

## Appendix

# Language Testing Research Centre 20<sup>th</sup> Anniversary Conference



THE UNIVERSITY OF  
MELBOURNE

**The University of Melbourne  
July 15<sup>th</sup> – 16<sup>th</sup>, 2010**

## Sponsors



FACULTY OF  
ARTS



UNIVERSITY of CAMBRIDGE  
ESOL Examinations

PEARSON



Australian Council for Educational Research



UniSA

Research Centre for  
Languages  
and Cultures

## **Workshops**

### **Workshop 1: Rasch measurement ... the very basics**

Tim McNamara  
The University of Melbourne

Thursday, 9:00am – 12:00pm  
Alice Hoy 242

This workshop will introduce fundamental concepts and procedures in Rasch measurement to those who are relatively new to Rasch measurement, or those who would like to refresh their understanding. Topics covered will include: how Rasch modelling relates candidate ability and the demands of items; the use of probabilistic modelling to generate general estimates based on samples of performance; how Rasch programs handle different data types, such as objectively and subjectively scored responses; estimating the impact on scores of task and rater characteristics; assumptions underlying the Rasch model and the question of dimensionality; validation research using Rasch measurement.

*Pre-reading:*

Participants are encouraged to read the relevant sections of McNamara, T.F. (1996) *Measuring second language performance*. London and New York: Longman.

*Presenter:*

Tim McNamara is a leading authority on the use of Rasch measurement in language testing research. He is a former director of the Language Testing Research Centre at The University of Melbourne, where he is currently Professor in the School of Languages and Linguistics. He has played a key role in the development of major international language tests such as IELTS, TOEFL iBT and the Occupational English Test (OET).

### **Workshop 2: Principles and practice of task design: listening assessment**

Luke Harding  
The University of Melbourne

Thursday, 9:00am – 12:00pm  
Alice Hoy 325

This workshop will set out the key principles of language assessment and their relationship to good task design, with a specific focus on the assessment of listening in an English for academic purposes context. The workshop will involve group discussion activities as well as hands-on task design and feedback. It is designed for language teachers and those involved in test development projects, but may also be of interest to researchers of listening comprehension or language testing more generally.

*Presenter:*

Luke Harding has worked as a test developer and researcher in the Language Testing Research Centre since 2006. He has been involved in a wide range of test development projects including the Occupational English Test (OET), the Diagnostic English Language Assessment (DELA), the University Test of English as a Second Language (UTESL), and the Oxford English Language Test (OELT). He has a particular interest in the assessment of listening ability.

**Workshop 3: Standards-based classroom assessment: Principles and practice**

Geoff Brindley  
Macquarie University

Thursday, 9:00am – 12:00pm  
Alice Hoy 225

This workshop will look at recent developments in standards-based assessment in a variety of language learning contexts. First, examples of some influential national curriculum and assessment frameworks will be examined and their implications for classroom assessment will be discussed. Next, ways of using statements of learning outcomes as a basis for creating assessment tools will be explored and illustrated.

*Presenter:*

Geoff Brindley is an Honorary Associate Professor in the Department of Linguistics at Macquarie University. He has written and edited a wide range of publications on language assessment, second language learning and language curriculum development. He has been involved in a number of test development projects and has also served as Chief Examiner (Australia) for IELTS and as a member of the TOEFL Committee of Examiners.

## **Official Welcome**

### **Reflections on 20 years of the Language Testing Research Centre**

Alan Davies  
The University of Edinburgh

*With responses from:*

Cathie Elder and Tim McNamara  
The University of Melbourne

Thursday, 4:45pm  
University House

**ABSTRACT:** Joe Lo Bianco's National Policy on Languages provided the argument for the establishment of the Language Testing Unit (later LTRC) at the University of Melbourne. In the paper I reflect on the beginnings of the LTU, our reception by the language departments reflecting different views of language, the key issues in language testing at that time and their influence on our early projects and our approach to staff training in the LTU.

**BIO STATEMENT:** Alan Davies was founder director of the Language Testing Research Centre. A former editor of the journals *Applied Linguistics* and *Language Testing*, he has taught and researched English and Applied Linguistics in Kenya, Nepal, Hong Kong and Australia as well as the UK. He is now Emeritus Professor of Applied Linguistics at the University of Edinburgh and is working on a book for CUP, *Native Speakers and Native Users*.

## Plenary

### **Reconciling construct validity with standards, frameworks, and tasks**

Micheline Chalhoub-Deville  
University of North Carolina at Greensboro

Friday, 9:00am  
Old Arts, Theatre D

**ABSTRACT:** In the presentation I explore the two interrelated issues of construct and validity. In terms of the construct, I discuss how language testers have defined and operationalized language proficiency predominantly as a psychological, cognitive construct internal to the test taker and make the argument that the social, interactional nature of language deserves increased attention. In terms of validity, it seems that the centrality of construct validity is in question once again. Renewed deliberations of the role of construct validity appeared in the articles published in the 2007 special issue of *Educational Researcher*, edited by Lissitz, and the related conference, organized by Lissitz, at the University of Maryland (Ninth Annual Maryland Assessment Conference, October, 2008). In the presentation, I explore the issues associated with content validity by focusing on standards, frameworks, and tasks. The present paper discusses the standards used for the development of tests for nonnative speakers of English in schools and frameworks commonly relied upon to develop foreign language tests. Additionally, the paper introduces critical arguments raised in explorations of context that have relevance for the discussion of tasks as a domain from which to sample for test development and validation.

**BIO STATEMENT:** Micheline Chalhoub-Deville has a Ph.D. in Foreign and Second Language Studies. In the field of language testing, she has presented and published widely on numerous topics, including: validity, socio-cultural context, raters' perceptions and behaviors, task-based assessment, computerized testing, and Title III and NCLB assessment practices in the U.S., among others.

## **Symposium 1**

### **Beyond the test score: The use of qualitative approaches in test validation research at the LTRC**

Cathie Elder (Chair), Luke Harding, Kellie Frost, Susy Macqueen & Ute Knoch, John Pill

Thursday, 1:00pm  
Old Arts, Theatre D

**OVERVIEW:** While test validation has traditionally been associated with quantitative analysis, the field has now embraced a broader range of methodologies including qualitative and mixed method approaches (Lumley & Brown, 2005; Xi 2009). In this symposium we illustrate a number of such approaches with reference to recent projects conducted by LTRC staff. These projects address validity issues pertinent to a range of different tests including the University Test of English as a Second Language (UTESL), the Oxford English Language Test (OELT) and the Occupational English Test for overseas trained health professionals (OET). The methodologies covered include stimulated recall of test taking processes, discourse analysis of test taker production on both spoken and written tasks and a content analysis of feedback on spoken performance. These diverse methodologies are used to address a range of validity issues, including the question of whether the accent of a speaker makes a difference to performance on test of listening comprehension (Harding), whether high and low scores on an integrated (listening then speaking) task reflect meaningful differences in the quality of test takers' speech (Frost), whether a lexicogrammatical measure can be developed to gauge construct variation across tasks (Macqueen & Knoch) and whether the linguistic criteria used to assess the communication skills on a health-specific English test are aligned to what doctors consider crucial for effective clinical interaction (Pill). For each study we show how qualitative data (in conjunction with quantitative analysis in some cases) can be used to either support or contest validity claims about the tests in question.

### **Accent-related difficulty and attitudinal response on an EAP listening test featuring L2 speakers: Evidence from verbal reports**

Luke Harding  
The University of Melbourne

**ABSTRACT:** Given the linguistically diverse nature of many academic institutions in English-speaking contexts, a strong rationale exists for the incorporation of L2 accents of English in academic listening assessment on the grounds of greater authenticity, enhanced construct representation, and the potential for positive washback. However language testing has tended to take an orthodox, native-speaker approach to accents which may be understood as stemming from concerns over intelligibility, threats to construct validity and acceptability among stakeholders. Although research has begun to address these issues (e.g., Major et al., 2002; 2005), there have been no studies to date which have investigated the extent to which test-takers experience accent-related difficulty during a listening test, or the nature of test-takers' "online" attitudes towards speakers with L2 accents. This paper reports on a study which used verbal report methods to collect data of how test-takers experience L2 accented speech during an English for Academic Purposes (EAP) listening test.

Specifically, the study investigated the degree, and nature, of accent-related difficulty experienced by listeners, and the degree, and nature, of attitudinal response towards the speakers. Findings from the study are discussed with respect to the interaction of cognitive and social factors in the comprehension of accented speech, and implications are drawn for the use of L2 accents in EAP listening assessment.

### **Investigating the validity of an integrated speaking task: A discourse-based analysis of test-takers' oral summary performances**

Kellie Frost

The University of Melbourne

**ABSTRACT:** Performance on integrated tasks requires candidates to engage skills and strategies beyond language proficiency alone, in ways that can be difficult to define and measure for testing purposes. While it has been widely recognised that stimulus materials impact test performance, our understanding of the way in which test-takers make use of these materials in their responses, particularly in the context of listening-speaking tasks, remains predominantly intuitive, with little or no base in empirical evidence. Using discourse data from a trial test administration of the new Oxford English Language Test, this paper investigates the way in which test takers integrate stimulus materials into their speaking performances on an integrated listening then speaking summary task, and examines if test scores reflect real differences in the quality of oral summaries produced.

### **Writing the wrongs: The detection of lexicogrammatical patterning in L2 written discourse**

Susy Macqueen & Ute Knoch

The University of Melbourne

**ABSTRACT:** The language of highly proficient users comprises recurring combinations of words such as idioms and collocations ('lexicogrammatical patterns'). There is a growing body of linguistic research and theorizing on the role of lexicogrammatical patterns, with some theorists contending that patterning is fundamental and pervasive in language use (e.g. Hoey, 2005). Thus, the extent to which L2 writers are able to use and manipulate native-like chunks of language is an important developmental question and potentially revealing in terms of both proficiency level and genre. However, operationalization in research is hampered by the difficulty of defining the construct and the range of pattern types (Wray, 2002) as well as methodological difficulties in detecting patterns. In order to compare the discourse produced at different score levels in both the integrated and independent writing tasks of an academic English test, we developed a method of analysing the lexicogrammatical patterning of written L2 discourse. This paper will explore the issues and challenges that arose in the development process.

### **Medical professionals' criteria for effective communication**

John Pill

The University of Melbourne

**ABSTRACT:** Overseas trained health professionals play an important role in filling health workforce shortages in Australia and elsewhere. Most registration authorities require language proficiency screening prior to registration. However, such screening tends to be undertaken by language specialists using traditional linguistic criteria; it is rarely from the perspective of health professionals

assessing whether the candidates possess the communicative competence to deliver safe and effective health care in a new cultural setting. This raises questions about whose standards and measures should be employed to screen for language proficiency and communicative competence.

This project aims to identify criteria for effective communication as employed in health professional assessment and reflected in supervisor feedback to trainees in order to review and validate the criteria used to assess the Speaking sub-test of the Occupational English Test (OET). Examples of this feedback and what it reveals about the professionals' views of what matters for effective communication will be presented, using data collected in a workshop with medical educators. The study's initial findings show discrepancies between the currently used linguistic criteria and 'indigenous criteria' (Jacoby and McNamara, 1999) which emerge implicitly from the supervisors' responses to trainees' performance.

**BIO STATEMENTS:** Cathie Elder is Associate Professor of Applied Linguistics and Director of the Language Testing Research Centre at the University of Melbourne. She is a former Chair of the TOEFL Committee of Examiners and currently co-editor of the journal *Language Testing*. Her research interests and publications are in the areas of language testing, language program evaluation and bilingualism. Over the years Cathie has been involved in a broad range of assessment related consultancies for government and international agencies. Amongst her co-authored publications are the *Dictionary of Language Testing* (CUP, 1999) and *Implicit and Explicit Knowledge in Second Language Learning Testing and Teaching* (Multilingual Matters, 2009). She is also co-editor (with Alan Davies) of the *Handbook of Applied Linguistics* (Blackwell, 2004).

Luke Harding is currently a research fellow in the Language Testing Research Centre at the University of Melbourne. He completed his PhD in 2008 with a thesis on the use of speakers with L2 accents in academic English listening assessment. His research interests are in listening assessment, accent and intelligibility, attitudes towards varieties of English, and assessor decision making.

Kellie Frost completed a BA (Hons) in French studies at the University of Melbourne in 2008. She recently started working in the Language Testing Research Centre, and is nearing completion of a Masters in Applied Linguistics, specialising in Language Testing.

Susy Macqueen joined the Language Testing Research Centre as a Research Fellow in 2010. Her doctoral research was on the development of lexicogrammatical patterning (e.g. collocations) in second language writing. Her other academic interests are communication in healthcare contexts and approaching research via Complex Systems Theory and Sociocultural theory.

Ute Knoch is a research fellow in the Language Testing Research Centre, University of Melbourne. Her research interests are in the areas of second language writing, writing assessment, rater training and assessing languages for specific purposes.

John Pill is in the first year of his PhD candidature working on a three-year ARC-funded project 'Towards improved healthcare communication: Development and validation of language proficiency standards for non-native English speaking health professionals'. John has worked as an English language teacher and teacher trainer in Italy, the UK, Ireland, Hong Kong and Australia. Before returning to study, he was the assessment manager for the Occupational English Test.

## **Symposium 2**

### **Educational reform: The roles of assessment**

Tim McNamara (Chair), Chris Davison, Geoff Brindley, Angela Scarino, Tom Lumley & Helen Moore (discussant)

Friday, 3:35pm  
Old Arts, Theatre D

**OVERVIEW:** This symposium considers the varying roles of assessment in educational reform. Assessment will be considered as

- a target of reform, in the case of initiatives to support teacher-based assessment;
- a tool of reform, via the mechanism of outcomes-based assessment tied to curriculum frameworks;
- a means of evaluating reform, where tests are used as part of evidence gathering on the effectiveness of an educational reform program.

Case studies from Australia, Brunei, Hong Kong, Singapore and Mexico will be used to support the arguments being made about the complex and varied roles of assessment in each case of educational reform. Implications will be drawn about the appropriateness and effectiveness of the role of assessment in each setting.

### **Assessment for Learning in English Language Teaching: Reconciling rhetoric and reality**

Chris Davison  
The University of New South Wales

**ABSTRACT:** Assessment for learning (AfL), a concept first used in the UK in the late 1980s, and widely promoted through the work of the Assessment Reform Group (Assessment Reform Group, 1999, 2001; Black & Wiliam, 1998), commonly refers to any assessment in which the primary purpose of the information being collected is to improve learning. Black & Wiliam (1998) convincingly demonstrate the learning gains that can be achieved through well-focused teacher-based formative assessment, claiming no other strategy has such potential for enhancing student learning across age levels and in different contexts. Increasingly, English language teachers are being called upon to plan and/or implement appropriate assessment procedures to monitor and evaluate student progress in their own classrooms, including designing and implementing their own classroom-based assessment tasks, actively engaging learners in self and peer assessment and incorporating critical but constructive feedback into the assessment cycle (Davison & Leung, 2009).

However, somewhat paradoxically, AfL has been embraced as the centrepiece of assessment reform in a range of traditional examination cultures, including Brunei, Hong Kong and Singapore, hence creating many conflicts between rhetoric and reality. Drawing on data from a range of questionnaires, interviews with key stakeholders and teachers, and the analysis of policy documents collected as part of a number of recent research and development studies undertaken by the author in Hong Kong, Singapore, and Brunei, this presentation will identify, describe and critically evaluate

the key assumptions underlying the top-down drive to implement assessment for learning in English language teaching, and explore some of the common misunderstandings and conceptual confusions that arise within educational communities when policy imperatives collide with practical realities. The implications of the findings for assessment reform and professional development will also be discussed.

### **Outcomes-based assessment and educational reform**

Geoff Brindley  
Macquarie University

**ABSTRACT:** The paper will look at recent examples of educational reform which have involved the introduction of outcomes-based curriculum and assessment systems and/or frameworks. The influence of social, political and educational factors on these reforms will be examined with reference to a number of different language learning contexts.

### **Curriculum and assessment frameworks for languages in Australia: The power of assessment without assessing**

Angela Scarino  
University of South Australia

**ABSTRACT:** In this paper I focus on the impact of the state and national curriculum and assessment frameworks developed for languages (other than English) in Australian school education. Despite the fact that the assessment dimensions of these frameworks have never been formally and systemically implemented, their discourse has shaped in powerful ways teachers' understandings of language assessment. It is this understanding, in turn, that influences their assessment practices and the possibilities and limits of reform. I provide a brief critical account of the 30 year history of framework development for languages to highlight the changing discourses of curriculum and assessment and the consequences for languages education. Data from a series of projects focussed on teacher assessment illustrates the influence.

### **Assessment and educational reform: The case of Mexico**

Tom Lumley  
Australian Council for Educational Research

**ABSTRACT:** There have been three major rounds of educational reform in Mexico since 1992. The most recent round of reform has been the gradual introduction, since 2006, of a new curriculum in junior secondary school (years 7 – 9). As in other countries, the consistent trend has also been towards an increasing amount of centralised assessment. This paper describes some of the reasons behind the use of assessment in secondary schools, and will examine the role in reform of three standardised assessments of reading (in Spanish): PISA, a sample assessment of reading, mathematics and science, administered to 15-year-olds every three years; ENLACE, Mexico's population assessment of Spanish and mathematics, administered annually at Year 3 and every subsequent year of compulsory education; and SEP-ISA, a sample assessment of reading and mathematics provided by ACER modelled on PISA, which aims to fill a monitoring function as well as contribute to an evaluation of the junior secondary curriculum reform that is currently under way.

Some of the challenges of this relationship, in this federation of 111 million people, where education is valued very highly, will also be considered.

**BIO STATEMENTS:** Tim McNamara is Professor in the School of Languages and Linguistics at the University of Melbourne and a former director of the Language Testing Research Centre. His recent research has focused on the policy and social contexts of language testing, particularly in relation to language testing for citizenship and the use of language analysis in the determination of origin of asylum seekers. He is the author (with Carsten Roever) of *Language testing: The social dimension* (Blackwell, 2006).

Chris Davison is Professor of Education and Head of the School of Education at the University of New South Wales. She was previously Associate Dean (Research) in the Faculty of Education at the University of Hong Kong, where she remains an Honorary Professor. Before going to Hong Kong in 1999, she worked in teacher education at the University of Melbourne and La Trobe University for fifteen years. Chris has researched and published extensively on the interface between English as a mother tongue and ESL development, integrating language and content curriculum, and English language assessment. With colleagues at the University of Hong Kong, she has just completed the research and development of a range of oral school-based assessment initiatives for the Hong Kong Examinations and Assessment Authority, and has also been working with the Ministries of Education in Singapore and in Brunei on integrating school-based assessment and assessment for learning into their new curricula.

Geoff Brindley is an Honorary Associate Professor in the Department of Linguistics at Macquarie University. He has written and edited a wide range of publications on language assessment, second language learning and language curriculum development and has also served as Chief Examiner (Australia) for IELTS and as a member of the TOEFL Committee of Examiners.

Angela Scarino is an Associate Professor of Applied Linguistics and Director of the Research Centre for Languages and Cultures at the University of South Australia. Her research is in the area of languages education, language and culture in education and language assessment. She has conducted a broad range of language related projects for government and other agencies in Australia as well as in Hong Kong, Singapore, Malaysia and New Zealand. She is the immediate past president of the Applied Linguistics Association of Australia.

Tom Lumley has been a Senior Research Fellow at ACER since 2004. During this period Tom has played a major role in the development of the reading materials for the 2009 cycle of PISA, for both the electronic reading assessment (ERA) and print versions of the test. He has delivered workshops on reading item development, and marker training for reading and writing assessments, in Australia, Europe, Mexico and Colombia. He has recently also directed development of the reading component of an online computer-adaptive assessment tool of literacy and numeracy for adults in New Zealand. Before joining ACER, Tom benefitted from being one of the earliest members of the LTRC, and returned there for a couple more years after a five-year stint at the Hong Kong Polytechnic University.

Helen Moore worked in TESOL teacher education for 17 years and has researched school and adult ESL policy issues in Australia. She is currently Vice President of the Australian Council of TESOL Associations (ACTA) and the Applied Linguistics Association of Australia (ALAA). She has recently been involved in advocacy work for school and adult ESL, this year drafting submissions for ACTA, ALAA and ALS (the Australian Linguistic Association) on Indigenous Education, Teacher Standards and the impact of NAPLAN testing on English language learners in schools.

## **Individual Papers**

**(In alphabetical order by presenter)**

### **Practical issues of online assessment**

Carol **Aeschliman** & Kathleen **Mendan**  
Swinburne College

Friday, 9:50am (Parallel session D)

Economics & Commerce, Theatre 3 (Language assessment and technologies)

**ABSTRACT:** As technology advances, online assessment is becoming increasingly popular. With the automation of routine administrative tasks, particularly in regard to marking, computer-based tests appear to have many advantages over the paper-based format. Online tests are, overall, user-friendly, relatively inexpensive and efficient. However, because of the differences in the administration and marking processes between paper-based and online assessments, ensuring reliability and validity is a particular challenge. Swinburne College has recently undertaken a process of developing a number of online tests for placement and academic purposes. This presentation will outline some of the issues we have experienced, as well as discussing the implications and outcomes both for Swinburne College and for other educational institutions wishing to design and develop online assessments.

**PRESENTERS' BIO STATEMENTS:** With many years experience teaching ESL and LOTE at both the secondary and tertiary levels in Australia, Germany and Hong Kong, Carol Aeschliman is the e-Learning Leader and Coordinator of the Independent Learning Centre at Swinburne College. She has a special interest in online learning, flexible delivery and e-assessment.

Kathleen Mendan's role at Swinburne has included a combination of College activities. These have included development of the online placement test and online listening test, creation of e-learning materials and resources and helping to further improve College processes with technology. Kathleen is also currently completing her PhD at Monash University.

## 21<sup>st</sup> Century accountability and assessment: Where is the data on language?

Denise Angelo & Sophie McIntosh

Far North Queensland Indigenous Schooling Support Unit & University of Sydney

Thursday, 3:05pm (Parallel session B)

Economics & Commerce, Theatre 1 (Accountability and assessment)

**ABSTRACT:** The educational agenda in Australia is becoming increasingly data-driven. This paper deconstructs the current trend towards obtaining statistical 'evidence', before problematising both the focus and the means of current data collection. In the Queensland context, the lack of awareness about Indigenous language situations has resulted in limited - and often inaccurate - information being documented about students' language backgrounds. This is an issue at: the local school-based level, such as enrolment forms; in state level educational contexts, such as ESL funding eligibility; as well as national non-educational contexts, such as the Australian Bureau of Statistics census data. Where students' language backgrounds are not acknowledged, their language proficiency in Standard Australian English (SAE) will also not be flagged. Many Indigenous students are *hidden* as learners of English as a Second/Subsequent Language (ESL), so their performance in standardised tests cannot be disaggregated from that of students with first language (L1) proficiency in SAE, including Indigenous L1 SAE speakers.

This paper highlights the need for national and state system responses. Collection of data about students' SAE proficiency needs to be mandated. Such a measure would enable students' assessment results to be interpreted more accurately (allowing overall trends for ESL learners to become apparent) and more appropriately responded to. Authentic information about Indigenous students' second/subsequent language acquisition of SAE can be collected and described by assigning an ESL Bandscale level via the *ESL Bandscales for Aboriginal and Torres Strait Islander Learners* (Department of Education, Queensland, 2002). Valid ESL data, such as this, will provide the opportunity for standardised performance data to be more insightfully analysed, as well as enabling educators to support students more effectively in their acquisition of SAE and its application to wider school-based learning.

**PRESENTERS' BIO STATEMENTS:** Ms Denise Angelo is the Language Perspectives manager of the Far North Queensland Indigenous Schooling Support Unit. She has worked with Indigenous ESL learners, and their teachers, throughout Queensland and the Northern Territory. In her current role, with the Far North Queensland Indigenous Schooling Support Unit, she undertakes ESL research projects, delivers professional development workshops (i.e. ESL EsSentialS), and supports teachers through the Adopt-a-School program.

Miss Sophie McIntosh recently completed an Early Childhood Honours degree at the University of Wollongong. After graduating, she was involved in collecting language samples from young Indigenous children across Far North Queensland, for the Indigenous Schooling Support Unit. Sophie currently works as a research assistant for the University of Sydney.

## **Integrated post-enrolment language assessment (PELA): A case study**

Alex **Barthel** & Liz **Craven**

University of Technology Sydney

Friday, 11:45pm (Parallel session G)

Economics & Commerce, Theatre 3 (EAP assessment)

**ABSTRACT:** Students enrolling at Australian universities come from a variety of cultural, language and socio-economic backgrounds. Over a quarter of all Australian university students are non-English language speaking background international students, mainly from China and South Asia. Many local students also come from language backgrounds other than English. Universities are thus faced with considerable educational challenges, in particular, with assessing the wide range of students' levels of English language proficiency early in their course of study and addressing their needs for either remedial English or developmental academic language assistance within their areas of study. The context of this presentation is a long-term project at a large Australian urban university which is aimed at enhancing all students' academic and professional communication skills. The university is gradually implementing a systematic approach to ensure that all its students have the necessary communication skills to complete their studies and be prepared for employment. This presentation focuses on a pan institution Post-Enrolment Language Assessment (PELA) project, which informs the development of all students' communication skills within their discipline area. Processes include the design, implementation and grading of language assessment tasks as well as the development and integration of language support within students' academic course structure and throughout the length of their university studies. Details of the project outcomes to date will be presented and discussed. Pedagogical and policy issues as well as resource implications, which are wide ranging for a project of this size, will also be discussed.

**PRESENTERS' BIO STATEMENTS:** Alex Barthel is the Director of the ELSSA Centre at the University of Technology, Sydney. Alex is also the inaugural President of the Association for Academic Language & Learning (AALL Inc.). His research areas include: Academic integrity; post-enrolment language assessment (PELA) of university students; and integration of academic language and learning into disciplines.

Dr Liz Craven is a lecturer at the ELSSA Centre at the University of Technology Sydney, where she teaches both support academic language workshops and faculty-based communication subjects for credit. Liz has been involved in IELTS examining since 1990. At UTS, she is also involved in post-enrolment language assessment (PELA)

## Identifying *hidden* ESL students: Lessons learned from a North Queensland high school

Nina Carter

Education Queensland, Indigenous Schooling Support Unit

Friday, 10:50am (Parallel session F)

Economics & Commerce, Theatre 2 (Language assessment and policy implications)

**ABSTRACT:** In Queensland, over recent years, many learners of English as a Second/Subsequent Language (ESL) have been effectively missing from ESL data sets. ESL learners have primarily been identified through criteria based on eligibility for specific funding programs and not through systematic policy, curriculum or even language-learning requirements. Funding eligibility has provided the impetus and purpose for collecting student language data, such as language backgrounds and ESL Bandscale levels. Those students who have *not* been eligible for any ESL funding, have generally *not* figured in ESL data. Indigenous ESL students constitute a major group of these *hidden* ESL learners. Their invisibility has been further compounded by the complex Indigenous language situation caused by widespread language shift. Speakers of creoles and related contact language varieties may be silenced if educational initiatives have not ensured that their language backgrounds are recognised and valued.

This paper reports on a study which developed and implemented a methodology to identify hidden ESL students. The study was conducted with Year 8 students at an urban high school in Far North Queensland. Students participated in a series of “language awareness” activities, which provided them with opportunities to discuss language situations and language varieties, before asking them to describe their own language backgrounds. Speaking and writing samples were collected from these students as sources of data from which approximate ESL Bandscale levels could be determined. Although school data did not identify any of the Year 8 students as learning ESL, over half were able to be identified as such through this process. Situations where large numbers of ESL learners are hidden have serious ramifications, as equitable access to - and participation in - classroom settings are predicated on proficiency in Standard Australian English.

**PRESENTER’S BIO STATEMENT:** Ms Nina Carter has had experience in diverse ESL contexts, ranging from pre-prep classrooms in Aboriginal communities to high schools in the Torres Strait. In her current role, with the Central and Southern Queensland Indigenous Schooling Support Unit, she undertakes ESL research projects, delivers professional development workshops, and supports teachers.

## Multilingual assessment as an alternative

Jasone **Cenoz**

University of the Basque Country

Friday, 2:30pm (Parallel session J)

Economics & Commerce, Theatrette 1 (Multilingual assessment)

**ABSTRACT:** Even when schools aim at multilingualism they have traditionally built strong boundaries between languages in teaching and assessment. In language testing, this approach has taken the monolingual native speaker as the reference for each of the languages involved and has assessed languages separately. Although a bilingual view considers that bilinguals are fully competent speaker-hearers (Grosjean, 1989) who develop a different complex type of competence (Cook, 1992), there have been few changes in the assessment of multilingualism in school contexts. The *Common European Framework of Reference for Languages* considers the native speaker not as the reference and “a given individual does not have a collection of distinct and separate competences” (CEFR, 2001: 168), but its scales refer to native speakers and only consider one language at the time.

Based on a holistic view of multilingualism, this paper proposes alternative ways of assessment in multilingual education that take into account the whole linguistic repertoire of multilingual learners and take other multilingual speakers and not monolingual native speakers as a reference. Furthermore, a multilingual approach needs to consider code-mixing and code-switching as distinctive behavior of multilingual speakers who use these strategies as a resource in everyday interaction. The last section of the paper discusses the possibility of incorporating these strategies in language teaching and assessment in multilingual school contexts.

**PRESENTER’S BIO STATEMENT:** Jasone Cenoz is Professor of Applied Linguistics in the Faculty of Education at the University of the Basque Country. Her research focuses on the acquisition of English as a third language, bilingualism, multilingualism and education. She is the editor (in collaboration with Ulrike Jessner) of the 'International Journal of Multilingualism'. She is on the board of AILA (International Association of Applied Linguistics). In collaboration with other colleagues, she has edited several books on bilingualism and multilingualism. One of her recent publications is 'Towards Multilingual Education: Basque educational research in international perspective'.

## **Self-assessment of English skills by examinees of an oral English proficiency test**

**Lixia Cheng**

Purdue University, US

Friday, 10:50am (Parallel session F)

Economics & Commerce, Theatre 1 (Assessing speaking)

**ABSTRACT:** Self-assessment of foreign language skills has been argued to have high level agreement with teacher ratings or test scores (LeBlanc & Painchaud 1981, 1985; Rea 1981). The factors that have been studied which affect self-assessment of language skills include gender, cultural background, self-esteem, proficiency levels and even wording of the self-assessment survey items. This study examined how prospective International Teaching Assistants (ITAs) at a large public university in Midwestern United States evaluate their own performance on a computer-based speaking proficiency test and their general skills in using the English language in various aspects. 200 questionnaires completed by potential ITAs from 34 countries were collected at the end of their tests. The questionnaire used a four-point Likert scale for the respondents to rate their own proficiency levels (poor, fair, good or excellent) in all the subskill areas including “reading in content area” and “listening to English televisions/movies”, and a five-point scale was used for their self-estimation of whether or not they would pass the test (strongly disagree, disagree, no opinion, agree, and strongly agree). Internal reliability of the survey items was assessed. Spearman correlation tests were performed to examine the relationship between examinees’ self-assessment and their actual performance on the oral English test, and chi-square tests were performed to investigate the effects of gender and native country on self-assessments. The research found out that the respondents’ self-estimation of test performance had a fairly high correlation with their actual test scores ( $r = .51710$ ). The respondents tended to associate the ability to speak fluently with the ability to speak without a lot of grammatical errors. There was no clear evidence that gender is related to the respondents’ self-estimation of test performance although their native country might have an effect on their self-estimation.

**PRESENTER’S BIO STATEMENT:** Lixia Cheng is a PhD student in the English as a Second Language Program at Purdue University. Her research interests include second language assessment and second language writing. She has worked as a testing coordinator in the Oral English Proficiency Program and taught an undergraduate introductory linguistics course at Purdue.

## **Assessment of listening & note-taking skills on EAP courses: Two test methods**

**Donal Crawford**

International Pacific College, New Zealand

Friday, 3:00pm (Parallel session K)

Economics & Commerce, Theatre 2 (Assessing listening)

**ABSTRACT:** Gaining skills in taking notes during lectures is commonly seen as an important component of EAP courses (Rost, 2002; Jordan, 1997; Flowerdew, 1994a). There is, however, little information on the assessment of this component (King, 1994). This paper describes and compares two different instruments that have been used to assess listening & note-taking skills in EAP examinations on pre-undergraduate Foundation year courses. In the first method - the retrieval method - assessment is made via marking responses to questions which students must answer using their notes. The second method - the direct method - assesses the notes themselves. The strengths and weaknesses of both methods are discussed and analysed in terms of practicality, authenticity, reliability and validity. Finally, the question of the need for traditional lecture note-taking skills in light of modern lecturing practices (Shin, 2008) is raised in terms of EAP course design and assessment.

**PRESENTER'S BIO STATEMENT:** I am Assistant Dean (ESOL) at International Pacific College, a private tertiary institution in New Zealand. I also currently act as External Examiner for EAP for Xian Jiaotong Liverpool University in China. I previously worked with the Northern Consortium of UK Universities (NCUK) as EAP Coordinator and Centre Principal.

## **EAP and high-stakes contexts: Tensions between teaching and assessment**

Russell Cross & Kieran O'Loughlin

The University of Melbourne

Friday, 12:45pm (Parallel session I)

Economics & Commerce, Theatre 3 (EAP assessment)

**ABSTRACT:** This presentation reports on the second phase of a larger study (AUTHORS, 2009) that examined a case study of teacher practice in a setting where 70% of the classroom-based assessment contributes to the students' final score in an ESL preparation/bridging course for university. Since the final score determines whether students have met the university's English language requirements for entry, the context is therefore high-stakes. The focus of the previous stage of analysis considered how the teacher viewed the utility of classroom based assessment within such a setting, and the implications that then raises for classroom practice (i.e., "what" classroom based assessment meant from the teachers' perspective for teaching and learning within high-stakes contexts). In particular, the findings of that analysis highlighted the tensions between the role of formative assessment (assessment for learning) and summative assessment (assessment of learning)

By way of contrast, the focus of the present analysis turns to an understanding of the construction of the teacher's own professional knowledge and learning within a high-stakes assessment context, given the demands placed upon individual choices about "internal" classroom practices by the existence of "externally" imposed requirements. In other words, the focus is on "how" and "why" the teacher came to ascribe the particular purposes and values to his own classroom based assessment practices in the ways that he did.

The analysis draws on a sociocultural/activity theoretical framework (Vygotsky, 1978; Engeström, 1981) to understand the ways in which the high-stakes assessment context mediated the teacher's understanding of their role – as both classroom teacher and assessor – in their position as the subject of that socioculturally constructed activity (Cross, in press). Data comprises classroom observation and field notes, document analyses, teacher interviews, and stimulus recall procedures (Gass & Mackey, 2000).

**PRESENTER'S BIO STATEMENT:** Dr Russell Cross's background is in bilingual education, and his research focuses on the social, cultural, and political nature of language teachers' work, knowledge, and practice. His research has appeared in *Language & Education*, *Language Problems & Language Planning*, *Modern Language Journal*, and he is current co-editor of *TESOL in Context*.

## **ESL Assessment in Vocational Education and Training: Issues for teachers and learners in the national policy context**

**Liz Davidson**

The University of Melbourne

Thursday, 3:35pm (Parallel session C)

Economics & Commerce, Theatre 3 (Assessing vocational English)

**ABSTRACT:** Adult ESL in Australia is taught in a number of settings (TAFE institutes, community organisations, private colleges and workplaces) in commonwealth and state funded programs. ESL teachers and their learners in these programs work in a national policy context which demands rigorous record keeping and reporting using nationally accredited courses and frameworks. A number of contradictions exist, for example:

- there is no single framework for ESL assessment nationally, (and no centralised ESL test)
- teachers use classroom-based assessment, locally designed and administered but linked to national quality standards for vocational education
- some programs require reporting which is linked to specific funding and others do not, yet students may be in the same classroom doing the same course
- in spite of requirements for sophisticated reporting and language assessment processes, many ESL teachers are employed casually or on short term contracts

This presentation will give an overview of the state of play in adult ESL and how learners are assessed in the different programs. Some of the key assessment issues which arise for second language teachers and learners because of the policy context will then be highlighted with reference to language assessment principles and practices.

**PRESENTER'S BIO STATEMENT:** Liz Davidson has taught and assessed secondary and adult ESL. Recently she has been involved writing accredited curriculum, in particular the Victorian ESL certificates for adults used in TAFE and community programs. This included conducting professional development for ESL teachers in assessment. She is completing post graduate studies in TESOL.

## **What are they thinking? Tracking raters' responses to an online rater training program**

Rosemary Erlam, Janet von Randow, John Read & Janna Fox  
The University of Auckland

Friday, 10:20am (Parallel session E)  
Economics & Commerce, Theatrette 3 (Language assessment and technologies)

**ABSTRACT:** The Diagnostic English Language Needs Assessment (DELNA) for first-year students at the University of Auckland includes a writing task that identifies students who need support in the area of academic literacy. Although this is a low-stakes assessment, it is important that the scripts are rated accurately so that students receive meaningful feedback on their writing. Face-to-face rater training takes place before each semester but refresher training is also required periodically. For this purpose an online rater training program was developed in 2004. Research carried out by Elder, Barkhuizen, Knoch and von Randow (2005, 2007) evaluated the effectiveness of the program, and, as a result, changes were made to the analytic descriptors and the online program itself. Raters who participated in that research were somewhat ambivalent about working online. However, as this has become more the norm, it seemed timely to investigate the use of the modified version of this program as a potential substitute for face-to-face training sessions to make rater training easier to manage. It would also enable universities in other parts of the world wishing to use the DELNA writing task to train their raters online. Phase one of this project, undertaken with Carleton University, Ottawa, had novice and experienced DELNA raters' give simultaneous verbal protocols as they rated the same six benchmark scripts online. This enabled a comparison of what the two groups of raters attended to in each script and how they arrived at the ratings. In this presentation their different responses to the DELNA writing descriptors and the program itself will be discussed, as will changes that need to be made in line with improvements in technology.

**PRESENTERS' BIO STATEMENTS:** Rosemary Erlam is Senior Lecturer in the Department of Applied Language Studies and Linguistics at The University of Auckland where she teaches language assessment. Her interest and experience in assessment dates back to her work with Cathie Elder in the initial trialling and design of the DELNA project.

Janet von Randow has managed the Diagnostic English Language Needs Assessment (DELNA) since its introduction at the University of Auckland in 2002 and has presented the related research at conferences both in New Zealand and overseas. She first joined the University teaching in the Graduate Diploma of Secondary Teaching after a long career in modern language teaching. Her research interests include language assessment, second language acquisition and the first-year university experience.

## **Targeting positive washback: A demand for more teacher education and teacher training in language assessment**

Masoomeh Estaji

Allameh Tabataba'i University, Iran

Friday, 3:00pm (Parallel session K)

Economics & Commerce, Theatre 3 (Language assessment and washback)

**ABSTRACT:** Assessment and evaluation play a pivotal role in teaching and learning. "Teachers, as the agents of assessment, need to ensure the validity and reliability of their classroom assessment practices and need to use these practices in ways that support their students' learning" (Brindley, 2001; Brindley, 2007; Cheng et. al. 2004; Rea-Dickins, 2004). This paper reports the findings of an investigation into teachers' beliefs and practices regarding the Academic Module of IELTS writing tasks. The study makes use of a classroom observation (10 sessions observed) and interview methods (8 teachers interviewed) to elicit: (1) teachers' views of students' abilities in different aspects of writing; (2) teachers' assessment and evaluation practices in the classroom; and (3) teachers' beliefs about the teaching and learning of IELTS Academic writing. The findings of the survey reveal that there exists a gap between teachers' beliefs and practices. Although most of the teachers think that task fulfillment and discourse coherence are essential to IELTS writing instruction, the findings suggest that they primarily attend to sentence structure and the use of a variety of appropriate vocabulary in their evaluation of students' writing and in their own teaching. In addition, it is found that most of the beliefs about good teaching of writing are overlooked in the test-preparation classroom. The results of this study have shown a multi-layered account of the link between assessment and instruction in terms of authenticity and congruence of assessment practices in relation to a program of study. The paper concludes that teachers need to be made aware of their role as writing teachers rather than language teachers and presents ways of how teachers can become more "assessment literate" and how testing experts and teacher trainers can help in this direction.

**PRESENTER'S BIO STATEMENT:** Masoomeh Estaji is Assistant Professor of TEFL at Allameh Tabataba'i University. She holds a Ph.D in TEFL from Allameh Tabataba'i University and has participated in numerous local, national workshops on educational issues, methodology, and testing. She has also presented numerous papers on methodology, testing, and education at a number of TEFL conferences. She has been actively involved in developing English language tests and materials at the Kish English Institute since 1999, and has taught different courses in ELT at B.A. levels at Allameh Tabataba'i University, Shahid Rajaee University and Media Studies Faculty. She has also taught ESP courses at Islamic Azad University (Engineering Faculty). Her research interests include Language Testing and Assessment, SLA and Psycholinguistics.

## Exploring strategies in multilingual assessment

Durk Gorter

University of the Basque Country

Friday, 3:00pm (Parallel session K)

Economics & Commerce, Theatrette 1 (Multilingual assessment)

**ABSTRACT:** This paper focuses on research carried out in Basque schools in Spain where Basque and Spanish are official languages and English is learned as a third language. This research is based on the multilingual approach to language assessment that considers multilingual speakers' whole linguistic repertoire and has other multilingual speakers and not monolingual speakers as a reference. This paper will explore specific ways in which this multilingual approach can be used in language assessment. First a trilingual receptive vocabulary test that uses a compound score will be discussed. Second, a writing test in Basque, Spanish and English will illustrate the differences between a monolingual and a multilingual approach in language assessment. Finally, the possibility of incorporating code-mixing and code-switching in assessment will be discussed as related to a productive multilingual vocabulary test. The multilingual approach to language assessment offers at least two advantages as compared to more traditional approaches. From a language ecological perspective a multilingual approach is more valid because it is closer to the way languages are used in real interaction in a social context where multilinguals use the languages in their repertoire as a resource. Furthermore, a holistic multilingual approach also offers a better picture of the languages being acquired and used by multilingual learners.

**PRESENTER'S BIO STATEMENT:** Durk Gorter is Ikerbasque research professor at the Faculty of Education of the University of the Basque Country in San Sebastian/Donostia, where he does work on multilingualism, European minority languages and linguistic landscapes. Before that he was a researcher in the sociology of language at the Fryske Akademy in Ljouwert/Leeuwarden, The Netherlands, and also part-time full professor at the University of Amsterdam in the sociolinguistics of Frisian. Two recent edited books are *Linguistic Landscape: Expanding the Scenery* (with Elana Shohamy) and *Multilingual Europe: Facts and Policies* (with Guus Extra).

## **IELTS and Community Welfare Workers – A strenuous exercise in standard-setting**

Elisabeth Grove

Thursday, 2:35pm (Parallel session A)

Economics & Commerce, Theatre 3 (Assessing vocational English)

**ABSTRACT:** The use of IELTS for gate-keeping – as a measure of the language competence of international professionals seeking to immigrate – has attracted recent critical attention (O’Loughlin 2009). Despite concern about its extension to purposes other than those it was designed for (e.g. Davies 2008), the test has been adopted by a number of government agencies and institutions, including professional associations. However, little is known about how such groups have gone about setting minimum band-scores for skills recognition. From the available evidence, the professions involved have generally accepted ‘official’ IELTS recommendations rather than establishing and verifying standards for themselves.

The present study of language standard-setting, carried out in 2009 on behalf of the Australian Institute of Welfare and Community Workers (AIWCW), sought to determine the minimum IELTS band levels to be required of international graduates by enlisting a range of stakeholder perspectives (employers, professional representatives, course providers, international graduates and current students of welfare work, and IELTS-IDP). Several investigative methods were used in the project, including analysis of community welfare communication tasks; stakeholder interviews; a survey questionnaire matching IELTS band-levels to a range of work tasks; and a focus-group analysis of IELTS benchmark performances.

The fact that AIWCW’s language requirements were to be applied to international graduates of Australian colleges on course completion was a major difference from the practice of other professional groups – whose requirements apply only to internationally-trained graduates – and also a source of some controversy. The complexities of the study and its mixed outcomes are the subject of this paper.

**PRESENTER’S BIO STATEMENT:** Lis Grove was a Research Fellow at the LTRC for 11 years, until mid-2004. Her particular interests are in second-language teaching and learning, and in professional communication, both written and spoken. She now works as an independent researcher and editor.

## **New media and technologies in second language assessment**

**Paul Gruba**

The University of Melbourne

Friday, 10:50am (Parallel session F)

Economics & Commerce, Theatre 3 (Language assessment and technologies)

**ABSTRACT:** New media and technologies - including digital videos, wikis, podcasts - are now established in innovative second language teaching, yet their integration into classroom assessment and test development continues to lag. In this paper, I first examine three emerging trends in educational technologies: social networking, convergence and mobility. I then look at these in the context of classroom assessment, for example, with regards to institutional resources, professional development and policy. In a third section, the focus shifts to language test development with consideration of construct definitions, test practicality and security, and the potential for washback in technology rich settings. To conclude, I propose an agenda for further research in the use of new media and technologies in second language assessment.

**PRESENTER'S BIO STATEMENT:** Paul Gruba's research areas concern the interplay of language, technology, media and assessment in second language contexts. He is currently a senior lecturer at the University of Melbourne.

## Continuity between primary and secondary school languages programs: What can classroom-based assessment tell us?

Kathryn Hill

The University of Melbourne

Friday, 11:45am (Parallel session G)

Economics & Commerce, Theatre 2 (Assessing languages other than English)

**ABSTRACT:** This paper uses an ethnographic study of classroom-based assessment (CBA) practices to investigate the impact of education and languages education policy on the issue of continuity between primary and secondary school languages (LOTE) programs.

The context for the study is the trend in Australia, as elsewhere, towards commencing languages at primary school level. While there are a number of additional benefits cited for this policy, there is a clear assumption that it will result in higher proficiency outcomes in the longer term (e.g., Lo Bianco 2001). However, research findings suggest that this is not necessarily the case (e.g., Johnstone, 1999). A number of researchers have identified '*discontinuity*' as one of the key factors responsible for undermining any proficiency advantages gained from an early start to language learning (e.g., Kubanek-German et al., 1998). Yet this area remains relatively under-researched (Rea-Dickins, 2009). A distinctive feature of the current study is that it uses classroom-based assessment (CBA) practices as a lens for investigating the continuity issue.

The research, which took place in Victoria, Australia, involved Indonesian language classrooms at two successive levels of schooling: the last year of primary (Year 6) and the first year of secondary school (Year 7) respectively.

Data collection took place in the final term of Year 6 and Terms 1 and 2 of Year 7, comprising a total of ten weeks in each classroom. The study was longitudinal in that the Year 7 cohort included a group of students who had also participated in Year 6.

Data comprised classroom observation and field notes, audio recordings (including classroom interactions, teacher and student focus group interviews, and meetings) and documents (including policy documents, reports, course-books, and work samples).

Key differences were identified in *how* learning was assessed as well as *what* was assessed in the respective classrooms with consequences for continuity between the two levels. The paper will explore how some of these differences can be related to aspects of education and languages education policy, including the introduction of a new assessment and reporting framework.

The findings have relevance for languages education policy, planning and practice.

**PRESENTER'S BIO STATEMENT:** Kathryn has a B.A., Dip.Ed., M.A. and PhD, all from the University of Melbourne. She was a research fellow in the Language Testing Research Centre (LTRC) for 15 years before moving to the Faculty of Medicine, Dentistry & Health Sciences where she lectures in clinical communication.

## **Before and after: Developing classroom teachers' language understandings through the collection and use of speaking and writing samples from learners of ESL.**

Natasha **Holzberger**  
Education Queensland

Thursday, 3:35pm (Parallel session C)  
Economics & Commerce, Theatre 1 (Accountability and assessment)

**ABSTRACT:** Educators who are supporting classroom teachers with learners of English as a Second/Subsequent Language (ESL) need effective tools and strategies to highlight the effects of English language acquisition levels on these students' access and participation in the classroom curriculum. This paper describes how "before and after" samples constitute a straightforward, flexible and effective strategy for this context. Oral and written language samples are collected both at the start and end of a unit of work, by interviewing students on the current class topic (and transcribing these interviews), and/or by asking students to write about the topic. Such measures provide concrete evidence of ESL learners' successes and challenges with the classroom content.

Speaking and writing samples constitute a highly productive strategy for working with classroom teachers, as they have a myriad of uses. Discussing and analysing them with teachers can influence planning approaches and teaching choices, as well as drive learning conversations and professional development offerings. The initial sample is particularly useful for comparing to assessment items in order to determine the language that will need to be taught. As data sources, "before and after" samples provide material to analyse ESL learners' linguistic development and to assist with assigning ESL Bandscale levels. The presenter of this paper is currently a Head of Curriculum at an all-Indigenous primary school in Far North Queensland and also provides "Adopt-a-School" services to two remote schools on Cape York. In this paper, she shows how she has used "before and after" speaking and writing samples to alert classroom teachers to their ESL students' language-learning needs, so as to influence their planning and teaching approaches.

**PRESENTER'S BIO STATEMENT:** I am currently working at Yarrabah State School in an indigenous community near Cairns, in an ESL coaching role. I have taught in several different contexts in Australia and overseas, as a classroom, learning support and ESL teacher.

## **Intelligibility of learner speech in academic speaking task performance**

Noriko Iwashita

The University of Queensland

Friday, 9:50am (Parallel session D)

Economics & Commerce, Theatre 1 (Assessing speaking)

**ABSTRACT:** Speaking tasks used in EAP tests increasingly seek to replicate the roles and demands of students in academic contexts. Integrated tasks in which test-takers are required to process and transform a cognitively complex stimulus (e.g., a written text or a lecture) and integrate information from this source into the speaking performance have been widely used to assess learner proficiency in academic context. Integrated tasks are more complex and demanding than traditional stand-alone tasks in which test-takers draw on their own knowledge or ideas to respond to a prompt. Researchers (e.g., Skehan, 1998; Robinson, 2001) have suggested the potential impact of cognitive demand on task performance especially in light of grammatical accuracy, complexity, vocabulary and fluency. Features which may affect the intelligibility of speech such as pronunciation, intonation and rhythm were three of a number of performance features to which assessors oriented when evaluating the test performances (Brown, et al. 2005), but to date these features are not fully investigated in relation to cognitive demands on task performances.

The study reported in this paper is an investigation of production features (i.e., pronunciation, intonation and rhythm) in academic speaking test performance in the context of a larger project to develop a rating scale for a new international test of English for Academic Purposes (Brown et al., 2005). Spoken test performances representing two different tasks and five different proficiency levels (79 performances in all) were analyzed using the speech data labeling application software *Xwaves* (Rommark 1995) according to pronunciation, intonation and rhythm. The results showed that the frequency of 'target-like' syllables showed some difference across task types; the intonation units in independent task performances were more frequently categorized as 'Many and English-like', compared with integrated task performances. Overall, the study contributes important insights into the nature of spoken proficiency as it develops and its measurement in rating scales for speaking.

**PRESENTER'S BIO STATEMENT:** Noriko Iwashita is a senior lecturer in Applied Linguistics at The University of Queensland. Her research interests include the interfaces of language assessment and SLA, task-based assessment, and cross-linguistic investigation of four major language traits. Her work has appeared in *Language Testing*, *Applied Linguistics*, *Language Learning* and *Studies in Second Language Acquisition*.

## **Insider perspectives on the impact of the aviation English testing policy in Korea**

**Hyejeong Kim**

The University of Melbourne

Friday, 12:15am (Parallel session H)

Economics & Commerce, Theatre 1 (Assessing English for the workplace)

**ABSTRACT:** Recent writings on the role of tests in policy contexts (Shohamy, 2009; McNamara, 2008; McNamara & Roever, 2006) have emphasized the importance of considering the use of language tests in their policy context as a way of understanding their broader social meaning and impact. This paper adopts a discourse analytic approach evaluating the impact of the International Civil Aviation Organisation (ICAO) language testing policy, designed to address concerns about the safety consequences of limited English proficiency among non-native aviation personnel. The policy has met with considerable resistance in many countries, including Korea, which is the focus of the current research. Although the Korean government developed and administered an English language test in compliance with the ICAO policy and most of those sitting the test achieved the minimum required standard, there have been concerns amongst users that the test was not valid for its purpose, and indeed that the policy itself was based on false premises.

This paper probes the reasons for this resistance by engaging feedback from a group of expert informants, Korean pilots and controllers, on the nature of and the factors contributing to miscommunication in radiotelephony found in 6 recorded discourse samples collected from 2 air control centres in Korea. Findings show that issues of language proficiency are perceived to be less central to miscommunication in radiotelephony than is sometimes claimed and that other factors relating to the way interlocutors orient to their professional role are critical to efficient and effective radiotelephony interaction. They also reveal that responsibility for miscommunications is shared between native and non-native speakers, suggesting that the focus of the ICAO policy and associated test may be too narrow. Interviews and questionnaire data also support the findings. Implications are drawn for the ICAO policy in particular and for the implementation of test-related policies more generally.

**PRESENTER'S BIO STATEMENT:** Hyejeong Kim is doing her PhD in Applied Linguistics at the University of Melbourne. Her research interests include English as a lingua franca, radiotelephony communication, English for specific purposes, and testing policy.

## **Reading for real: Authenticity in the assessment of academic reading**

Tim **Moore**, Janne **Morton** & Steve **Price**

The University of Melbourne & Swinburne University of Technology

Friday, 12:15pm (Parallel session H)

Economics & Commerce, Theatre 3 (EAP assessment)

**ABSTRACT:** Test authenticity, a permanent goal of language assessment, can be defined as that situation where there is a strong link between test tasks and TLU tasks (Bachman and Palmer, 1996). We look at the challenges involved in achieving greater authenticity in the assessment of academic reading, using IELTS as an example. Results are presented from a study that compared tasks set on the IELTS academic reading and the forms of reading students typically need to engage in the first year of undergraduate study. A number of key differences are noted between the two domains, some of which can be explained in relation to Barton and Hamilton's (1998) useful notion of the 'literacy event'. Implications for test development are discussed, along with the study's broader implications for EAP syllabus design.

**PRESENTERS' BIO STATEMENTS:** Tim Moore is a lecturer at Swinburne University of Technology. He is mainly interested in the teaching and researching of academic literacies. He has done a small amount of research with colleagues into the IELTS test.

Janne Morton teaches EAP and communication skills subjects in Linguistics and Applied Linguistics at the University of Melbourne. She is currently completing her PhD on academic discourse socialisation.

## **School-based evaluation: A possible remedy to the ills of school accountability**

**David Nevo**

Tel Aviv University, Israel

Thursday, 2:35pm (Parallel session A)

Economics & Commerce, Theatre 1 (Accountability and assessment)

**ABSTRACT:** The presentation will start with a short review of the history of school accountability and its basic assumptions since its initiation in the USA in the early seventies of the previous century and its spread to many countries around the world, including Australia. Research findings will be then presented to explore its impacts on the educational system pointing to its benefits and shortcomings. The presentation will conclude with a proposal to use school-based evaluation as a possible way to benefit from the advantages of accountability and avoid as much as possible its shortcomings.

**PRESENTER'S BIO STATEMENT:** Expert in educational evaluation with special interest in *Dialogue Evaluation*, combining internal and external evaluation of schools and educational programs. Among his publications: *Evaluation in decision making* (Kluwer, 1988), *School-based evaluation: A dialogue for school improvement* (Pergamon, 1995), and *School-based evaluation: An international perspective* (Elsevier, 2002). Was Editor-in-Chief of *Studies in Educational Evaluation*. Served as Head of School of Education, Tel Aviv University.

## **An investigation of validity of two EFL listening tests: IELTS and TOEFL iBT**

Thi Nhan Hoa **Nguyen**

Hanoi National University, Vietnam

Friday, 2:30pm (Parallel session J)

Economics & Commerce, Theatre 2 (Assessing listening)

**ABSTRACT:** IELTS and TOEFL iBT tests have been used widely in the world as EFL English language proficiency tests for admission for tertiary education in English speaking countries. Generally scores of the two tests can be used interchangeably as the number of institutions in the world accepting both scores (of the IELTS and TOEFL iBT) as evidence about English language proficiency has increased greatly. It is, therefore, of great importance to investigate whether the two tests measure the same listening ability of test takers.

The study drew on the validity frameworks of Messick (1989) and Kane (1992, 2001) in comparing the construct validity of the two listening tests. It investigated both the judgmental justifications and empirical evidence as advocated by Messick (1989) and the five types of validity inferences outlined by Kane (1992, 2001) as elaborated by McNamara and Roever (2006). As a result, the method used in the study to compare the validity of the two tests is a multi-faceted approach, covering (i) the test content, (ii) the test measurement properties, (iii) the degree to which the two tests measured the same listening ability, (iv) the effects of test preparation on test performance, and (v) the strategies used while doing the tests.

Findings reveal that the two tests shared some common ground (level of difficulty, features of listening stimulus, candidates' use of listening strategies) but they differed markedly in many other aspects (test content, test construct and test preparation effect on test scores).

**PRESENTER'S BIO STATEMENT:** Hoa Nguyen is the Vice-Dean of the Foreign Languages Department of the International School, Hanoi National University, Vietnam. She is interested in language teaching and language testing, especially EFL (English as a Foreign Language) teaching and testing.

## **Student language as a factor in variability and agreement amongst assessors of undergraduate essays: An issue for quality assessment in higher education?**

**Sally O'Hagan**

The University of Melbourne

Thursday, 3:05pm (Parallel session B)

Economics & Commerce, Theatre 2 (Assessing Writing)

**ABSTRACT:** In recent years, the quality of student assessment in higher education (HE) has been subject to widespread criticism. The reliability and validity of HE assessment has been questioned in the scholarly literature, public debate has considered issues of bias and falling academic standards, and in some cases, the growing reliance on international student revenue has been implicated in concerns over assessment quality. While the reality or extent of such problems may be uncertain, the imperative to investigate them in the interests of student equity is underscored by the growing recognition of the role of assessment in shaping the quality of student learning.

This paper considers the problem of assessment quality in terms of the assessment of student performance in university disciplines and the influence of student language on this. A study investigated the responses of 10 assessors who each marked the same 5 English native-speaker (NS) and 5 non-native speaker (NNS) undergraduate essays while producing think-aloud verbal reports. Assessors focussed more on grammar when marking NNS essays, and more on content when marking NS essays. The most significant finding however was that interpretations of assessment criteria were highly varied. As a result, inter-assessor reliability was low, with many essays receiving the whole range of possible grades. Matters of essay content were found to be most subject to differing interpretations. Owing to the greater focus on the content of the NS essays, variability in grades was found to be a greater issue for the NS essays. The think-aloud data explore the reasons for this, and the implications for student equity and the validity of current assessment practices are discussed. The discussion will show that, despite moves to implement greater accountability and transparency in HE assessment, difficulties persist in the articulation and sharing of the meanings of criteria.

**PRESENTER'S BIO STATEMENT:** Sally recently completed her PhD on the assessment of undergraduate disciplinary writing at university. Her research interests include performance assessment and rater behaviour, academic literacy, and the impact of student diversity on assessment practices in higher education.

## **Construct definition and ethical issues in aviation English assessment**

**John Read**

The University of Auckland

Friday, 11:45am (Parallel session G)

Economics & Commerce, Theatrette 1 (Assessing English for the workplace)

**ABSTRACT:** One of the primary functions of the Language Testing Research Centre over the last 20 years has been to undertake projects on behalf of governmental agencies to assess the language proficiency of various professional groups in their areas of expertise. Such work involves challenges not only in defining the nature of the construct to be assessed but also in addressing a range of ethical issues that can arise, particularly in maintaining high quality standards. A current area of great concern to many language testers is the assessment of pilots and air traffic controllers to meet the Language Proficiency Requirements set by the International Civil Aviation Organization (ICAO). In this case, construct definition involves clarifying the respective roles of the standard phraseology and so-called plain language in aviation discourse, as well as the larger question of whether language competence per se or proficiency in English as a Lingua Franca (ELF) should be the major concern. With regard to the ethical dimension, aviation English assessment represents an unusual situation. Documents like the ILTA Code of Ethics implicitly assume that language testers primarily work on the development and administration of particular tests but, with the assessment of English for international aviation, there are multiple tests in the marketplace and three overlapping levels of potential engagement: with test publishers, with national civil aviation authorities and with ICAO at the international level. In this context, policy advice and advocacy can take on a more prominent role than the application of technical skills. In this paper, I will give an overview of the issues in each of these areas of construct definition and the ethical dimension, drawing on my own experience and those of colleagues in the field at these various levels of engagement.

**PRESENTER'S BIO STATEMENT:** John Read is an associate professor in Applied Language Studies at the University of Auckland. His research interests focus on assessing English for academic and professional purposes and L2 vocabulary testing. He is Vice-President of the International Language Testing Association and a former co-editor of *Language Testing*.

## Language tests and Australian citizenship

Kerry Ryan

Swinburne University of Technology

Friday, 10:20am (Parallel session E)

Economics & Commerce, Theatre 2 (Language assessment and policy implications)

**ABSTRACT:** This paper discusses the long history of the use of language tests as a means of enforcing government policy in Australia. It will focus primarily on the latest example of such policy-making with the introduction in October of 2007 of the Australian citizenship test, a test, in English, of the knowledge of the traditions, history and values of Australian society. Prospective citizens are required to pass the test as one of the conditions required in order to gain Australian citizenship.

The paper will chart the implementation and administration of the Australian citizenship test as well as detail the review of the test commissioned by the Department of Immigration and Citizenship in 2008. It is contended that the review process failed to address some of the chief concerns raised by academics, refugee advocate groups and organisations, among others, and really only served to further legitimise the test in the eyes of the voting Australian public.

**PRESENTER'S BIO STATEMENT:** Kerry Ryan has been an employee of the Language Testing Research Centre since 2005. He is currently doing his PhD on the Australian Citizenship Test at the Institute for Social Research at Swinburne University of Technology.

## **Issues in eliciting and analysing national student achievement data in Chinese and Indonesian**

**Andrew Scrimgeour & Michelle Kohler**

University of South Australia

Friday, 12:15pm (Parallel session H)

Economics & Commerce, Theatre 2 (Assessing languages other than English)

**ABSTRACT:** In 2009 the DEEWR commissioned the LTRC and RCLC to conduct a research project to investigate student achievement in Chinese, Indonesian, Japanese and Korean as targeted under the National Asian Languages and Studies in Australian Schools Project (NALSSP). The project builds on a number of previous projects in Australian languages education that aim to understand the nature of students' language learning (e.g. Scarino, McKay & Clark 1989; Scarino et al 1998; Scarino et al 1997; Hill et al 2004). The project was designed to capture samples of students' performance at Year 6/7, 10 and 12 in order to develop descriptors of performance in the specific languages.

A key concern for this project was to capture both quantitative and qualitative data including information related to student background and intensity of learning in order to investigate the relationship between these variables and performance. The experience of gathering such data has raised a number of issues of particular significance for each of these languages and this paper focuses in particular on the issues for Chinese and Indonesian. These issues relate to the nature of the student groups, their performances and the program contexts in these particular languages. It is hoped that the experience and insights developed in this project will further understandings of the nature of these languages, the students who study them and the contexts in which they are studied. Such understandings may also inform developments in the languages field such as the development of the Heritage Speakers syllabus implementation (BOS NSW 2009) and the national curriculum for Languages.

**PRESENTERS' BIO STATEMENTS:** Andrew is a Research Fellow at the Research Centre for Languages and Cultures (RCLC) at the University of South Australia. He has been involved in research on aspects of languages teaching and learning, including assessing student outcomes in schools, with particular focus on character-based languages. He has been involved in a number of collaborative research projects including the current Student Achievement in Asian Languages Education (SAALE) Project.

Michelle Kohler is a researcher at the Research Centre for Languages and Cultures, University of South Australia. She has contributed to a number of national projects in languages and is currently the panel chair for Indonesian for the Student Achievement in Asian Languages project. Her research interests include intercultural perspectives on language curriculum and pedagogy, language teacher education, and classroom based research.

## **Assessing postgraduate ESL learners' writing after one semester of study**

Neomy Storch

The University of Melbourne

Thursday, 2:35 (Parallel session A)

Economics and Commerce, Theatrette 2 (Assessing Writing)

**ABSTRACT:** In recent years, there has been a dramatic increase in the number of international (ESL) students enrolling in Australian universities, particularly in postgraduate courses. One of the assumed benefits of studying in an L2 medium university (in this case English) is that of improved L2 language skills. However, there have been relatively few studies which have investigated what happens to learners' L2 whilst studying in an L2 medium university. The paper reports on a study (N=69) which investigated the changes (if any) in the writing of students enrolled in an EAP course specifically designed for postgraduate students. The study used a pre – post test design, comparing the writing produced in-class by the students at the beginning and at the end of semester. The students' texts were analysed using a range of measures for fluency, linguistic accuracy and use of academic vocabulary. Texts were also analysed for overall structure, cohesion and coherence. Two cohorts were identified: conditional entry students with an IELTS score of 6.0 (or equivalent) and those with a score of 6.5. The study found improvements for the entire group and for the two cohorts mainly in terms of accuracy, use of academic vocabulary, and structure. These findings are discussed in relation to key pedagogical features of the course.

**PRESENTER'S BIO STATEMENT:** Dr Neomy Storch teaches a range of ESL and Applied Linguistics subjects at the School of Languages & Linguistics. Her research has focused on issues related to second language pedagogy, particularly the nature of pair interaction in classroom contexts and the development of academic writing, and the role of feedback in that development.

## **Adults' language proficiency in the light of the National Certificates: The case of Finnish and Swedish as a L2 and English as a FL in Finland**

Mirja Tarnanen & Marita Härmälä  
University of Jyväskylä, Finland

Friday, 12:45pm (Parallel session I)  
Economics & Commerce, Theatrette 2 (Assessing languages other than English)

**ABSTRACT:** Finland is a bilingual country with 94 % of the population speaking Finnish as their mother tongue. The role of Swedish as the second domestic language is secured by legislation, which means that Swedish is a compulsory subject in basic and upper secondary education. Working life sets language skills requirements on civil servants in the two national languages. Immigrants applying for Finnish citizenship need to prove their language skills, e.g., by taking the National Certificates (NC) in Finnish or Swedish at the Intermediate level. The NC is a national language test for adults assessing speaking, writing, and reading and listening comprehension in nine languages. The certificate is used for different purposes, e.g., applying for work, study or citizenship.

The data of the study consist of a background questionnaire (e.g., age, mother tongue, the use of the target language) and test scores of five subskills in English (n=1 500), Swedish (n=615) and Finnish (n=8 678). In our presentation we will discuss NC participants' background, their language skills profiles, and the relationship between language use and test performance at the Intermediate and Advanced level. What kinds of differences are there across languages, and what do these differences tell us about the roles of these three languages in Finland? In addition to this, we will also present some data showing which domains of language use are most significantly correlated to the test scores. Finally, we will discuss how the skill profiles in these three languages reflect language policy and integration issues in Finland.

**PRESENTER'S BIO STATEMENT:** Dr Marita Härmälä works as a researcher at the University of Jyväskylä, in Finland. Her work tasks include compiling tests of Swedish, French, and Italian in the National Certificates of Language Proficiency, and training item writers and raters. Her other field of expertise is language skills assessment in competence-based qualifications.

## **The medical educator, the discourse analyst and the phonetician: An interdisciplinary approach to diagnostic assessment in the workplace**

Robyn **Woodward-Kron**, Mary **Stevens** & Eleanor **Flynn**  
The University of Melbourne

Friday, 12:45pm (Parallel session I)  
Economics & Commerce, Theatre 1 (Assessing English for the workplace)

**ABSTRACT:** Most International Medical Graduates (IMGs) are required to undertake the Australian Medical Council's clinical examination for general registration. This high stakes examination consists of 16 Objective Structured Clinical Examinations (OSCE), in which the candidate conducts a clinical interview with a simulated patient. This paper reports on a collaborative project to develop a diagnostic assessment methodology incorporating feedback on language, communication and clinical aspects of a candidate's performance. The methodology included development of communication feedback guidelines (CFGs), which were informed by a model of language in context and observation of IMG communication training. IMG participants were video-recorded during practice OSCEs. Language oriented feedback was provided by the discourse analyst and phonetician. Participants received a diagnostic feedback package consisting of their practice video and the completed communication feedback proforma.

**PRESENTER'S BIO STATEMENT:** Dr Robyn Woodward-Kron is a senior lecturer in the Medical Education Unit, Melbourne Medical School, University of Melbourne. Her research interests include healthcare communication, intercultural communication, genre and academic discourse. Robyn is also involved in developing multimedia resources to enhance the teaching and learning of healthcare communication.

## **Investigating the Validity of the GEPT Level Framework**

Jessica Wu

The Language Training and Testing Center, Taiwan

Friday, 9:50am (Parallel session D)

Economics & Commerce, Theatre 2 (Language assessment and policy implications)

**ABSTRACT:** This paper is a report of three studies which investigated the validity of the GEPT Level Framework. The GEPT (General English Proficiency Test) is a criterion-referenced testing system which has been implemented widely in Taiwan to assess the general English proficiency of Taiwanese EFL learners since 2000. Each of the studies employed a different approach. Study 1 investigated the validity of the GEPT framework by mapping different levels of the GEPT against the CEFR (CoE, 2001). The study adopted the methods proposed in the Manual for Relating Language Examinations to the CEFR (CoE, 2003) and the procedures outlined in the Dutch project (Alderson et al, 2006). The results show that the first four levels of the GEPT, from Elementary Level to Advanced Level, correspond to the CEFR A2 to C1 levels, providing external validation evidence to support the level framework. Study 2 investigated the validity of the framework by linking different levels of the GEPT vertically onto a common score scale on the basis of Item Response Theory. The results show a pattern of increasing difficulty across the levels, providing internal validation evidence to support the level framework. To supplement Study 2 which focused on the psychometric aspect, Study 3 looked into the language aspect by comparing texts used in the reading tests at different levels on the basis of contextual features including type-token ratio, vocabulary range, sentence complexity, rhetorical organization, conceptual cohesion, and text abstractness. Significant differences were found between the different levels of the GEPT based on the results of analysis of various contextual features. By triangulating evidence from multiple sources, the paper concludes that the validity of the level framework is generally supported. However, to strengthen the validity of the GEPT Level Framework, the paper also recommends a possible direction for improvement.

**PRESENTER'S BIO STATEMENT:** Dr Jessica Wu is Program Director of the Research and Development Office at the Language Training and Testing Center (LTTC), Taiwan. Her research interests include oral assessment and the impact of large-scale language testing. She is currently a member of the Editorial Board of *Language Assessment Quarterly*.

## How good should a taxi driver's English be?

Elaine Wylie, David Ingram & Hilda Maclean

ISLPR Language Services Pty Ltd & The University of Melbourne

Thursday, 3:05pm (Parallel session B)

Economics & Commerce, Theatre 3 (Assessing vocational English)

**ABSTRACT:** This presentation will report on a recent (2009/10) project to develop National Minimum English Standards for taxi driver recruits in Australia. The project arose from the concerns of state and territory authorities responsible for regulating the taxi industry about the English skills of drivers from a non-English-speaking background. The national regulators decided in 2009 that the International Second Language Proficiency Ratings (ISLPR) scale would be used in establishing national standards. One state, Victoria, had been using the ISLPR for some years, although there was disagreement among stakeholders about what the acceptable skills profile should be.

The presentation will outline the process that the project team went through in making their recommendations about the minimum levels for Speaking, Listening, Reading and Writing. In surveying the language needs of trainee and licensed drivers, they used reported data (e.g. from interviews) and direct observations (e.g. from the radio control rooms of taxi companies). Documents used as data sources included the national driver training package and messages from control rooms to drivers' computers. The perspectives of regulatory authorities, managers of taxi companies and trainers were sought, and consumer views (e.g. as expressed in mainstream media) were noted.

The challenge of synthesising the data to arrive at defensible compromises (within the broad context of recruitment needs for a relatively low-paid job) will be discussed. One issue was the gap between the levels that would be sufficient for drivers to cope with routine tasks and the levels that they would need if faced by a situation such as an escalating threat of violence against them. Another was the interplay, often cited during interviews, between language skills and cultural and personality factors (cf McNamara's 1990 distinction between weak and strong task performance).

**PRESENTERS' BIO STATEMENTS:** Elaine Wylie is a Director of ISLPR Language Services. Her previous position was Senior Research Fellow at Griffith University. She is co-author, with David Ingram, of the International Second Language Proficiency Ratings (ISLPR®). She conducts training programs around Australia and overseas in the use of the ISLPR system.

Professor David Ingram, AM is an Honorary Professorial Fellow, University of Melbourne. He recently prepared a report on English standards for the taxi industry; he has published widely, is the co-author of the ISLPR, was a developer and then Chief Examiner (Australia) of IELTS and, in "retirement", is Co-Director of ISLPR Language Services.

## **Washback of CET4 design and uses on test preparation and performance: A SEM study**

**Qin Xie**

Hong Kong University

Friday, 2:30pm (Parallel session J)

Economics & Commerce, Theatre 3 (Language assessment and washback)

**ABSTRACT:** College English Band 4 (CET4) is an English proficiency test given to all university students in China; it is widely used as a qualification for university graduation and job application. This paper reports a recent study which investigated the influences of CET4 design and uses on students' test preparation and performance, and their implications for test validity.

The study comprises two phases in three waves. Based on a qualitative study, a perception and a test preparation questionnaire were developed and piloted with 157 students. Items with sound psychometric properties were selected, and a SEM model was specified. Phase 2 was quantitative, involving over 1,000 students at each of the four data collection points. The study was conducted two months before the CET4 administration. The perception questionnaire was given at the beginning of test preparation and the test preparation questionnaire was given near the end. Meanwhile, a pre and a post test of CET4 performance were administered. Major data analyses include Exploratory Factor Analysis, Confirmatory Factor Analysis, and Structural Equation Modeling. AMOS was employed to assess the SEM model.

The study found that the CET4 had primarily negative washback effects; both CET4 design and uses affected test preparation. Negative washback, however, did not seem to threaten test validity. Post-test performance showed large improvement, but test preparation only contributed a very small part to the improvement. Post-test performance was influenced primarily by prior language proficiency. Compared with the pretest, the post-test model had better convergent and discriminate validity. Following the argumentative validation approach (Mislevy 2003, Bachman 2005), I argue that negative washback does not necessarily damage test validity. The significance and limitations of this study are discussed at the end.

**PRESENTER'S BIO STATEMENT:** Dr. Xie Qin recently completed her PhD study in the Faculty of Education, University of Hong Kong. Her PhD thesis adopts Structural Equation Modelling to examine the consequential validity of high-stakes English proficiency tests. She also held a MA with distinction in English Language Teaching from the University of Warwick, UK.

## Comparing Native and Non-native teachers' ratings of English in a group oral proficiency test

Ying Zhang  
Monash University

Friday, 10:20am (Parallel session E)  
Economics & Commerce, Theatre 1 (Assessing speaking)

**ABSTRACT:** This paper reports findings from an empirical study on ESL/EFL teachers' evaluation and latent interpretation of oral English proficiency in the nation-wide College English Test-Spoken English Test (CET-SET) of China. Enlightened by the debate on the NES/NNES issue which has come to professional attention in recent years, this study answers the question of whether the non-native English-speaking (NNES) teachers' judgments correspond to those of native English-speaking (NES) teachers. The present study probes the potential effect of language background on teachers' judgments of oral proficiency by eliciting data from 20 NNES and 19 NES teachers in relation to a sample of 30 CET-SET test-takers. It addresses the following sub-questions:

- a. Do different teacher groups differ in consistency and severity of rating students' oral test performance?
- b. How do different teacher groups differ in defining the oral proficiency construct as manifest in CET-SET task performance?

Data for the study are derived from three sources: CET-SET scores given by teachers in a holistic approach, and written comments and survey responses from teacher participants. Findings, drawn from both quantitative (FACETS) and qualitative (content analysis) analyses of teacher data, reveal differences in the way each teacher group defines and measures the oral proficiency construct, suggesting that teachers' language backgrounds may have a biasing effect on the judgement of test performance. Implications are drawn for CET-SET test validity in particular and for oral proficiency testing more generally.

**PRESENTER'S BIO STATEMENT:** Ying Zhang is a doctoral candidate in TESOL (Language Testing) at Monash University, Australia. She is an associate professor at the College of Foreign Languages, Tianjin Normal University, China, teaching English there for 13 years. Her research interests include language teacher education, EFL teaching, and language testing.