishi and Tono describe the work that has been done since the introduction of the CEFR-J and then look into its impact by tapping into 'big data'. They find that the CEFR and the CEFR-J still have not made a significant impact in the Japanese context in terms of internet searches. This paper highlights several issues related to the use of the CEFR in educational contexts. That is, it may not be appropriate to simply adopt the CEFR as it is into some educational contexts, but that it is important to have a coherent and integrated solution to ensure the CEFR is appropriate or modified for the specific learning reality and that teachers understand the purpose behind its adoption and its use.

The next paper by **Eli Moe**, **Marita Härmälä** and **Paula Lee Kristmanson** looks more closely at the language of schooling for young learners with an immigrant or minority language background. They investigate the minimum language levels needed for learning mathematics and history in a second language in order to demonstrate that young language learners need to have a certain threshold of language before it becomes possible to learn content in these subjects. The authors then develop a tool to help raise awareness of the challenges faced by language learners when they are learning both content and language. This paper emphasises the need for the primary and secondary educational systems to consider the needs of immigrant learners and the challenges they face in the mainstream classroom. As migration rates in Europe increase, there is a need to develop a clear and flexible approach to addressing the needs of the children who find themselves in the mainstream classroom. This issue is one that faces many European educational systems and is an area of concern in particular with education budgets being cut across the continent.

4

After having looked at more specific interventions that address language learning, **Enrica Piccardo** deals with the issue of how to promote the development of non-dominant languages in Canada. Even though Canada is home to a linguistically diverse population, Piccardo points out that languages other than French and English are frequently not recognised and undervalued. She describes the development of an e-portfolio that not only supports language learning in general but was designed to encourage the learning of aboriginal languages in particular, which are disappearing at an alarming rate. She hopes that this e-portfolio can act as a catalyst for introducing ideological and pedagogical change in language learning and assessment. This is one of the only papers in this volume that looks specifically at learning languages other than English in an English-dominant environment and the challenges that are faced in supporting multilingualism in these contexts.

The Italian *Progetto Lingue 2000* (*PL2000*) is then discussed, which looks at the impact of a 10-year project to improve the English language levels of Italian school-aged pupils. **Roger Hawkey** and **Sarah Ellis** describe the origins of this project, started in 2000, which aimed to encourage communicative language teaching approaches in schools and to introduce international language certification at all levels. Hawkey and Ellis, who conducted an impact study of this project in 2002, then go on to describe how they retraced their steps 10 years later to investigate longer-term effects of the programme. The iterative Cambridge English impact study of *PL2000* involves language learners, teachers, school heads, parents and exam managers in north, central and southern Italy. Using mixed methods and capturing their data collection processes visually with pictures and images, they identify how the project has developed over the last decade. This paper highlights the need to revisit and monitor large-scale projects to investigate their impact over time, which can be used to provide feedback that can lead to well-intentioned improvements or used by other educational authorities to learn lessons for their own language programmes.

Nowadays a growing number of countries are using formal language testing as a requirement for residency and citizenship. The next two papers address this controversial policy decision. Paola Masillo looks at the situation in Italy while Catarina Gaspar, Maria José Grosso and Heinz-Peter Gerhardt focus on Portugal.

The paper by **Paola Masillo** looks at the fairness, validity and reliability of tests used to test migrants' Italian language ability. Although she finds that the language tests and the criteria for marking written interaction are flawed, the main issue is the language policy which has generated both. The paper deals with a problematic issue of setting language requirements for migrants without considering the reality of their language use in society. In addition, she points out that using the CEFR as the framework to evaluate the language ability of migrants creates further issues as it was not designed for

this population. Masillo emphasises the need for language policies and tests which recognise and support the plurilingual context that many migrants find themselves in when they move to a new country.

The next paper deals with migration but from the perspective of the role of non-formal education in supporting immigrants' integration into society. **Catarina Gaspar**, **Maria José Grosso** and **Heinz-Peter Gerhardt** describe a non-formal educational project in Portugal which aimed to address the issue of adult migrants who have had little experience with formal education in their home countries and find it challenging to participate in adult language programmes because of limited literacy and lack of familiarity with the expectations of formal education. If learners are not literate in their L1, the authors wonder how they can be expected to sit a standardised exam which assumes these skills have been developed in the L2. In response to this question, the authors describe the development of a non-formal educational project which aims at attracting learners who traditionally do not enrol in language courses because they lack formal educational experiences and therefore feel uncomfortable in these settings. The paper goes on to investigate whether standardised assessments are appropriate for this population. They conclude, not unlike Masillo's conclusion, that language policies need to take into consideration the reality of migrants and not select a standard that is not achievable for many who have not necessarily had access to formal education. The solution to this issue is not an easy one and the appropriacy of the tests used to test migrants' language ability is a recurring topic at ALTE conferences.

Having looked at the issue of children and migrants who are both groups that are vulnerable in the sense that their language development can easily be affected by the language policies set by governments, the last two papers in this section look at more typical concerns in language assessment, one of which is the identification of standards.

The first paper, by **Beate Zeidler**, focuses on how test developers decide on the characteristics of the minimally competent person (MCP). This is an important question as determining the pass score on a test is often related to this conceptualisation and as Zeidler points out, there is very little agreement about how to best arrive at an understanding of the MCP. This has serious implications for understanding and setting the standard needed to demonstrate competency for a particular purpose. After the author reviews the strengths and weaknesses of different methods to arrive at the MCP, she compares the results of several studies using these methods to try to identify the optimal one. This paper is then followed by **Vivien Berry** and **Barry O'Sullivan** who tackle the same issue of the MCP but for a specific purpose. Their study deals with a topic of current public interest: what English language level should doctors have when seeking to practise medicine in the UK? By leading a standard-setting exercise with healthcare providers, the authors identified the *IELTS* level that would be appropriate for doctors

6

practising in the UK. An important outcome of their research is the insights made by healthcare providers on the language needs of doctors, which were gathered during focus group discussions.

Moving on from the focus on international migration, the next section focuses on teacher competencies and assessment literacy more specifically.

Section 3 deals with aspects of teachers' professional development in relation to assessment in specific contexts. In the opening paper, **Brian North** addresses on profiling teacher competences, describing a study that encompassed multilingual checking of the European Profiling Grid (EPG), which is an online scale containing descriptors of teaching competences across developmental phases. North addresses the questions of whether there is a definable best practice shared between pedagogies for different languages and whether context-free descriptors can be developed for teacher competences in line with the CEFR's generic language learner competences. North describes the validation of the EPG through pilot and main studies, each consisting of a qualitative descriptor sorting and commentary phase together with a quantitative phase. The study's results indicate that the teachers' self-assessments using the EPG helped to produce an effective criterion-referenced measurement scale and that the teachers of the different languages surveyed shared a common view of language instruction, with minor differences by educational sector, as expected. There are, North notes, some disadvantages to using self-assessment data, however the overall impact of the validation study is positive, with the EPG extending its reach and further surveys planned for different contexts. This paper links the EPG with the general aims, design and application of the CEFR, also pertaining to the ALTE conference theme of social mobility for teachers.

Continuing the theme of teacher competences in the broader sense, the following two papers focus on teachers' assessment literacy, which is a combination of all of their competences and a key factor in their ability to engage with assessment both theoretically and in practical ways. The next paper, by **Daniel Xerri** and **Patricia Vella Briffa**, considers teachers' involvement in high-stakes testing in Malta, advocating direct involvement as a route to a better and more equitable testing system for students. The case study reported goes some way towards filling a gap in the assessment literature by showing how teacher involvement in high-stakes test development and delivery can enhance their understanding of assessment and can benefit students by providing a bridge between the local learning context, student cohort and the subject matter, one or more of which, it is argued, may be lacking when teachers are not given the opportunity to be involved in test development activities. This paper affirms the benefits for all stakeholders in direct involvement in assessment and reflects the stages involved in producing test specifications, specimen materials and assessment scales, demonstrating a practical example of assessment literacy happening from grass-roots upwards.

The following paper by **Anthony Manning** reports on a mixed methods study that explored aspects of assessment literacy for English for Academic Purposes (EAP) teachers internationally, based on a desire to enhance EAP teachers' professional assessment practices concerning the admissions of international students into higher education, using theories and principles from language testing and the broader assessment literacy literature. Manning set out to explore whether EAP teachers' views on testing reflect language testing and assessment (LTA) practices around assessment literacy and whether EAP assessment literacy can be sustained or enhanced. Overall, Manning reports a number of examples of good practice in EAP in relation to assessment literacy which reflect theory and research from LTA more widely, adding that there are key areas needing improvement. Manning concludes with a series of actions to be taken including the development of an online framework, bibliography and statistical analysis training for in-service EAP assessment literacy training; provision of collaborative EAP opportunities within and across institutions; and raising awareness of ethics in relation to EAP. This paper shows the way ahead to engaging EAP teachers in higher education which relates to their ongoing training and development and which will enhance their mobility and expertise, and also that of their students who will benefit from their instructors' increased knowledge of assessment for EAP.

The final paper in this section by **Tomoyasu Akiyama** focuses on the concept of test fairness through rater behaviour, looking at prospective English teachers who undertake a microteaching session as part of their teacher employment examinations (TEEs) in Japan. These examinations are high stakes and failing on one occasion means the test taker has to wait another year to retake, which can have negative implications for unsuccessful candidates' employment prospects. Despite this, Akiyama notes the lack of research into 68 regionally produced TEEs, specifically into the micro-teaching sessions, which are 5–10-minute lessons intended for use in teacher training. Akiyama's paper explores whether the ratings given by employers to prospective candidates of this high-stakes test are fair. Using a mixed methods approach to analyse rating data, Akiyama found that raters, whilst consistent, rated candidates with different degrees of severity and seemed to interpret assessment criteria differentially. This study has key implications for rater cognition research and reiterates the importance of piloting rating scales and rater training, as discussed in Xerri and Briffa's earlier paper.

Section 4 of this volume considers the important issues of fairness, quality and validation of assessments across a wide range of contexts in order to ensure that tests are relevant, accessible and fit for purpose. In this section, the first three papers concern language testing in Asia, one of an English test and two of Chinese tests. **Jessica Wu**'s paper provides an overview of the testing of English as a Foreign Language (EFL) in the region, focusing on

8

ensuring that tests of English are designed with quality and fairness in mind and identifying current challenges and opportunities in this area. This paper links to the ALTE conference theme of fairness and quality standards in language test administration. Wu notes that although objective testing started in China, contemporary English language testing in Asia tends to conform to Western testing theories, although EFL tests in Taiwan, Korea, Japan and China are tailored to the educational systems and contexts of test use in these countries. For example, Wu points out that some regionally produced English language tests have up to 18 million test takers per annum. The testing of such large cohorts of test takers provides many challenges to delivering unbiased and high-quality tests, particularly in relation to test design and quality control. The conclusion is that a set of professional standards for test developers and test users is needed for Asia, which take into account context-specific principles along with universal principles of LTA.

The following two papers focus on tests of Chinese, which allows us to explore similarities between testing English and other languages. Firstly, **Xiangdong Gu**, **Yuwen Shen** and **Jian Xu** focus on test takers' mental processes in their reporting of a study that used Think Aloud Protocols (TAPs) to capture Thai test takers' processes when sitting a Chinese vocabulary test, so linking to the ALTE conference theme of language assessment for migration and social integration as well as aiming to better understand test takers' cognitive processes at different proficiency levels. Gu, Shen and Xu focus on the vocabulary component of the reading section of the highest of six levels of the Chinese Proficiency Test (Hanyu Shuiping Kaoshi, HSK), seeking to identify through TAP analysis what level and type of information test takers use to process the reading texts (i.e. text, sentence, extra-textual), their test-taking strategies (i.e. cognitive, metacognitive and test-wiseness) and whether proficiency affects these two aspects of test taking. Their results indicate that clause-level information is the most commonly used by all levels of test taker, with inferring, contextualisation and comprehension monitoring being the most used cognitive and meta-cognitive strategies. Avoidance is the most used test-wiseness strategy, although the study only revealed significant differences in this category amongst higher and lower-scoring groups. This paper suggests implications for teaching and testing reading for Chinese and by extension for other languages, whilst providing an accessible account of aspects of testing reading in a character language.

Next, **Shujiao Wang** considers the broader washback effects of the Chinese Proficiency Test (Hanyu Shuiping Kaoshi, HSK), revised in 2010, on test takers' learning strategies and beliefs, which also links to the ALTE conference theme of language assessment for migration and social integration. This study investigated learner strategies and beliefs using a mixed methods explanatory approach consisting of a survey, interviews and document analysis of related HSK materials. Wang's study reveals that HSK has significant

washback on learning Chinese, with both positive (i.e. increased motivation for learning Chinese) and negative impacts (i.e. encouraging memorisation of grammar rules over communicative strategies) on the learners surveyed and differences in learning strategies between regular learning and test preparation being predicted by items including language proficiency, motivation and nationality. An important feature of this paper is the indication of aspects of the validity and reliability of HSK that need improvement, namely more subjective question forms and emphasising the measurement of the productive skills. Further implications involve the teaching and learning of Chinese as a heritage language and using HSK for admissions purposes to Chinese institutions.

The next group of four papers are concerned with assessing speaking, a key communicative skill that presents particular challenges to teachers and language testers. The first pair of papers relate to improving the quality of teaching and assessing speaking in Spain and Italy. The paper by **Jesús García Laborda** and **Mary Frances Litzler** focuses on improving speaking skills through testing in Spanish schools, reporting on the first year of the international OPENPAU Project which applies low-cost technology to the testing of 15–17-year-old Spanish students of English. This project is part of a movement to address the perceived negative washback effects of language assessment in secondary classrooms which do not include listening or speaking components, for example, which means that these skills are not focused on in the classroom. The authors adopt a social-constructivist approach which emphasises test taker interaction and improved teacher training, both of which should be supported by suitable technology. This new approach to testing speaking and listening uses face-to-face group interviews and also delivery via various devices (mobile phones and tablets) and web videoconferencing technologies such as Skype, which were recommended to the relevant Ministry of Education. The OPENPAU project runs until the end of 2015 and further work is envisaged by the authors to design collaborative tasks for all four skills and to find ways of using social networking in testing. Laborda and Litzler's paper has wider implications for test design and delivery, demonstrating the importance of principled test design that is well suited to its context and purpose, adopts technology where appropriate to do so, and has positive impacts on teaching and learning.

In the following paper, **Giuliana Grego Bolli**, **Jane Lloyd** and **Danilo Rini** discuss a project run by Centro per la Valutazione e le Certificazioni linguistiche (CVCL) at the University for Foreigners in Perugia that aimed to improve the quality of rating of high-stakes Italian speaking assessments through examiner monitoring. This paper relates to the quality management of language test administration which is a key a challenge for all language testing professionals; CVCL's project encompasses rater management and training and reports on quantitative data analysis of the speaking component

10