Section 1
History and background

1 Introduction

Rosemary Wilson
Consultant, Cambridge English Language Assessment
Monica Poulter
Cambridge English Language Assessment

The current emphasis internationally on the importance of English language skills has had a significant impact on educational systems worldwide and, by extension, on the numbers of English language teachers required. This increased demand for teachers has in turn led to a related increase in teacher training programmes of all kinds linked to internationally recognised specifications of English language teacher competencies (Burns and Richards 2009). From the broader educational perspective, large-scale meta-analyses of studies of student achievement in all subject areas (Hattie 2003) highlight the key role that teachers play in student achievement and in particular the part that instructional quality plays in student success. A useful starting point for a discussion of teacher assessment is a related study that aimed to identify the qualities of expert teachers (Hattie 2003). The study drew on evidence from a series of activities designed to quantify the influence of specific attributes on student achievement. The detailed battery of tests included interviews with the teachers before and after each of a series of observed lessons, lesson transcripts, coded lesson observations by pairs of observers providing evidence such as student engagement as measured by time on task, interviews and surveys with students, artefacts such as teaching materials and responses to a number of scenarios about teaching and learning. Each of the pieces of evidence was then coded independently by a team of researchers, who were themselves assessed for inter-rater reliability.

The aim of the study was to identify specific characteristics of excellent teachers rather than to assess their performance but closer examination of the procedures used suggests the scope of evidence that is ideally needed in order to reach a judgement about teacher performance. Lesson observations may seem the most straightforward way to assess teacher performance but judgements will tend to be influenced by the subjective view of the observer. The use of a coding system or protocol promotes objectivity and encourages more standardised responses to teaching events. The presence of two observers also provides a balanced view; where this is not possible for logistical reasons, a comparison of assessments by different observers of a number of lessons taught by a particular teacher can provide a fuller picture of the

teacher's performance. Teaching events are shaped by teachers' beliefs about teaching and learning as well as their thoughts about a particular lesson (Borg 2009), and their teaching will reflect these beliefs to a greater or lesser extent (Basturkmen 2012). The research study described above carried out interviews before and after lessons in order to capture teachers' planning processes before the lesson as well as their reactions to its effectiveness afterwards, while the use of scenarios elicited more evidence about their beliefs. In-class performances were further supported by artefacts in terms of teaching materials illustrating teachers' pedagogical knowledge through the way that information was selected, presented and tested. A more unusual aspect of the study was the inclusion of feedback from students both about particular lessons and teachers' general qualities.

This brief account of an in-depth research study demonstrates the challenges of assessing an activity as complex as teaching but also highlights the need for integrating *multiple sources of evidence* in order to provide a more rounded view of teachers' knowledge and skills. What can be observed is compared to 'the tip of an iceberg' by Turner-Bisset (2001:xii) in her study of the knowledge bases underpinning expert teaching: 'Under the surface of a seemingly effortless act of teaching is the other nine-tenths of the iceberg: a wealth of different kinds of knowledge on which the teacher has drawn for that particular teaching performance.' In the context of any programmes which provide formal, often high-stakes certification, a valid and reliable teacher assessment procedure needs standardised methods of collecting and collating evidence. A volume of essays documenting the development of an assessment framework for the National Board for Professional Teaching Standards for the US educational system describes the complexities of collecting and assessing multiple sources of evidence (Ingvarson and Hattie (Eds) 2008). The assessment was complex as it needed to demonstrate achievement of national standards; it soon became clear that rigorous training was needed for assessors tasked with reviewing the evidence and assigning scores to define what constituted 'accomplished teaching'. In her contribution to the Ingvarson and Hattie volume, Pearlman (2008:181) notes that 'no-one had thought about what might be the differences between training (of assessors) for the purposes of learning and development, and training for the purposes of legally defensible, operationally feasible scoring'; she describes the considerable challenges that were encountered in developing a reliable scoring system, with the need to develop manageable analytical criteria and holistic rubrics, and the difficulties of training a sufficient number of assessors to review the multiple sources of evidence without assessor bias or preference based on their own experience of teaching and teaching methods. While the collection of multiple sources of evidence may be seen as a solution to 'the tip of the iceberg' issue, the lengthy and detailed processes and procedures described in the volume illustrate that the multiple sources of evidence

approach complexifies the assessment process and, without intensive training and monitoring, may increase rather than solve problems of ensuring reliability. The sources of evidence selected are perhaps best described as 'a union of insufficiencies' (Schulman 1988) designed to compensate for the inherent shortcomings in each of them. In this context of both theoretical complexity and practical limitations, this volume aims to illustrate one approach to teacher assessment in English language teaching. In doing so, we hope to contribute to the field of language teacher development and assessment by initiating a discussion on a topic of fundamental importance, but limited academic discussion.

The Cambridge English Teaching Qualifications encompass a series of certificates and diplomas designed for English language teachers at different stages of their professional careers. The development of the different qualifications is described in an account of their history in this volume by Alan Pulverness (Chapter 2), who makes clear that they had their origins in two specific teaching contexts in which the medium of instruction was of necessity English: multilingual classes in the UK and monolingual classes in other countries where the English-speaking teacher did not speak the students' language. It would be naive to ignore the fact that the majority of English language teaching worldwide is carried out by teachers who share their students' language and who may use that language almost exclusively as a medium of instruction. The renewed interest in bilingual teaching (Cook 2010) as well as in the use of specific translation activities in English-language classes (Kerr 2014) should also be considered. However, the historical context of the development of the qualifications is undoubtedly a key factor in their continued emphasis on using the language, in this case English, to teach the language.

Linked to the role of English as the medium of instruction, the account of the history of the qualifications makes clear that the first qualifications were aimed at prospective teachers whose first language was English and therefore competence in the language was assumed. The changing nature of applicants for the qualifications is demonstrated by the case studies in this volume by Peter Watkins, Bill Harris and Alan Pulverness (Chapter 14). It reflects the changing nature of English language users worldwide and raises the question of the level of linguistic competence required to teach a language effectively, as discussed by Jenny Johnson and Monica Poulter (Chapter 8). Assessment frameworks that are used internationally such as those in the Cambridge English Teaching Qualifications need to apply the same criteria to teachers and teachers-in-training from a wide range of language backgrounds while at the same time acknowledging the different strengths that individuals bring to their teaching. The interplay of the influence of linguistic competence, awareness of the target language and teaching knowledge and skills, as well as interpersonal qualities, reflects the holistic nature of language teacher assessment central to the discussion throughout this volume. Although all

the assessment protocols in the Cambridge English Teaching Qualifications provide a list of analytical criteria, usually grouped together in domains or categories, the inter-relationship between criteria and categories is apparent. The discussion of language awareness by Martin Parrott (Chapter 9) illustrates how the teacher's language awareness underpins most decisions that a teacher makes from planning through to teaching to reflecting on the lesson.

The study of expert teachers described previously highlighted the need not only for multiple sources of evidence but also for procedures for reliability of assessment. When the source of evidence is ephemeral, as in the case of a teaching event, the risk of subjectivity on the part of the observer is clearly a factor. The approach to standardisation of assessment in the Cambridge English Teaching Qualifications is addressed in detail by Evelina Galaczi and Marie Therese Swabey (Chapter 6). When observations are scheduled as part of a training programme, an observer familiar with the teacher-in-training may tend to base the assessment as much on their progress as on their achievement as well as other interpersonal factors. The inherent tensions between training, development and assessment are explored from different perspectives by Simon Borg and David Albery (Chapter 3) and by Simon Phipps (Chapter 15). Despite these tensions, in a standards-based era, the need for formal summative assessment has taken on greater importance in training programmes and requires procedures which do not always sit easily with the development process, for example the need to advise a teacher-in-training that a lesson is below the required standard. Indeed, anecdotal evidence suggests that such statements often begin 'With my assessor hat on', thus making a clear distinction between the roles of trainer and assessor. Transparency of assessment procedures is also increasingly demanded by the teachers themselves, in many cases in the interest of achieving their full potential but in others in order to play 'the rules of the game'; for example, choosing to teach more straightforward content for assessment purposes rather than more challenging topics. The decisions that teachers make at the preparation and planning stages of lessons can also make them more or less equipped to address any unanticipated interventions in a lesson, as discussed by Rosemary Wilson (Chapter 10).

A number of chapters in this volume refer to the complexity of the assessment process and draw attention to the unequal relationship between the parties involved in the assessment as well as the inter-personal factors which may impact on the behaviour of the students, the teacher being assessed and the observer making an assessment of the lesson. To ascertain teachers' impact on student achievement, evidence is needed of learning or progression in relation to the stated goals. Disengaged learners working through under-challenging tasks are evidence of ineffective teaching performance. However, it may be the case that students behave in supportive ways and co-operate with the teacher, while the teacher's behaviour may be influenced by what

they think the observer expects. Observers may also be influenced by their own beliefs about teaching, which may conflict with the teacher's beliefs and learner expectations about what constitutes good teaching. The complicated interpersonal dynamics of the teaching/training room are highlighted by Jo-Ann Delaney (Chapter 5) and David M Palfreyman (Chapter 16), while Marie Morgan (Chapter 7) discusses an approach to training and standardisation procedures for English language teacher trainers.

The point has been previously made of the need for multiple sources of evidence in order to provide a rounded picture of teachers' knowledge and skills. One way of collecting and collating evidence is through the use of portfolios, in which teachers can include narrative accounts of their planning process, examples of artefacts such as teaching materials or samples of student work and comments on the effectiveness of lessons taught. Their use for different purposes and their specific content and role in the Cambridge English Teaching Qualifications assessment procedures are discussed by Neil Anderson (Chapter 4). The portfolios in question are highly detailed documents that are part of a resource-intensive teacher development programme. For other contexts with limited resources, the design of methods of assessment needs to be fit for purpose. The discussion of the development of the *Teaching Knowledge Test* (*TKT*) by Mary Spratt (Chapter 11) and the related test of Content and Language Integrated Learning (CLIL) described by Kay Bentley (Chapter 13) illustrate an approach to assessing teachers' personal practical knowledge through paper-based tests when resources are limited. The impact of the test in one particular country, as described by Gerardo Valazza (Chapter 12) indicates the role that formal assessment can play in promoting professional development.

This volume was proposed by the Series Editors of *Studies in Language Testing* and is the first to address assessment in the Cambridge English Teaching Qualifications. Indeed, it is one of the few volumes with a focus on English language teacher *assessment* as opposed to education or development; assessment seems to be the elephant in the room in language teacher education. Rather than issuing a general call for papers, the editors approached potential authors with substantial experience of the different qualifications and who were preferably actively involved as trainers and/or assessors. This decision reflects the complex nature of the qualifications in terms of approach, procedures and terminology. Readers unfamiliar with the Cambridge English Teaching Qualifications will find full details of each qualification on the website (www.cambridgeenglish.org) but may find it more relevant to consider the principles of assessment involved than the precise details of the procedures. The format proposed for each of the chapters is a discussion of key issues involved in a particular aspect of language teacher assessment supported by an account of how those issues are addressed in one or more of the qualifications. Most of the chapters include

small-scale qualitative investigations that draw on assessment reports, evidence from assessed portfolios, examination papers or a combination of these. One aim was to draw authors from different countries in order to reflect the international nature of the qualifications. That was partially successful in that a range of different contexts are described in several chapters but one regret shared by the editors is that most of the authors are first language speakers of English, a fact that does not reflect the composition of many course teams working with the Cambridge English Teaching Qualifications as well as the cohort of assessors. The organising principle for the volume developed along with the chapters themselves. The final decision was to organise the sections around the themes that linked the various chapters: development and assessment; the assessment of specific criteria; assessment in context. To foreground these themes, each section is introduced with an overview of the chapters within it. Organising the chapters into themes served to highlight the inter-related nature of the subject matter: is a chapter that discusses self- and peer-assessment linked to specific criteria more about development or more about specific criteria? In almost every case, each chapter could fit into more than one section and indeed the sections themselves could be reconfigured and renamed. It was assumed that most readers will select chapters at random and so key factual material about the relevant qualifications is included in each chapter as appropriate, despite the resulting repetition.

This volume has explored and discussed a limited number of themes drawing on descriptions and data from the Cambridge English Teaching Qualifications. The editors are aware of gaps both in the range of topics explored and the contexts discussed. We hope that these chapters will prompt others to identify additional areas for research or to undertake further research into some of the areas explored in this volume in relation to their own teaching context and assessment frameworks.

References

Basturkmen, H (2012) Review of research into the correspondence between language teachers' stated beliefs and practices, *System* 40 (2), 282–295.

Borg, S (2009) Language teacher cognition, in Burns, A and Richards, J C (Eds) *The Cambridge Guide to Second Language Teacher Education*, Cambridge: Cambridge University Press, 153–171.

Burns, A and Richards, J C (2009) Second language teacher education, in Burns, A and Richards, J C (Eds) *The Cambridge Guide to Second Language Teacher Education*, Cambridge: Cambridge University Press, 1–8.

Cook, V (2010) *Translation in Language Teaching*, Oxford University Press.

Hattie, J (2003) *Teachers make a difference: What is the research evidence?,* paper presented at the Australian Council for Educational Research, October 2003, available online: cdn.auckland.ac.nz/assets/education/hattie/docs/teachers-make-a-difference-ACER-(2003).pdf

Ingvarson, L and Hattie, J (Eds) (2008) *Assessing Teachers for Professional Certification: The First Decade of the National Board for Professional Teaching Standards, Advances in Program Evaluation*, Amsterdam: Elsevier Press.

Kerr, P (2014) *Translation and Own-language Activities*, Cambridge: Cambridge University Press.

Pearlman, M (2008) The evolution of the scoring system for NBPTS assessments, in Ingvarson, L and Hattie, J (Eds) *Assessing Teachers for Professional Certification: The First Decade of the National Board for Professional Teaching Standards, Advances in Program Evaluation*, Amsterdam: Elsevier Press, 177–209.

Schulman, L S (1988) A union of insufficiencies: Strategies for teacher assessment in a period of educational reform, *Educational Leadership* 46 (3), 36–41.

Turner-Bisset, R (2001) *Expert Teaching, Knowledge and Pedagogy to Lead the Profession*, London: David Fulton Publishers Ltd.

Preface to Chapter 2

This volume presented an ideal opportunity to present a detailed account of the development of the Cambridge English Teaching Qualifications. The account of the history of the Cambridge English Teaching Qualifications by Alan Pulverness (Chapter 2) owes much to the paper by Hazel Orchard, former Deputy Director at the Royal Society of Arts (RSA), whose powers of recall of the minutiae of the decision-making process are legendary. Pulverness accessed the archives at Cambridge Assessment to trawl through documentation as well as interviewing key stakeholders and has pulled the somewhat motley sources together into a coherent story. The word 'story' is used advisedly because the theme that emerges strongly from the chapter is of the vision and energy of committed individuals. Pulverness begins the chapter with a quote from the late John Haycraft, the founder of International House and the originator of the intensive, classroom-based training courses from which the Cambridge English Teaching Qualifications have developed. Assessment was not formalised in those early days but Haycraft's 'frank grades' of 'Outstanding, Good, Moderate and Below Average' gave clear feedback to trainees as well as guidance to prospective employers. The foundations for more formal assessment were laid with the introduction of the RSA Certificate in Teaching English as a Second or Foreign Language, a notoriously difficult test that created the myth in English language teaching circles that it was impossible to pass 'The RSA' but just as impossible to fail a Post-Graduate Certificate in Education (PGCE). Pulverness refers to the Chief Examiner's report on the first examination for the RSA Certificate in 1967, with one of the main reasons for failure being 'Clarity and Limitation of Aims' and notes that achievement of aims remains a key issue. Assessment procedures needed to change as the number of candidates, courses and institutions offering the courses continued to grow. Pulverness also notes the more recent changes that technology has made to the process of standardisation throughout the Cambridge English Teaching Qualifications, enabling tutors and assessors anywhere in the world to take part in online training and updating.