Learner Interfaces with Language Testing and Assessment
LTRC
2003

25th Language Testing Research Colloquium

Learner Interfaces with Language Testing and Assessment

Organizing Committee

Barry O'Sullivan (CRTEC, The University of Surrey Roehampton)

Pauline Rea-Dickins (The University of Bristol)

Jayanti Banerjee (Lancaster University)

The University of Reading
22-25 July 2003
Hosts

When the bid to hold this LTRC at The University of Reading was first discussed, the plan was that the Testing and Evaluation Research Unit (TERU) based in Reading would host the event. As ever, things change and that unit is no longer in existence. In its time a major influence on language testing around the world (the Language Testing Journal was founded there by Don Porter and Arthur Hughes in 1984) TERU finally closed its doors early in 2003.

So, the 25th annual LTRC is based at the University of Reading, but is jointly 'hosted' by the institutions represented on the organizing committee. These are

The Centre for Research in Testing Evaluation & Curriculum (CRTEC)
The University of Surrey Roehampton

The Graduate School of Education
The University of Bristol

The University of Lancaster

26th Annual Language Testing Colloquium
March 25 – 28 2004
Temecula Creek Inn, Temecula, California
Acknowledgements

We are extremely grateful to the following for their support:

University of Cambridge ESOL Examinations
ILTA (International Language Testing Association)
TOEFL Program, Educational Testing Service
CRTEC, The University of Surrey Roehampton
CALS, The University of Reading

We particularly wish to acknowledge the support of

Dr Mike Milanovic, Cambridge ESOL
The University of Reading Commercial Services Office
Thanks are due to the following people who served as proposal reviewers

Lyle Bachman, University of California at Los Angeles
Vivien Berry, The University of Hong Kong
Annie Brown, The University of Melbourne
Frances Butler, CRESST/UCLA
Micheline Chalhoub-Deville, University of Iowa
Caroline Clapham, Cambridge ESOL
Christine Coombe, Higher Colleges of Technology, UAE
Alan Davies, University of Edinburgh
Barbara Dobson, The University of Michigan
Dan Douglas, Iowa State University
Catherine Elder, University of Auckland
Hossein Farhady, Iran University of Science & Technology
Janna Fox, Carleton University
Ari Huhta, University of Jyväskylä
John de Jong, Language Testing Services, The Netherlands
Antony Kunnan, California State University
Jo Lewkowicz, The University of Hong Kong
Sari Luoma, Defense Language Institute FLC
Meg Malone, Center for Applied Linguistics
Tim McNamara, The University of Melbourne
Barry O'Sullivan, CTERC The University of Surrey Roehampton
Adrian Palmer, University of Utah
John Read, Victoria University of Wellington
Pauline Rea-Dickins, University of Bristol
Steven Ross, Kwansei Gakuin
Mary Schedl, Educational Testing Service
Rob Schoonen, University of Amsterdam
Bernard Spolsky, Bar-Ilan University
Charles W. Stansfield, Second Language Testing, Inc.
Peter Storey, Hong Kong Institute of Education
Diane Strong-Krause, Brigham Young University
Lynda B. Taylor, Cambridge ESOL
Randy Thrasher, Okinawa Christian Junior College
Carolyn E. Turner, McGill University
Wu Wei Ping, Chinese University of Hong Kong
As ever, a conference of this size cannot happen without the help of a large group of people

Ms Bobby Davies for taking care of the financial side of things. This conference could not have happened without her untiring professionalism.

Ms Pat O’Brien (University of Bristol), who worked on the management of the programme and on fielding questions and requests over the past months.

Dr Mike Milanovic of Cambridge ESOL who made available the resources of his organization and offered unstinting support from the time of the initial bid to the conference itself.

Adrian Kearney and Jason Hurren from Cambridge ESOL who went beyond the call of duty in their role of logistics and planning advisors – a special thanks to Jason for the green logo, it must have hurt!

Finally, a huge thanks to the volunteers from The University of Reading. Though I don’t like to single out individuals, some of them have been working on the conference for some time and really do deserve special recognition:

Ms Rana Shaaban, The University of Reading, who operated the Enquiries line and managed to stay sane.

Ms Astrid Lopez and Ms Veronique Dell’acqua who dealt with the registration database, keeping it up to date and accurate. I’m not so sure if they managed to stay sane but they did a great job.

Ms Li Chen, who took care of confirmations. Sanity was never an issue.

You will see many of the following group of volunteers over the days of the conference:

Ms Abeer Shabarek
Mr David Megaw
Ms Gabriella Righetti
Ms Hanan Al-Maghout
Ms Manal Shoshara
Ms Mariko Tamada
Ms Nisreen Aal Sayyed Hassan
Ms Yasuko Okabe

Ms Lu Yang (The University of Reading)
Ms Soubakeavathi Rethinasamy (CRTEC, The University of Surrey Roehampton)
Ms Saidatul Zainal abidin (CRTEC, The University of Surrey Roehampton)
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site Map</td>
<td>1</td>
</tr>
<tr>
<td>Campus Map</td>
<td>2</td>
</tr>
<tr>
<td>Reading Map</td>
<td>3</td>
</tr>
<tr>
<td><strong>LTRC 2003 Overview</strong></td>
<td>4</td>
</tr>
<tr>
<td><strong>LTRC 2002 Programme Abstracts</strong></td>
<td></td>
</tr>
<tr>
<td>Samuel Messick Memorial Lecture</td>
<td>7</td>
</tr>
<tr>
<td>Papers</td>
<td>8</td>
</tr>
<tr>
<td>Symposia</td>
<td>22</td>
</tr>
<tr>
<td>Posters</td>
<td>25</td>
</tr>
<tr>
<td>Research network</td>
<td>31</td>
</tr>
<tr>
<td><strong>Conference Participants</strong></td>
<td>36</td>
</tr>
<tr>
<td><strong>Contact information</strong></td>
<td>39</td>
</tr>
<tr>
<td><strong>Notes</strong></td>
<td>40</td>
</tr>
</tbody>
</table>
Site Maps
Reading Map
**LTRC 2003 Overview**

### 22 July, Tuesday

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00 am – 4:00 pm</td>
<td>Pre-Conference Workshop: Guoxing Yu, University of Bristol - <em>Using winMAX to Analyse Qualitative Data in Language Testing Research</em></td>
</tr>
<tr>
<td>4:00 pm - 6:30 pm</td>
<td>Registration Desk Open (Whiteknights Hall)</td>
</tr>
<tr>
<td>6:30 pm - 8:30 pm</td>
<td>Opening Reception Sponsored by Cambridge ESOL (Whiteknights Hall)</td>
</tr>
</tbody>
</table>

### 23 July, Wednesday

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30 am – 5:00 pm</td>
<td>Registration Desk Open</td>
</tr>
<tr>
<td>8:45 am - 9:00 am</td>
<td>Opening Ceremony</td>
</tr>
<tr>
<td>9:00 am - 9:10 am</td>
<td>Samuel Messick Award Presentation</td>
</tr>
<tr>
<td>9:10 am - 10:10 am</td>
<td>Samuel Messick Memorial Lecture: Prof Patricia Broadfoot, University of Bristol – <em>Dark Alleys and Blind Bends: Testing the Language of Learning</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:10 am - 10:30 am</td>
<td><strong>Coffee &amp; Tea break</strong></td>
</tr>
<tr>
<td>10:30 am – 12:30 pm</td>
<td>Paper Sessions 1-4</td>
</tr>
<tr>
<td>1.</td>
<td><em>The impact of new TOEFL on Central and Eastern Europe: baseline study</em> – Dianne Wall &amp; Tania Horak</td>
</tr>
<tr>
<td>2.</td>
<td><em>Forward planning for washback impact research</em> – Elana Shohamy, Liz Hamp-Lyons, Matt Poehner &amp; Stefanie Rehn-Jordan</td>
</tr>
<tr>
<td>3.</td>
<td><em>Goals and gains: washback from IELTS writing on teaching, learning and outcomes</em> – Tony Green</td>
</tr>
<tr>
<td>12:30 pm - 1:30 pm</td>
<td>a) Lunch Sponsored by ETS/TOEFL</td>
</tr>
<tr>
<td></td>
<td>b) ILTA Board Meeting</td>
</tr>
<tr>
<td>1:30 pm – 4:00 pm</td>
<td>Paper Sessions 5-9</td>
</tr>
<tr>
<td>5.</td>
<td><em>Test-takers' 1st Languages and their discoursal performance in oral proficiency tests</em> – Yang Lu</td>
</tr>
<tr>
<td>6.</td>
<td><em>Turn-taking and the number of speakers in a small group oral</em> – Gergely A Dávid</td>
</tr>
<tr>
<td>7.</td>
<td><em>An investigation of the interactions in paired oral proficiency testing</em> – Lindsay Brooks</td>
</tr>
<tr>
<td>8.</td>
<td><em>Interaction between test-taker characteristics, task facets and L2 oral proficiency test performance</em> – Noor Lide Abu Kassim &amp; Ainol Zubairi</td>
</tr>
<tr>
<td>9.</td>
<td><em>Effects of task design on candidate performance in the IELTS writing module</em> – Kieran O'Loughlin and Gillian Wigglesworth</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:00 - 4:15 pm</td>
<td><strong>Coffee &amp; Tea break</strong></td>
</tr>
<tr>
<td>4:15 pm - 5:45 pm</td>
<td>Symposium 1</td>
</tr>
<tr>
<td>1.</td>
<td><em>Exploring the relationship between language corpora and language testing</em> – Lynda B Taylor, Paul Thompson, Michael McCarthy &amp; Fiona Barker</td>
</tr>
</tbody>
</table>
### 24 July, Thursday

**8:45 am - 10:45 am**  
10. Learner response to new TOEFL integrated reading/writing tasks – Tom Lumley, Annie Brown & Bonnie Zhang Wenxia  
11. A study of the impacts of external language tests on learners in the Lingue 2000 project in Italy – Roger Hawkey  
12. Gender and reading anxiety in IELTS and TOEFL – Behzad Ghosnooly  
13. Exploring the speed element in computer-based reading assessment – Toshihiko Shiotsu

**10:45 am - 11:15 am**  
**Coffee& Tea break**

**11:15 am - 12:45 am**  
14. What’s the pay-off? Outcome measures in school foreign language learning – Kathryn Hill, Noriko Iwashita & Tim McNamara  
15. Defining the construct ‘academic language’ through empirical research – Gary Buck, Alison Bailey, Yeonsuk Cho & Sari Luoma  
16. An approach to strengthen the assumption of representativeness, generalisability and predictability of language assessment tasks – Sara Gysen

**12:45 pm - 1:45 pm**  
- **a) Lunch**  
- **b) Language Testing Board Meeting**

**1:45 pm - 3:15 pm**  
17. Linking results on a new language test to national performance level descriptors – Dorry M Kenyon  
18. C-tests as measures of general language proficiency – Thomas Eckes & Rüdiger Grotjahn  
19. Students’ and parents’ perceptions of their rights as test consumers – Pavlos Pavlou

**3:15 pm - 4:30 pm**  
- Poster Session (Tea/coffee will be available)  
  1. Assessing writing for academic purposes: a comparative study of IELTS and TOEFL – Mark Beitel  
  3. What’s so hard about that? Investigating item/task difficulty across two examinations – Barbara K Dobson, Mary C Spaan & Amy D Yamashiro  
  4. An ethical use of IELTS scores in the recruitment of international students – Kevin Haines & Angela Ashworth  
  5. The validation study on the first achievement test in Japanese university entrance examinations – Akihiro Ito  
  6. Developing a test of lexical choice – Tomoko Ishii  
  7. Vocabulary levels test: an issue of item independence in clusters – Tadamitsu Kamimoto  
  8. Norwegian CATS say "yes" to Europe: using CEF in school English tests – Eli Moe & Cecilie Carlsen  
  9. A many-facet Rasch analysis of three-dimensional oral proficiency test data – Yuji Nakamura  
  10. The effects of different training procedures on rating accuracy: an experimental study – Soubakeavathi Retinhasamy  
  11. Developing a vocabulary test for Malaysian primary school students – A Affendi Shabdin & Norbert Schmitt

**4:30 pm - 6:00 pm**  
2. Identifying and comparing models of language and pedagogy in ESL assessment scales for linguistic minorities in three countries – Constant Leung, Hugh South, Catriona Scott, Sibel Erduran and Pauline Rea-Dickins.

**19:00 pm – 22:00**  
**Conference Banquet**
25 July, Friday

9:15 am - 10:15 am  Paper Sessions 20-21
20. Conflicting perspectives? Judging clinical communication skills in a test of bilingual proficiency – Elisabeth Grove & Anna Whelan
21. Validities: multidimensionality in educational measurement – Hameed Esmaeili

10:15 am - 11:00 am  Lifetime Achievement Award
Paper presented by award recipient
(Presentation to be made at the Conference Banquet)

11:00 am - 11:15 am  Coffee & Tea break

11:15 am - 12:45 pm  Paper Sessions 22-24
22. Diagnosing foreign language proficiency – J Charles Alderson

12:45 pm - 2:15 pm  Lunch
(incorporating the ILTA Annual Business Meeting)

2:15 pm - 3:30 pm  Research Network (Tea/coffee will be available)
1. A construct-based approach into the nature of reading: understanding, identifying and validating reading skills – Hossein Barati
2. Developing an accountability system for measuring annual progress of ESL students – Gary Buck, Yeonsuk Cho & Sari Luoma
3. Managing the transition: the role of assessment at the primary/secondary school interface – Kathryn Hill
5. The underachievement of ethnic minority students in Wales – Christine Pegg
6. A case of washback with primary EAL learners – Catriona Scott
7. From the candidate’s point of view: diary-keeping in a test context – Constance Tsagari
8. Summarisation as a measure of reading comprehension: expert and popular dialogues? - Guoxing Yu

3:30 pm - 5:00 pm  Paper Sessions 25-27
25. International oral language proficiency: construct investigation – Micheline Chalhoub-Deville & Gillian Wigglesworth
26. Planning, narrative task structure and performance – Parvaneh Tavakoli & Peter Skehan
27. Use of multiple-choice completion test for Chinese EFL learners – Jianda Liu

5:00 pm - 5:15 pm  Thanks and closing of Conference
This year's conference theme: 'Learner Interfaces with Language Testing and Assessment' puts the spotlight firmly on learning and the learner in the business of assessment—a focus that is as welcome as it is overdue, both in the field of language testing in particular and in the world of assessment more generally. Samuel Messick's great contribution to the study of assessment was to emphasise the importance of validity—crucially, the extent to which an assessment instrument addresses a given domain of learning and its overall impact of the learning process.

In this lecture I use Messick's concern with validity to argue four central points.

Firstly, that a great deal of assessment and testing in the past has been effectively located on 'blind bends'. Assessments have been initiated without consideration of their link to the domain it is desired to test and they have been initiated without sufficient consideration of their 'wash-back' effect on the learning process. This is as potentially damaging as it is unhelpful.

Secondly I argue that even more assessment and testing practice has found itself in 'dark alleys' where there has been little or no illumination as to the central purpose of the activity and even less illumination about what the results mean.

The overall result of this situation, I suggest, is that a great deal of educational assessment in the past has not been fit for purpose and therefore not useful. Because the link between means and ends has typically been implicit rather than explicit, assumptions have been made about the meaning of assessment data and the effects of assessment on learners and the learning process, which do not bear scrutiny. In short, despite the sophisticated technologies underpinning much of the testing industry, their application in practice has not been matched by a similarly sophisticated understanding of the learning process itself. As a result, we have been using what are in practice a very blunt set of instruments to intervene in a highly sensitive and complex business of learning.

The second part of my presentation therefore explores the nature of the challenge that currently confronts the testing and assessment world. From the perspective of a non-language-testing specialist, I address four central questions about language learning namely:

1. What is it to learn a language?
2. Why is the language being learned in this particular context?
3. What constitutes success in this context -short-term and long-term?
4. What is the purpose of the assessment being used –formative or summative?

I explore the implications of these questions in relation to emerging new research insights into learning. Drawing on recent ground-breaking research on 'assessment for learning' and on the identification of elements of 'the learning genome', I identify some of the key factors that need to be taken into account in the consideration of 'learner interfaces' with language testing and assessment. These are identified as context, collaboration, confidence, communication and coercion which together, I suggest, constitute a potential lexicon for understanding the language of testing.
## Paper Abstracts

<table>
<thead>
<tr>
<th>#</th>
<th>Day/Time</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>July 23</td>
<td><em>The impact of new TOEFL on Central and Eastern Europe:</em> baseline study</td>
</tr>
<tr>
<td></td>
<td>10:30 – 12:30</td>
<td><strong>Dianne Wall &amp; Tania Horak</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>It is now accepted that the agencies responsible for high-stakes examinations should seriously consider the impact of their exams on teaching and learning. This issue has only recently gained prominence in language testing, but it has now been explored both theoretically (Alderson and Wall 1993, Bailey 1996) and practically (see Wall 1997 for examples). It is still difficult to predict the impact of specific examinations, but we are gradually building a picture of the types of factors which may facilitate or hinder what is intended. One of the goals of New TOEFL is to create positive impact on the classroom. The impact of the current exam is seen as problematic (see Cumming et al 2000 for references) and each of the monographs in the TOEFL 2000 Framework declares a wish to rectify this situation. Some examples are given of desirable impact, but there is no comprehensive statement of how the new examination would ideally change practice. There are no details of how impact should be investigated or of how research results can be used to improve teaching and learning. This paper describes the first phase of an investigation into the impact of New TOEFL. It will discuss attempts to discover what impact is envisaged and to find out what TOEFL preparation classes look like before the ELT community becomes aware of the content and format of the new examination. It will report on progress in instrument design and validation, and the collection of classroom and contextual data.</td>
</tr>
<tr>
<td>2</td>
<td>July 23</td>
<td><em>Forward planning for washback/impact research</em></td>
</tr>
<tr>
<td></td>
<td>10:30 – 12:30</td>
<td><strong>Elana Shohamy, Liz Hamp-Lyons, Matt Poehner &amp; Stefanie Rehn-Jordan</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Messick (1996) suggests that a well-designed test can positively impact teaching and learning. The new format of the TOEFL, to be launched in 2005, aims to do this through a more communicative orientation. This project will ultimately investigate whether the new TOEFL affects English language instruction, how far, and in what ways, particularly whether it promotes communicative language teaching. Importantly, the present study builds on and greatly extends previous washback research. The project design is: 1) longitudinal (to examine the test's impact in various phases); 2) multinational, and includes different kinds of institutions within each country (to understand washback from the TOEFL in a variety of cultural and educational settings); and 3) methodologically integrated (using quantitative and qualitative methodology to document the TOEFL's effects from a variety of complementary perspectives). The first stage of the project has been development and validation of the instruments for use in the baseline and later project stages. Challenges and issues faced by the researchers have included: 1) developing construct definitions for the variables hypothesized to play a role in washback; 2) operationalizing these variables into data collection instruments that will eventually allow for data triangulation (e.g., classroom observations, questionnaires, textbook analyses, etc.); 3) determining stakeholders and ensuring equivalent instruments for each group; 4) attempting to make the instruments 'ideology-free' so as not to bias respondents; 5) gaining access to data collection sites; and 6) training researchers in each country. Studying test washback before, during and after implementation provides unique opportunities for language testing research.</td>
</tr>
<tr>
<td>#</td>
<td>Day/Time</td>
<td>Paper</td>
</tr>
<tr>
<td>----</td>
<td>------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| 3  | July 23    | **Goals and gains: washback from IELTS writing on teaching, learning and outcomes**  
Tony Green  
The International English Language Testing System (IELTS) test plays a gatekeeping role for students intending to enter higher education in the UK. The IELTS includes a direct test of writing; but how far does preparation for this test encourage development of academic writing skills required by students?  
In the context of debates about consequential validity, it has been suggested that preparing for the IELTS writing test, to the extent that the test reflects specific academic literacy needs, will help international students to learn the range of academic writing skills they will need for university study. Conversely, where the test fails to reflect these needs, it is said to have damaging effects on learning.  
This presentation describes research into the influence of the IELTS Academic Writing module in academic writing instruction. Key questions addressed include;  
1) How far the IELTS writing test reflects current theories of academic literacy and whether IELTS preparation might constitute a narrowed form of instruction.  
2) Whether dedicated test preparation courses substantially improve students' chances of success on the test relative to more broadly focussed courses in academic writing skills. 
The study included questionnaires, interviews with teachers and students, classroom observation and consideration of the score gains made by students on the IELTS writing test under different conditions of study (presessional courses in academic writing provided by universities or intensive IELTS preparation). Implications are suggested for the use of the IELTS writing test and for the relationship between construct validation and washback theory. |
| 4  | July 23    | **Using language proficiency test scores: the admissions process in a British university**  
Jayanti Banerjee  
This paper is 'situated' in the context of the often problematic relationship between students' language proficiency test scores and their performance on university degree courses. It takes the view that one route towards insight into this relationship lies in better understanding the role of language proficiency test scores in the admissions process. It therefore discusses the process of making admissions decisions at one British University. The particular issue that it addresses can be expressed in two parts:  
1. How do admissions personnel use language proficiency test scores to make admissions decisions?  
2. What does this score use reveal about how admissions personnel view the relationship between language proficiency and academic performance?  
In order to address these questions, interviews have been conducted with admissions personnel responsible for two degree schemes that attract and admit large numbers of international students. The paper describes how the data was collected and its analysis. The findings indicate that admissions personnel take admissions decisions on the basis of a combination of 'success indicators' that make them feel confident that the student will be able to cope with the linguistic and academic demands of the degree course for which they have applied. They also reveal interesting attitudes towards language proficiency test scores as well as views of language proficiency in general. |
5 July 23 13:30 – 16:00

**Test-takers' 1st Languages and their discoursal performance in oral proficiency tests**

Yang Lu

This paper focuses on the interface between test-takers' 1st languages and their discoursal performance in Oral Proficiency Tests (OPTs). Previous studies have suggested that test-takers of different 1st languages at different proficiency levels tend to have different conversational styles or discoursal behaviors in terms of initiating topics, elaborating and dealing with confirmation requests (Lebra 1987; Young & Milianovic 1992; Young 1995; van Betterajj, Kellerman, and Schils 1996; Young & Hallect 1998; Ross 1998).

The present research hypothesizes that (1) test-takers of different 1st languages initiate and sustain topics differently in discourse in OPTs in relation to frequency, location and the sub-categories of initiating and sustaining; (2) patterns of pairing test-takers according to their 1st languages affect their discoursal performance in the above-mentioned three aspects.

Thirty UCLES First Certificate English (FCE) Speaking Test events were analysed to test the hypotheses using a framework based on a modification of the systemic and functional approaches to spoken discourse analysis. The data were divided into three groups according to the test-taker's 1st languages in the pairs: Italian Data, Chinese Data and Mixed 1st Languages Data.

The resulting corpus of transcribed data was tagged for six discourse features under the two main types of discoursal behaviors: initiating and sustaining. Frequencies of the two main types and the sub-categories were counted and locations of the produced discourse features were examined. The results show that test-takers' spoken discourse performance tend to be culture-dependent and the pairings could affect types and frequencies of the discourse features they produce.

6 July 23 13:30 – 16:00

**Turn-taking and the number of speakers in a small group oral**

Gergely A Dávid

The debate about the validity of the OPI in general, about its conversational nature, dynamism and power structure in particular is one that heats up from time to time. Apart from paired orals, group testing is one of the alternatives to the more traditional interview types of tests, with the difference that it has perhaps been less well researched to date.

The relevant educational context for the small group oral is the Test of Language Competence (TOLC), an in-service filter test at the Centre of English Teacher Training, Budapest, Hungary. In the interactive component of the test, students are tested on self-contained tasks with very little interlocution in groups of three.

This piece of research originated in an interest in the characteristics of communication that set the group oral apart from either the OPI or paired orals. The researcher's attention was directed towards its turn-taking mechanisms and the possible effect of the number of speakers on communication. The work of the conversation analyst school of discourse analysis provided a strong rationale for the analysis of the TOLC oral. The paper focusses on the characteristics of turn-taking and the recurrence of short conversational bias sequences, as evidenced in videoed performances of the test. Findings suggest that it may have been a happy choice to examine students in threes, rather than twos or in groups of four or more, because conversation has a tendency to stabilize between the same two speakers, always leaving a third party who wants to take the floor.
<table>
<thead>
<tr>
<th>#</th>
<th>Day/Time</th>
<th>Paper</th>
</tr>
</thead>
</table>
| 7 | July 23      | **An investigation of the interactions in paired oral proficiency testing**  
Lindsay Brooks  
This paper presents some of the results of a larger thesis study comparing paired oral proficiency testing, in which pairs of adult international ESL students interact with each other, to individual proficiency testing, in which each student interacts with an examiner.  
The data were collected in the context of the exit testing from an Academic Preparation Program at a large Canadian university. Because successful completion of the program with an appropriate grade, together with passing the exit testing, satisfies the language proficiency requirements of the university, the testing is high stakes for the students and for the program.  
The participants (N=80) each had two speaking tests: one in the individual format and one in the paired format. Comparable speaking prompts were used in each test. After the exit testing, participants responded to a questionnaire surveying their views on the two testing formats and 12 students participated in semi-structured interviews. The focus of this paper is on quantitative and qualitative analyses of the multiple interactions involved in the paired oral proficiency testing. In addition to statistical results, analyses of the dialogue of students as they interacted in the paired tests will be presented.  
The results suggest that there was a person by format interaction with some students performing differently depending on whether they were tested in pairs or individually. The findings have implications for test development, implementation, and rating and most importantly, on the decision whether to test in pairs.  
| 8 | July 23      | **Interaction between test-taker characteristics, task facets and L2 oral proficiency test performance**  
Noor Lide Abu Kassim & Aino Zubairi  
The nature of test-taker characteristics and their impact on second language test performance has drawn a considerable amount of research interest among language testers in the last two decades (for example, Purpura, 1999; Clapham, 1996 and Kunnan, 1995). This is hardly surprising given the potential influence of test-taker attributes on L2 learning and assessment. However, examining the influence of these attributes on test performance alone is insufficient. There are other elements in the assessment framework which not only influence the variability of test scores (Bachman, 1990) but which also yield a significant amount of influence on test-taker characteristics (for example, Fulcher, 1996). It is therefore necessary to further investigate the nature of these other factors in the assessment framework and examine how they relate to test-taker characteristics and impact test performance. This will provide further insights and a better understanding of how these factors behave in relation to one another and how they each contribute to variability in test performance.  
In this study the interaction between three main factors in the assessment framework (as proposed by Bachman, 1990) – test task facets, test-taker personal attributes, and language ability – and their simultaneous effect on oral proficiency test performance are examined through the use of structural equation modelling. In this approach the effects of the different factors in the assessment framework can be directly observed and modeled, and their individual effect on test performance estimated. |
# Day/Time  
**Paper**

9  
July 23  
13:30 – 16:00  
*Effects of task design on candidate performance in the IELTS writing module*  
Kieran O’Loughlin and Gillian Wigglesworth

This paper reports on a study into task difficulty in the IELTS Academic writing task 1. A series of tasks were developed which differed in terms of the amount of information candidates were required to process to complete the task, and the manner of presentation (e.g. table, bar chart, line graph). Control tasks were included as benchmark tasks and administered to all candidates. The experimental tasks were administered to subgroups of the cohort.

English for Academic Purposes students (N=210) completed four writing tasks (two control and two experimental tasks). All scripts were double rated by trained and qualified IELTS raters.

Test score analyses indicated no substantial differences in difficulty between tasks as a function of the amount of information or differences in presentation. These analyses were then used to divide the students into proficiency levels (high, middle and low). The written texts were then subjected to a series of discourse analyses to investigate the levels of complexity and accuracy of the texts. All proficiency groups demonstrated greater complexity overall where less information had to be processed. However, the trend was less clear for accuracy, with the high proficiency group showing a strong tendency to display greater accuracy in response to the task with more information. Since tasks providing less information actually elicit more complex language, and the goal of these tasks is to produce as high a performance from the candidate as possible, we argue that this is best achieved through the use of simpler tasks.

10  
July 24  
08:45 – 10:45  
*Learner response to new TOEFL integrated reading/writing tasks*  
Tom Lumley, Annie Brown & Bonnie Zhang Wenxia

EAP tests have tended to employ separate assessments of the four macro-skills, reading, writing, listening and speaking, although most authentic academic language use requires integration of two or more of these macro-skills. The writing component of the revised TOEFL therefore aims to include integrated reading / writing as well as independent writing tasks. Integrated tasks require synthesis, summary and/or evaluation of the content of the input passage; the resulting texts therefore differ from independent argumentative tasks where the learner is required to generate all or most of the content.

Learners require training in appropriate task interpretation and demands. If they do not respond appropriately, the test will not offer a valid indication of their ability to produce the kinds of writing required. Key issues include the selection of propositional content, the potentially inappropriate use of input material in learners’ written texts (e.g., copying) and academic conventions (e.g., referencing), the impact of comprehension on production, and rhetorical organisation of learner texts.

This paper addresses issues of the validity of learners’ responses to integrated reading / writing tasks, with reference to four complementary sets of data:

1) texts produced by 60 Mandarin-, Cantonese- and Korean-speaking learners of English for both integrated and independent tasks,
2) retrospective verbal reports by these learners,
3) verbal reports produced by six raters whilst rating the performances, and
4) ratings of both task types.

The results have potential for refining procedures for task development, improving the provision of information to learners and managing the training of raters.
<table>
<thead>
<tr>
<th>#</th>
<th>Day/Time</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>July 24</td>
<td><strong>A study of the impacts of external language tests on learners in the Lingue 2000 project in Italy</strong></td>
</tr>
<tr>
<td></td>
<td>08:45 – 10:45</td>
<td></td>
</tr>
</tbody>
</table>
Roger Hawkey  
Lingue 2000 (PL2000) is a radical language education reform project of the Ministry of Education, Universities and Research in Italy. PL2000 aims to facilitate communicative, technologically informed language learning appropriate to the needs of students as defined by the Common European Framework of Reference for Languages. One of the Project's key measures is to encourage certification of learners' language proficiency through the examinations of external providers such as Cambridge ESOL.  
With the encouragement of the Ministry, Cambridge ESOL has carried out a study of the impact of the PL2000 on key stakeholders, including students, parents, teachers, education managers and language testers. Participants were a case-study selection of state elementary, middle and secondary schools in North, Central and Southern Italy. The focus was the impacts of PL2000 on language teaching pedagogy, materials, media and assessment, and the impacts of changes in these on student and teacher attitudes and performance. Data were collected during school visits at the beginning and near the end of the 2001-2002 school year, using background and attitude questionnaires, interviews, focus groups and video-recorded classroom observations, supported by informal contacts with the schools throughout the period of the study. Cambridge ESOL language test results for other PL2000 and for non-PL2000 classes provide broad comparative performance data for the case-study groups.  
The findings of the study suggest that despite positive learner attitudes to international high-stakes foreign language tests, careful attention is needed to classroom language teaching : testing relationships, and to gaps between communicative teaching theory and practice. |
| 12 | July 24   | **Gender and reading anxiety in IELTS and TOEFL**                                                                                  |
|    | 08:45 – 10:45 |  
Behzad Ghonsooly  
Research in language anxiety suggests that it may negatively influence performance. While most research has dealt with speaking as the most anxiety producing experience, few studies have tackled the role of EFL proficiency text in producing anxiety and the degree to which male and female L2 readers react to it. Based on the present researcher's intuition and testing experiences, it was hypothesized that the IELTS reading texts would trigger more stress than the TOEFL.  
To test the hypothesis, two randomly selected reading comprehension texts were chosen from the IELTS and the TOEFL published test samples. The texts were administered to three groups of 150 Iranian beginner, intermediate, and advanced university students who were randomly selected on the basis of a language proficiency test. Half of the students in each group were male and the other half female. The students were asked to read each individual sentence and express their feelings through the Spielberger anxiety test. To support the data, a retrospective study was done to find the cause of anxiety. The results showed that novice readers had more stress in reading and understanding the texts than the other two groups. They also suffered from problems at word level which seems to be the cause of anxiety. Moreover, mean comparison of the data revealed that most students showed more anxiety with the IELTS text than the TOEFL one. This was particularly more evident in the case of female readers. |
Paper

**Exploring the speed element in computer-based reading assessment**

**Toshihiko Shiotsu**

As facing the text on computer becomes as ordinary to many as on printed pages, research efforts for better understanding of computer based (CB) reading and its assessment must be gaining significance.

Meanwhile, one of the aspects of reading that deserves more work within the context of L2 ability research is text processing speed, which can be critical for fully participating in today's expeditious CB text exchanges. Compared to its paper-based (PB) counterpart, CB reading assessment has a distinct advantage of being able to control, measure, and record the temporal elements with greater accuracy.

The study to be reported has utilised this feature of CB testing to explore what types of skills or knowledge might account for the individual differences in the speed of reading text on computer.

219 adult EFL learners participated in both CB and PB testing, which included measures of word recognition latencies, vocabulary breadth, syntactic knowledge, and working memory capacity along with the measures of sentence reading speed and passage reading comprehension.

The data were submitted to multiple regression analyses, exploratory factor analysis, and structural equation modelling. Results indicated the participants' CB sentence reading speed is best accounted for by the efficiency in lexical semantic access while variables representing breadth of L2 knowledge and strictly visual word recognition latency contribute only marginally to the explanation of the CB reading speed. The implications of the findings will be discussed with particular reference to learner-to-text interface in reading ability assessment.

---

**What's the pay-off? Outcome measures in school foreign language learning**

**Kathryn Hill, Noriko Iwashita & Tim McNamara**

Internationally, demand for accountability has been increasing in all areas of education, including language education. Governments often turn to outcome measures; determining what form they will take is crucial, because of their potential washback effect. Reaching agreement on formats becomes more difficult in devolved systems, where negotiation across multiple curriculum frameworks is required. But data are needed, as their absence limits discussion regarding languages policy in schools.

This paper reports on an Australian project to resolve these difficulties, involving the development of a set of nationally agreed performance measures and procedures for assessing and reporting student learning outcomes at the end of primary school (years 6/7) and year 10 for two Asian languages (Indonesian and Japanese), which are among the most popular foreign languages in Australian schools.

The project involved a review of the various curriculum and standards documents used in Australia, and of systems established nationally and internationally for assessing and reporting school language learning performance. Subsequent analysis (following Mislevy, 2003) classified the material from the review into claims about what learners can do, evidence to support these claims, and the reporting of learner achievement. This formed the basis for a proposed set of common measures, whose acceptability was then negotiated with key stakeholders.

The paper presents an account of the agreed outcome measures, and discusses the problems encountered of diverse understandings of curriculum frameworks; developing an analytic framework of claims and evidence; cross-language differences; the role of textbooks; and the politics of negotiation across diverse educational settings.
Defining the construct 'academic language' through empirical research

Gary Buck, Alison Bailey, Yeonsuk Cho & Sari Luoma

In response to the demands of the 'No Child Left Behind' legislation, Ballard & Tighe has decided to create a new English proficiency test for use in US K-12 schools. One of the main purposes of this test will be to determine whether English language learners have sufficient English to function successfully in mainstream classrooms. A strong criticism of current tests used for that purpose has been that they test general social language rather than the language actually used in classrooms in the service of learning (see e.g., Bailey, in press). The purpose of this study is to determine exactly what language knowledge students need (vocabulary, structures, and discourse skills), what functions they need to carry out, and what tasks they must perform, in each of the grades within US education.

The exact nature of academic language is something on which scholars differ. This study will attempt to answer the question through empirical research. The design takes a multi-pronged approach including review of current research, textbook analysis, test analysis, classroom observation, and questionnaires to teachers at each grade level. The work will be synthesized into an operational definition of academic language suitable for use in designing test specifications.

The paper presents and discusses the results of the study. The theoretical implications from the paper concern the relationship between social and academic language, while practical implications are relevant for the new generation of tests that will follow from the NCLB legislation.

An approach to strengthen the assumption of representativeness, generalisability and predictability of language assessment tasks

Sara Gysen

The language examination system of the Certificate Dutch as a Foreign Language addresses language assessment from a needs related, functional and contextualised language proficiency perspective. This approach is closely linked to the principles of task-based language assessment (Long & Norris, 2000; Van den Branden, Gysen & Depauw, 2002).

Based on language learners' needs, Dutch language examinations for four different societal domains were developed. Each of these Target Language Use (TLU, Bachman, 2002) domains has been described by means of 'typical tasks' (Van Avermaet et al., 2003) which were generated from an exhaustive list of language use situations or 'real life' tasks of each TLU domain. These 'typical tasks' are generic and their level of difficulty is determined by a set of linguistic and non-linguistic parameters. The use of 'typical tasks' provides a description of TLU domains that is neither too specific, nor too abstract.

Bachman (2002), in his reflection on task-based language performance assessment, expresses the concern of the extrapolation and generalisation of performance on assessment tasks to the TLU domain. The approach to language assessment presented in this paper will try to provide a possible answer to this critique. We think that through the translation of specific tasks for a TLU domain, derived from needs analysis research, into more generic 'typical tasks', as a basis for language assessment, the claim of representativeness (cf. content and difficulty), generalisability and predictability (extrapolation) of the examination tasks with regard to the TLU domain can be made.
Linking results on a new language test to national performance level descriptors

Dorry M Kenyon

In language testing, several scales of proficiency have appeared as "standards", including those of ACTFL, ALTE, and the Common European Framework. In the psychometric literature, various approaches to linking test performance to "standards" have been developed. These standard setting procedures, through which judges make decisions regarding classifications of test performance, produce achievement levels that interpret performance on specific tests.

In language testing, however, a problem exists when attempting to link results on new performance-based assessments to extant proficiency-based descriptors of performance levels. How can test results be defensibly and validly related to the levels delineated in the descriptors? This paper explores the pros and cons of utilizing three newer approaches to standard setting from the psychometric tradition. These approaches are the Body of Work Method (Kingston, Kahl, Sweeney and Bay, 2001), the Bookmark Procedure (Miztel, Lewis, Patz and Green, 2001), and the Analytic Judgment Method (Plake and Hambleton, 2001).

The paper then describes an eclectic approach that adopts features of all three methods. The approach was used to link results on a new assessment of listening and speaking skills for adult ESL learners in the United States to the ten Student Performance Levels (commonly used in adult ESL education) and, more importantly, to the six levels of the National Reporting System (NRS), now used as a part of the federal accountability system. The outcome of this eclectic methodology, conducted in December of 2002, will be presented, as well as the results of studies to independently cross-validate that outcome.

C-tests as measures of general language proficiency

Thomas Eckes & Rüdiger Grotjahn

The construct validity of C-tests has been an issue of debate for many years. While some scholars criticized C-tests for tapping only highly specific linguistic abilities, others argued in favor of the notion that these tests are well-suited to assess general language proficiency.

To take a closer look at this issue, the present research studied the relationships between a carefully constructed C-test, composed of 4 texts with 20 blanks each, and the subtests of the recent "Test Deutsch als Fremdsprache" (TestDaF; Test of German as a Foreign Language). Advanced learners of the German language from 4 independent samples (total N = 470) completed the TestDaF and the C-test. The items of the C-test were inserted between the Reading subtest and the remaining 3 TestDaF subtests (Listening, Writing, and Speaking). Using a Rasch measurement approach, examinee proficiency measures were computed. Fit indices showed that each of the TestDaF subtests and the C-test reliably measured a single dimension. To examine whether these tests all measured the same dimension, a structural equation modeling analysis was performed. In each sample, the one-factor congeneric measurement model was strongly supported. Moreover, the C-test had particularly high loadings on the latent variable. Linear regression analysis with the C-test as a criterion and the TestDaF subtests as predictors yielded particularly high beta coefficients for Writing and Speaking.

Taken together, these results clearly attest to the validity of the C-test as an efficient instrument for measuring general language proficiency.
19  July 24
13:45 – 15:15

Students’ and parents’ perceptions of their rights as test consumers
Pavlos Pavlou

This paper is a part of a research project on consumer’s rights and language testing, which builds on the discussion of democratic perspectives on assessment Shohamy (2001). Shohamy mentions that one necessary principle in assessment is the need to protect the rights of test-takers. Moreover, she cites this need as well as the need to conduct and administer testing in collaboration and in cooperation with those tested. This paper relies upon these principles and elaborates upon the rights of the test taker as consumer. Such basic rights and the responsibilities of testers with regard to possible violations are briefly discussed.

The main focus of the paper presents cases of possible rights violations that have been revealed through analysis of the results of a survey among test-takers and their parents.

The elicited information relates to the following consumer rights: the right to take test of choice, the right to be informed about the content and possible uses of the test results, and the right to protest in cases where they are not pleased with the test.

20  July 25
09:15 – 10:15

Conflicting perspectives? Judging clinical communication skills in a test of bilingual proficiency
Elisabeth Grove & Anna Whelan

Quality of care is compromised when communication between health professional and patient is impeded by a lack of shared language. In multi-ethnic communities with large populations of immigrants from diverse language backgrounds, this situation is both familiar and problematic.

Specialist interpreter services provide invaluable assistance, but cannot meet the huge demand for their services. However, many health professionals are immigrants, or the children of immigrants, possessing linguistically and culturally diverse skills that may be used to enhance clinical communication. But because there is considerable variation in proficiency, these resources are often misused through the absence of valid and reliable procedures for determining what bilingual staff are actually able to do with their mother tongues.

An innovative program to assess the bilingual communication skills of health professionals for accreditation purposes is the subject of this paper. We focus on the differing perspectives of the language-trained and occupational experts who collaborated to develop a test of the ability of English-medium-trained bilingual health professionals (Cantonese, Mandarin, Spanish and Vietnamese background-speakers) to use their background language in interactions with patients. The paper explores some of the issues that had to be resolved in test design and in the subsequent setting of standards of performance, since the groups and individuals involved had different understandings of what such a test instrument could or should be expected to measure. Discussing the source of these differences, and resolving them as far as possible, was essential to deciding what constituted adequate performance and to establishing test validity.
<table>
<thead>
<tr>
<th>#</th>
<th>Day/Time</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>July 25</td>
<td><strong>Validities: multidimensionality in educational measurement</strong></td>
</tr>
<tr>
<td></td>
<td>09:15 – 10:15</td>
<td>Hameed Esmaeili</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Messick (1988) justifiably questioned the tangibility of applying validity evidence in testing practice. Messick (1989), Angoff (1988), Cumming (1996), Bachman and Palmer (1996), and Braun (2000) all portray numerous sources for collecting validities evidence. The question arises as to why and how all this evidence led to a belief of a unitary concept. This article is an attempt to penetrate into this fundamental issue, to introduce a framework that acknowledges the multidimensionality of validities and views authenticity and real-life situations as its prime criteria and foundation. Furthermore, the proposed framework includes all validity indices significant in educational measurement and presents principles and a checklist of excellence, appropriateness, usefulness, effectiveness, and fairness of tests to be employed in test development and test use. The framework is based on the beliefs that (a) explanation and understanding is the ultimate goal of validation in assessment (Cronbach, 1989), (b) the process of validation requires a continuous accumulation from a variety of sources (Clineham, 1998), (c) the rights and needs of client populations should be taken into account (Hamp-Lyons &amp; Lynch, 2000), (d) since constructs are 'open concepts' (Messick, 1988), real-life situations should be the bases of our conceptualization of constructs, (e) examinees should be empowered (Shohamy, 2000; Cook-Sather, 2002), specifically by their involvement in scoring, and (f) new tools and constant change in human life require revisiting of applied frameworks at any point of time and executing necessary adjustments.</td>
</tr>
<tr>
<td>22</td>
<td>July 25</td>
<td><strong>Diagnosing foreign language proficiency</strong></td>
</tr>
<tr>
<td></td>
<td>11:15 – 12:45</td>
<td>J Charles Alderson</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Whilst procedures for the certification of foreign language proficiency are relatively familiar, little is known about how such proficiency develops, or about how strengths and weaknesses can be diagnosed. This paper reports on the work of a project to develop diagnostic tests in 14 European languages, to be delivered over the Internet, known as DIALANG. DIALANG offers learners the opportunity to assess their own proficiency, using instruments based on the scales of the Council of Europe's Common European Framework (CEF), and to compare their performance with their self-assessment. The system also offers learners the opportunity to explore feedback on their performance and advice on how they might progress to a higher level of proficiency. The project presents a unique opportunity to study in detail the nature of foreign language proficiency across different European languages. The data consist of learners' responses to pilot test items, their self-assessments, and their responses to a set of pilot questionnaires. The paper reports on an extensive analysis of this data in order to develop and test hypotheses about how learners' proficiency develops across a number of languages. Learners' performances on items and whole tests are compared with the theoretical statements of the CEF in order 1) to compare the development of proficiency in different foreign languages; 2) to explore, for any given language, how this developing proficiency differs among learners with different first languages; and 3) to compare insights into proficiency provided by learners' self-assessments with the assessment provided by the tests.</td>
</tr>
</tbody>
</table>
# 23  
**Day/Time**  
July 25  
11:15 – 12:45  
**Paper**  

*Self-assessment in independent learning programmes: task-based performance*

**Annie Brown**

The availability of multi-media technologies in education has made the option of independent learning increasingly attractive. However, independent learning presents new challenges in assessment, in that the onus is placed upon the learners themselves to monitor and evaluate their own learning.

This paper reports on the development of a novel approach to learner self-assessment in the context of a CD-Rom-based independent learning course in Chinese. Originally developed as a distance course (Orton, 2002), it was based on communicative principles, and summative end-of-unit assessment was text- and task-based. The move to independent learning naturally led to a loss of teacher involvement, so the challenge was to find a way to help learners to effectively evaluate their own performance on these extended performance tasks.

A review of approaches to self-assessment revealed no precedent for learner self-assessment of task-based performance; self-assessment tended to be limited to tasks where students could evaluate their performance as right or wrong. An approach was therefore developed which used annotated samples of student texts, where the annotations would elucidate both the criteria by which performance should be judged and the standards expected, thus enabling learners to assess their own performances by comparing their own work with the annotated samples.

This paper reports on the trialling of the self-assessment methodology on a small sample of students. The methodology was found to be both extremely reliable and useful, allowing students to accurately evaluate their own performance and, importantly, to learn new language skills from the samples.

---

# 24  
**Day/Time**  
July 25  
11:15 – 12:45  
**Paper**  

*Responding to diversity: issues in assessing second language learners with disabilities*

**Lynda B Taylor**

Language test providers are increasingly expected to offer special arrangements for test-takers with disabilities (learning disabilities as well as physical impairments); this usually involves departing from the established testing protocol and modifying test content/administration to minimize the impact of test-taker attributes that are not relevant to the ability being measured. The use of such modifications or ‘accommodations’ (e.g. braille/lip-reading format, assistive technology) raises important theoretical, practical and ethical considerations; in particular, how to determine a disability requiring a test accommodation, which accommodation to provide, and how to interpret scores produced under a test-accommodated condition.

This paper outlines some of the test accommodations offered by one major international provider of English language tests and describes the recent growth in demand for such special arrangements. The practical implications and challenges for test developers, test-takers, testing centres and other test users are discussed.

Professional judgement plays a key role in decisions about the nature and extent of test accommodations; but the role of professional judgement is complicated by the fact that empirical studies on test accommodations are rare due to the constraints on research in this field, e.g. small sample size, non-random selection. With this in mind, the presentation will report on a recent study into the L2 writing performance of test-takers with dyslexia. Findings from this and similar studies, however small-scale, help test providers design accommodations which are suited to the diverse needs of test-takers with disabilities and which increase their chances of participating in mainstream academic, professional and public life.
**International oral language proficiency: construct investigation**

**Micheline Chalhoub-Deville & Gillian Wigglesworth**

TOEFL developers indicate in the test specifications that the purpose of the test is to assess the English proficiency of international students who are interested in studying at colleges and universities in the USA and Canada. The TOEFL results, however, are increasingly being used by various educational and professional institutions around the world. Because TOEFL is administered and used worldwide, it is important to investigate whether test users agree on how to interpret the test scores.

The focus of the proposed project is to determine the extent to which English language learning (ELL) experts, who come from different countries where English is the native language, agree when rating speech samples of international English. The ELL experts targeted include educators residing in Australia, Canada, the UK, and the USA who are likely to teach a diverse group of international students. Educators were asked to rate samples from international students who took the test of spoken English (TSE), the oral component of TOEFL. These speech samples represent the range of tasks found in the TSE, including narration, description, and presentation that target different language functions.

Building on previous research which has demonstrated that different tasks and rater groups affect results obtained from learner performance on oral tests, this project investigated both rating variation as a result of country of origin and variations due to task effects. Multivariate analyses were used to analyze the ratings. The presentation will report the results of these investigations.

---

**Planning, narrative task structure and performance**

**Parvaneh Tavakoli & Peter Skehan**

Task-based language assessment has become a popular vehicle for assessing second language ability since tasks are believed to be appropriate devices to stimulate demands of the real world, elicit "authentic" samples of language and reflect underlying language ability. Research on task-based pedagogy indicates that task characteristics and conditions affect task difficulty and would therefore influence performance on tasks (Skehan, 1998, 2001). However, in assessment settings, the effect of task characteristics and performance conditions on task difficulty, learner performance and learner perceptions of task difficulty have been neither thoroughly researched nor explored. In effect, language testing research concedes that the oral ability continuum is not well defined by differences resulting from task difficulty and tasks are not selected on the basis of a hierarchy of difficulty. Since different task characteristics might have unforeseen influences on language performance, which would in turn intrude into the assessment of such performances, measuring task difficulty remains a significant challenge for performance-based testing.

The present research has attempted to explore how planning, language proficiency and degree of narrative task structure affect task difficulty, learner perceptions of task difficulty and learner performance on tasks. In a study with a 2 x 2 x 4 factorial design, 80 adult Iranian language learners at two proficiency levels performed four narrative tasks of varying degrees of structure in an assessment setting. Fluency, accuracy and complexity of the data were measured. Retrospective questionnaires were employed to investigate learner perceptions of task difficulty. Although some recent studies of task difficulty (Elder et al., 2001; Iwashita et al., 2001) have reported that task characteristics and performance conditions do not influence task difficulty and task performance, results of the data analyses in the present study provide clear indications that the independent variables researched do have a significant impact on task performance.
<table>
<thead>
<tr>
<th>#</th>
<th>Day/Time</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>July 25</td>
<td>Use of multiple-choice completion test for Chinese EFL learners</td>
</tr>
<tr>
<td></td>
<td>15:30 – 17:00</td>
<td>Jianda Liu</td>
</tr>
</tbody>
</table>

Only a few studies to date have focused on assessing the interlanguage pragmatic knowledge of learners of English as a second/foreign language (ESL/EFL). All of these studies included Multiple-choice Discourse Completion Test (MDCT) and showed that MDCT was problematic in terms of reliability and validity. This study attempts to investigate the reliability and validity of MDCT in testing interlanguage pragmatic knowledge of Chinese EFL learners. Two speech acts (requests and apologies) with three variables (imposition, power, and familiarity) were included in this study.

Much effort was devoted to the development of the test paper in this study. The development procedure followed several steps: exemplar generation, to collect situations from the Chinese EFL learners and investigate the likelihood of each situation; metapragmatic assessment, to elicit participants’ assessments of the context variables involved in each scenario; generation of options, to collect distractors and generate the key for each scenario; and pilot testing (both quantitative and qualitative), to validate preliminarily the test paper.

Altogether 212 subjects (both native Chinese speakers and native English speakers) participated in the development of the test. An additional 310 Chinese university EFL learners took part in the actual data collection. The validation was based mainly on Messick (1989), which included a range of statistical procedures and analyses of test-taking process by means of verbal protocols. Quantitative and qualitative data were presented with discussion on the reliability and validity of MDCT in assessing the interlanguage pragmatic knowledge of Chinese EFL learners.
Symposia Abstracts

<table>
<thead>
<tr>
<th>#</th>
<th>Day/Time</th>
<th>Symposium</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>July 23</td>
<td>Exploring the relationship between language corpora and language testing</td>
</tr>
<tr>
<td></td>
<td>16:15 – 17:45</td>
<td>Lynda B Taylor, Paul Thompson, Michael McCarthy &amp; Fiona Barker</td>
</tr>
</tbody>
</table>

This symposium brings together applied linguists and language testers to consider the ways in which language corpora - both native-speaker and second learner - are becoming increasingly useful to the language testing community. Corpora of written language have been available for some decades now and have been instrumental in helping to describe more fully aspects of lexis, grammar, syntax, semantics and pragmatics; more recently, as technology has advanced, it has become easier to construct and investigate spoken language corpora and so gain a better understanding of how speech differs markedly in its use from the conventions of written language.

Being able to define levels and features of language learner performance and match these to the norms of 'expert' or 'proficient' users is a fundamental requirement for language assessment so findings from corpus linguistics research have a direct relevance for language test design and construction. Beyond that, corpus studies have an important contribution to make in areas such as: test validation; the monitoring of standards over time; examiner training and performance; and the investigation of domain specificity.

The speakers in this symposium will highlight a wide range of theoretical and practical applications of written and spoken corpora to different aspects of language proficiency assessment. The session will include plenty of time for questions and for general discussion of possible future directions for corpus research and applications in relation to language testing.

Bringing corpora to bear on the testing of writing and reading

Paul Thompson

In this paper I will review and discuss the uses, both actual and potential, of corpora in the testing of writing and reading.

The presentation will begin with a taxonomy of the different types of corpus that may be of interest to the language tester: the general and the specific 'native speaker' corpus, the general learner corpus, or the examination script corpus, the corpus of readings. These corpora can be in raw form, or they can be annotated for certain features, such as errors. Next, the various uses that such corpora can be put to, and the types of users (for example, item writers, test developers, materials writers), will be considered.

Some practical examples of the value of corpus-derived insights will be given through a review of relevant studies. In relation to writing, for example, Kennedy and Thorp (2001) investigated features of writing across three different bands on IELTS writing task two scripts. Of relevance to reading, Coxhead (1998) has derived the Academic Word List from the study of a corpus of textbook and articles in a wide range of subject areas.

Finally, I will conclude by speculating on the future of corpus research and applications in language testing, and list issues that need to be addressed in the development and interpretation of language corpora in relation to language testing.
Spoken corpora and testing applications
Michael McCarthy
In this presentation I consider how oral testing may gain useful insights from the special kinds of information brought to light by recent work in spoken corpus linguistics. Based on the 5-million word CANCODE corpus and the (10-million word) spoken element of the British National corpus, I examine first the question of lexical levels and in what key ways the typical spoken lexis of native speakers differs from typical written lexis. Although English is the benchmark language, there is reason to believe that some of the broader principles of lexical usage will be valid across languages. Next I consider whether spoken corpus evidence can be used to delineate broad bands of lexical proficiency for oral assessment, from beginning to advanced levels. Questions here include: what constitutes the basic, core spoken vocabulary, what characterises an 'advanced' vocabulary, etc. I then ask similar comparative questions concerning spoken and written grammar, and give evidence from spoken corpora of the pervasiveness of certain structures and the rarity of others, and suggest that our unconscious application of 'written' norms when assessing oral performance may be a hindrance to the desired degree of objectivity. Finally, I look at what spoken corpora tell us about interaction in different genres. I look at the behaviour of listeners, until recently an area neglected in favour of a focus on primary speakers, and consider the implications for our understanding of interaction in the oral testing situation.

Learner Corpora and testing applications
Fiona Barker
Although many corpora exist, including a growing number of learner corpora, they are not yet exploited fully by the language testing community. This presentation considers some of the ways in which learner corpora can benefit language testing in terms of the evidence they provide and how they can inform and validate examinations. These issues are discussed in the Cambridge ESOL context although the insights to be gained are applicable to other contexts.

Collections of examination scripts or speaking tests are a unique record of what learners of English actually produce in live testing events, a very specific context that may incorporate additional features not reflected in other corpora. This evidence can inform many aspects of language testing including describing different levels of performance or the effect of revised mark schemes on examiners. Some evidence from a corpus of speaking tests will be provided to compare children’s speech against teacher-oriented wordlists and the language of examiners against the prescribed interlocutor format. Other ways of using spoken learner corpus data will be suggested. The second part of this presentation suggests how a corpus-informed approach can aid validation activities such as comparing test and version performance over time and comparing candidate performance across different types of examination. Two such activities are described: producing genre specific wordlists for item-writers and developing a common rating scale for writing using corpus data.

Finally, this presentation suggests other potential applications of corpora to language testing and considers how language testers should use corpora to greater effect in future.
# Day/Time
16:30 – 18:00

## Symposium

**Identifying and comparing models of language and pedagogy in ESL assessment scales for linguistic minorities in three countries**

**Constant Leung, Hugh South, Catriona Scott, Sibel Erduran and Pauline Rea-Dickins**

In recent years a number of ESL assessment scales for linguistic minority students in mainstream education contexts have been developed. These claim to be informed by specific theories of language or notions of communicative competence. Some are explicitly pedagogically- and formatively-oriented. All seem to be amenable to being used, *inter alia*, for classroom-based teacher assessment. Some of the epistemological and operational complexities in assessment scales of this kind have been discussed (see Brindley 2002; Rea-Dickins and Gardner, 2000; Teasdale and Leung, 2000, among others). The focus of this symposium is on four such scales developed in Australia, Britain, Canada and the United States. The presenters, drawing on the data of an ongoing research project supported by the Paul Hamlyn Foundation and NALDIC (National Association for Language Development in the Curriculum), will provide an account of the assessment scales in question in respect of their conceptualisation and operationalisation of:

- the construct of assessment
- language performance, in relation to grammar, communicative use, cultural knowledge, and strategies of language use.

These dimensions will be analysed with reference to the pedagogic contexts and purposes for which they have been designed.

The symposium is organized as five 15-minute papers, each followed by 5-minute follow-up for clarification. The first four papers will each focus on a particular assessment scale. The final paper will offer a comparative view of all the four scales discussed, and their contributions to summative and formative assessment. There will be an introduction (5 minutes) and a final 15-minute plenary discussion.
Poster Abstracts

1
July 24
15:15 – 16:30

Assessing writing for academic purposes: a comparative study of IELTS and TOEFL

Mark Beittel

The purpose of the writing components of both the TOEFL and IELTS is to assess academic writing ability, but many differences exist in how they prompt and score writing samples. The aim of this study was to investigate whether the two tests perform similarly. Forty-five student volunteers from the University of Trento (northern Italy) were recruited to take both the IELTS Academic Writing Module and the TOEFL Writing and Structure sections. One important aspect of the research design was to have both tests corrected authentically.

The main hypothesis was that scores on one test serve as a good predictor of scores on the other. For the purpose of the study, predictive strength was defined as the degree of association between test-takers’ scores, with “good” predictive strength specified as a range of correlation coefficients that has as its upper limit the coefficient for the scores given by the two raters of the TOEFL Writing section and as its lower limit the coefficient between the scores of the TOEFL Writing and Structure sections.

The main conclusion was that there is a strong degree of association between the overall scores \( r = .77 \), a figure which is indeed comparable to the correlations between the raters of the TOEFL Writing section. Additional analysis of the data indicated that these writing tests are surprisingly good predictors of how participants scored on the other. The implications of this study for further comparative research on the measurement of writing ability in large-scale proficiency exams are also presented.

2
July 24
15:15 – 16:30

Validating TESOL standards in US education

Gary Buck, Meg Malone, Yeonsuk Cho & Sari Luoma

One important trend in US education over the last decade has been the move to Standards-based instruction and testing. In fact, the new education act known as ‘No Child Left Behind’ explicitly requires that assessments be aligned to English language proficiency standards. There are number of these; the most common being the TESOL Standards. Based on the act, tests derived from these standards will be used to determine whether students have attained sufficient English language proficiency to function in mainstream classes, with native speaker children. In a sense then, the standards define the construct that such tests should measure.

However, most standards commonly used, including the TESOL standards, reflect teachers’ classrooms experience rather than empirical research. The question then arises whether these standards define an appropriate construct. The purpose of this session is to address that question. As part of the development of a new test of academic English for K-12 children, an empirical needs analysis is planned to determine exactly what language proficiency students need to have mastered to function successfully in mainstream classrooms. We propose to compare the results of that needs analysis with the standards. Or more specifically, we will attempt to determine whether the construct that such a test needs to measure as determined by empirical research, would be the same as the construct based on the TESOL standards. The results of the comparison will be presented and discussed with the participants. We will also discuss implications for test design and development.
<table>
<thead>
<tr>
<th>#</th>
<th>Day/Time</th>
<th>Poster</th>
</tr>
</thead>
</table>
| 3 | July 24  | *What’s so hard about that? Investigating item/task difficulty across two examinations*  
Barbara K Dobson, Mary C Spaan & Amy D Yamashiro  
As part of a larger test validation research study, two different EFL examinations, one targeted at the advanced proficiency level and the other at an intermediate level, were compared. The advanced-level examination has five sections: speaking, writing, listening, grammar/vocabulary/reading (GVR), and cloze. The intermediate-level examination has just four sections: speaking, writing, listening, and GVR. Some of the task types are different in the two exams, e.g., listening prompts with picture alternatives versus listening prompts with text alternatives. Other task types are the same across the two exams, e.g., grammar.  
In this study, the participants (N=146) took both examinations and their test scores and individual item responses were analyzed to answer the following research questions:  
1. To what extent do the two exams represent different levels of EFL proficiency?  
2. To what extent do the individual sections of the two exams represent different levels of EFL proficiency?  
3. To what extent are different difficulty levels represented by different task types?  
4. Can features be identified that characterize tasks at different proficiency levels?  
5. Where the task is the same across the two tests, to what extent do individual items reflect different difficulty levels?  
6. Can features be identified that characterize items at different proficiency levels?  
Results of this analysis can be used to improve item specifications and can provide helpful interpretive information to test users and test takers. |
| 4 | July 24  | *An ethical use of IELTS scores in the recruitment of international students*  
Kevin Haines & Angela Ashworth  
With commercial pressures in Dutch higher education encouraging universities to become more international, high-stakes language tests such as IELTS play an increasingly critical role in the recruitment of students. We are interested in the ways that scores derived from these tests are used to decide the educational future of candidates. From an ethical perspective, we would like to find ways of avoiding such tests being used, in the words of Spolsky (1995), "to play Russian roulette with the test takers".  
Firstly, we will describe the difficulties in ensuring the validity of interpretations drawn from such test scores, highlighting the inherent ethical consequences of using such tests outside the contexts for which they were designed. Our aim is to define the limitations of these test scores, while recognising that the information provided still has great value when used appropriately in combination with other assessment tools.  
Secondly, we are investigating the practical opportunities for IELTS as a formative tool in the educational process. Gibb and Fellowes (2002) have questioned whether IELTS is currently being used to its strengths when used by institutions solely as an entrance test. We feel there may be scope for positive washback when we combine IELTS-oriented teaching and assessment with our teaching programmes for pre-sessional, undergraduate and Masters students.  
Our thesis, therefore, is that, despite its limitations when used as a normative entry tool, IELTS may provide a foundation that can be used by local testers, developers and teachers to ensure quality in formative assessment practices. |
<table>
<thead>
<tr>
<th>#</th>
<th>Day/Time</th>
<th>Poster</th>
</tr>
</thead>
</table>
| 5  | July 24   | *The validation study on the first achievement test in Japanese university entrance examinations*  
Akihiro Ito  
The purpose of the present study is to report on the validation study of the English language test in the First Achievement Test (FAT) constructed and administered by the National Center for University Entrance Examinations in Japan. In order to apply for universities in Japan, especially for national universities, examinees must take the FAT to show their achievement levels in a variety of subjects. The English language test consists of five subtests: (1) paper-pencil pronunciation test; (2) paper-pencil stress test; (3) grammar and usage test; (4) spoken English expression; (5) sequencing; and (6) reading comprehension. The test was administered to 100 Japanese university students learning English as a foreign language. The validity of the test will be examined through (1) a concurrent validation study (correlating with the results of a general English language proficiency test), (2) an internal construct validation study described by Alderson, Clapham & Wall (1995) and (3) a multivariate analysis study. The results indicate that the English test is somewhat reliable and valid for measuring the examinees' English language proficiency but there is a severe internal problem with the distribution and number of test items. |
| 6  | July 24   | *Developing a test of lexical choice*  
Tomoko Ishii  
This poster will present a part of the author's PhD project, which is the development of a battery of vocabulary tests for Japanese learners of English. The battery will consist of four tests which address different aspects of vocabulary knowledge and will aim to draw a profile of types of vocabulary knowledge learners have. One of the four tests, which assesses the learners' ability to choose an appropriate word from two or more semantically close words, is to be presented in this poster.  
Learners often confuse words which are similar in meaning, or which are translated in their L1 in a similar way and make errors such as below. (Erroneously used words in capitals, intended meanings between parentheses.)  
- My schedule is very STRICT. (tight)  
- I had to EXCHANGE the tires because they were worn out. (change/replace)  
Although L1 translation can be an efficient way to convey the meaning when introducing new words, it also causes problems as shown in the examples above, and learners have to develop a deeper understanding of meaning of words and their collocational behaviour at some stage of their learning. The test to be presented in this poster is designed to capture such a development. The poster will include a discussion of the problems encountered in the development and validation of the test and their solutions. |
Poster

Vocabulary levels test: an issue of item independence in clusters

Tadamitsu Kamimoto

The Vocabulary Levels Test (Nation, 1990) has been used widely as a test to measure learners' vocabulary size. While some validation studies have been carried out (Beglar & Hunt, 1999; Read, 1988; Schmitt, Schmitt, & Clapham, 2001), there is one particular problem which has not been addressed satisfactorily. That is the issue of item independence, which means that success or failure of one test item should not influence the success or failure of any other items (Brown & Hudson, 2002). However, from the way words are presented in the Levels Test, there is a risk of this principle being violated. The test items are presented in clusters which consist of 6 words each (3 target words and 3 distractors) and 3 definitions; the learners' task is to match each definition with a word. The purpose of this matching-format is to involve less L2 reading and to reduce the chances of guessing, but this may violate local independence within a cluster. To make the matter more complicated, this principle is also one of the assumptions of Item Response Theory (IRT), which has become a standard way of analyses of the tests. Beglar and Hunt (1999) and Schmitt, et al. (2001) both used IRT as a measure of analysis and, although well aware of a possible violation of item independence, they have not dealt with this issue squarely.

This study addresses this issue with a research design that allows us to investigate directly whether items in a cluster can be considered independent.

Norwegian CATS say "yes" to Europe: using CEF in school English tests

Eli Moe & Cecilie Carlsen

This poster presentation will highlight the issues and problems encountered so far in an attempt to test the English of pupils in Norwegian schools, based on the CEF levels and using branched computer adaptive testing.

A first step in the work has been to establish which levels on the CEF best describe pupils' performance at four stages in schooling across a range of skills. These levels are defined as target levels for the particular stage. Sets of "can-dos" have been developed for each CEF level, which form the basis of test items.

By allowing pupils to begin with a test set below the target level and then preceding to more advanced sets if appropriate, it is hoped to establish whether a pupil is "at", "below" or "above" the target level in any skill.

The project is very demanding - not only on those who work for it, but on the computer skills and equipment in schools across Norway. For this reason we are beginning "gently" with a full internet delivered reading test planned for the spring of 2004 and a cd of practice tests for reading, writing and listening. Additionally CEF-based rating criteria will be developed for the assessment of writing and speaking. This material will prepare the ground for the next round which will extend the range of skills tested.
9  July 24  
15:15 – 16:30  
Yuji Nakamura  

A many-facet Rasch analysis of three-dimensional oral proficiency test data  

In a real life situation or even in academic settings, oral language proficiency is not an isolated skill. It is related to sociocultural factors, personality factors, interpersonal skills etc. Furthermore, oral language skill itself is not a single unit. Asking questions in lectures is different from participating in discussions or making oral presentations in class.  

This research examines the effectiveness of students’ English communication abilities by investigating the target language use in monologue, dialogue and “multilogue” settings. It explores Japanese college students’ oral English proficiency by focusing on the linguistic and pragmatic aspects in three types of speaking tests: 1) monologue (speech-making test), 2) dialogue (tape-mediated sociolinguistic test), and 3) multilogue (small-group discussion test).  

50 Japanese students’ speaking test data were evaluated by four raters using a four-point scale (1: poor; 4: good), and data analysis was conducted using the FACETS model so that the three facets (students, raters, and evaluation items) can be shown on the same continuum.  

The results demonstrate that some students who are good at monologue tests or even dialogue tests, but not good at discussion-oriented tests. Students should be urged to develop oral communication-oriented learning habits since fluent oral communication in the target language is the ultimate goal for most, if not all students. Future research should be carried out by considering: 1) Is there any cultural oriented influence on a lesser ability to communicate in multilogue?, 2) What is the construct of practical, basic speaking ability common to all those tests?

10  July 24  
15:15 – 16:30  
Soubakeavathi Rethinasamy  

The effects of different training procedures on rating accuracy: an experimental study  

Rater training which was once scorned upon is now widely accepted and implemented by many testing organisations. During rater training, various procedures are employed to ensure that each test-taker's score is not dependent on which rater have rated his/her performance. Although, it is clearly an important element within the whole rating procedure and as such major has major validation concerns, a review of literature shows that there is a serious shortage of empirical evidence relating to the impact of different training procedures on raters rating performance.  

In addition to outlining the major issues related to rater training, this poster will describe in some detail the design of a pre-post experimental study which uses qualitative (based on questionnaire and semi-structured interview) and Multi Faceted Rasch (MFR) based quantitative method to shed light on the effects of some commonly used rater training procedures on rater accuracy. Since the poster is therefore primarily methodology focused, issues related to the potential implication of the study will not be highlighted in great detail.
Poster

Developing a vocabulary test for Malaysian primary school students

A Affendi Shabdin & Norbert Schmitt

The Ministry of Education Malaysia has admitted that the poor command of English especially among students in the primary school was due to, among others, lack of vocabulary. Since "words are the basic building blocks of language" (Read, 2000:1) insufficient amount of vocabulary will not only hamper students' ability to read but will also affect their ability to apply various vocabulary learning strategies (Laufer, 1997; Ryan, 1997; Schmitt, 2000). It was suggested that learners need to have a good knowledge of at least 3000 high frequency words before learning other vocabulary (Nation and Waring, 1997).

This study developed a new test (following the frameworks proposed by Bachman and Palmer (1996) and Read and Chapelle (2001)) for assessing the vocabulary knowledge of primary school students in Malaysia. All of the items in the test, based on the Vocabulary Levels Test format, were visually represented (i.e. pictures) since most represented basic words were "almost impossible to define or explain using words of a similar frequency" (Read, 2002: 168). In a series of pilots (including a final pilot with 1077 students), the validation procedure used by Schmitt, Schmitt and Clapham (2001) was replicated to obtain validity evidence for the test. The results strongly suggest that the test has acceptable validity and reliability characteristics when used with the target students.

This poster summarises the stages that were involved in constructing and validating the test, highlighting some of the problems encountered and their solutions. Non-native speaker assessment, item analysis and students' interview results are presented.
## Research Network Abstracts

<table>
<thead>
<tr>
<th>#</th>
<th>Day/Time</th>
<th>Research Network</th>
</tr>
</thead>
</table>
| 1 | July 25  | **A construct-based approach into the nature of reading: understanding, identifying and validating reading skills**  
Hossein Barati  
The notion of componentiality of reading has been considered by many practitioners in both fields of teaching and testing English as a foreign/second language (e.g., Alderson 2000, Weir et al 2000, Urquhart and Weir 1998). However, despite all the efforts, it is still a question whether reading ability is unitary, bi-divisible or multi-divisible. Moreover, the fact that almost all available reading skill taxonomies have been developed by native speakers (NS) of English, who might use different strategies in reading their L1 compared with non-native speakers (NNS) reading in their L2 (Alderson 1995), has put the issue under further questions. Also the employment of NS expert judges in various investigations to decide about item/skill correspondence in tests of L2 reading, and the effect of L1 reading and cognitive development on L2 reading comprehension have been debatable areas. Very little attention has also been paid to the interaction between EFL learners and the written text.  
This study aims at investigating further the nature of reading. It uses both qualitative and quantitative methods and will basically concern the interaction between non-native readers and the text in identifying reading skills. The main research questions include 1) Do native and non-native expert judges assign reading skills to test items similarly? 2) Will 'think aloud' procedures help identifying reading skills? 3) Do test-takers with different L1’s and similar FL proficiency perform differently on the same test of reading skills? |
| 2 | July 25  | **Developing an accountability system for measuring annual progress of ESL students**  
Gary Buck, Yeonsuk Cho & Sari Luoma  
The main focus of the US 'No Child Left Behind' act is accountability. All schools must show that their students are making adequate annual progress. In the case of mainstream content classes, the schools are also required to show that more and more of their students are reaching the 'proficient level'. By 2013/14, 100% of their students must be 'proficient' at grade level in reading and mathematics. However, there are 4.5 million English language learners in the system, and the act also requires that schools show them making adequate progress in their language learning. The federal government have offered no guidelines on how that should be done—what is clear is that the method used to determine school progress in content knowledge, described above, cannot be used for ESL students who, by definition, cannot be 100% proficient.  
The purpose of this study is to explore one way of developing an accountability system for ESL students in K-12 education. The idea is simple in principle: we will go back into student records to retrieve standardized proficiency test scores on the Idea Proficiency Test for students who have been tested annually for at least three years, calculate mean gains, and use the data to set standards for annual progress. We will then explore a variety of ways of aggregating the individual data for school accountability.  
The study will be described, the results will be presented and discussed, and implications drawn for the development of accountability system for ESL learners in US schools. |
3  July 25
14:15 – 15:30

Research Network

Managing the transition: the role of assessment at the primary/secondary school interface

Kathryn Hill

In Australia, as elsewhere, targets have been set to increase the number of students studying a foreign language from early primary through to the end of high school and beyond (Rudd 1994). Clearly, the main objective of this policy is higher proficiency outcomes. However, a number of studies have concluded that, despite their longer exposure, students who commence foreign language study at primary level do not necessarily outperform those commencing in later years of schooling (Burstall et al 1994, Johnstone 1994, Genelot 1996, Hill et al 1997, Hill 1998). Nevertheless, studies investigating the transition from primary to secondary foreign language learning suggest that this disappointing outcome is not necessarily inevitable (Hill, 2001, Ward & Hill, forthcoming). One of the issues these studies have identified is the need for more appropriate assessment and reporting to ensure that what students have learned in primary school is both recognised and capitalised upon at high school.

The proposed research will investigate the role of assessment and reporting in improving learning outcomes for students continuing the study of a foreign language from primary school. A pilot study involving one primary school and one high school will investigate students’, parents’ and teachers’ perspectives on how well existing assessment and reporting procedures capture what students have achieved by the end of a primary foreign language program. It will also explore how this information is used by high school foreign language teachers. The results will be used to produce revised assessment and reporting procedures for the main study.

4  July 25
14:15 – 15:30

Validating Japanese language can-do-statements

Sukero Ito, Soichi Aoki, Mako Aoyama, Kaori Asami, Masayo Hirokawa, Mikio Kaawarazaki, Kenichi Koide, Kaoru Koyanagi, Hiroyuki Noguchi, Kazuo Ohtsubo, Reiko Saegusa, Megumi Shimada, Yuan Sun and Yoshiko Tsuruo

Japanese language Can-do-statements where developed for the purpose of providing an external criterion for evaluating the validity of the Japanese-Language Proficiency Test (JLPT). It was indicated by Saegusa, et al., (1999) that, while Can-do-statements themselves have a high level of reliability, the correlation coefficient of Can-do-statements to the JLPT is not quite so high. Based on the above results, the Japanese language Can-do-statements were re-examined and the number of questions were increased to 60, while at the same time the content of the questions was limited to situations familiar to the respondents.

In addition to the above, both the object and method of investigation were improved in the following manner.

1) The object of the investigation was broadened from the previous focus on examinees who sat for the 1st Level JLPT, to include examinees who sat for the 2nd Level JLPT.

2) In order to broaden the range of the object of this investigation to cover respondents from beginner to advanced level, as well as to refine the ability of the Can-do-statements to act as an external standard of measure, the correlation will be examined between Can-do-statements and other tests which measure Japanese language proficiency.

3) Due to the difficulty of investigating actual behavior, teacher evaluations are being carried out to check the validity of self-evaluation in Can-do-statements. Based on the results of the analysis of the above points, it has been confirmed that Can-do-statements function sufficiently as an external criterion for evaluating the validity of the JLPT.
5  July 25  
14:15 – 15:30  
**Research Network**

*The underachievement of ethnic minority students in Wales*

**Christine Pegg**

The underachievement of ethnic minority students in Wales is currently a subject of local concern. A National Assembly of Wales grant was awarded last year to investigate this area. The results of this extensive survey are indicative of problematic areas within the field of language teaching provision and subsequent assessment. The lack of achievement of ethnic minority students is possibly the result of inappropriate assessment criteria relevant to the needs and provision for this group of students.

Fieldwork based on interviews and the collection of personal data from teachers, parents and pupils has provided detailed quantitative evidence of the current situation. In so far as ethnic minority students are concerned, the learner interface with assessment requirements is one of great concern. Suggestions for a change in policy to reflect the needs of these young students will be presented.

6  July 25  
14:15 – 15:30  
**A case of washback with primary EAL learners**

**Catriona Scott**

The research network proposed is to present research in progress and provide an opportunity to discuss the research methodology. The research comprises a longitudinal case study investigating the washback effects of summative NC assessment at Key Stages 1 and 2 on: (a) primary school learners with English as an Additional Language (EAL), (b) the 'programme' (Bailey 1996), and (c) classroom language assessment practices.

There is currently a dearth of research into washback in mainstream education from national testing in the UK and the implications for EAL learners. A number of issues arise in the assessment of EAL learners, such as whether the Level Descriptions of the English (subject) National Curriculum are sufficient for language assessment (e.g. Leung & Harris 1997, OfSTED 1997, QCA 2000). Furthermore, special provisions for language support in Key Stage tests are not applied to all subjects, raising issues of 'fairness' (Kirkland 1971 and Shohamy 1997), particularly as there is a potential mismatch for EAL learners between language proficiency and curricular ability (Gardner & Rea-Dickins 1999).

A key issue in this study is the relationship between summative, formative and language assessments. The study comprises a longitudinal examination of washback using qualitative data covering a six-year period, from 1998 to 2003. It involves analysis of key policy documents, assessments and secondary data from Rea-Dickins (1999-2002) research on Classroom Assessment of English as an Additional Language in Key Stage 1 Contexts, interviews with stakeholders in the assessment process, namely pupils, teaching staff, and parents, and classroom observations.
# Day/Time | Research Network
---|---
7 | July 25 | **From the candidate's point of view: diary-keeping in a test context**
14:15 – 15:30 | **Constance Tsagari**

Keeping a diary has been a pervasive narrative form in many people's private reflections in diverse fields such as sociology, anthropology, history, (Cort, 1993), psychotherapy (Progroff, 1975; Assagioi, 1976, etc. In English language teaching and learning in particular, diaries have become increasingly significant as an introspective technique over the past twenty years, contributing to our understanding of the processes underlying language learning and teaching in formal classroom settings (Schumann and Schumann, 1977; Schumann, 1980; Jones, 1977; 1994; Bailey, 1980; 1983; Schmidt and Frota, 1986; Brown, 1983; 1985; Gillete, 1987; Howell-Richardson and Parkinson, 1988; Halbach, 2000; Myers, 2001; inter alia).

The present paper will report on a diary study that was set in a Greek private language school involving teenage students preparing for an EFL test, namely the First Certificate in English (UCELS). This explorative, qualitative piece of work, being part of research in progress, has brought into light many interesting findings concerning teaching and learning English in the present context of inquiry and has shed light onto issues such as test-preparation and examination impact on candidates' feelings and motivation towards learning English. The research data mainly involved the diarists' entries but an attempt was also made to triangulate these with teachers' perceptions and classroom observations. The approach to the analysis undertaken was essentially heuristic and one of discovery (Seliger and Shohamy, 1989) yielding rich and exciting findings. The role of diary-keeping both as a classroom and research instrument, the process of setting up a diary project, the pitfalls and "rewarding moments" will also be discussed. It is hoped that the present study will relate and contribute to the work of other researchers who undertake investigations into the impact of language tests or use diaries as a research instrument.

8 | July 25 | **Summarisation as a measure of reading comprehension: expert and popular dialogues?**
14:15 – 15:30 | **Guoxing Yu**

Summarization has been extensively researched in psychology and discourse comprehension (Hidi & Anderson 1986), especially since the seminal works of Kintsch and van Dijk (1978) and Rumelhart (1977). As the hub of reading comprehension (Axelrod 1975), summarization involves full understanding (Nuttall 1996) and a "full range of organizational characteristics" (Bachman 1990: 139) of a text, and has a natural appeal in communicative language testing (Brown & Smiley 1978; Cohen 1993, 1994). However, in language testing, traditional summarization tasks are compromised (see concerns of Alderson 1996; Head et al 1989; Weir 1993; and proposed compromises of Bensoussan 1993; Hughes 1989; Huhta & Randell 1996; Taylor 1996). Existing research on summarization in reading comprehension measurement focused on short texts to be summarized in either target language or L1 at paper-and-pencil contexts (e.g. Kobayashi 2002; Riley & Lee 1996; Taylor 1996). This presentation reports part of a study investigating undergraduates performances of traditional summarization tasks in both L1 (Chinese) and EFL after reading lengthy English texts either computer or paper presented. The students are randomly assigned to one of the six experiments (3 text types X 2 presentation modes). Their written summaries (English and Chinese) are scored against both the "popular" and the "expert" templates generated from among the students themselves and from language experts and an auto-summarization programme respectively.

Of central focus in this presentation are the "dialogues" between the "expert" and the "popular", and their differential impacts on summarization score variances across the experiments with particular references to (1) source text discourse features, (2) presentation modes, (3) summarization languages, (4) participants' computer familiarity, and (5) participants' reading and other language abilities.
9    July 25    14:15 – 15:30   Research Network

_Alistair Van Moere & Miyoko Kobayashi_

*Who speaks most in this group? Does that matter?*

This study explores the effects of group configuration and characteristics on learners' performance in group oral tests. Performance testing by group discussion is a promising test paradigm in that it is time- and cost-effective and that it reflects real-world and classroom tasks, providing rich samples of learners' natural language use in interaction. However, the amount and quality of such samples largely depends on task conditions such as raters, rating scales, candidate characteristics and interlocutors as well as tasks themselves (McNamara 1996, Skehan 1998). A relatively small number of research studies has investigated the effects of interlocutor variables on oral performance (e.g. O'Sullivan 2002), but most research conducted so far has examined pair-tasks and very few studies have investigated group discussions. The current study examines the performance of individuals within approximately 130 separate small-group discussion tests which form part of an institutional proficiency test in a Japanese university. It investigates whether there is any interaction with the number of turns an individual takes, the size of turns, the learner's relation within the test group, the overall language proficiency level of the test group, and the group gender configuration. All performances are video-recorded and transcribed for analysis. The performances are analysed in terms of their accuracy and complexity, following Skehan and Foster's (1995) framework, and compared with the rater's holistic and analytic grades of the performance. It is hoped that the findings will provide useful insight into the construct validity of group oral tests.
# Conference Participants

<table>
<thead>
<tr>
<th>Family Name</th>
<th>Given Name</th>
<th>Institution</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALDERSON</td>
<td>J Charles</td>
<td>Lancaster University</td>
<td><a href="mailto:c.alderson@lancaster.ac.uk">c.alderson@lancaster.ac.uk</a></td>
</tr>
<tr>
<td>ALMARZA</td>
<td>Miguel</td>
<td>Complutense University, Madrid</td>
<td><a href="mailto:mkalmar@mi.madritel.es">mkalmar@mi.madritel.es</a></td>
</tr>
<tr>
<td>ASAMI</td>
<td>Kaori</td>
<td>Tokyo Medical &amp; Dental University (International Student Centre)</td>
<td><a href="mailto:asakair@clias.net">asakair@clias.net</a></td>
</tr>
<tr>
<td>ASHWORTH</td>
<td>Angela</td>
<td>University of Groningen</td>
<td><a href="mailto:a.ashworth-kadwell@let.rug.nl">a.ashworth-kadwell@let.rug.nl</a></td>
</tr>
<tr>
<td>BAGNA</td>
<td>Carla</td>
<td>Universita per Stranieri-Siena</td>
<td><a href="mailto:bagna@unistrassi.it">bagna@unistrassi.it</a></td>
</tr>
<tr>
<td>BAILEY</td>
<td>Alison</td>
<td>UCLA</td>
<td><a href="mailto:aimbens@ucla.edu">aimbens@ucla.edu</a></td>
</tr>
<tr>
<td>BANERJEE</td>
<td>Jayanti</td>
<td>The University of Lancaster</td>
<td><a href="mailto:J.Banerjee@lancaster.ac.uk">J.Banerjee@lancaster.ac.uk</a></td>
</tr>
<tr>
<td>BANOCZY</td>
<td>Erika</td>
<td>Budapest Business School</td>
<td><a href="mailto:eriban5@yahoo.com">eriban5@yahoo.com</a></td>
</tr>
<tr>
<td>BARATI</td>
<td>Hossein</td>
<td>University of Bristol</td>
<td><a href="mailto:H.Barati@bristol.ac.uk">H.Barati@bristol.ac.uk</a></td>
</tr>
<tr>
<td>BARENFAenger</td>
<td>Olaf</td>
<td>TestDaF Institute, University of Hagen</td>
<td><a href="mailto:olaf.barenfanger@testdaf.de">olaf.barenfanger@testdaf.de</a></td>
</tr>
<tr>
<td>BARKER</td>
<td>Fiona</td>
<td>Cambridge ESOL</td>
<td><a href="mailto:Turner.r@ucles.org.uk">Turner.r@ucles.org.uk</a></td>
</tr>
<tr>
<td>BARNI</td>
<td>Monica</td>
<td>Universita per Stranieri-Siena</td>
<td><a href="mailto:barni@unistrassi.it">barni@unistrassi.it</a></td>
</tr>
<tr>
<td>BEITTEL</td>
<td>Mark</td>
<td>Universita di Trento</td>
<td><a href="mailto:mark.beittel@lett.unir.it">mark.beittel@lett.unir.it</a></td>
</tr>
<tr>
<td>BENKE</td>
<td>Eszter</td>
<td>Budapest Business School</td>
<td><a href="mailto:e.benke@axelero.hu">e.benke@axelero.hu</a></td>
</tr>
<tr>
<td>BERRY</td>
<td>Vivien</td>
<td>The University of Hong Kong</td>
<td><a href="mailto:vberry@hkucc.hku.hk">vberry@hkucc.hku.hk</a></td>
</tr>
<tr>
<td>BOLTON</td>
<td>Sillyle</td>
<td></td>
<td><a href="mailto:bolton@goethe.de">bolton@goethe.de</a></td>
</tr>
<tr>
<td>BRADSHAW</td>
<td>Jenny</td>
<td>National Foundation for Educational Research</td>
<td><a href="mailto:j.bradshaw@mfer.ac.uk">j.bradshaw@mfer.ac.uk</a></td>
</tr>
<tr>
<td>BROOKS</td>
<td>Lindsay</td>
<td>OISE/University of Toronto</td>
<td><a href="mailto:lbrooks@oise.utoronto.ca">lbrooks@oise.utoronto.ca</a></td>
</tr>
<tr>
<td>BROWN</td>
<td>Annie</td>
<td>University of Melbourne</td>
<td><a href="mailto:a.brown@unimelb.edu.au">a.brown@unimelb.edu.au</a></td>
</tr>
<tr>
<td>BUCK</td>
<td>Gary</td>
<td>Ballard &amp; Tighe, Publishers</td>
<td><a href="mailto:gbuck@ballard-tighe.com">gbuck@ballard-tighe.com</a></td>
</tr>
<tr>
<td>CARLSEN</td>
<td>Cecilie</td>
<td>University of Bergen, Norway</td>
<td><a href="mailto:cecilie.carlsen@nor.uib.no">cecilie.carlsen@nor.uib.no</a></td>
</tr>
<tr>
<td>CHALHOUDEVILLE</td>
<td>Micheline</td>
<td>University of Iowa</td>
<td><a href="mailto:m-chalhoub-deville@uiowa.edu">m-chalhoub-deville@uiowa.edu</a></td>
</tr>
<tr>
<td>CHENG</td>
<td>Lying</td>
<td>Queen's University</td>
<td><a href="mailto:cheng@educ.queensu.ca">cheng@educ.queensu.ca</a></td>
</tr>
<tr>
<td>CUMBRENO</td>
<td>Ana Belen</td>
<td>Universidad de Extremadura</td>
<td><a href="mailto:belencespada@hotmail.com">belencespada@hotmail.com</a></td>
</tr>
<tr>
<td>DALLER</td>
<td>Thomas</td>
<td>University of the West of England</td>
<td><a href="mailto:helmut.daller@uwe.ac.uk">helmut.daller@uwe.ac.uk</a></td>
</tr>
<tr>
<td>DAVID</td>
<td>Gergely</td>
<td>Centre for English Teacher Training, Eotvos University, Budapest, Hungary</td>
<td><a href="mailto:david.soprorni@axelero.hu">david.soprorni@axelero.hu</a></td>
</tr>
<tr>
<td>DAVIES</td>
<td>Alan</td>
<td>University of Edinburgh</td>
<td><a href="mailto:a.davies@ed.ac.uk">a.davies@ed.ac.uk</a></td>
</tr>
<tr>
<td>DEVILLE</td>
<td>Craig</td>
<td>University of Iowa</td>
<td><a href="mailto:craig-deville@uiowa.edu">craig-deville@uiowa.edu</a></td>
</tr>
<tr>
<td>DOBSON</td>
<td>Barbara</td>
<td>University of Michigan</td>
<td><a href="mailto:bdobson@umich.edu">bdobson@umich.edu</a></td>
</tr>
<tr>
<td>DOUGLAS</td>
<td>Dan</td>
<td>Iowa State University</td>
<td><a href="mailto:dandoug@iastate.edu">dandoug@iastate.edu</a></td>
</tr>
<tr>
<td>ECKES</td>
<td>Thomas</td>
<td>TestDaF Institute, University of Hagen</td>
<td><a href="mailto:thomas.eckes@testdaf.de">thomas.eckes@testdaf.de</a></td>
</tr>
<tr>
<td>ERICKSON</td>
<td>Gudrun</td>
<td>Goteborg University, Sweden</td>
<td><a href="mailto:gudrun.erickson@iped.gu.se">gudrun.erickson@iped.gu.se</a></td>
</tr>
<tr>
<td>ESMAELLI</td>
<td>Hameed</td>
<td>Seneca College</td>
<td><a href="mailto:hesmaelii@oise.utoronto.ca">hesmaelii@oise.utoronto.ca</a></td>
</tr>
<tr>
<td>FARHADY</td>
<td>Hossein</td>
<td>Iran University of Science and Technology</td>
<td><a href="mailto:farhady@hotmail.com">farhady@hotmail.com</a></td>
</tr>
<tr>
<td>FEKETE</td>
<td>Hajnal</td>
<td>ITK</td>
<td><a href="mailto:hfekete@itk.hu">hfekete@itk.hu</a></td>
</tr>
<tr>
<td>FIGUERAS</td>
<td>Neus</td>
<td>Department d'Ensenyament-Generallitat de Catalunya</td>
<td><a href="mailto:nfiguera@pie.xtec.es">nfiguera@pie.xtec.es</a></td>
</tr>
<tr>
<td>FOX</td>
<td>Janna</td>
<td>Carleton University</td>
<td><a href="mailto:jfox@ccs.carleton.ca">jfox@ccs.carleton.ca</a></td>
</tr>
<tr>
<td>FRASER</td>
<td>Wendy</td>
<td>Carleton University</td>
<td><a href="mailto:wfraser@ccs.carleton.ca">wfraser@ccs.carleton.ca</a></td>
</tr>
<tr>
<td>FULCHER</td>
<td>Glenn</td>
<td>University of Dundee</td>
<td><a href="mailto:g.fulcher@dundee.ac.uk">g.fulcher@dundee.ac.uk</a></td>
</tr>
<tr>
<td>GERANPAYEH</td>
<td>Aradeshir</td>
<td>Cambridge ESOL</td>
<td><a href="mailto:Turner.r@ucles.org.uk">Turner.r@ucles.org.uk</a></td>
</tr>
<tr>
<td>GLOVER</td>
<td>Philip</td>
<td>British Council Turkey</td>
<td><a href="mailto:philipglover@superonline.com">philipglover@superonline.com</a></td>
</tr>
</tbody>
</table>

36
<table>
<thead>
<tr>
<th>Family Name</th>
<th>Given Name</th>
<th>Institution</th>
<th>email</th>
</tr>
</thead>
<tbody>
<tr>
<td>GORGENYI</td>
<td>Istvan</td>
<td>Budapest Business School</td>
<td><a href="mailto:e.benke@axelero.hu">e.benke@axelero.hu</a></td>
</tr>
<tr>
<td>GREEN</td>
<td>Rita</td>
<td>Freelance</td>
<td><a href="mailto:rikagreen@compuserve.com">rikagreen@compuserve.com</a></td>
</tr>
<tr>
<td>GREEN</td>
<td>Anthony</td>
<td>Cambridge ESOL</td>
<td><a href="mailto:Turner.r@uclcs.org.uk">Turner.r@uclcs.org.uk</a></td>
</tr>
<tr>
<td>GREGO BOLLI</td>
<td>Guilia</td>
<td>Universita per Stranieri di Perugia</td>
<td><a href="mailto:nicciboll@tin.it">nicciboll@tin.it</a></td>
</tr>
<tr>
<td>GRIFFIN</td>
<td>Valerie</td>
<td>TOEIC UK</td>
<td><a href="mailto:toiecpta@eurobell.co.uk">toiecpta@eurobell.co.uk</a></td>
</tr>
<tr>
<td>GROTEJAHN</td>
<td>Ruediger</td>
<td>Ruhr-Universitaet Bochum</td>
<td><a href="mailto:ruediger.grotejahn@ruhr-unibochem.de">ruediger.grotejahn@ruhr-unibochem.de</a></td>
</tr>
<tr>
<td>GROVE</td>
<td>Lis</td>
<td>University of Melbourne</td>
<td><a href="mailto:efg@unimelb.edu.au">efg@unimelb.edu.au</a></td>
</tr>
<tr>
<td>GYSSEN</td>
<td>Sara</td>
<td>Certificate Dutch as a Foreign Language, University of Leuven</td>
<td><a href="mailto:sara.gysen@arts.kuleuven.ac.be">sara.gysen@arts.kuleuven.ac.be</a></td>
</tr>
<tr>
<td>HAINES</td>
<td>Kevin</td>
<td>Hanze University, Groningen, The Netherlands</td>
<td><a href="mailto:k.b.j.haines@pl.hanze.nl">k.b.j.haines@pl.hanze.nl</a></td>
</tr>
<tr>
<td>HAMP-LYONS</td>
<td>Liz</td>
<td>Dept. of English, The Hong Kong Polytechnic University</td>
<td><a href="mailto:egihl@polyu.edu.h">egihl@polyu.edu.h</a>/c</td>
</tr>
<tr>
<td>HAWKEY</td>
<td>Roger</td>
<td>Cambridge ESOL</td>
<td><a href="mailto:Turner.r@uclcs.org.uk">Turner.r@uclcs.org.uk</a></td>
</tr>
<tr>
<td>HILL</td>
<td>Kathryn</td>
<td>University of Melbourne</td>
<td><a href="mailto:kmhill@unimelb.edu.au">kmhill@unimelb.edu.au</a></td>
</tr>
<tr>
<td>HOOSHMAND</td>
<td>Dariush</td>
<td>Defense Language Institute, Foreign Language Center</td>
<td><a href="mailto:hooshmandd@monterey.army.mil">hooshmandd@monterey.army.mil</a></td>
</tr>
<tr>
<td>HORAI</td>
<td>Tomoko</td>
<td>University of Surrey Roehampton</td>
<td><a href="mailto:thorai2003@hotmail.com">thorai2003@hotmail.com</a></td>
</tr>
<tr>
<td>HORAK</td>
<td>Tania</td>
<td>Lancaster University</td>
<td><a href="mailto:t.horak@lancaster.ac.uk">t.horak@lancaster.ac.uk</a></td>
</tr>
<tr>
<td>ISHII</td>
<td>Tomoko</td>
<td>University of Nottingham</td>
<td><a href="mailto:tmk_ishii@mac.com">tmk_ishii@mac.com</a></td>
</tr>
<tr>
<td>ITO</td>
<td>Akihiro</td>
<td>Aichi Gakuin University, Japan</td>
<td><a href="mailto:akito@dpc.aichi-gakuin.ac.jp">akito@dpc.aichi-gakuin.ac.jp</a></td>
</tr>
<tr>
<td>KALNBERZINA</td>
<td>Vita</td>
<td>Stockholm School of Economics</td>
<td><a href="mailto:vitak@bluewin.ch">vitak@bluewin.ch</a></td>
</tr>
<tr>
<td>KAMIMOTO</td>
<td>Tadamitsu</td>
<td>Kumamoto Gakuin University/University of Wales Swansea</td>
<td><a href="mailto:tad_kamimoto@hotmail.com">tad_kamimoto@hotmail.com</a></td>
</tr>
<tr>
<td>KATONA</td>
<td>Lucia</td>
<td>ITK</td>
<td><a href="mailto:ikatona@itk.hu">ikatona@itk.hu</a></td>
</tr>
<tr>
<td>KENNYON</td>
<td>Dorry</td>
<td>Center for Applied Linguistics</td>
<td><a href="mailto:dorry@cal.org">dorry@cal.org</a></td>
</tr>
<tr>
<td>KUNNAN</td>
<td>Antony</td>
<td>California State University, Los Angeles</td>
<td><a href="mailto:akunnan@calstatela.edu">akunnan@calstatela.edu</a></td>
</tr>
<tr>
<td>KIM</td>
<td>Jee Wha</td>
<td>Teachers College, Columbia University</td>
<td><a href="mailto:jeewha@yahoo.com">jeewha@yahoo.com</a></td>
</tr>
<tr>
<td>KOBAYASHI</td>
<td>Mlyoko</td>
<td>Graduate School of Language Sciences, Kanda University of</td>
<td><a href="mailto:m.kobay@kanda.kuis.ac.jp">m.kobay@kanda.kuis.ac.jp</a></td>
</tr>
<tr>
<td>KOLLATH</td>
<td>Katalin</td>
<td>Budapest Business School</td>
<td><a href="mailto:e.benke@axelero.hu">e.benke@axelero.hu</a></td>
</tr>
<tr>
<td>KOYANAGI</td>
<td>Kaoru</td>
<td>Sophia University</td>
<td><a href="mailto:kaoru-k@sophia.ac.jp">kaoru-k@sophia.ac.jp</a></td>
</tr>
<tr>
<td>LIU</td>
<td>Jianda</td>
<td>City University of Hong Kong</td>
<td><a href="mailto:enjiu@cityu.edu.hk">enjiu@cityu.edu.hk</a></td>
</tr>
<tr>
<td>LOCH</td>
<td>Agnes</td>
<td>Budapest Business School</td>
<td><a href="mailto:lochagnes@yahoo.com">lochagnes@yahoo.com</a></td>
</tr>
<tr>
<td>LOPRIORE</td>
<td>Lucilla</td>
<td>University of Cassino, Cassino, Italy</td>
<td><a href="mailto:lloprio@tin.it">lloprio@tin.it</a></td>
</tr>
<tr>
<td>LU</td>
<td>Yang</td>
<td>The University of Reading</td>
<td><a href="mailto:lir00yl@reading.ac.uk">lir00yl@reading.ac.uk</a></td>
</tr>
<tr>
<td>LUMLEY</td>
<td>Tom</td>
<td>University of Melbourne</td>
<td><a href="mailto:tulumley@unimelb.edu.au">tulumley@unimelb.edu.au</a></td>
</tr>
<tr>
<td>LUOMA</td>
<td>Sari</td>
<td>Ballard &amp; Tighe, Publishers</td>
<td><a href="mailto:sluoma@ballard-tighe.com">sluoma@ballard-tighe.com</a></td>
</tr>
<tr>
<td>MOE</td>
<td>Eli</td>
<td>University of Bergen, Norway</td>
<td><a href="mailto:eli.moe@nor.uib.no">eli.moe@nor.uib.no</a></td>
</tr>
<tr>
<td>MOUSA</td>
<td>Nawal</td>
<td>Canadian Foreign Language Institute</td>
<td><a href="mailto:nawal.mousa@dfait-maei.gc.ca">nawal.mousa@dfait-maei.gc.ca</a></td>
</tr>
<tr>
<td>NAKAMURA</td>
<td>Yuji</td>
<td>Tokyo Keizai University</td>
<td><a href="mailto:nkyj@tku.ac.jp">nkyj@tku.ac.jp</a></td>
</tr>
<tr>
<td>NOGUCHI</td>
<td>Hiroyuki</td>
<td>Nagoya University, JAPAN</td>
<td><a href="mailto:HiroNoguchi@aol.com">HiroNoguchi@aol.com</a></td>
</tr>
<tr>
<td>NTELIOU</td>
<td>Eleftheria</td>
<td>Charokopio University</td>
<td><a href="mailto:elfie1903@yahoo.com">elfie1903@yahoo.com</a></td>
</tr>
<tr>
<td>O'SULLIVAN</td>
<td>Barry</td>
<td>CRTEC, The University of Surrey Roehampton</td>
<td><a href="mailto:b.osullivan@roehampton.ac.uk">b.osullivan@roehampton.ac.uk</a></td>
</tr>
<tr>
<td>PAPACONSTANTINOU</td>
<td>Melina</td>
<td>Technological Educational Institute (TEI) of Kavala, Greece</td>
<td><a href="mailto:mpapacon@teikav.edu.gr">mpapacon@teikav.edu.gr</a></td>
</tr>
<tr>
<td>PARIZZI</td>
<td>Francesca</td>
<td>University of Pisa</td>
<td><a href="mailto:francesca.parizzi@libero.it">francesca.parizzi@libero.it</a></td>
</tr>
<tr>
<td>PAVLOU</td>
<td>Pavlos</td>
<td>University of Cyprus</td>
<td><a href="mailto:ppavlou@ucy.ac.cy">ppavlou@ucy.ac.cy</a></td>
</tr>
<tr>
<td>Family Name</td>
<td>Given Name</td>
<td>Institution</td>
<td>email</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------</td>
</tr>
<tr>
<td>PEGG</td>
<td>Christine</td>
<td>Cardiff University</td>
<td><a href="mailto:PeggCA@cardiff.ac.uk">PeggCA@cardiff.ac.uk</a></td>
</tr>
<tr>
<td>PERLMANN-BALME</td>
<td>Michaela</td>
<td>Goethe Institut Inter Nationes</td>
<td><a href="mailto:perlmann-balme@goethe.de">perlmann-balme@goethe.de</a></td>
</tr>
<tr>
<td>PHELAN</td>
<td>David</td>
<td>The University of the West of England, Bristol</td>
<td><a href="mailto:david.pheLAN@uwe.ac.uk">david.pheLAN@uwe.ac.uk</a></td>
</tr>
<tr>
<td>PURPURU</td>
<td>James E.</td>
<td>Teachers College, Columbia University</td>
<td><a href="mailto:jp248@columbia.edu">jp248@columbia.edu</a></td>
</tr>
<tr>
<td>READ</td>
<td>John</td>
<td>Victoria University of Wellington</td>
<td><a href="mailto:John.Read@vuw.ac.nz">John.Read@vuw.ac.nz</a></td>
</tr>
<tr>
<td>REA-DICKINS</td>
<td>Pauline</td>
<td>The University of Bristol</td>
<td><a href="mailto:P.Rea-Dickins@bris.ac.uk">P.Rea-Dickins@bris.ac.uk</a></td>
</tr>
<tr>
<td>RETHINASAMY</td>
<td>Soubakeavathi</td>
<td>CRTEC, The University of Surrey Roehampton</td>
<td><a href="mailto:rsouba@email.com">rsouba@email.com</a></td>
</tr>
<tr>
<td>RICO</td>
<td>Mercedes</td>
<td>Universidad de Extremadura</td>
<td><a href="mailto:mricogar@unex.es">mricogar@unex.es</a></td>
</tr>
<tr>
<td>ROBINSON</td>
<td>Martin</td>
<td>Cambridge ESOL</td>
<td><a href="mailto:Turner.r@udes.org.uk">Turner.r@udes.org.uk</a></td>
</tr>
<tr>
<td>SAEGUSA</td>
<td>Reiko</td>
<td>Hitotsubashi University</td>
<td><a href="mailto:saegusa.reiko@srv.cc.hit-u.ac.jp">saegusa.reiko@srv.cc.hit-u.ac.jp</a></td>
</tr>
<tr>
<td>SAVILLE</td>
<td>Nick</td>
<td>Cambridge ESOL</td>
<td><a href="mailto:Turner.r@udes.org.uk">Turner.r@udes.org.uk</a></td>
</tr>
<tr>
<td>SCHMITT</td>
<td>Diane</td>
<td>Nottingham Trent University</td>
<td><a href="mailto:diane.schmitt@ntu.ac.uk">diane.schmitt@ntu.ac.uk</a></td>
</tr>
<tr>
<td>SCOTT</td>
<td>Catriona (Katie)</td>
<td>University of Bristol</td>
<td><a href="mailto:Katie.Scott@bris.ac.uk">Katie.Scott@bris.ac.uk</a></td>
</tr>
<tr>
<td>SEDGWICK</td>
<td>Carole</td>
<td>University of Surrey Roehampton</td>
<td><a href="mailto:carolesedwick@hotmail.com">carolesedwick@hotmail.com</a></td>
</tr>
<tr>
<td>SHABBIN</td>
<td>A. Affendi</td>
<td>University of Nottingham</td>
<td><a href="mailto:AEXAAS@HERMES.Nottingham.ac.uk">AEXAAS@HERMES.Nottingham.ac.uk</a></td>
</tr>
<tr>
<td>SHAW</td>
<td>Stratt</td>
<td>Cambridge ESOL</td>
<td><a href="mailto:Turner.r@udes.org.uk">Turner.r@udes.org.uk</a></td>
</tr>
<tr>
<td>SHIMADAY</td>
<td>Megumi</td>
<td>Tokyo Gakugei University</td>
<td><a href="mailto:shimada@u-gakugei.ac.jp">shimada@u-gakugei.ac.jp</a></td>
</tr>
<tr>
<td>SHIOTSU</td>
<td>Toshikiko</td>
<td>Kurume University, Japan</td>
<td><a href="mailto:ts@jp.org">ts@jp.org</a></td>
</tr>
<tr>
<td>SHOHAMY</td>
<td>Elana</td>
<td>Tel Aviv University</td>
<td><a href="mailto:elana@post.tau.ac.il">elana@post.tau.ac.il</a></td>
</tr>
<tr>
<td>SIGOTT</td>
<td>Guenther</td>
<td>English Dept., University of Klagenfurt, Austria</td>
<td><a href="mailto:guenther.sigott@uniklu.ac.at">guenther.sigott@uniklu.ac.at</a></td>
</tr>
<tr>
<td>SPAAN</td>
<td>Mary</td>
<td>University of Michigan, English Language Institute</td>
<td><a href="mailto:mcspaan@umich.edu">mcspaan@umich.edu</a></td>
</tr>
<tr>
<td>TAVAKOLI</td>
<td>Parvaneh</td>
<td>King's College London</td>
<td><a href="mailto:parvaneh.tavakoli@kcl.ac.uk">parvaneh.tavakoli@kcl.ac.uk</a></td>
</tr>
<tr>
<td>TAYLOR</td>
<td>Linda</td>
<td>Cambridge ESOL</td>
<td><a href="mailto:Turner.r@udes.org.uk">Turner.r@udes.org.uk</a></td>
</tr>
<tr>
<td>THOMPSON</td>
<td>Paul</td>
<td>University of Reading</td>
<td><a href="mailto:P.A.Thompson@rdg.ac.uk">P.A.Thompson@rdg.ac.uk</a></td>
</tr>
<tr>
<td>TSAGARI</td>
<td>Constance</td>
<td>Lancaster University</td>
<td><a href="mailto:ctsagari@otenet.gr">ctsagari@otenet.gr</a></td>
</tr>
<tr>
<td>VAN MOERE</td>
<td>Alistair</td>
<td>Kanda University of International Studies, Japan</td>
<td><a href="mailto:alimoere@hotmail.com">alimoere@hotmail.com</a></td>
</tr>
<tr>
<td>WAGNER</td>
<td>Elvis</td>
<td>Teachers College, Columbia University</td>
<td><a href="mailto:mjw32@columbia.edu">mjw32@columbia.edu</a></td>
</tr>
<tr>
<td>WALL</td>
<td>Dianne</td>
<td>Lancaster University</td>
<td><a href="mailto:d.wall@lancaster.ac.uk">d.wall@lancaster.ac.uk</a></td>
</tr>
<tr>
<td>WEIR</td>
<td>Cyril</td>
<td>CRTEC, The University of Surrey Roehampton</td>
<td><a href="mailto:c.weir@roehampton.ac.uk">c.weir@roehampton.ac.uk</a></td>
</tr>
<tr>
<td>WIGGLESWORTH</td>
<td>Gillian</td>
<td>University of Melbourne</td>
<td><a href="mailto:gillianw@unimelb.edu.au">gillianw@unimelb.edu.au</a></td>
</tr>
<tr>
<td>WRIGHT</td>
<td>Anne</td>
<td>U.S. Department of Defense/ Defense Language Institute</td>
<td><a href="mailto:anne.wright@monterey.army.mil">anne.wright@monterey.army.mil</a></td>
</tr>
<tr>
<td>YAGHOUBI GOODARZI</td>
<td>Shahindokht</td>
<td>Iran University of Science and Technology</td>
<td><a href="mailto:goodarzish@yahoo.com">goodarzish@yahoo.com</a></td>
</tr>
<tr>
<td>YAGHOUBI GOODARZI</td>
<td>Parviz</td>
<td>University of Michigan</td>
<td><a href="mailto:goodarzish@yahoo.com">goodarzish@yahoo.com</a></td>
</tr>
<tr>
<td>YAMASHIRO</td>
<td>Amy D.</td>
<td>University of Michigan</td>
<td><a href="mailto:anydyama@umich.edu">anydyama@umich.edu</a></td>
</tr>
<tr>
<td>YI</td>
<td>Jiy-Yeon</td>
<td>The University of Edinburgh</td>
<td><a href="mailto:jiyeyeon@ling.ed.ac.uk">jiyeyeon@ling.ed.ac.uk</a></td>
</tr>
<tr>
<td>YU</td>
<td>Guoxing</td>
<td>University of Bristol</td>
<td><a href="mailto:Guoxing.Yu@bris.ac.uk">Guoxing.Yu@bris.ac.uk</a></td>
</tr>
<tr>
<td>ZAINAL ABIDIN</td>
<td>Saitdul</td>
<td>CRTEC, The University of Surrey Roehampton</td>
<td><a href="mailto:akmarza@msm.com">akmarza@msm.com</a></td>
</tr>
</tbody>
</table>

**Leung** | Constant  |
**Howell** | Bruce     |
**Bernstein** | Tarei    |
**Milanovic** | Michael  |
Contact Information

Dr Barry O’Sullivan

Centre for Research in Testing, Evaluation and Curriculum in ELT
University of Surrey Roehampton
Erasmus House
Roehampton Lane
London
SW15 5PU

Tel: +44 (0)20 8392 3797
Fax: +44-(0)20-8392-3031

E-mail: b.osullivan@roehampton.ac.uk

Dr Pauline Rea-Dickins

Graduate School of Education
35 Berkeley Square
Clifton
BRISTOL BS8 1JA

Tel: + 44 (0)117 928 7031
Fax: + 44 (0)117 929 1941

E-mail: P.Rea-Dickins@bris.ac.uk

Ms Jayanti Banerjee

Department of Linguistics and Modern English Language
Lancaster University
Lancaster LA1 4YT
United Kingdom

Tel: + 44 (0)1524 893039
Fax: + 44 (0)1524 843085

E-mail: J.Banerjee@lancaster.ac.uk