

## **ALTAANZ Conference 2016**

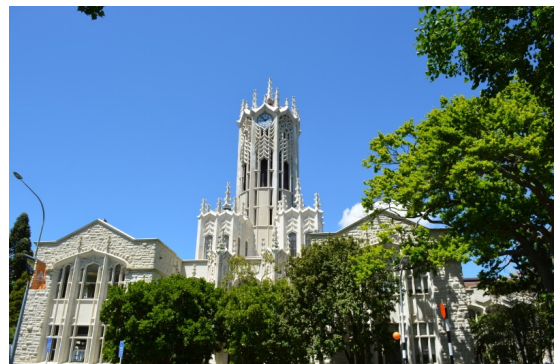
In the classroom and beyond:  
Assessing language ability in different contexts



**November 17–19, 2016**  
**The University of Auckland**  
**New Zealand**

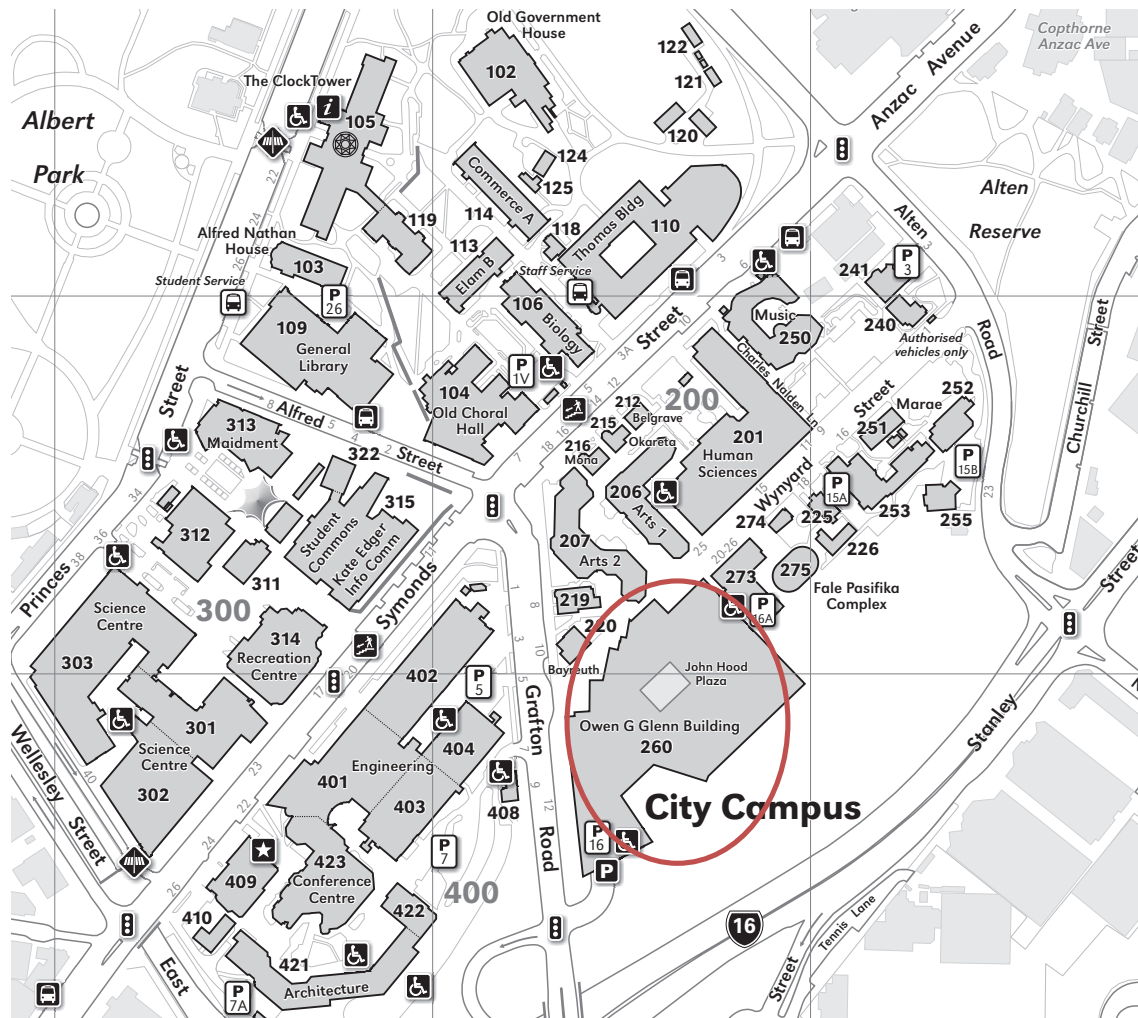
# TABLE OF CONTENTS

1.	Map Showing the Conference Venue	3
2.	Welcome Message from the ALTAANZ Presidents	4
3.	About ALTAANZ	5
4.	ALTAANZ Conference Committees and Groups	6
5.	Sponsors and Supporters	8
6.	Important Information	9
7.	Pre-Conference Workshops	11
8.	Conference Day 1	13
9.	Conference Day 2	15
10.	Teachers' Forum	18
11.	Plenaries	19
12.	Individual paper abstracts	24
13.	Papers in Language Testing and Assessment	41
14.	PLTA Best Paper Awards 2013–2015	41
15.	DELNA	42
16.	PTE Academic	43
17.	IELTS Research Grants	44
18.	TOEFL	45
19.	The British Council	46



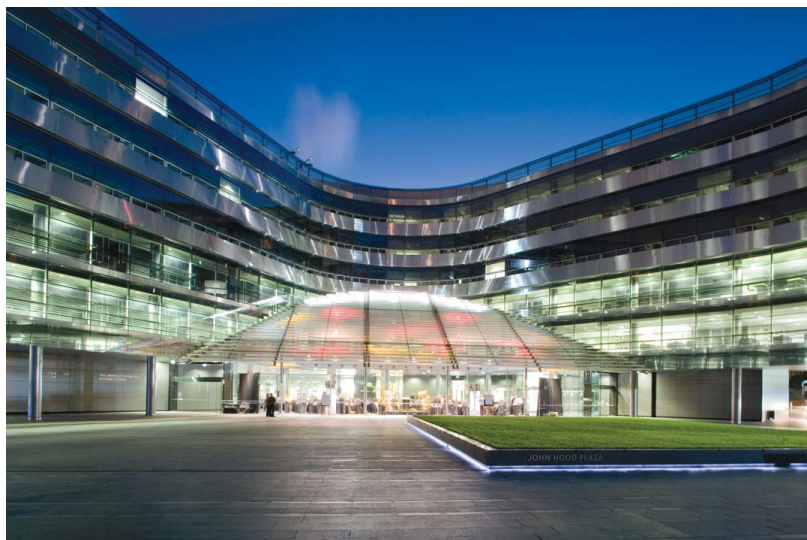


# MAP SHOWING THE CONFERENCE VENUE



Directions: The Conference will be held in the Owen G Glenn Building (260) on Grafton Road. This is the University's Business School. You can access paid parking from Grafton Road as seen on the map. Weekday earlybird (entry before 10.30am, exit by 6.30pm) is \$12; weekend flat rate is \$6.

For a larger map, please visit the University of Auckland website, [www.auckland.ac.nz](http://www.auckland.ac.nz) and search for maps.



# WELCOME

We would like to warmly welcome everyone to the ALTAANZ conference 2016 in beautiful Auckland, New Zealand. We are particularly excited as this is the first time that this key event in the calendar of the association will be held in New Zealand. We are looking forward to sharing ideas, furthering our mission of research, training and policy formation and continuing to build a community of language assessment specialists in our region, and in particular, to learn more about the assessment work of colleagues in New Zealand. We are excited to see that this year's conference has attracted delegates from many regions of the world, including countries in Asia, Europe, Africa and North America.

The conference theme 'In the classroom and beyond: assessing language ability in different contexts' reflects ALTAANZ's aim of connecting classroom teachers and researchers. This is reflected in a special day scheduled for teachers. We are hoping that these two-way conversations between teachers and researchers will continue to inform the practices of both groups. We are also happy to see, for the first time at this conference, a student-organised event and we are hoping that many students will use this opportunity to network and make lasting friendships.

The program promises high quality presentations around a large range of language assessment-related topics, including both high-stakes standardised assessments, and formative assessment practices in the classroom. Our four keynote speakers, Associate Professor Matt Poehner (Penn State University), Professor Barry O'Sullivan (British Council), Dr Peter Keegan (University of Auckland) and Dr Ute Knoch (University of Melbourne), bring a wealth of combined experience in many contexts of language assessment.

We owe special thanks to the sponsors, IDP IELTS, TOEFL, British Council, Pearson, the University of Auckland, Cactuslab, and DELNA for their generous sponsorship of the conference. These contributions are extremely important and make it possible for a young organisation such as ALTAANZ to continue to grow and to keep registration fees for students and teachers reasonable.

We would like to thank the conference organising committee in Auckland, in particular the two co-chairs John Read and Janet von Randow. Organising such an event is challenging and takes many months of planning.

We hope that you find the experience fruitful professionally and personally and that you have a pleasant stay in Auckland. We are also hoping you will be inspired to return to Auckland in 2017 where ALTAANZ will be co-organising a conference strand in the joint ALAA/ALANZ/ALTAANZ conference and again in 2018 for the prestigious Language Testing Research Colloquium, the annual conference of the International Language Testing Association with which ALTAANZ is affiliated.

Angela Scarino & Ute Knoch  
ALTAANZ Co-Presidents

# ABOUT ALTAANZ

The purpose of the Association for Language Testing and Assessment of Australia and New Zealand (ALTAANZ) is to promote best practice in language assessment in educational and professional settings in these two countries and to foster collaboration between academia, schools and other agencies responsible for language testing or assessment. Its goals are listed under three broad headings below:

## Training:

Stimulate professional growth and best practice in language testing and assessment through workshops and conferences.

## Research:

Promote research in language testing and assessment through seminars, conferences and/or publications (ALTAANZ publishes a web-based journal and a newsletter).

## Policy formation/advice:

Provide advice on assessment to public and other relevant agencies on assessment-related issues, and advocate on behalf of test-takers, students and other stakeholders whose life chances may be affected by assessment-related decisions.

For further information about the organisation, please visit the website at: <http://www.altaanz.org/>.

To become a member of ALTAANZ, please download a membership form from the website and email it to [altaanz@gmail.com](mailto:altaanz@gmail.com).

## ALTAANZ Committee

### Co-Presidents

Dr Ute Knoch (University of Melbourne)

Associate Professor Angela Scarino (University of South Australia)

### Vice President

Associate Professor Aek Phakiti (University of Sydney)

### Secretary

Dr Katherine Quigley (Victoria University of Wellington)

### Treasurer

Denise Angelo (The Australian National University)

### PLTA Editors

Dr Sally O'Hagan (University of Melbourne) and Dr Lyn May (Queensland University of Technology)

### Communications officer (website and newsletter)

Dr Johanna Motteram (University of Adelaide)

### Student representatives

Xiaohua Liu (University of Auckland) and Megan Yucel (University of Queensland)

# ALTAANZ CONFERENCE COMMITTEES & GROUPS

## Conference Organising Committee

Co-chairs:

John Read and Janet von Randow, University of Auckland

Karen Ashton	Massey University
Jenni Bedford	University of Auckland
Morena Botelho de Magalhães	University of Auckland
Darren Conway	Languages International, Auckland
Rosemary Erlam	University of Auckland
Peter Gu	Victoria University of Wellington
Shellee Hall	
Maree Jeurissen	University of Auckland
Peter Keegan	University of Auckland
Margaret Kitchen	University of Auckland
Ute Knoch	University of Melbourne
Xiaohua Liu	University of Auckland
Katherine Quigley	Victoria University of Wellington

Assisted by:

Amy Edwards	Event Services, University of Auckland
Annemiek Huisman	LTRC, University of Melbourne
Johanna Motteram	University of Adelaide
Martin von Randow	University of Auckland

## Best Student Presentation Award Committee

Katherine Quigley (Chair)	Victoria University of Wellington
Rosemary Erlam	University of Auckland
Peter Gu	Victoria University of Wellington
Margaret Kitchen	University of Auckland

## Abstract Reviewers

Karen Ashton	Massey University
Ana Maria Ducasse	RMIT University
Martin East	University of Auckland
Cathie Elder	University of Melbourne
Rosemary Erlam	University of Auckland
Kellie Frost	University of Melbourne
Peter Gu	Victoria University of Wellington
Michael Harrington	University of Queensland
Kathryn Hill	University of Melbourne
Noriko Iwashita	University of Melbourne
Peter Keegan	University of Auckland
Ute Knoch	University of Melbourne
Susy Macqueen	Australian National University
Tim McNamara	University of Melbourne
Sally O'Hagan	University of Melbourne
Aek Phakiti	University of Sydney
John Pill	American University of Beirut
Katherine Quigley	Victoria University of Wellington
John Read	University of Auckland
Carsten Roever	University of Melbourne

**Best PLTA Paper Selection Committee**

Aek Phakiti (Chair)	University of Sydney
Susy Macqueen	Australian National University
Angela Scarino	University of South Australia
Rosemary Wette	University of Auckland

**Winner, Best PLTA Paper 2013–2015**

Knoch, U., & Elder, C. (2013). A framework for validating post-entry language assessments (PELAs). *Papers in Language Testing and Assessment*, 2(2), 48–66.

**Student Travel Award Committee**

Aek Phakiti (Chair)	University of Sydney
Loc Nguyen (former award winner)	Victoria University of Wellington
Sally O'Hagan	University of Melbourne

**Winners, Student Travel Award 2016**

Simon Davidson	University of Melbourne
De Phung	University of New South Wales

**Student Volunteers**

Doctoral candidates from Applied Language Studies and Linguistics, University of Auckland.

Vincent Greenier  
Priscilla Shak  
Vivian Qiong Wang  
Miyoun Song  
Xiaoming Xun



School of Cultures,  
Languages and Linguistics

**ARTS**

**IELTS**<sup>TM</sup>

 **TOEFL**<sup>®</sup>

 **BRITISH  
COUNCIL**

**PTE** ACADEMIC<sup>TM</sup>



 **Cactuslab**

*Delna*  
Diagnostic English Language Needs Assessment



# IMPORTANT INFORMATION

## Registration

Thursday	17th November 8am–5pm in the foyer, Level 0, OGGB
Friday	18th November 8am onwards in the foyer, Level 0, OGGB
Saturday	19th November 8am onwards in the foyer, Level 0, OGGB

## Workshops

Thursday	17th November 9am–12pm and 1–4 pm, Level 0, OGGB, Rooms 040B and Computer Lab 5
----------	---

## Conference

Friday 18th – Saturday 19th November 8:50am–5:30pm, Level 0 OGGB

- **Opening Reception** – Thursday 17th November 5pm in the foyer, Level 0, OGGB
- **AGM** – Friday 18th November 1:25–2:25pm, Level 0, OGGB5
- **Conference Dinner** – Friday 18th November 6:45pm, Level 3, OGGB, Decima Glenn Room
- **Students' lunch** – Saturday 19th November 12–1pm, Level 0, OGGB, Room 008

## Guidelines for presenters

- 1) **Paper presentations** – These cover a range of topics and focus on both research and assessment. Presenters are researchers and practitioners from diverse sectors. Presentations will be 20 minutes, followed by 10 minutes' discussion. Please save your slides on a USB stick and copy the file to the computer in your room well before your talk.
- 2) **Work in Progress (WIP)** Saturday 19<sup>th</sup> November 11am–12pm  
This session gives researchers and teachers the opportunity to share and discuss aspects of research projects in which they are currently involved, or classroom assessment tasks they have developed. In doing so, they will receive feedback from conference attendees. The one-hour session will be divided into three lots of 20 minutes, so that each presenter has the opportunity to discuss their project/task with 3 groups of interested conference attendees. Each room will have several presenters, and each presenter will be stationed at a table in a different area of the room.

## Facilities on campus

**ATMs** : Level 1, OGGB, by ASB, or Student Commons, Buildings 315 and 322 on the map, page 3 of conference handbook: ANZ. ASB. BNZ, Westpac.

**Post Office & Pharmacy** – 8am–6pm weekdays, Kate Edger Commons, Building 315

**Recreation Centre** (Open to public) Building 314, Symonds St. Opening hours: Monday – Thursday 6am – 9.30pm, Friday 6am to 8.30pm, Weekend 7am – 6.30pm Please take your conference name tag to show to get the special casual rate for conference delegates of \$10 a session.

**Car park** (Open to public) Under OGGB, entrance from Grafton Rd. Weekday earlybird (entry before 10:30am, exit by 6:30pm) is \$12; weekend flat rate is \$6.

**Internet access:** information will be given to you at the registration desk as you collect your conference bag.

***Please note that lunch and morning/afternoon tea will be served at the conference venue during the conference on Friday and Saturday.***

## Cafes on campus –

### Excel Café, Level 1, OGGB

Shaky Isles (7:30am–8pm weekdays and 9am–4:30pm Saturday) in the Kate Edger Commons, Building 315  
Tank (8am–7pm) weekdays, Kate Edger Commons, Building 315

## Our main City Campus food court caters for a range of tastes.

Hello Chinese; Uni Sushi; Uni Kebab; Jewel of India.

**Location:** AUSA Quad, Building 322, City Campus.

**Opening hours:** Monday–Friday, 8:30am–7pm.

**Restaurants near the campus**

There are a number of restaurants with a variety of cuisines near the campus (approx. 15 mins walk from the conference venue) as well as on campus. A variety can be found in:

- The Viaduct, on the harbour approximately 15 minutes' walk from the conference venue
- Britomart, near the harbour, a 10 minute walk from the conference venue
- The Stables, Elliot St, behind Smith and Caugheys on Queen St

**Food Store on Campus** – Level 2 Kate Edger Commons, Building 315

**Supermarkets in the city–**

Countdown, 19–25 Victoria St West, (7am–10pm)

New World, 125 Queen St (8am–10pm)

Countdown, 76 Quay St (24 hours)

**Devonport walk**

Sunday 20th November, 2pm ferry from Ferry Building, Quay St

A walk to discover Devonport, Auckland's historic seaside town, bustling with cafés, shops and galleries. Rosemary, a Devonport resident and ALTAANZ attendee, will meet you at the ferry at 2:15 (yes, it is a short 12 minute boat ride). Plan to spend around an hour discovering historic Devonport. The walk will include scaling the summit (by road) of Mt Victoria for those who are stout hearted (walking shoes recommended). Others can explore the village. Ferries return to Auckland at a quarter past and quarter to the hour.

## PRE-CONFERENCE WORKSHOPS – THURSDAY 17TH NOVEMBER

Registration	8:30am–5pm	Level 0, Owen G Glenn Building (OGGB)
Workshops	9am–12pm	<b>Workshop 1 (040C)</b>
		Dynamic Assessment: Leveraging classroom activities to understand and support learner language development Matthew Poehner
		<b>Workshop 2 (Computer Lab 5)</b>
		Introduction to Rasch measurement using Winsteps Ute Knoch
Lunch break	12–1pm	Food and drinks can be purchased from the café on Level 1, OGGB, or across Symonds Street in the Student Commons
Workshops	1–4pm	<b>Workshop 3 (040C)</b>
		Using free online resources to develop reading texts for classroom assessment Barry O’Sullivan
		<b>Workshop 4 (Computer Lab 5)</b>
		Introduction to many-facet Rasch measurement Ute Knoch
Welcome reception	5–7pm	Level 0, OGGB

## ABSTRACTS

### **Workshop 1: Dynamic Assessment: Leveraging classroom activities to understand and support learner language development** **Matt Poehner – The University of Pennsylvania**

Second/foreign language (L2) teachers are routinely charged with promoting learners’ development in the target language while at the same time conducting formal and informal assessments of progress. These sets of responsibilities are often characterized by different practices and sets of assumptions, e.g. offering support when learners encounter difficulties may be understood quite differently during instruction versus assessment.

In Dynamic Assessment (DA), teaching and assessing are understood as interrelated features of the same activity, one that seeks to promote the development of learner L2 abilities. By jointly engaging with learners in activities that are beyond their current independent functioning, it is possible to reach a diagnosis of their emerging abilities; that is, one can glimpse abilities that have not yet fully developed but that are still ripening. At the same time, the instructional quality of this interaction may serve to continue guiding their development.

This workshop is concerned with how both a teaching and an assessment function may be pursued during a given activity. Participants will examine instances of DA interactions to identify some of the ways in which learner abilities may manifest. Principles of probing and prompting through interaction, systematicity, and shifting focus from individual to group, will be highlighted. In addition, participants will be invited to begin to plan how activities from their teaching contexts might be reorganized according to DA principles.

### **Workshop 2: Introduction to Rasch measurement using Winsteps** **Ute Knoch – The University of Melbourne**

This workshop aims to provide participants with an introduction to the basic Rasch model. The session will include a mixture of theory and hands-on practice. Differences between classical and modern test theory will be explored. The use of the statistical software Winsteps will be demonstrated, and participants will have plenty of opportunity to gain hands-on experience with the software using data provided in the workshop.

The interpretation of the output of a Rasch analysis will be one of the key foci of the session. Participants may bring their own data sets to the workshop. No prior knowledge is required and participants are not required to have an understanding of statistics or mathematics.

**Workshop 3: Using free online resources to develop reading texts for classroom assessment**  
**Barry O’Sullivan – The British Council**

Selecting appropriate texts for use in tests and for classroom activities is a critical part of a teacher’s work. It’s pretty easy to think about some of the issues that need to be considered: length, topic, difficulty. The first two of these are easily dealt with. Count the words. Know your students. The third, however, causes teachers the most grief. How can we know in advance how difficult the students are likely to find the text? Unless we can predict, even roughly, the difficulty it becomes extremely problematic to use a text meaningfully in a test.

Some language testing companies have been dealing with this problem in a systematic way for years, others still take a “we’re the experts, we just know” approach. In my own work, I don’t like to leave anything to chance and like to have as much information about a text as I possibly can before I even consider using it in a test. I very much believe that we should work hard to establish some measures of texts to help build a picture of its likely difficulty. Luckily, there are a number of resources available to the teacher and tester that are free and generally quite easy to use.

In this workshop, we will work with a number of these resources to build a useful picture of what an ideal text should look like. Using materials supplied by participants, we will create a specification template that is targeted at specific classes or groups of learners. Participants are asked to bring along a number of texts that have been successfully used for a specific class or level in the past (4 to 6 would be good) as these will be used to build the specification.

**Workshop 4: Introduction to many-facet Rasch measurement**  
**Ute Knoch – The University of Melbourne**

This workshop aims to provide participants with an introduction to many-facet Rasch measurement using Facets. The session will include a mixture of theory and hands-on practice. Differences between classical and modern test theory will be explored, with a particular focus on understanding the effects raters can have on the outcomes of an assessment. The use of the statistical software Facets will be demonstrated and the interpretation of the output will be explored using data provided in the workshop. Participants are also able to bring their own data sets to the workshop. No prior knowledge is required and participants are not required to have an understanding of statistics or mathematics.

## CONFERENCE DAY 1 – FRIDAY 18TH NOVEMBER

Registration open	8–8:50am	Level 0, Owen G Glenn Building (OGGB)			
Welcome	8:50–9:15am	OGGB 051			
Plenary address 1, 9:15–10:15am, <b>Matt Poehner: Dynamic Assessment and Vygotsky’s unrealized vision of developmental education (OGGB5)</b>					
		Stream A: Case Rm 2	Stream B: Case Rm 3	Stream C: Case Rm 4	Stream D: OGGB 5
		<b>Second language proficiency models and implications for developing assessment tasks</b>	<b>Standardised tests: development, implementation and/or use by institutional stakeholders</b>	<b>Assessment for the learning of indigenous languages</b>	<b>Assessing language for academic purposes</b>
Parallel session A	10:20–10:50am	<b>Karen Huang</b> Developing a Chinese placement test for heritage students in tertiary education: Issues and concerns	<b>Jessica Wu</b> Evaluating score reporting practice for two large-scale EFL tests: Intended goal and actual use	<b>Peter Keegan</b> Trialling a Māori language pronunciation tool based on a Māori speaker database	<b>Pamela Humphreys</b> Theoretical and conceptual models of academic English language proficiency in higher education: Considerations for principled assessment in EAP
Morning tea, 10:50–11:15am					
Parallel session B	11:20–11:50am	<b>Ruslan Suvorov</b> Test-taking strategies during the completion of multiple-choice items from the Michigan English Test: Evidence from eye tracking and verbal reports	<b>Paul Moore</b> Cohesion in oral language test performance	<b>Jeanette King</b> Tuhinga Māhorahora: a corpus of children’s writing in Māori	<b>Kellie Frost, Ute Knoch, Annemiek Huisman</b> Setting standards on a post-entry language assessment: Exploring differences in values of content lecturers and academic skills staff
		<b>Second language proficiency models and implications for developing assessment tasks</b>	<b>Standardised tests: development, implementation and/or use by institutional stakeholders</b>	<b>Classroom-based assessment: issues and practice</b>	<b>Assessing language for academic purposes</b>
Parallel session C	11:55am–12:25pm	<b>Lin Lin</b> Investigating relationships between second language test takers’ strategy use and Chinese reading comprehension test performance	<b>Fawzi Al Ghazali</b> Investigating the Washback Effect on Language Proficiency: A Case Study from an Arab Context	<b>Anne Moir Scott</b> Locating the Learning: Measuring the impact of L1 reflection on L2 development	<b>Leila Iranmanesh</b> Evolution of formative assessment in an English academic writing class: The role of emotion and power

		<b>Second language proficiency models and implications for developing assessment tasks</b>	<b>Standardised tests: development, implementation and/or use by institutional stakeholders</b>	<b>Classroom-based assessment: issues and practice</b>	<b>Assessing language for academic purposes</b>
Parallel session D	12:30–1pm	<b>Megan Yucel</b> Narrative inquiry in language assessment research		<b>Rosemary Erlam</b> Using evaluation to promote change in language teacher practice	<b>Naoki Ikeda</b> Assessing L2 learners' oral pragmatic and interactional abilities for university settings: Implications for classroom assessment
Lunch, 1–2:25pm, Level 0; <b>ALTAANZ AGM</b> , 1:25–2:25pm, OGGB 5					
Parallel session E	2:30–3pm	<b>Miki Tokunaga</b> Effect of time pressure on grammaticality judgment tests with L1 translation	<b>Jan Eyre</b> Starting Points Listening: An online assessment for beginning English language learners	<b>Peter Davidson</b> Assessing EAP: The case for authentic assessment	<b>Xiaohua Liu</b> Analysing existing reading test tasks: Implications for developing tasks to measure different reading abilities
Parallel session F	3:05–3:35pm		<b>Jinsong Fan</b> Factor structure and factorial invariance of a university-based English test: A longitudinal study	<b>Peter Gu</b> Creating and validating the Classroom Assessment Confidence Index among Chinese EFL teachers	<b>Michael Mersiadès</b> Validity of EAP reading test inference items: A pilot study
			<b>Standardised tests: development, implementation and/or use by institutional stakeholders</b>	<b>Assessing language for professional purposes</b>	<b>Assessing language for academic purposes</b>
Parallel session G	3:40–4:10pm		<b>Michelle Czajkowski</b> Judgements of writing proficiency by non-native and native English speaking teachers: Comparing holistic and analytical scoring	<b>Susy Macqueen</b> Profession-specific language standards: Perspectives from professional bodies on the use of language tests	<b>Shahrazad Saif, Zahra Mahdavi</b> Language needs of international graduate students working as teaching assistants (ITAs) in Canadian Francophone universities: Implications for Assessment
Afternoon tea, 4:10–4:25pm					
Plenary address 2, 4:30–5:30pm, <b>Peter Keegan: Māori language testing and assessment in Aotearoa: past, present and future prospects (OGGB 5)</b>					
Group photos, 5:30–5:45pm					
Conference dinner, 6:45 onwards, Level 3, OGGB, Decima Glenn Room					



## CONFERENCE DAY 2 – SATURDAY 19TH NOVEMBER

Registration open	8–8:50am	Level 0, Owen G Glenn Building (OGGB)			
Welcome	8:50–9am	Housekeeping			
Plenary address 3, 9:00–10:00am, <b>Ute Knoch: Measuring writing development: Implications for research and pedagogy (OGGB 5)</b>					
		Stream A: Case Rm 1	Stream B: Case Rm 2	Stream C: Case Rm 3	Stream D: Case Rm 4
		<b>Classroom-based assessment: Issues and practice</b>	<b>Standardised tests: development, implementation and/or use by institutional stakeholders</b>	<b>Expanding existing, and creating new, validity frameworks for language assessment</b>	<b>Classroom-based assessment: Issues and practice</b>
Parallel session H	10:05–10:35am	<b>Laurie Lu</b> Coming to grips with technical issues in developing and implementing EAP Unit standards assessments	<b>Matthews M Makgamatha, Kathleen Heugh</b> Multilingual Assessment: opportunities for teacher development and equitable learning	<b>Sheryl Cooke</b> Selecting the gatekeepers: the fairness dimension of language proficiency requirements for language assessors	<b>De Phung</b> What did EAL/D teachers actually think and do when marking oral performances?
Morning tea, 10:35–10:55am					
		Stream A: Case Rm 1	Works in Progress I: Seminar 40B	Works in Progress II: Seminar 40C	Room D Case Rm 4
Parallel session I	11–11:30am	<b>Bernadette Barker</b> Assessing Language through tasks in the classroom: A process for determining whether a student has additional needs to language learning in English	<b>Apisak Sukying</b> An Investigation of Receptive and Productive Affix Knowledge and its Relatedness to Vocabulary Size in Thai EFL Learners  <b>Wenjing Yao</b> Language assessment versus language testing-a comparative study of language assessment courses in New Zealand and China	<b>Miki Tokunaga</b> Comparing grammar knowledge and production of Japanese EFL learners  <b>Morena Dias Botelho de Magalhães</b> Investigating use of a screening tool for recommendations regarding compulsory academic English language courses	<b>Karen Ashton</b> It's a juggling act: Assessing learners in the multi-level language classroom
Parallel session J	11:35am–12:05pm	<b>Sue Edwards</b> Issues for New Zealand primary mainstream teachers assessing English language learners			<b>Faisal Faisal</b> Assessing writing: a certified teacher's perception (a preliminary research finding)
Lunch and students' meeting, 12:05–1pm					

		Stream A: Case Rm 1	Stream B: Case Rm 2	Stream C: Case Rm 3	Teachers' Forum: OGGB 5
		Classroom-based assessment: Issues and practice	Assessing language for professional purposes	Expanding existing, and creating new, validity frameworks for language assessment	See Page 17 for details
Parallel session K	1:05–1:35pm	Keiko Nakao Reflective assessment tasks for intercultural language learning in a beginner foreign language course	Simon Davidson Can doctors set valid standards on an ESP test for health professionals?	Albert Weideman The refinement of the idea of consequential validity within an alternative framework for responsible test design	
Parallel session L	1:40–2:10pm		Cate Gribble English language proficiency and employment, migration and professional registration outcomes in healthcare and early childcare education in Australia	Christina Judy Fernandez Test takers' speaking strategies: "I do, I think and I think about my thoughts because..."	
Parallel session M	2:15–2:45pm		Paul Crump Fitness to practise: Revising the Occupational English Test Listening component	Kazuo Amma Partial scoring of sequencing tasks with distance penalty	
Afternoon tea, 2:45–3:10pm					
Closing plenary (address 4), 3:15–4:15pm, Barry O'Sullivan: Making consequence happen (OGGB 5)					
Conference closing, 4:15–4:35pm					

# TEACHERS' FORUM – SATURDAY 19TH NOVEMBER

This session on Saturday, 1:05–2:45pm in OGGB 5 will see practising primary and secondary school teachers raising assessment issues they face in their day-to-day work with English Language Learners.

Jan Tagaloa (Primary Bilingual)

*How should standardized assessments be carried out in bilingual settings?*

Jacqui Lindsay and Inge Millard (Mainstream Primary)

*How can we access or conduct bilingual or other assessments to identify specific learning needs of English Language Learners (ELLs) in mainstream settings?*

Simon Crosby (Secondary ESOL)

*What assessment accommodations for ELLs in mainstream settings are valid and practicable?*

Rosemary Gillies (Primary ESOL)

*What recommendations would you make for initial placement assessment for newly arrived ELLs?*

Jenni Bedford (in consultation with primary and secondary teachers)

*What needs to be taken into consideration when creating reading assessments for ELLs?*

Researchers, language assessment specialists and Ministry of Education representatives will respond to these issues from a variety of theoretical and practical perspectives, and there will be the opportunity for follow-up discussion from the floor.

The forum will be chaired by Maree Jeurissen, President of TESOLANZ, Jenni Bedford and Margaret Kitchen.

In addition, there will be two strands of parallel sessions for teachers on Saturday from 10:05am until 12:05pm: **Classroom-based assessment: Issues and practice** (Room A, Case Rm 2; and Room D, OGGB 5).

# PLENARIES

## Plenary 1: “Dynamic Assessment and Vygotsky’s unrealized vision of developmental education”

Matthew E. Poehner, The Pennsylvania State University



Day: Friday 18th November, 2016

Time: 9:15–10:15am

Room: OGGB 5

Dynamic Assessment (DA) refers to the administration of an assessment in which the conventional approach of observing learners as they independently complete tasks is abandoned and the assessor, or mediator, intervenes when learners experience difficulties to offer prompts, feedback, leading questions, and other forms of support. The rationale behind this departure from accepted assessment practice is that the degree of external support learners require to overcome problems reveals the extent to which relevant abilities have begun to develop. In short, learners who fail independently but are successful with minimal intervention are developmentally more advanced than those requiring more intensive support. Proponents of DA argue that it thus provides a more nuanced picture of learner abilities while also pointing to the forms of support that were most beneficial to individuals, thereby offering a starting point for subsequent instruction (e.g. Feuerstein, Falik & Feuerstein 2015).

For nearly half a century, DA has been pursued in psychology and cognitive education with a wide range of populations (Lidz & Elliott 2000; Sternberg & Grigorenko 2002), and for more than a decade it has been undertaken in L2 educational contexts (Lantolf & Poehner 2014). Despite its considerable promise and extensive research literature, DA has yet to become a fixture of mainstream education. In this paper, I propose that two issues in particular have impeded realization of DA’s potential and must be addressed. The first derives from traditional divisions between formal testing and day-to-day classroom teaching and learning. Outside of the L2 field, DA has primarily been applied by assessment specialists, with the result that insights gained from procedures frequently do not lead to changes to teaching practice (see Haywood & Lidz 2007; Tzuriel 2011).

A second problem, which pertains equally to general education and L2 teaching, concerns the use of DA to target development of learner abilities in contexts where the curriculum is not guided by a theory of development but instead emphasizes memorization and skills. Following an overview of DA’s theoretical origins in L. S. Vygotsky’s writings (1987, 1998), I argue that engagement with the Zone of Proximal Development as a framework for cooperative educational activity offers a way forward. Examples are presented of DA conducted in both L2 formal testing and classroom learning situations, with discussion of how these may function in tandem to continually monitor learner progress. In addition, recent research in the area of L2 Mediated Development (Poehner & Infante 2015, 2016) is highlighted to capture how curricular revisions might further learner appropriation of knowledge about the language in an effort to enhance their capacity to regulate their L2 use.

### Biography

*Dr Matthew E. Poehner is Associate Professor of World Languages Education and Applied Linguistics at The Pennsylvania State University (Penn State). He has taught French as a foreign language and English as a second language in U.S. K-12 schools as well as at private institutions and universities. After completing his PhD in Applied Linguistics at Penn State, he has continued to work at that university. In his current position (since 2008) he directs the teacher education program for candidates pursuing certification to teach a world language in the K-12 school system and also contributes to the doctoral programs in Curriculum and Instruction and Applied Linguistics.*

*Dr Poehner’s research examines the use of Sociocultural Theory, as conceived by Russian psychologist L. S. Vygotsky, as a basis for second language educational practices, including Dynamic Assessment, Mediated Development, and Systemic-theoretical Instruction. Much of Dr Poehner’s work has focused specifically on*

*Dynamic Assessment as a framework for organizing interactions with learners in order to simultaneously diagnose their abilities and promote their continued development. His research has involved partnerships with language teachers, learners, and program directors and has been supported through grant awards, particularly through the Center for Advanced Language Proficiency Education and Research (CALPER) at Penn State, which is funded through the U.S. Department of Education. More recently, Dr Poehner was co-principal investigator for a project funded by the U.S. Department of Education to develop computerized tests of comprehension in Chinese, French, and Russian using principles of Dynamic Assessment.*

*Dr Poehner is widely published in the areas of second and foreign language teaching, language assessment, theories of second language acquisition, and applied linguistics. He is the author, co-author, or editor of four books, including Sociocultural Theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice divide (2014), which received the 2015 Kenneth W. Mildenberger Prize from the Modern Language Association. An earlier book, Dynamic Assessment: A Vygotskian approach to understanding and promoting second language development (2008), was a finalist for the Outstanding Book Award through the British Association for Applied Linguistics. Dr Poehner has published more than thirty scholarly book chapters and journal articles. His work has appeared in venues including TESOL Quarterly, Language Teaching Research, The Modern Language Journal, Language Testing, and the International Journal of Applied Linguistics. In 2008, he received the Pimsleur Award for Outstanding Research Contribution from the American Council on the Teaching of Foreign Languages.*

## Plenary 2: “Māori language testing and assessment in Aotearoa: past, present and future prospects”

Peter Keegan, University of Auckland

Day: Friday 18th November, 2016

Time: 4:30–5:30pm

Room: OGGB 5



Despite a long history of teaching Māori as a subject and re-introducing Māori as a medium of education since the late 1980s, there have been few developments of robust tools for assessing Māori language. The only standardized instrument is the e-asTTle Māori numeracy and literacy online assessment tool for Māori-medium students in the compulsory school sector. A recent development is the Ministry of Education sponsored Kaiaka Reo Māori oral language proficiency tool. However, most projects, including the University of Auckland’s longitudinal study ‘Growing Up in New Zealand’, have had to adapt existing tools for measuring the proficiency of younger speakers of Māori.

This presentation will provide an overview of recent Māori language testing and assessment in Aotearoa/ New Zealand. Despite government and community efforts to increase the numbers of speakers of Māori, Census results clearly indicate that the language is declining. For many Māori-medium students, the school remains the only domain where Māori is used exclusively; home and community activities for most tend to be conducted in English. This means that it is difficult to define what represents first (or “native”) language proficiency in Māori for younger learners. Although Māori dialects show very little variation linguistically, many second language learners have begun to infuse their pronunciation and written Māori with features that are characteristic of a particular tribe or region. However, most of the Māori materials produced tend to follow a de facto standardized Māori. The presentation describes the tools that have been developed for assessing Māori, including work in progress. It concludes with a discussion of ongoing issues, such as a lack of developers/practitioners with appropriate technical knowledge, and suggests priorities for future development.

### Biography

*Dr Peter J. Keegan (Waikato-Maniapoto, Ngāti Porou) is a senior lecturer in Te Puna Wānanga (school of Māori Education), the Faculty of Education and Social Work, the University of Auckland, New Zealand. His research interests include the structure, changes and current use of Māori language, assessment/ measurement and language testing especially in indigenous language contexts, Māori/indigenous medium education and indigenous literatures. He worked as a project manager for asTTle (assessment tools for teaching and learning), a New Zealand computer based online numeracy and literacy assessment tool in English and Māori (<http://e-asttle.tki.org.nz/>). He teaches courses on assessment for teaching and learning, and serves on government advisory panels on assessment. Peter is a co-editor of *Teachers voyaging in plurilingual seas: Young children learning through more than one language* (New Zealand Council for Educational Research, 2016). Current research projects include developing an online Māori language speech pronunciation aid (MPai) and trying to visualize Māori language census data (<http://peterjkeegan.github.io/>). Other interests and activities include spending time with whānau, reading, computers, kayaks, hiking and trying to keep fit.*



### Plenary 3: “Measuring L2 writing development: implications for research and pedagogy”

Ute Knoch, University of Melbourne

Day: Saturday 19th November, 2016

Time: 9–10am

Room: OGGB 5



L2 writing development has received both implicit and explicit attention in different areas of second language research such as second language acquisition and L2 writing pedagogy for many years, although the different research strands often do not overlap much in terms of the definitions used and the methodological choices made. Many studies have narrowly focussed on linguistic variables, such as the development of accuracy, fluency and complexity. In a recent edited volume, Manchon (2012) calls for a broader conceptualisation of writing development, examining broader aspects in writing such as discourse structures, content and genre knowledge.

In this presentation, I will focus on the kind of work that has been undertaken in the area of L2 writing development both in research and in classroom contexts. By drawing on a range of studies, I will show that there are several possible spheres in which writing can develop. There are also a number of purposes for measuring writing development. I argue that unless the methodology chosen matches the sphere of writing development and the purpose of measuring development, the measurement will have limitations for the stakeholders. I propose that conceptualizing writing development in this way, will help clarify operational definitions applied and tighten measurement designs employed and ultimately broaden the type of investigations undertaken in both research and educational settings.

#### **Biography**

*Dr Ute Knoch is the Director of the Language Testing Research Centre at the University of Melbourne. She has published widely with over 40 peer-reviewed publications which have been published in journals such as Language Testing, Language Assessment Quarterly, TESOL Quarterly, Applied Linguistics, Assessing Writing, Journal of Second Language Writing and English for Specific Purposes. Her research interests are in the area of writing assessment, rating processes, assessing languages for academic and professional purposes, and placement testing. She is currently the Co-President of the Association for Language Testing and Assessment of Australian and New Zealand (ALTAANZ) and has served on the Executive Board of the International Language Testing Association (ILTA) from 2012 to 2015. In 2014, Dr Knoch was awarded the TOEFL Outstanding Young Scholar Award by the Educational Testing Service (Princeton, US), recognizing her contribution to language assessment.*

#### Plenary 4: “Making Consequence Happen”

Barry O’Sullivan, The British Council

Day: Saturday 19th November, 2016

Time: 3:15–4:15pm

Room: OGGB 5



Consideration of the social consequences of test use has been a central theme in validation theory since Messick (1989) brought the idea into his model of validity. While the negative impact of test use has quite often been stressed, little meaningful attention has been paid to how test developers might operationalise the concept of consequence in the test development process. Where consequence has been addressed, it has tended to be as an a posteriori evidence source, primarily concerned with test impact. The reality is we do not know what consequence means to test development.

In this paper I will first outline how the social cognitive validation model has been developed over the past decade or more, describing how it has informed test conceptualisation, development and validation. While the earlier versions of the model proved to be of practical use to test developers, it failed to recognise the importance and place of consequence in the process. This is particularly clear in the way in which Weir (2005) conceptualised what he, and others, referred to as consequential validity as one of the final elements to be brought into play in development and validation.

Over time both Weir and O’Sullivan have revisited the model, and in the latter’s most recent interpretation (2014, 2016) finally attempted to operationalise consequence in a meaningful way. This version of the model sees consequence as being specifically related to the context of test use, which itself is defined by the key stakeholder groups who comprise that context. In order to understand how the contexts impact on the test, it is necessary to take relevant stakeholders into account when conceptualising the test itself. This has the effect of informing us how test construct is to be operationalised. It will also inform all of the decision-making that is made in the process of test development.

Finally, it will impact on how validation evidence is presented. This latter is critically important, since traditionally validation arguments have been written with no specific audience in mind or were aimed at an academic audience — or, since Kane (1992), at a legal one.

By conceptualising consequence in the way suggested here we must accept that validation arguments should be targeted squarely at a whole range of specific stakeholder groups. This will impact on structure, content and delivery mode. Examples of how this is dealt with in an operational way will be presented and discussed.

#### Biography

*Professor Barry O’Sullivan is the Head of Assessment Research & Development at the British Council, London. His recent work includes the development and validation of a new business to business language test called Aptis. He is the founding president of the UK Association of Language Testing and Assessment and holds honorary and visiting chairs at the Universities of Reading and Roehampton in the UK, the University of Technology MARA (Kuala Lumpur) and at the University of Lisbon. Barry is particularly interested in issues related to performance testing, test validation, test-data management and analysis and scaling and calibration; he has conducted research into factors affecting spoken performance, assessing rater behaviour, assessing speaking and writing, specific purpose assessment, benchmarking English language tests to the Common European Framework of Reference for Languages and standard setting in professional contexts.*

*Barry’s publications have appeared in a number of international journals and he has presented his work at international conferences around the world. His books include: Issues in Business English Testing (Cambridge, 2006); Modelling Performance in Oral Language Testing (Peter Lang, 2008); Language Testing: Theories and Practices (Palgrave Macmillan, 2011) and The Cambridge Guide to Second Language Assessment (with C. Coombe, P. Davidson, and S. Stoyanoff, eds.) (Cambridge, 2012). He is currently finalising (with Cyril Weir) a major project documenting a history of language testing within the British Council to be published by Equinox in 2017, and is working on two volumes (on validity and localisation) with Micheline Chalhoub-Deville for the British Council Monographs Series (also Equinox).*

# INDIVIDUAL PAPER ABSTRACTS

**Fawzi Al Ghazali** (Abu Dhabi University, United Arab Emirates)

## ***Investigating the washback effect on language proficiency: A case study from an Arab context***

**Day:** Friday

**Time:** 11:55am–12:25pm

**Room:** Case Rm 3

Language tests are necessary to gauge learners' progress and to set future plans for improving language skills. Tests sometimes have great influence on teaching and learning that are tailored to match with the way they are drafted through a phenomenon known as the "Washback effect" (Wall & Alderson 1993). However, what are the most optimal ways to minimize the negative outcomes of the washback effect on language development?

This paper reveals that diligent analysis of learners' performance in language courses shows very striking results related to the influence of testing on teaching. The case study presented in this paper investigates the performance of some Arab learners in language courses after having graduated from high schools. The analysis shows they get high scores on school-level-related exams on which they have extensive training throughout the whole years; nonetheless, they achieve very poorly on proficiency tests like IELTS or TOEFL. This is simply because achievement tests measure how much a student learnt in a particular course; whereas doing well in proficiency tests depends on learners' general mastery of language.

Achievement tests possibly have face validity when they measure what they aimed to measure. However, in terms of reliability, there is no guarantee that learners would be able to achieve the same results even if they take the same tests after a period of time. This surface approach to learning could hardly lead to permanent knowledge and learners are expected to stumble behind when fluent communication in a foreign language is needed.

The presenter will review these issues and the presentation will provide insights and pedagogical implications into how to use tests in the most optimal ways. The audience will be invited to share their ideas and reflect on their experience of language assessment.

**Kazuo Amma** (Dokkyo University, Japan)

## ***Partial scoring of sequencing tasks with distance penalty***

**Day:** Saturday

**Time:** 2:15–2:45pm

**Room:** Case Rm 4

In language testing, reordering items (words, sentences, and paragraphs) is one of the common techniques of confirming learners' comprehension as well as reproduction of text/discourse. Yet a valid partial scoring method has not been established to date. The 'all or nothing' method, which gives a full score only when all elements are correct, is widely used at school and by nationwide tests, but it unduly diminishes the construct validity when part of the sequence is correctly arranged. A valid scoring method should reflect the complexity of the cognitive manipulation involved in the problem-solving process.

The presenter's method Maximal Relative Sequence (MRS) calculates the number of relative transitions of elements in the response to give longest sequence of transitions (Amma, 2007; 2010a; 2010b). For example, in a response "A\_C\_B\_D" the longest transition is either "A\_C\_D" or "A\_B\_D" (score = 2). This is equivalent to a technique known as Minimal Edit Distance (MED), or counting the elements of digressions (B or C) from the correct sequence, subtracted from the perfect transition size (3).

Compared with MED, the algorithm of MRS is far simpler, thus saving time to calculate a large number of samples. However, MRS does not consider the factor of distance between the right position and initial incorrect position. As a result, two sequences, "B\_A\_C\_D" and "B\_C\_D\_A" are both scored as 2 (taking up "B\_C\_D" as the MRS), even though one has to retrieve the correct answer by moving A one slot ahead in the first sequence and three slots in the second.

The present report demonstrates a new computer program that generalises the distance penalty by counting the shifts for retrieval, thus achieving a more accurate and fair measurement than MRS. An exhaustive list of scores comparing various methods will also be presented.

**Karen Ashton** (Massey University, New Zealand)

***It's a juggling act: Assessing learners in the multi-level language classroom***

**Day:** Saturday

**Time:** 11–11:30am

**Room:** Case Rm 4

Multi-level classes, where learners from different year levels working towards different National Certificate of Educational Achievement (NCEA) levels and assessments are taught together within a single language classroom, are becoming increasingly common in New Zealand secondary schools. This is cited as one of the most significant challenges currently facing teachers. The literature on differentiation suggests that while teachers understand the principles, they find it difficult to implement them effectively in the classroom.

An additional challenge is that the majority of research in this area has focused on teaching and assessing classes of mixed-ability learners within the same year level working towards the same assessment. There is a gap in the field looking at mixed-level classes, where within a single class, in addition to teaching learners across a range of abilities, teachers are expected to teach different curriculum content to learners at each year level in order to prepare them for different NCEA assessments.

This paper presents the findings from a teacher survey and interviews looking at how teachers prepare students for NCEA assessments in the multi-level language classroom. It starts by summarising teachers' feelings about teaching multi-level classes, their level of confidence in teaching them, and the main challenges and benefits experienced by teachers and learners. The remainder of the paper focuses on how teachers plan for teaching multi-level classes and explores the range of strategies that are used together with teachers' views on which strategies are the most successful in preparing learners for assessment, the least successful, and why.

**Bernadette Barker** (Brisbane Catholic Education, Australia)

***Assessing language through tasks in the classroom: A process for determining whether a student has additional needs to language learning in English***

**Day:** Saturday

**Time:** 11–11:30am

**Room:** Case Rm 1

The progress and rate of learning English can vary significantly between young EAL/D learners. The diverse experiences and backgrounds of these learners influence their English language development and learning outcomes. Teachers seeking to understand those students who are not progressing may refer students for assessment outside the classroom context, usually to Speech Pathologists or Guidance Officers. In response to referral, standardised tests, normed on mainstream school populations are used to determine whether an EAL/D student has learning or speech problems. These investigations centred on standardised testing can lead to unreliable results with students incorrectly diagnosed with a learning problem. This paper outlines a process of investigation used by Brisbane Catholic Education that utilises an action plan for teaching and learning in the classroom using task based assessment before consideration of referral to another professional.

**Sheryl Cooke** (The British Council, China)

***Selecting the gatekeepers: The fairness dimension of language proficiency requirements for language assessors***

**Day:** Saturday

**Time:** 10:05–10:35am

**Room:** Case Rm 3

A compromise in test fairness can be construed as a weakness in overall test validity (Xi, 2010). One fairness-compromising construct-irrelevant factor is rater bias, a factor that is likely affected by biased or inconsistent selection criteria for performance-based language test raters. Traditionally, first-language English speakers have been assigned to the rater role, but large-scale testing, testing in a variety of locations around the world outside of the Inner Circle countries, and a shift towards reflecting a range of World Englishes in standardised testing has seen an increase in the use of raters for whom English is not a first language. This has raised questions about rater consistency and suitability for international tests and has sparked research comparing the rating of assessors from different L1 backgrounds and levels of English proficiency (Hill 1996; Xi & Moullan 2011; Zhang & Elder 2011; Harding & Griffiths 2014).

While a certain level of proficiency in the language being assessed is clearly necessary, how this requirement is reflected in the recruitment criteria for language assessors has received little attention. This paper provides an overview of the minimum proficiency requirements for raters across a range of English language performance-based tests and considers two questions related to test fairness:

- a) Are language proficiency criteria transparent and equally applied to all rater applicants or is there bias?
- b) What are the implications of the recruitment practices?

A consideration of the possible impact of these practices on both the rater applicants and, subsequently, on the test-takers is presented in a fairness argument framework with the aim of encouraging scrutiny of the potentially exacerbating effect of selection criteria for raters as a contributing factor to test fairness and thereby test validity.

**Paul Crump** (Cambridge Assessment, United Kingdom)

***Fitness to practise: Revising the Occupational English Test Listening component***

**Day:** Saturday

**Time:** 2:15 – 2:45pm

**Room:** Case Rm 2

The Occupational English Test (OET) is an international English language test that assesses the language communication skills of healthcare professionals looking to register and practise in an English-speaking environment. It seeks to ensure that candidates are prepared, in language terms, for work in their profession.

The test is currently undergoing a revision process to ensure it is fit for purpose and reflects the latest developments in language testing research. The revised test is due to launch in 2018.

The purpose of the OET Listening sub-test is to provide test users with a valid and reliable assessment of candidates' listening abilities in a health-related context. This presentation will confine itself to changes planned to the Listening component of the test and will focus on the following areas:

1. An overview of the development of the new Listening specification and the rationale behind it.  
The test has been revised to include a wider range of patient: professional and professional: professional interactions among healthcare professionals from the same and different disciplines as well as from different L1 backgrounds. This will ensure the test adequately reflects the realities of medical practice, as it is able to include a wider range of testing focuses.
2. The introduction of scripted dialogues to the listening test.  
The new specification includes the use of scripted dialogues, based on authentic sources, rather than recordings of extemporised, semi-structured conversations. The presentation will explore how the new test aims to balance authenticity with reliability, present the rationale and research behind the move to scripted dialogues, and outline how authenticity can be maintained in a scripted task.

**Michelle Czajkowski** (University of Melbourne, Australia)

***Judgements of writing proficiency by non-native and native English speaking teachers:  
Comparing holistic and analytical scoring***

**Day:** Friday

**Time:** 3:40–4:10pm

**Room:** Case Rm 3

International English language proficiency tests have started to consider how to best reflect the status of English as a global language. However, it is still somewhat unusual for high-proficiency non-native speakers to perform in the role of productive language rater. Whether L1 background has an effect on how different raters approach the rating of L2 writing has been the subject of previous research, much of which investigates how the raters define the construct of proficiency for themselves through unguided holistic rating.

This study builds on previous research by first following the methodology of these holistic rating studies, and by secondly extending this methodology by asking participants to rate the same learner essays using detailed analytic rating scales, simulating the role of a novice rater in high stakes tests. Holistic and analytic ratings were collected from both native (n = 19) and non-native (n = 20) speaker teachers of ESL/EFL, as were comments justifying holistic scores.

The results show that rating behaviour between the two groups showed only a slight difference in either scoring method, though non-native speakers were consistently harsher in their ratings and focused more on negative aspects when rating. Analyses using FACETS showed a range of rating behaviours within the two groups with a great deal of overlap. These results suggest that native speaker teachers may not be innately advantaged. They also suggest that descriptive analytic rating scales can reduce variance between individual novice raters, regardless of L1.

**Peter Davidson** (Zayed University, United Arab Emirates)

***Assessing EAP: The Case for Authentic Assessment***

**Day:** Friday

**Time:** 2:30–3pm

**Room:** Case Rm 4

Authentic assessment is one of the key principles of language testing that is becoming increasingly prominent in the field of EAP and language assessment. In this paper I will outline exactly what authentic assessment is, contrasting it with traditional assessment. I will then look at a needs analysis to determine what is that university students actually do in the university context. This needs analysis informed the development of a range of authentic EAP assessment tasks that I will showcase. These authentic EAP tasks consist of listening and retelling a lecture, writing an essay based on reading semi-academic texts, participating in an academic discussion, and completing an information literacy project. This will be followed by a brief discussion on how to deal with the inevitable resistance you are likely to get when implementing authentic assessment for the first time. By the end of this talk I hope to have convinced participants that authentic assessment has the potential to produce more accurate assessments of a wider range of EAP constructs than traditional type testing. Furthermore, because authentic assessment utilizes assessment tasks that replicate the types of tasks that students will be required to actually do in their actual university contexts (Davidson, 2009), it is more likely to have a positive washback effect, and it is a better predictor of academic success, than the typical tests we currently implement.

**Simon Davidson** (University of Melbourne, Australia)

***Can doctors set valid standards on an ESP test for health professionals?***

**Day:** Saturday

**Time:** 1:05–1:35pm

**Room:** Case Rm 2

As part of the prerequisite to obtain professional registration and practice in Australia, overseas-qualified doctors need to demonstrate satisfactory English language proficiency. Concerns have been raised that the specified minimum level of proficiency needed on language tests used for this purpose (including the Occupational English Test (OET), a specific-purpose language test for health professionals), might be inadequate for functioning successfully in the workplace. In answer to these concerns, a study was carried out to set the minimum standards of performance on the writing sub-test of the OET via the procedure of 'standard setting'. Standard setting is a methodical process of eliciting insights from relevant stakeholders about levels of proficiency that are regarded as adequate for a particular purpose. While there is a clear need to establish defensible standards on the OET, there is a lack of research thus far on what informs judgements about such standards.

The study sought to answer the question: What aspects of OET writing performance do health professionals attend to in making their judgements and to what extent are they language based? To investigate the basis for doctors' perceptions, verbal reports in the form of a think aloud protocol (TAP) were utilised. Five doctors, all with experience of working as GPs, specialists and medical educators were asked to say 'out loud' what they think or notice while reading and judging the adequacy of 10 OET writing responses selected to represent different levels of writing ability. The doctors' comments were coded thematically and intercoder reliability checks were carried out.

The findings showed that each of the five doctors attended to similar aspects, however there was some discrepancy between what was remarked on. Some of the features mentioned appeared to have more to do with clinical competence than with the dimensions of communicative competence which the OET is designed to assess. This has implications for the construct validity of the OET standards in particular and for standard-setting in ESP testing more generally.



**Sue Edwards** (Waikato Institute of Technology, New Zealand)

***Issues for New Zealand primary mainstream teachers assessing English language learners***

**Day:** Saturday

**Time:** 11:35am–12:05pm

**Room:** Case Rm 1

A significant proportion of the NZ primary school population now comprises English language learners (ELLs). If schools wish to apply for additional funding to support ELLs' learning, mainstream teachers are required to assess their ELLs twice yearly and place learners at their 'achieved level' of the English Language Learning Progressions (ELLP) (2008). Key guidelines for teachers as they assess ELLs' language skills are that they are to "use a wide range of assessment tasks, activities and observations to make an OTJ (overall teacher judgment) with reference to the various descriptors on the ELLP matrices" and that the assessment "should not be seen as additional to the school's normal assessment schedule but as an integral part of it" (Ministry of Education 2015).

This paper describes a pilot study involving a small number of primary mainstream teachers, who were interviewed using a semi-structured interview schedule, and asked to report their practices and perceptions in regard to forming their OTJs of learners' English levels for funding eligibility. Specific questions were also asked regarding several potential issues for mainstream teachers as they assess ELLs for funding eligibility. The first of these issues was how teachers determine what the "wide range" of assessment tasks should consist of, and which of their normal assessment practices can provide appropriate evidence for rating learners' skills. A second question was how easily teachers are able to interpret the ELLP descriptors, as these include language-focussed terminology which mainstream teachers may not be familiar with. A third question was how teachers try to ensure assessment reliability, or consistency of their OTJs with those of other teachers. Teachers were also asked to provide their opinions about using the funding eligibility system, and their confidence and competence in using the system.

**Rosemary Erlam** (The University of Auckland, New Zealand)

***Using evaluation to promote change in language teacher practice***

**Day:** Friday

**Time:** 12:30–1pm

**Room:** Case Rm 4

Recent literature in teacher education has argued for a shift away from the development of teacher cognitions as a goal of teacher education to the development of core practices which would make a difference to students' lives in the classroom (Ball & Forzani 2009; Kubanyiova & Feryok 2015; Zeichner 2012). Heibert and Morris (2012) propose that these key practices would be embedded into instructional contexts and preserved as lesson plans and as common assessments.

This paper focuses on evaluation tools developed for an in-service professional development programme for language teachers (Teacher Professional Development Languages (TPDL, <http://www.tpd.ac.nz>)). TPDL is a year-long programme for teachers of foreign languages in NZ schools. Programme participants are visited by TPDL In-School support facilitators four times during the course of the year. The facilitators observe their teaching practice and then use two key documents, the "Evidence of Principles and Strategies (EPS) portfolio" and the "Progress Standards" to assist teachers to evaluate their practice against key criteria. As the year progresses the teachers are increasingly encouraged to take ownership and control of the use of these tools, so that by Visit 4, the evaluation is conducted as a self-assessment. This presentation evaluates these tools and considers evidence for their validity. Data is presented from the case study of one teacher, to further demonstrate how the tools are used and to document evidence for any change in teaching practice.

**Jan Eyre** (New Zealand Council for Educational Research, New Zealand)

***Starting Points Listening: an online assessment for beginning English language learners***

**Day:** Friday

**Time:** 2.30–3pm

**Room:** Case Rm 3

This paper will explore the development of an online vocabulary assessment for beginning English language learners. This assessment, Starting Points Listening, was recently released in trial mode as part of the Adult Literacy and Numeracy Assessment Tool (ALNAT).

ALNAT was introduced to the tertiary sector in New Zealand in 2010. It is part of a coordinated system of resources developed by the New Zealand Tertiary Education Commission to support literacy and numeracy skills development for adults. All these resources are based on the Learning Progressions for Adult Literacy and Numeracy, which provide a framework that shows stages that learners typically move through as they develop expertise in literacy and numeracy.

The purpose of the Assessment Tool is to provide robust and reliable information on the literacy and numeracy skills of adults, including English language learners. This information can be used in various ways, such as to inform the development of teaching and learning programmes, as a means of tracking learners' progress, and to enable organisations to collate and report on learner progress.

Since the introduction of the Tool, educators have called for more tailored options to meet the specific needs of their learners. In response to feedback from the ESOL sector, work began on Starting Points options in 2014. This session will explore the development of one of these options: Starting Points Listening.

We will explore the development of the Starting Points Listening from the initial consultation through to release to the sector: a period of almost two years. In particular, we will consider the challenges of designing an online language assessment for learners who may be unfamiliar with computers and who have very low levels of literacy in English.

**Faisal Faisal** (Purwokerto Muhammadiyah University, Indonesia & University of Auckland, New Zealand)

***Assessing writing: a certified teacher's perception (a preliminary research finding)***

**Day:** Saturday

**Time:** 11:35am–12.05pm

**Room:** Case Rm 4

Assessment is an important part of the teaching and learning process and plays an important role in knowing learners' problems as well as measuring their progress and outcomes after a period of learning time. Thus, the Indonesian curricula have mandated teachers to assess their learners' performance appropriately. For learners, a suitably administered assessment will encourage them to take part more actively and optimize their learning improvement. For teachers, it will help them reflect and improve their practice of instruction and measurement which later contribute to the development of a more effective classroom management.

This paper presents preliminary research findings of a certified English teacher's perception about assessment, particularly in assessing the learners' writing skill. The data are obtained from semi-structured interviews as one of the techniques to collect data in this qualitative case study involving fifteen English teachers who were randomly selected and voluntarily took part. The data are analysed by following the principles of conversation analysis and primarily to reveal how the teacher perceives the curricula's notions of assessment of writing skill and will bring them into practice.

**Jinsong Fan** (Fudan University, Australia)

***Factor structure and factorial invariance of a university-based English test: A longitudinal study***

**Day:** Friday

**Time:** 3:05–3:35pm

**Room:** Case Rm 3

Understanding the factor structure of a language test is crucial to the establishment of its construct validity (e.g. AERA, APA, & NCME, 2014). Despite the growing number of factor structure studies in the field of language testing (e.g. Gu 2014; In'nami & Koizumi 2011; Sawaki, Stricker & Oranje 2009), almost none of them used multi-year test data to investigate the factor structure and factorial invariance of language tests (see Sims & Kunnan 2016 for an exception).

Adopting longitudinal design and multi-sample Confirmatory Factor Analysis (CFA) as its analytic methodology, this study investigated the factor structure and factorial invariance of a high-stakes university-based English proficiency test. The data of this study were students' section-level scores on the test across two years (2014 and 2015). Based on relevant theories of language ability and previous research of this test, four theoretical models were specified, including a correlated four-factor model, a higher-order factor model, and two correlated two-factor models. These four models were then tested against the test data of the two years respectively, and compared with each other to determine the best-fitting model. Finally, multi-sample CFA analysis was performed to investigate whether the final best-fitting model had the same configurations on the tests across two years.

Results indicated that the higher-order factor model best fit the test data; multi-sample analysis demonstrated that this model had basically the same configurations on the tests across two years, thereby supporting the principle of measurement invariance. The results of this study are generally consistent with previous factor structure studies (e.g. Sawaki, et al. 2009), suggesting the complicated and hierarchical structure of language ability. While lending crucial empirical support to the construct validity and score-reporting policy of this English test, this study also has methodological implications for other testing agencies in their test validation and evaluation endeavours.

**Christina Judy Fernandez** (University of Sydney, Australia)

***Test takers' speaking strategies: "I do, I think and I think about my thoughts because..."***

**Day:** Saturday

**Time:** 1:40–2.10pm

**Room:** Case Rm 3

Although many studies have been undertaken to validate Bachman and Palmer's (1996) notion of strategic competence, these have mostly focused on reading, listening and writing tests and not many on the speaking test domain. It is this gap in knowledge that this qualitative study fills and contributes to. This presentation reports on a study exploring test takers' strategy use in a two-way discussion (or Part 3) of the International English Language Testing System (IELTS) speaking test. The study sought answers to the following questions:

1. What cognitive, metacognitive and communication strategies do test takers employ to enhance their test performance?
2. How do test takers perceive the usefulness of their cognitive, metacognitive and communicative strategy use to address the two-way IELTS discussion?
3. What are individual and contextual factors affecting their use of strategies during the test task?

Data were collected from 12 international students in Sydney, Australia. The participants consented to a 5-minute two-way discussion which was video recorded. This was followed by a stimulated recall session of their speaking test experience. The stimulated recalls were transcribed and coded.

Detailed analysis of strategies used by a few participants will be provided and discussed in the presentation. Some key findings and implications of this study on the two-way discussion of the IELTS speaking test will also be highlighted.

**Kellie Frost, Ute Knoch & Annemiek Huisman** (University of Melbourne, Australia)

***Setting standards on a post-entry language assessment: Exploring differences in values of content lecturers and academic skills staff***

**Day:** Friday

**Time:** 11:20–11:50am

**Room:** OGGB 5

Linguistically diverse student populations within Australian tertiary institutions have meant that a large number of tertiary education providers now assess the English language ability of all or certain targeted student groups post-entry. This is designed to ensure that students who are likely to experience difficulties are identified early and guided towards English language support opportunities available on campus. While several research studies and books have focussed on issues surrounding these post-entry language assessments (PELAs) (see e.g. Read 2014; Knoch & Elder 2013 for summaries of such research), it is not always clear how students taking such assessments are categorised in terms of requiring support. Furthermore, the documentation on these assessments rarely specifies whose standards of English language proficiency are used to categorise students, content lecturers or the academic skills staff typically responsible for providing language support across campuses.

This paper reports on a project aimed at setting meaningful standards on the Diagnostic English Language Assessment (DELA), a PELA used at the University of Melbourne since the early 1990s. The project was prompted by a recently proposed policy change to broaden the student base that would be assessed post-entry, which sparked discussions about standards on the assessment. To ensure the standards used are current and set empirically, standard-setting workshops were convened to establish the minimum required standards of English proficiency expected by:

- (1) Content lecturers from a range of disciplines, and
- (2) Academic skills advisory staff.

Standards were set by both groups for students requiring compulsory support, recommended support and no language support. The results show that the standards are remarkably similar across the two groups of stakeholders. A qualitative analysis of the reasons given for judgements during the workshops, however, showed some differences between the aspects of language valued. The findings are discussed in light of standard-setting in academic contexts.

**Cate Gribble** (Deakin University, Australia)

***English language proficiency and employment, migration and professional registration outcomes in healthcare and early childcare education in Australia***

**Day:** Saturday

**Time:** 1:40–2:10pm

**Room:** Case Rm 2

This presentation reports on a British Council, Cambridge English Language Assessment, and IDP: IELTS funded project that investigated the role of IELTS in shaping the supply of skilled labour and determining employment outcomes of international graduates and overseas trained professionals in healthcare and early childhood professions in Australia.

Focusing on two professional fields that are critical to Australia's future – health and early childhood education – the study provides key insights into the many challenges facing overseas-trained professionals and international graduates transitioning into the Australian labour market.

The study incorporated qualitative research methods to investigate the role of IELTS in determining employment and migration outcomes of overseas-trained graduates in health and early childhood. Interviews (n = 77) with professional bodies, employers, academics, overseas-trained graduates and Australian international graduates were conducted in Melbourne and two regional sites.

The results of the study highlight the complex language requirements in the professions of medicine, nursing and early childhood education and reveal key implications for stakeholders. Employers require graduates to have high-level English language skills, and universities are increasingly expected to ensure international students graduate with the required English language proficiency.

The study also reveals highly differentiated labour markets. While metropolitan hospitals are turning away both domestic and international graduates, some rural hospitals are predominantly staffed by international doctors and nurses. Early childhood education also faces skill shortages in rural areas and some areas are struggling with how to stem the flow of graduates into the primary sector.

Finally, the study highlights the many challenges facing overseas-trained/international graduates transitioning into the labour market. For participants, the challenges of working in their profession in Australia are many and varied. These challenges include workplace discrimination, isolation and extreme frustration when unable to work in their area of qualification.

**Peter Gu** (Victoria University of Wellington, New Zealand)

***Creating and validating the Classroom Assessment Confidence Index among Chinese EFL teachers***

**Day:** Friday

**Time:** 3:05–3:35pm

**Room:** Case Rm 4

This presentation outlines the process in creating and validating the Classroom Assessment Confidence Index (CACI) for EFL teachers. We will first present a framework for conceptualising 'classroom assessment competence', followed by descriptions of the design and validation process.

The conceptual framework was developed based on a comprehensive review of theories and research on assessment literacy and classroom assessment, from classic conceptualisations of assessment literacy such as the Standards for teacher competence in educational assessment of students by the American Federation of Teachers, the National Council on Measurement in Education and the National Education Association (1990) and Stiggins (1995), to more recent reformulations such as Heritage (2007), Brookhart (2011), Willis, Adie, and Klenowski (2013), and DeLuca, LaPointe-McEwan, and Luhanga (2015). The framework also refers to latest applied linguistics research on classroom assessment practices (e.g. Davison & Leung 2009; Rea-Dickins 2001; Rea-Dickins & Gardner 2000; Taylor 2009).

A 40-item Likert-scale CACI was created based upon the conceptual framework and upon an analysis of existing assessment literacy measures (Gotch & French 2014), aiming to elicit EFL teachers' self-assessment of their own confidence in classroom assessment practices. One hundred and twenty secondary school teachers teaching English as a foreign language in China participated in the validation survey. Exploratory factor analysis revealed 8 factors:

- 1) Believing in classroom assessment,
- 2) Having clear standards before assessment,
- 3) Ability to choose and design assessment tools,
- 4) Ability to do flexible online monitoring,

- 5) Ability to evaluate assessment results and diagnose problems,
- 6) Ability to adjust teaching and learning,
- 7) Ability to report to stakeholders, and
- 8) Ability in ethical assessment.

Confirmatory factor analysis was next performed to obtain model-fit statistics for the 8-factor model. Discussion will focus on the use and limitations of a measure such as the CACI for teacher education programmes.

**Karen Huang** (University of Auckland, New Zealand)

***Developing a Chinese placement test for heritage students in tertiary education: Issues and concerns***

**Day:** Friday

**Time:** 10:20–10:50am

**Room:** Case Rm 2

An increasing number of students who speak Chinese at home intend to study Chinese in tertiary institutions. However, these Chinese Heritage Language Learners (CHLLs) show a wide range of Mandarin proficiency levels. Some speak Mandarin fluently but cannot read or write; some can understand but rarely speak the language; some can read/write Chinese characters, but do not speak any Mandarin—they speak Cantonese. Although it is extensively accepted that CHLLs need to receive separate instructions, the budgetary constraints and diverse student profiles make it impossible for the programme to accommodate all of the CHLLs. Chinese programmes are forced to either place students in one of the foreign language classrooms, or turn them down from studying Chinese.

This study aims to analyse the issues and difficulties in current placement practices and propose a framework for CHLL placement. The current practices include a biographic background questionnaire, an oral interview and a written test that might be an essay or reading comprehension questions. However, these tests often fail to capture the true proficiency of these CHLLs. Their grammar knowledge might be overestimated because of their fluency and the topic that happened to be asked; while they might be underestimated due to their lower reading and writing ability. This study identified features in CHLLs' phonology and morphosyntax, and further designed a series of questions that assess CHLLs' grammatical knowledge such as the use of classifiers, relative clauses, complex word order, grammatical aspects, passive and the disposal construction, as well as simple questions that test their knowledge on literary stratum and vocabulary ranges. By developing an objective measurement, this study intends to find a more practical placement solution.

**Pamela Humphreys** (Griffith University, Australia)

***Theoretical and conceptual models of academic English language proficiency in higher education: Considerations for principled assessment in EAP***

**Day:** Friday

**Time:** 10:20–10:50am

**Room:** OGGB 5

Models of communicative competence and English language proficiency are well established in the fields of TESOL and Applied Linguistics (e.g. Bachman 1990; Canale 1984; Canale & Swain 1980; Celce-Murcia 2007; Celce-Murcia & Dornyei 1995; Purpura 2004). But, given the high number of international students with English as an additional language (EAL) preparing for and undertaking degree studies in our universities, how well do we understand the construct of academic English language proficiency? As a keystone attribute for academic success (Humphreys 2015; Sawir, et al. 2012; Woodrow 2006), it is argued that we need to better understand this construct.

This presentation will provide an overview of the extant conceptual frameworks related to academic English language proficiency in the higher education context (Arkoudis & O'Loughlin 2012; Harper, Prentice & Wilson 2011; Mahboob 2014; Murray 2010). A heuristic of English Language Proficiency for this context will also be proposed (Humphreys 2015), which combines the aforementioned frameworks in a new way to aid conceptualisation. This session will be of particular interest to those interested in the implications of such frameworks for principled assessment in EAP.

**Naoki Ikeda** (University of Melbourne, Australia)

***Assessing L2 learners' oral pragmatic and interactional abilities for university settings: Implications for classroom assessment***

**Day:** Friday

**Time:** 12:30–1pm

**Room:** OGGB 5

Pragmatic and interactional competence are underrepresented in the construct of large-scale proficiency tests for university admission (Roever 2011). Furthermore, the assessment of pragmatic and interactional competence is rarely undertaken in L2 classrooms (Youn 2013). Therefore, limited information has been provided about what features of pragmatics are challenging for L2 students, how these features can be differentiated and, what specific features could be targeted so that students can successfully participate in university activities.

The present study, by integrating multiple sources of data, aims to identify measurable features of oral pragmatics and interaction to construct instruments for task-based classroom assessments of oral pragmatic abilities for university activities.

Oral discourse data were collected from 67 L2 students in Australia (current university students with IELTS equivalent 6.0 to 8.5 and pre-entry students with IELTS 6.0 to 6.5) who completed three dialogue and three monologue role-play tasks simulating university situations. The data also include their self-assessed task performances, interlocutors' perspectives and raters' judgments. The data were analysed both quantitatively and qualitatively through use of a range of methods: descriptive statistics, discourse-based analyses and Rasch analyses.

The findings suggest that there is a clear difference between L2 students' performances in terms of their diversity of linguistic resources to handle pragmatic demands (Bardovi-Harling 2013) and abilities to tailor their social actions (Kasper 2006) for the context. Interactional features (e.g. repair, turn-taking) unique to some pre-entry students were also identified. In addition, Rasch analyses have revealed to what extent and how demanding each feature is for the students.

By drawing on interview data that investigates learners' perceptions about the designed instruments and their task performances, this study will discuss the measurable constructs of pragmatics and interaction with relevance to both the design and implementation of classroom assessment instruments in order to better assess L2 learners' readiness for university.

**Leila Iranmanesh** (University of New South Wales, Australia)

***Evolution of formative assessment in an English academic writing class: The role of emotion and power***

**Day:** Friday

**Time:** 11:55am–12:25pm

**Room:** OGGB 5

Looking through the lens of critical emotion theories in this presentation, I will discuss a part of an in-depth inquiry into the integration and evolution of formative assessment in an English academic writing class in Iran. Accordingly, I will also explore in what ways formative assessment evolved and impacted teaching and assessment practices as well as participants' understanding of assessment practices in a critical action research. The data comprise the systematic observations, interviews, reflections, detailed field notes and sample of participants' writings. The cycles of planned actions, reflective evaluation and revised actions shed light on different dimensions of evolved formative assessment, shifts in participants' emotions, learning and understanding and the challenges encountered by 9 participants; learners and I as the teacher and researcher. Taking the participants' voices and emotions into account from critical perspectives, I will discuss the implications for policy, program development, academic writing pedagogy and future research in the assessment context.

**Peter Keegan** (University of Auckland, New Zealand)

***Trialling a Māori language pronunciation tool based on a Māori speaker database***

**Day:** Friday

**Time:** 10:20–10:50am

**Room:** Case Rm 4

The MAONZE project (Maori and New Zealand English) uses recordings from three sets of speakers to track changes in the pronunciation of Māori (the indigenous language of New Zealand) and evaluate influences from English. The first group of speakers were born in the late nineteenth century and recorded mostly in 1946–48. The second group of speakers are kaumātua/kuia (elders) born between 1920 and 1940, and the third group are young speakers born between 1970 and 1990.



Results from the project show changes in both vowel quality and vowel duration (for all age groups and both genders) and evidence of diphthong mergers especially amongst the younger speakers. Female speakers from all three age groups were ahead of the male speakers in terms of raising the short vowels /e/ and /o/ and in glide weakening in the diphthongs. The young women are also in advance of the young men in /u/ fronting.

In this presentation we describe the results of developing and trialling several phases of a computer-based aid that assists learners to improve their own pronunciation of Māori. Several prototypes of the aid have been developed. The first trial (n = 11) took place in late 2015, the second trial (n = 35) on an improved version, took place in the first half of 2016. Those trialling the aid were positive, found it easy to use, and that it provided informative feedback. The aid allows users to get real time feedback on their own pronunciation of individual vowels, diphthongs and commonly mispronounced Māori words. It also allows users to listen to and compare their pronunciations with 'gold standard' pronunciations of kaumātua (elder males) or kuia (elder females) by drawing on the speaker database developed by the MAOZNE project. We conclude with thoughts on possible future directions for the aid.

**Jeanette King** (University of Canterbury, New Zealand)  
***Tuhinga Māhorahora: a corpus of children's writing in Māori***

**Day:** Friday **Time:** 11.20–11.50am **Room:** Case Rm 4  
Each week, children in Māori medium classrooms are encouraged to write in Māori in their Tuhituhi Māhorahora writing book. The aim of this writing programme is to "help children develop their own personal writing voice" (Ministry of Education 2008, p5). The Tuhituhi Māhorahora programme also provides teachers with an opportunity to assess their students' productive output in Māori and devise interventions and strategies to support the writing development of their students.

The 67,168 word Tuhinga Māhorahora corpus was designed and compiled to trial analyses which can assist teachers in their language enrichment strategies. The corpus comprises 1,329 pieces of writing collected in 2013 from 69 year 1–8 children at a Māori-medium school in Christchurch. The children's writing has been transcribed and marked up in oXygen and entered into LaBB-CAT, a browser-based searchable linguistic analysis tool (Fromont & Hay 2012). We use Paul Nation's (<http://www.victoria.ac.nz/lals/about/staff/paul-nation>) Range program and WordSmith concordance tools to conduct some of the analyses.

Here we report on the results of the analyses which reveal the vocabulary that the 5–13 year old students are using, and, crucially, the vocabulary not yet within the productive language output of the learner. We also demonstrate how this information can be used to inform teaching practice.

At present we know too little about how children are using te reo Māori in immersion classrooms. The Tuhinga Māhorahora project can add to our understanding of what children's immediate vocabulary learning needs are, and can help ensure that programmes provide both quantity and quality of exposure in the target language, thus enabling children to express themselves adequately.

**Lin Lin** (University of Hong Kong, Hong Kong)  
***Investigating the relationships between second language test takers' strategy use and Chinese reading comprehension test performance***

**Day:** Friday **Time:** 11:55am–12:25pm **Room:** Case Rm 2  
The important roles of metacognitive and cognitive strategies in language test performance have been recognized in the theoretical models (Bachman 1990; Bachman & Palmer 1996) and empirical studies (Purpura 1999; Phakiti 2003, 2008; Zhang & Zhang 2013; Zhang, Goh & Kunnan 2014). However, there is rare consensus on the relationships between second language (L2) test takers' metacognitive and cognitive strategy use and their test performance. In addition, compared with metacognitive and cognitive strategy use, the nature of affective strategy use, another essential type of strategy discussed in the previous literature, and its relationships with L2 test performance are poorly understood.

The current study examines L2 test takers' strategy use through a questionnaire and the relationships between their strategy use and their Chinese reading test performance on a large-scale standardised Chinese proficiency test: Hanyu Shuiping Kaoshi (HSK). It particularly investigates the nature of metacognitive

(planning, evaluating, and monitoring), cognitive (comprehending, memory, and retrieval) and affective (activating supportive emotions, and generating and maintaining motivation) strategy use and their relationships with reading test performance assessed by literal and inferential comprehension questions. The study involves 562 L2 intermediate-level test takers who studied Chinese language at university in mainland China. The test takers reported on a 56-item strategy use questionnaire immediately after they completed a 45-item HSK reading subtest.

Results showed metacognitive strategy use and affective strategy use were significantly correlated and metacognitive strategy use had a significant effect on cognitive strategy use in the test context. Comprehending and retrieval strategies had positive effects on literal comprehension. It was also found that literal comprehension had a significant effect on inferential comprehension in the reading test. Findings from the study provide practical implications for language teachers to teach reading comprehension strategies in classroom and for HSK test designers to design reading test items.

**Xiaohua Liu** (University of Auckland, New Zealand)

***Analysing existing reading test tasks: Implications for developing tasks to measure different reading abilities – development and validation of diagnostic language assessment tasks***

**Day:** Friday

**Time:** 2:30–3pm

**Room:** OGGB 5

The recent growing interest in diagnostic language assessment has brought about a number of studies attempting to retrofit the interpretation of test results of existing reading tests developed for other purposes (e.g. selection, placement), through analysing reading tasks and test takers' performances. Findings of these studies show a promising picture: although a few studies did not find clear evidence for the divisibility of a general reading construct, most of them identified multiple reading subskills underlying their test data. Nevertheless, the subskills found vary from study to study in terms of number and nature, and some of them reflect more of a task processing model than a reading model. Moreover, most of them stopped at identifying subskills in existing tasks, without exploring the factors that may facilitate or inhibit their assessment.

By summarising those subskills previously identified and comparing and contrasting them with reading theories, as well as by drawing on findings from studies investigating the impact of test task features on reading test performance, I developed a framework (consisting of reading subskills and task features) for subjective analysis of reading tasks. Using this framework, eight language experts were invited to independently analyse a set of reading tasks from a post-entry English language test administered by an English-medium university, meanwhile verbalising their thoughts during the process. The results were compared with the verbal reports of a group of students doing those tasks. Findings of these procedures will be presented and their implications for developing tasks to measure different types of reading ability will be discussed.

**Laurie Lu** (Nelson Marlborough Institute of Technology, New Zealand)

***Coming to grips with technical issues in developing and implementing EAP Unit standards assessments***

**Day:** Saturday

**Time:** 10:05–10:35am

**Room:** Case Rm 1

This paper examines some of the key technical issues that are either intrinsically true to EAP unit standards assessments or possibly acquired "iatrogenically" in the process of developing and implementing these standard-based assessments. It focuses its discussions on challenges imposed by the English Language Unit Standards, fairness and consistency issues, sufficiency issues re achievement evidence and task quantity and types, difficulty level of assessment tasks, assessment designing and the summative and formative use of the assessments. All the challenging features are explored in the context of ensuring assessments to be fair, valid and reliable. Overall, the research seeks a better understanding of these teacher-made and internally administered assessments, makes sense of certain challenging features and attempts an identification of some feasible solutions for the improvement of such an assessment system.

**Susy Macqueen** (The Australian National University, Australia)

***Profession-specific language standards: Perspectives from professional bodies on the use of language tests***

**Day:** Friday

**Time:** 3:40–4:10pm

**Room:** Case Rm 4

Standards are physical or behavioural mechanisms which make society orderly (Busch 2011; Lampland & Star 2009). They may arise organically in societies and/or they may be imposed so that order is forced. Language itself is a kind of standard, a social organiser par excellence; it evolves organically through social interaction but it is also imposed via policy and sources of power.

In this paper, we explore the notion of language tests as imposed standards, which are used to bring order to an aspect of society, namely human migration. In order to better understand how standardised tests are used as filtering mechanisms for skilled migration, we interviewed those who implement the standard: members of the six professional registration bodies for the Australian accounting, engineering, medical and nursing professions (11 interviewees).

A thematic analysis of the interview data offers insights into how profession-specific language standards are used and understood in the Australian context. In particular, findings show that the meanings of standardised test scores are constructed in relation to other standard indicators such as applicants' professional narratives and professional examination processes. Language test standards are trusted measures (to a certain degree), which operate differently for different professions; a test standard might act as a first filter in the registration process for one profession, but a later stage check for another. In general, the board representatives perceived that the implementation of a language test standard involves responsibility to different social worlds: test-takers, employers, the public and the Australian government. We discuss the implications of these responsibilities in relation to those of the test providers, and more generally, in relation to assessment literacy.

**Matthews M Makgamatha** (Human Sciences Research Council, South Africa)

**& Kathleen Heugh** (University of South Australia)

***Multilingual assessment: Opportunities for teacher development and equitable learning***

**Day:** Saturday

**Time:** 10:05–10:35am

**Room:** Case Rm 2

This paper draws attention to the potential for system-wide multilingual assessment to have a positive washback effect on teacher development and enriched teaching and learning in classrooms with linguistically diverse students. The paper is based on a first system-wide multilingual assessment of students attending the Western Cape Department of Education schools in South Africa.

This study was conducted in 2006 by the Human Sciences Research Council (HSRC). The tests administered to students comprised three versions of the languages instruments (Afrikaans, English and isiXhosa) and two language versions of mathematics (in Afrikaans and English). The format of items in the tests included selected response questions (SRQs) and constructed response questions (CRQs) that were crafted to provide diagnostic information. Although we used only two language versions of the mathematics tests, they included trilingual glosses for 40% of the items in each test. This enabled students to draw on their translanguaging expertise when answering these items. Teachers were contracted to undertake diagnostic marking and grading of extended response items.

In this paper, our attention is on the benefits which multilingual assessment may have for teacher development and enriched classroom learning. We provide a detailed account of a system we designed for marking of the mathematics and language tests administered. The paper will address the following issues: (a) training of teachers in preparation for marking of tests, (b) management of the marking process with special reference to marking learner responses to CRQs, (c) the process put in place to ensure quality assurance during marking, (d) gathering of qualitative diagnostic information from learners' responses to CRQs for the purpose of informing classroom teaching and learning.

The paper concludes by highlighting and reflecting on the challenges and opportunities related to deriving value (and relevance) from qualitative information obtained through diagnostic marking and grading of multilingual assessment.

**Michael Mersiades** (University of Queensland, Australia)  
***Validity of EAP reading test inference items: A pilot study***

**Day:** Friday

**Time:** 3:05–3:35pm

**Room:** OGGB5

This paper presents a pilot study into the validity of inference items in EAP reading tests. There is not much in the language testing literature about inference items, but test developers commonly ensure fairness and construct validity in inference items using a framework based on Chikalanga's (1992) distinction between propositional and pragmatic inferences. In this framework, pragmatic inferences are excluded from EAP tests because they require test takers to draw on construct-irrelevant external knowledge (Khalifa & Weir 2009; Taylor 2014). However, it has been acknowledged that in practice it is not always easy to distinguish between pragmatic and propositional inferences (Hughes 2003; Urquhart & Weir 1998).

A study of the validity of inference items requires a tool for reliably analysing inference items, so this paper asks if it is possible to elaborate on Chikalanga's distinction to produce a procedure that enables test developers to reliably distinguish between propositional and pragmatic inference items. A procedure is proposed that encourages test developers to systematically identify the information needed for a test taker to arrive at the correct response on a particular item. These pieces of information are then classified according to how the test taker accesses them (by decoding the text, or drawing on topical knowledge, for example). After these classifications, the item can be identified as a propositional or pragmatic inference item.

Two expert raters apply the procedure to 160 test items across two IELTS and TOEFL past test papers in order to classify the items as propositional inference, pragmatic inference, or non-inference items. The results give an initial indication of the utility of this inference item identification procedure as a tool to assess the validity of inference items, and also provide a snapshot of the types of inference items that appear in IELTS and TOEFL test papers.

**Ann Moir-Scott** (University of Auckland, New Zealand)  
***Locating the Learning: Measuring the impact of L1 reflection on L2 development***

**Day:** Friday

**Time:** 11:55–12:25pm

**Room:** Case Rm 4

This paper details a variety of interactionist and sociocultural tools used to measure the impact of L1 written reflection on L2 development. They were employed in Doctor of Education research to collect and analyse data in an intervention study conducted in five French Foreign Language classrooms in four New Zealand co-educational secondary schools. The Year 11 participants (n = 71) were commencing their third year of French and their first of three years of the National Certificate of Educational Achievement (NCEA). In this high-stakes assessment, Foreign Language (FL) writing skills are showcased in internally-assessed portfolios.

The data for the research came from the first and second drafts of two tasks (four drafts per student) produced during two three-week process writing units three months apart. Collecting data from two writing units enabled a counter-balance research design. The first drafts in both writing units acted as pre-tests in this reflection intervention while the second drafts acted as post-tests.

The measuring tools detailed in this presentation include:

- The NCEA Level 1 French assessment schedule used to gather baseline data.
- East's (2008) Analytic Scoring Rubric, CAF analysis, Error Analysis and Aljaafreh and Lantolf's (1994) Regulatory Scale used to measure and analyse L2 development.
- Oxford's (2011) Strategic Self-regulated Model of L2 Learning (S2R), Biggs and Collis's (1982) SOLO Taxonomy, and Coding supported by Microsoft (2013) OneNote, Microsoft (2010) Word Count and Wientjes & Hakuta's (2010) Wordsift to analyse the L1 reflection data.

A brief discussion of the findings and usefulness of the tools is included in this presentation.

**Paul Moore**(University of Queensland, Australia)

***Cohesion in oral language test performance***

**Day:** Friday

**Time:** 11:20–11:55am

**Room:** Case Rm3

Cohesion, alongside other aspects of discourse competence, is a construct which is variably defined, but which is commonplace in language test scoring rubrics. Drawing on data from a larger study into discourse competence in spoken language test performance, this paper focuses on the role of cohesion in distinguishing candidates' performance across tasks and levels on the Aptis Test.

After reviewing interpretations of the theoretical construct of cohesion and how these are operationalized in language testing research, we review recent research into the role of cohesion in language performance on different tasks and at different levels of proficiency. We then report on quantitative and qualitative analyses of aspects of cohesion identified in 83 test-taker performances on four tasks across the six levels of the Aptis Speaking Test. The method of data analysis employed discourse measures used by Iwashita and Vasquez (2015) in their analysis of discourse competence in IELTS Speaking Task 2 performances including conjunction, reference and lexical cohesion. We operationalised discourse competence in terms of the textual features of cohesion and coherence. Quantitative analysis was performed using the computational tool Coh-Metrix.

As with other recent research into L2 speaking, measures of cohesion in this study were not, on the whole, clearly distinguishable in performances across tasks and levels. The presentation concludes with a discussion of the utility of current measures of cohesion, whether these may be adjusted to provide a more nuanced understanding of the role of cohesion, or whether cohesion as an indicator of test-taker performance is best understood in conjunction with other aspects of discourse competence.

**Keiko Nakao** (University of South Australia)

***Reflective assessment tasks for intercultural language learning in a beginner foreign language course***

**Day:** Saturday

**Time:** 1:05–1:35pm

**Room:** Case Rm 1

Since intercultural language learning (ILL) has gained greater emphasis in conceptualising about language teaching and learning, foreign language programs have implemented learning and assessment tasks focusing on eliciting the intercultural. However, many teachers have found difficulty in implementing the assessment of ILL, because it's not simply a case of assessing learners' linguistic proficiency or knowledge of language and culture, but it also involves learners' intercultural capability, understanding, reflectivity and values, etc. (Liddicoat & Scarino 2013).

Firstly, this paper will discuss two assessment tasks designed to elicit intercultural capabilities in a beginner Japanese language course at university level. These tasks were reflective assessment tasks, one of which involved participation in an online discussion board and the other a reflection paper on students' intercultural experience, language learning and the tasks.

The two tasks were developed based on the frameworks and pedagogical models of ILL (Byram 1997; Liddicoat 2008; Tomita 2013). The paper will then present a content analysis of the responses of 40 students to the reflection paper. Finally, the paper will discuss the findings focused on how students reflect on their intercultural understanding and their language and culture learning. These findings will contribute to understanding features of ILL, and development of pedagogy, and assessment of ILL.

**De Phung** (University of New South Wales, Australia)

***What did EAL/D teachers actually think and do when marking oral performances?***

**Day:** Saturday

**Time:** 10:05–10:35am

**Room:** Case Rm 4

A great deal of recent research attention has been drawn to examining and improving validity and reliability of English language teacher assessment in the Australian main stream schooling system. However, little has been done to enhance the trustworthiness of teacher assessment of English as a second or additional language or dialect (EAL/D). This paper reports on findings from a research study aimed at:

- (1) Examining to what extent EAL/D teachers' oral assessments are consistent, and
- (2) Exploring factors influencing their assessments.

This study adapted materials developed by a larger project building tools to enhance assessment literacy for teachers of English as an additional language (TEAL) in Victoria into the context of EAL/D instruction in New South Wales. Twelve EAL/D specialists were invited to first mark three students' sample works and then be followed up. Findings revealed that teachers were significantly different from each other in perception of student performances and in judgment decisions and that their assessments were driven by a number of factors related to teachers themselves, students and tasks. From these findings, educational implications will also be discussed.

**Shahrzad Saif, Zahra Mahdavi** (Université Laval, Canada)

***Language needs of international graduate students working as Teaching Assistants (ITAs) in Canadian Francophone universities: Implications for assessment***

**Day:** Friday

**Time:** 3.40 – 4.10pm

**Room:** OGGB 5

Following a rapid global progress and a reorientation of educational policies, each year, North American universities hire an increasing number of foreign graduate students to teach undergraduate courses. To better understand the nature of ITAs' challenges, several studies have been conducted over the past three decades. The existing research (Reinhardt 2010; Chiang & Mi 2008; Gorsuch 2006; Hoekje & Williams 1992; Briggs & Hofer 1991), however, has mainly been conducted in the English-speaking universities in the US and has neglected bilingual contexts such as that of Canada, where foreign graduate students often speak neither of the official languages as their native language.

This study explores the language proficiency issues of ITAs enrolled in the engineering programs in francophone universities where, in addition to proficiency in French, an advanced proficiency in English is integral to ITAs' academic and professional success. Adopting Bachman & Palmer's model of Language Knowledge and Framework of Task Characteristics (2010), as well as Long's (2005) model of needs assessment, the study uses a mixed-methods approach to data collection and analysis.

Data is gathered from 84 stakeholders (ITAs, their supervisors, and undergraduate students) using questionnaires, interviews and observations in a major francophone university in Canada. The results of the statistical analysis of data (MANOVA) point to the insufficient French language proficiency of the ITAs. The results further confirm that, compared with the French language, the ITAs possess a higher level of proficiency in English. The results of the qualitative analysis of the data, on the other hand, show a mismatch between the language admission requirements of the graduate programs and the level of (French and English) language proficiency required of the ITAs. Based on these findings, the characteristics of the tasks and constructs to be measured by potential admission tests for ITAs to francophone universities are defined.

**Ruslan Suvorov** (University of Hawai'i Mānoa, United States of America)

***Test-taking strategies during the completion of multiple-choice items from the Michigan English Test: Evidence from eye tracking and verbal reports***

**Day:** Friday

**Time:** 11.20–11.50am

**Room:** Case Rm 2

The past decades have witnessed a surge of interest in research on test-taking strategies in second language assessment (e.g. Cohen 1998; Kashkouli & Barati 2013). Understanding strategies used by L2 test-takers can play a critical role in validation research (Bachman 1990; Schmitt, Ng & Garras 2011) that has been traditionally restricted to the use of statistical methods (O'Sullivan & Weir 2011). To investigate test-taking strategies, researchers usually employ concurrent or retrospective verbal reports (e.g. Cohen & Upton 2007; Plakans 2009) that are prone to reactivity and veridicality risks (Bowles, 2010) and should be supplemented with behavioural data that can provide information about test-takers' actual engagement with L2 tasks (Brunfaut & McCray 2015).

This study aimed at leveraging emergent methodology that combines eye tracking and retrospective verbal reports to investigate strategies used by test-takers during their completion of 58 multiple-choice items from the Michigan English Test (MET). Using the convergence model of the data triangulation design, it entailed:

- (a) Gathering eye-movement data from 15 non-native speakers of English while they were completing the MET, and
- (b) Using eye-movement recordings as a stimulus for participants to describe test-taking strategies they employed for answering each item.

Descriptive statistics for global processing and task processing eye-tracking measures were converged with themes identified through retrospective verbal data analysis to provide evidence of the types of test-taking strategies used by L2 learners. Results reveal a variety of test-taking strategies for answering multiple-choice items and indicate that test-takers:

- (a) Differ in terms of strategies they employ to answer such items, and
- (b) Rely on test-wiseness strategies that tend to inflate test scores.

Implications of the study suggest that the multiple-choice format appears to encourage the use of test-wiseness strategies that may introduce construct-irrelevant variance and pose threats to the validity of proposed interpretations and uses of test scores.

**Miki Tokunaga** (Fukuoka University, Japan)

***Effect of time pressure on grammaticality judgment tests with L1 translation***

**Day:** Friday

**Time:** 2.30-3pm

**Room:** Case Rm 2

Grammaticality Judgment Tests (GJTs) have often been used in SLA research (e.g. Ellis 2009; Godfroid, et al. 2015; Green & Hecht 1992; Roeher 2008; Sakai 2008; Shimada 2010). While timed GJTs are thought to measure constructs related to implicit knowledge of the target language, untimed GJTs are often presumed to measure constructs of explicit knowledge.

In this study, timed and untimed GJTs with L1 (Japanese) translations were given to Japanese university EFL learners (n = 219) to examine whether time pressure in GJTs would significantly affect the performance, thus indicating that timed and untimed GJTs possibly measure different factors of learners' L2 knowledge or ability. Although GJTs in previous studies did not include L1 translations, this study attempted to minimise the effect of learners' vocabulary knowledge and reading ability, and measure their understanding of target grammar structures by adding Japanese translations.

Rasch analysis using the Winsteps® software package was conducted on the data from the tests. The results of t-tests and factor analysis indicated that, for this group of participants, time pressure did not significantly affect the results. The effect of grammaticality was more significant than that of time pressure, indicating that grammatical and ungrammatical items on GJTs may measure different factors of learners' L2. Among grammatical items, untimed grammatical items stood out to be different from other items in correlation and factor analysis. Further analysis is required to find out whether these items are in fact measuring something different, or judging grammatical items, which was found to be easier than judging ungrammatical items, with unlimited time, is simply much easier than other item types, making them appear to be a different component. The results of the study, along with additional data to be collected this year, will be presented at the conference.

**Albert Weideman** (University of the Free State, South Africa)

***The refinement of the idea of consequential validity within an alternative framework for responsible test design***

**Day:** Saturday

**Time:** 1:05–1:35pm

**Room:** Case Rm 3

National level language assessments pose a particular challenge to those who design and administer them. To illustrate that challenge, this contribution will consider a set of secondary school exit-level examinations for home languages in South Africa. These examinations illustrate a dilemma with such high-stakes assessments that may be informative in other cases as well. In order to resolve it, the refinement of the idea of consequential validity (Messick) will be considered from the vantage point of an alternative conceptualisation of the principles that inform the design of language tests.

The contribution will outline how a number of constitutive or necessary conditions for assessment design (their instrumental power, their consistency and their theoretical defensibility) relate to other, more recently articulated ideas and principles of test design. The more recently identified issues concern test accessibility, acceptability, utility, alignment, transparency, impact, accountability, and care for those taking tests. This latter set of ideas may be defined as regulative or sufficient conditions for language assessments. An emphasis on fairness and justice in language testing (Kunnan) makes an important contribution to the regulative conditions for test design.

These constitutive concepts and regulative ideas, and the design principles they reflect, will be illustrated with reference to work on the potential redesign of a set of home language exit examinations in South Africa that not only lack accountability and integrity in the public mind, but are also indefensible in terms of

adequacy and equivalence. While trade-offs and compromises may still have to be made, a responsible approach to the design of language assessment will mitigate the potentially negative social and economic impact of high-stakes language tests.

**Jessica Wu** (Language Testing and Training Centre, Taiwan)

***Evaluating score reporting practice for two large-scale EFL tests: Intended goal and actual use***

**Day:** Friday

**Time:** 10:20–10:50am

**Room:** Case Rm 3

In response to the call for the incorporation of diagnostic feedback into achievement and proficiency testing (e.g. Kunnan & Jang 2009; Sawaki & Koizumi 2015), some large-scale English language tests have started to report more detailed information about learners' test performance. Yet, the success of the new score reporting practice depends on effective communication between the test developer and the stakeholders.

This small-scale qualitative study demonstrates how one testing body re-examined the effectiveness of score reports for two large-scale EFL tests in order to better bridge assessment and learning. The study investigated the current score reporting practice and stakeholders' perception and use of the detailed feedback of test results reported for the General English Proficiency Test (GEPT) and the General English Proficiency English Test for Kids (GEPT Kids). Three types of qualitative analyses were conducted. First, a content analysis of the score reports was conducted by employing Roberts and Gierl's (2010) test score report analysis framework. Second, the score reporting practice of the two tests was compared with that of other large-scale EFL tests which are also popular in Taiwan. Third, interviews were conducted with 20 students, 10 teachers, and 6 parents, who were from the schools where the two focal tests were used. Key findings include:

1. The current score reporting practice of both tests is generally consistent with good score reporting practice identified in previous studies in educational assessment.
2. Stakeholders' perception of the content and format of the score reports for both tests was generally favourable.
3. Stakeholders attended to and used only limited parts of the reported information, indicating that there is a gap between the intended goal and the actual use of the information provided for subsequent learning and instruction.

Implications for improving language test score report design and communication of test results are also discussed.

**Megan Yucel** (University of Queensland, Australia)

***Narrative inquiry in language assessment research***

**Day:** Friday

**Time:** 12:30–1pm

**Room:** Case Rm 2

The paper presents two narrative inquiry studies undertaken in different language assessment contexts, Vietnam and Australia, and which feature a diverse range of participants needing to demonstrate English language proficiency for study, immigration, or professional purposes. The studies investigate the issue of test impact from their perspective. Narrative inquiry is a qualitative methodology based on the premise that we make sense of our lives through narratives (Bruner 1990). This approach emphasises the collaborative aspect of research between researcher and participant over time in a sociocultural context, and allows for rich description and an exploration of the meanings that participants derive from their experiences. Although a substantial number of studies employing qualitative methods in various areas of language assessment research are available, narrative inquiry studies are rare in the field. Considering its benefits of providing a holistic picture with rich information that may not be captured in other qualitative methods, narrative inquiry is particularly useful for investigating perspectives of various assessment stakeholders.

The studies were conducted in two different contexts focusing on teachers and learners respectively. Study 1 investigated the impact of a government-mandated English language proficiency standard on Vietnamese EFL teachers and explored participants' perceptions of their English language proficiency in comparison with the government standard. Study 2, in Australia, investigated the beliefs of test candidates about English language testing with an aim of providing further insights into test impact, including candidates' perceptions of IELTS and its use in making decisions about international education and employment. The paper reports on the studies' findings, which provide valuable evidence of test validity and use from the perspective of the test-taker, and demonstrate how participants' narratives can provide comprehensive information essential to interpret test results. The paper also discusses how the narrative approach could be used in future language assessment research.



# PAPERS IN LANGUAGE TESTING AND ASSESSMENT (PLTA)

*Papers in Language Testing and Assessment* (PLTA) is published by the Association for Language Testing and Assessment of Australia and New Zealand (ALTAANZ). It offers an opportunity for both new and experienced researchers to publish original research papers, essays/discussion papers on theory, research digests, and book and test reviews on language testing and assessment issues.

PLTA is a peer reviewed international journal and is one of only six journals dedicated exclusively to language testing and assessment. PLTA is freely available online at ALTAANZ <http://www.altaanz.org/> and at LTRC <http://ltrc.unimelb.edu.au/>.

Prior to 2012, PLTA was published by the Language Testing Research Centre at the University of Melbourne under the name Melbourne Papers in Language Testing. The full catalogue of back issues of Melbourne Papers in Language Testing can be found at LTRC <http://ltrc.unimelb.edu.au/> where individual papers can be downloaded.

PLTA is published annually or biannually. Enquiries can be directed to the Editorial Assistant, Annemiek Huisman at [plta.editor@gmail.com](mailto:plta.editor@gmail.com). Correspondence on editorial matters should be addressed to the Editors, Sally O'Hagan and Lyn May at [plta.editor@gmail.com](mailto:plta.editor@gmail.com). The book reviews editor is Amanda Muller of Flinders University.

## PLTA BEST PAPER AWARDS 2013–15

**Winner:** Knoch, U., & Elder, C. (2013). A framework for validating post-entry language assessments (PELAs). *Papers in Language Testing and Assessment*, 2(2), 48–66.

**Citation:** This paper presents a significant and substantive step in the development of PELAs (Post-entry English Language Assessments) in Australia and New Zealand over the past 20 years. This paper is an extremely useful adaptation of the validity argument conceptualisation into a practical framework for validating PELAs. It provides broader applicability in terms of the discussion of validation/evaluation distinction. The framework will no doubt be influential in time for many institutions developing PELAs.

**Runner-up:** Hudson, C., & Angelo, D. (2014). Concepts underpinning innovations to second language proficiency scales inclusive of Aboriginal and Torres Strait Islander learners: a dynamic process in progress. *Papers in Language Testing and Assessment*, 3(1), 44–85.

**Citation:** This paper not only has an invaluable contribution to language assessment in the Australian context, but also practical implications for other similar contexts. It documents the development of an instrument which has arisen out of social and pedagogical need with considerable input from classroom teachers. This paper is an excellent example of how a rating scale can serve a professional development role and how assessment instruments might fit in the nexus of second language acquisition, descriptive linguistics, policy and education.

### Other Finalists:

Clark, M. (2014). The use of semi-scripted speech in a listening placement test for university students. *Papers in Language Testing and Assessment*, 3(2), 1–26.

Ruegg, R. (2014). The effect of assessment of peer feedback on the quantity and quality of feedback given. *Papers in Language Testing and Assessment*, 3(1), 24–43.

Kokhan, K. & Lin, C-K. (2014) Test of English as a Foreign Language (TOEFL): Interpretation of multiple score reports for ESL placement. *Papers in Language Testing and Assessment*, 3(1), 1–23.

# DELNA

## Diagnostic English Language Needs Assessment

The University of Auckland's post-entry language assessment



**Students say DELNA  
is time well spent!**

**DELNA assesses academic English skills  
to help students succeed at university.**

**For more information about DELNA or to  
license the assessment:  
[www.delna.auckland.ac.nz](http://www.delna.auckland.ac.nz);  
[delna@auckland.ac.nz](mailto:delna@auckland.ac.nz)**





## The English test for international study

### PTE ACADEMIC™

The most accurate and objective test of academic English available.

PTE Academic is a computer-based language test that offers international students the fastest, fairest and most flexible way of proving their English language proficiency for university admission.

- ✓ **Secure** testing worldwide ensures test score validity
- ✓ **Recognised and accepted** as proof of English proficiency by thousands of institutions worldwide, including 100% of Australian and New Zealand universities
- ✓ **Approved** for all Australian and New Zealand visa categories with an English language requirement
- ✓ **Testing globally** every week in over 200 test centres around the world
- ✓ **Results** typically within five business days

[www.pearsonpte.com](http://www.pearsonpte.com)





# IELTS™

## IELTS Research Grants

**Educational institutions and suitably qualified individuals are invited to apply for funding to undertake applied research projects in relation to the International English Language Testing System (IELTS) for a period of one or two years.**

Areas of interest include:

- Test development and validation issues
- Issues relating to contexts of test use
- Issues of test impact

The IELTS partners are making this grant funding available for IELTS-related research projects commencing next year.

The amount available for individual projects is AU\$70,000/£45,000.

**Applications must be submitted by 30 June.**

For application details and free access to published research visit  
**[www.ielts.org/research\\_grants](http://www.ielts.org/research_grants)**



# The *TOEFL*® Test for Student Success

We know the effort you put in to prepare your students for the future and we're here to support you as you help your students realize their full potential.

Get the newest and most up-to-date *TOEFL*® brochures, flyers and posters in our free *TOEFL* Advisor Toolkit.

**Designed for advising prospective test takers about the *TOEFL* test, the Advisor Toolkit contains resources for you to share directly with your students to help them prepare for the test.**

Order your **FREE** materials today at **[www.ets.org/toefl/advisorkit](http://www.ets.org/toefl/advisorkit)**

Copyright © 2016 by Educational Testing Service. All rights reserved. ETS, the ETS logo and TOEFL are registered trademarks of Educational Testing Service (ETS) in the United States and other countries. 33197



*Quality Beyond Measure.*



## The British Council Assessment Research Awards and Grants Results for 2016

The British Council Assessment Research Awards and Grants recognise achievement and innovation within the field of language assessment and form part of the British Council's extensive support of research activities across the world.

### Assessment Research Awards

These awards are designed to assist research students in their studies or in presenting their work at an international conference. The maximum award given is £2,500. Winners for 2016 are:

**Maria Georgina Fernandez Sesma** (University of Southampton, UK, supervisor Dr Ying Zheng)

**Iftikhar Haider** (University of Illinois at Urbana-Champaign, USA, supervisors Professor Emeritus Fred Davidson and Professor Melissa Bowles)

**Benjamin Kremmel** (University of Nottingham, UK, supervisor Professor Norbert Schmitt)

**Suh Keong Kwon** (University of Bristol, UK, supervisor Dr Guoxing Yu)

**Heidi Han-Ting Liu** (Teachers College, Columbia University, USA supervisor Professor James E. Purpura)

**Yueting Xu** (The University of Hong Kong, SARPRC, supervisor Professor David R. Carless)

### Assessment Research Grants

This grant scheme is designed to support projects that are directly focused on Aptis, the British Council's English assessment tool. The maximum grant given is £17,500. Winners for 2016 are:

**Stephen Bax & Prithvi Shrestha** (Open University, UK) for their project to explore lexical thresholds and lexical profiles across the Common European Framework of Reference for Language (CEFR) levels assessed in the Aptis test.

**Nguyen Thi Thuy Minh & Ardi Marwan** (National Institute of Education, Nanyang Technological University, Singapore) for their project which seeks to analyse test-takers' pragmatic performance and cognitive processing in the Aptis General Writing Test, Task 4.

**Sally O'Hagan & Kellie Frost** (University of Melbourne) for their project which will examine test-taker processes and strategies and stakeholder perceptions of relevance of the Aptis for Teachers Speaking Test in the Australian context.

**Parvaneh Tavakoli & Fumiyo Nakatsuhara** (University of Reading) for their project which looks at the scoring validity of the Aptis Speaking Test: Investigating fluency across tasks and levels of proficiency.

**Xun Yan, Ha Ram Kim & Ji Young Kim** (University of Illinois at Urbana-Champaign) for their project which explores the complexity, accuracy and fluency features of speaking performances on Aptis across different CEFR levels.

### Innovation in Assessment Prize

The Innovation in Assessment Prize celebrates innovation in the area of language testing and assessment. The winner for 2016 is the:

**Language Academy, Universiti Teknologi Malaysia**, Johor Bahru, Malaysia, for their approach to developing an assessment instrument, the Test of English Communication Skills for graduating students (UTM-TECS), which measures university graduates' readiness to communicate in English in the workplace. The tasks and assessment criteria of the test were derived from, and developed through, collaboration with industries at various stages of test development and implementation, including the validation and revision phases.

### International Assessment Award

This award recognises an individual working for the promotion of excellence in language assessment internationally. This year's award is presented to **Professor Emeritus Sauli Takala**.

**Sauli Takala** received his PhD from the University of Illinois at Urbana-Champaign in 1984. Sauli is now Professor Emeritus in Applied Linguistics at the University of Jyväskylä, Finland. He has extensive experience of research in language testing and assessment, in language policy and planning, curriculum development and teachers' in-service education. For 15 years, he was on the Finnish Matriculation Examination Board. Sauli has published a large number of research reports and articles in Finnish, Scandinavian and international journals. He was editor of the *Finnish Journal of Educational Research*, co-editor of the *Scandinavian Journal of Educational Research* and is on the Editorial Board of *Language Testing*.

Sauli coordinated the IEA International Study of Writing in the 1980s and helped plan the EU-funded internet-based DIALANG diagnostic assessment project in the 1990s. For many years, he has been associated with the Council of Europe's work on modern languages, most recently with the *Common European Framework of Reference for Languages* (CEFR), in particular the *Manual for Relating Language Examinations to the CEFR* (2009). He is a founding member of the European Association for Language Testing and Assessment (EALTA), served on its Executive Committee and was its second President in 2007. He is a consultant for the European Centre for Modern Languages (ECML).



THE ASSOCIATION FOR LANGUAGE TESTING AND  
ASSESSMENT OF AUSTRALIA AND NEW ZEALAND

Email: [altaanz@gmail.com](mailto:altaanz@gmail.com)

Website: <http://www.altaanz.org>