Cross-cultural and Inter-cultural Communication in Second Language Assessment Conversations. Peter Storey, Hong Kong Institute of Education

Opening the Hong Kong language classroom to global influences was an implicit goal of the Hong Kong SAR Government’s Native English Speaking Teacher Scheme. This report explores the effects of an initiative in the context of oral assessment. Under the Scheme, 400 native speakers of English were recruited to join an English teaching force of over 14,000 local teachers comprising largely Hong Kong Chinese speakers for whom English is a second language. A comparative study of interlocutor effects was conducted drawing on data gathered in the context of the conversation component of a set of tests used in evaluating the effectiveness of the Scheme. Investigation of interlocutor effects in interviews conducted by local Chinese and expatriate teachers revealed qualitative differences in communicative style. These stylistic differences resulted in different degrees of student participation in the oral interview. Statistically significant differences in the performance of the students under the two conditions are therefore hypothesised to have resulted from the different interlocutor styles. Since the oral interview in question has a communicative orientation and is based around a conversation between students and interlocutor, Conversation Analysis provided an appropriate means of generating a rich description of the two interlocutor styles and facilitated identification and comparison of salient features. Validity issues relating to oral assessment techniques in the newly global context of the Hong Kong school classroom are explored.
Revising Assessment Criteria and Scales: Insights from the Rater Community. Lynda B. Taylor and Stuart Shaw, University of Cambridge Local Examinations Syndicate

The choice of appropriate rating criteria and the consistent application of rating scales by trained raters are regarded as key factors in the valid and reliable assessment of second language performance (Alderson, Clapham and Wall, 1995; McNamara, 1996; Bachman and Palmer, 1996). Criteria and rating scales, as well as test content and format, need to be kept under review during the operational life of any test and are likely to require revision from time to time. The process of revising the criteria and scales for the IELTS (International English Language Testing System) Speaking Test was reported at LTRC 2001 (see also UCLES EFL Research Notes, Issues 4, 5 and 6); a similar approach is being adopted to revise the criteria and scales for the IELTS Writing Modules. The revision project aims to enhance the validity and reliability of the writing assessment process for IELTS by redeveloping the criteria and scales, and by developing a more comprehensive set of materials and procedures for (re)training IELTS raters.

Initial discussion by the revision team was informed by a review of recent commissioned and non-commissioned studies relating to IELTS Writing, and also by a survey of the literature on holistic and analytic approaches to writing assessment. This was followed by a consultation phase to explore current practice among IELTS Writing assessors, to gauge their attitudes towards their respective assessment practice, and to highlight factors which might help shape the redevelopment of the writing assessment criteria and scales.

The consultation phase began with a series of semi-structured interviews with groups of IELTS Academic and General Training Writing assessors in the UK and Australia. This interaction led to the construction of a survey questionnaire which was sent out to a sample of several hundred IELTS assessors based at a range of test centres worldwide. The function of the interviews and questionnaires was to elicit from assessors individual approaches and attitudes to the assessment of IELTS writing tests, especially in relation to differing domains (Academic and General Training) and differing task genres (Task 1 and Task 2). This presentation reports on the findings from the semi-structured interviews and survey questionnaires and shows how the rich insights offered by the rating community can be instrumental in helping to develop assessment criteria and scales which are valid, reliable and practical.
Influence of Item Position on CAT Test Results. Lin Wang, Educational Testing Service

This paper presents the findings from a recent study that investigates the influence of item positions on the test results for a CAT-delivered English testing program. CAT-delivered tests are gaining increasing interest in language assessments and have been adopted in large-scale testing programs such as TOEFL. Well-designed CAT tests have several key advantages over conventional paper-pencil tests and offer a promising future in language assessment development in the global context. On the other hand, because CAT applications are very new in language assessment, many technical issues remain unsolved, unexplored, or even aware of at present. One important issue is the possible influence of item administration position on the test results.

In the CAT test for this study, each examinee may see a different test form in which items are not administered according to any pre-determined order. What items are administered to an examinee depends on three factors: content requirements, exposure control, and item information. Assuming adequate control over content and exposure, item information is then the only factor that plays a crucial role in item selection and the result of a CAT test. This paper discusses (1) how item positions in a CAT test are related to the test results, (2) where good (highly discriminating) items are typically administered, and (3) how effectively the test items are used.

The data from about 10,000 examinees were analyzed. The analyses included distribution of item characteristics with item positions in the tests and relative efficiency as a ratio of item information to item infomax values. Key findings of the study are: (1) items at different positions contribute different amount of information to the final determination of test results, (2) highly discriminating items tend to be given in the earlier part of tests where not much is known about examinees, and (3) highly discriminating items tend to have lower relative efficiency. This study has at least these four significant implications to the research in CAT-delivered language assessments. First, the study uses real data from a well-established language assessment program to demonstrate how item position may affect test results in terms of item and test information. Second, the study illustrates a variety of analytical methods that can benefit researchers who engage in CAT studies. Third, the findings of the study pose a challenge to the CAT researchers: how to resolve and control the problem that item position may affect test results. Finally, the study shows a practical approach for an objective evaluation of implementation and outcomes of CAT-delivered language assessments.
Factors Affecting Non-native Speakers’ Graph Comprehension and Spontaneous Interpretation. Xiaoming Xi, University of California at Los Angeles, and Irvin R. Katz, Educational Testing Service

This study investigated the fundamental issue of the relationship between visual information processing and language processing and production and the more practical issue of the validity of using visuals in language assessments in the context of the graph interpretation task in the TSE (Test of Spoken English), a widely used measure to screen international teaching assistants in American universities.

The study looked at how the number of visual chunks to be described in bar graphs and the prompt type affected the communicative quality of elicited responses, comparing examinees’ scores across bar graphs tasks with varying numbers of visual chunks and different prompt types. The few-chunks and many-chunks graphs were created by reversing the variables on the x axis and the prompts included the generic prompt (“Describe the information in the graph”) and the direct and explicit prompt (“Tell us the changes from year x to year y”).

It was found that when the bars were grouped by time and the major points were packaged in relatively fewer visual chunks, more sophisticated language functions such as global comparisons and trend descriptions were produced, which in turn were perceived as more developed responses and superior performances by the raters. However, the manipulation of the prompt type did not have significant influence on the performance on the graph tasks.

At the theoretical level, this research extended our understanding of working memory, attention, spatial cognition and language processing and production. The qualitative analysis of the responses to the graph tasks shed light on the relationship between models of visual information processing and spoken language production, demonstrating that graph display features such as the number of visual chunks, the number of variables and the saliency of trend patterns have influence on the information processed and encoded by the graph viewer and types of spontaneous interpretations elicited. At a more practical level, the results had important implications for tightening up the specifications, providing guidelines for designing graph tasks and improving rater training.
Sample Free Item Parameter and Test Free Score within CTT. Zhongen Xi, Shaoting University

Classical Testing Theory (CTT) has been with us for about a hundred years. But in the past few decades, it has been severely criticized for many of its shortcomings, amongst which are its sample dependent item parameters and test dependent scores. But through our research we found that these two so-called liabilities of classical testing theory are not what cannot be overcome from within. I want to take this opportunity to show our colleagues how item parameter can be calibrated sample-independently, how to derive test free scores and how item parameter can be linked. All these are accomplished within the theoretical framework of classical testing theory and under the guidance of what we call the Law of Invariant True Score. The greatest advantage of our scoring procedure over modern item response theory is its simplicity: simple arithmetic operation is all that is needed. The score thus derived is on the ratio scale rather than on the ordinal scale as the score derived by the conventional classical testing theory is or on the interval scale as the score derived by the modern item response theory is.
This study investigates the nature of language proficiency in interactive task performance on the basis of sociocultural theory by examining whether successful task performance results from effective use of L1 discourse markers uttered as private speech (PS). PS is differentiated from social speech by its lower voice pitch and by flat intonation, or formal properties, such as the use of informal style or the choice of sentence-final particles, in the case of Japanese. Specifically, this study examines the size (VS) and lexical density (LD) of PS obtained from the concordance analyses of five word-forms that signal interlocutors' awareness of their task performance either as automatic or non-problematic. LD indicates a degree of shared knowledge (Stubbs, 1996). Data were collected from sixty pairs of Japanese college students who had diverse education backgrounds. To examine communicative ability, a pair of maps were used to fill knowledge gaps between the dyads, or pairs of interlocutors, who were unable to see the map. One interlocutor had to draw a route guided by the partner whose map had marked out. According to the accuracy of the route drawn, the dyads were divided into groups.

The results showed that the effective use of L1 PS led to successful task performance. Though the standardized number of PS word-forms uttered in each group was almost the same, VS increased progressively from the least successful group to the most successful regardless of their English proficiency level. The LD indicated that only the most successful group co-constructed shared knowledge in problematic situations; surprisingly, the least successful group co-constructed shared knowledge in non-problematic situations much in problematic situations; the least successful group performed in the same manner of the situation. The study concludes that 'language proficiency in interaction' is in nature and an integral part of the task performance; it must be distinguished from proficiency measured independent of the task. The findings have implications for instruction and task-based pedagogy and research.
25th Annual Language Testing Research Colloquium

July 22nd to 25th 2003

The University of Reading

Learner Interfaces with Language Testing & Assessment

Proposal Submission
Deadline: January 31st 2003

Contact: Dr Barry O'Sullivan
School of Linguistics & Applied Language Studies
The University of Reading
Whiteknights
Reading RG6 6WB
Berkshire United Kingdom

www.rdg.ac.uk/AcaDepts/Ili/teru/ltc2003/
teru@reading.ac.uk

ASSESSING WRITING

New Editor from 2002!
L. Hamp-Lyons,
The Hong Kong Polytechnic University, Hong Kong

Professor Liz Hamp-Lyons, ably assisted by a distinguished and newly internationalised editorial board, has broadened the scope of Assessing Writing to reflect the concerns of teachers, researchers and writing assessment specialists from around the world.

Assessing Writing is a refereed international journal providing a forum for ideas, research and practice on the assessment of written language. Assessing Writing publishes articles, book reviews, conference reports, and academic exchanges concerning writing assessments of all kinds, including traditional (‘direct’ and standardised forms of) testing of writing, alternative performance assessments (such as portfolios), workplace sampling and classroom assessment.

The scope of the journal is wide, and embraces all work in the field at all age levels, in large-scale (international, national and state) as well as classroom, educational and non-educational institutional contexts, writing and programme evaluation, writing and critical literacy, and the role of technology in the assessment of writing. Through this scholarly exchange, Assessing Writing contributes to the development of excellence in the assessment of writing in all contexts, and, in so doing, to the teaching and appreciation of writing.

2003 • Volume 9 • 3 issues • ISSN 1075-2915

www.socscinet.com/linguistics/asw

SocSciNet.com

SocSciNet is a new online gateway from Elsevier Science. This single website opens up a world of research material from famous imprints such as Pergamon, JAI, Academic Press and North Holland.

Helping you stay ahead
From one homepage you can navigate through a host of interdisciplinary fields within the social sciences. Visit the Language and Linguistics Section of SocSciNet.com - www.socscinet.com/linguistics - and ensure you take full advantage of all the benefits and special offers available.

SocSciNet offers:
- Access to thousands of FREE abstracts
- Online access to full text articles via ScienceDirect *
- Details on new and forthcoming book publications
- News and announcements
- Exclusive special offers
- Instructions for authors
- Links to key websites
- Search facilities across a journal or across the web
- Alerts in your subject field

* Available to ScienceDirect subscribers
POSTERS

Abstracts are listed alphabetically by name of presenter


In this paper the speakers will present the work of the Association of Language Testers in Europe in developing their Code of Practice and will outline an approach now being implemented to manage the quality of the examinations provided by the twenty seven member organizations. A code of practice of this kind must be based on sound principles of good practice in assessment which allow high standards of quality and fairness to be achieved. Between 1991 and 1993, ALTE developed its first Code of Practice which set out the standards that the members of the association would aim to meet in producing their language exams (ALTE, 1994). As such it served as a broad statement of what the users of the examinations should expect and the roles and responsibilities of stakeholders in striving for fairness. However, while the ALTE Code of Practice outlines the principles in general terms, it provides no practical guidance to the practitioner on the implementation of these principles or how standards can be set and guaranteed. A supplementary document entitled Principles of Good Practice for ALTE Examinations was drafted (Milanovic and Saville) and discussed at ALTE meetings in 1992/1993; this document was intended to set out in more detail the principles which ALTE members should adopt in order to achieve their goals of high professional standards. It was influenced by a number of sources within the ALTE membership and from the field of assessment at large, (e.g. Messick, Bachman, the AERA/APA/NCME Standards, 1985 etc). While it was not published in its entirety, parts of the document were later incorporated into the Users Guide for Examiners produced by ALTE on behalf of the Council of Europe (1997).

In 1999/2000 ALTE re-established a working group to take this project forward. So far two outcomes have been achieved: a) a revised version of Principles of Good Practice for ALTE Examinations (2001) has been produced; b) the Code of Practice itself has been redesigned and expanded as a checklist which is being used by Members as part of a Quality Management System (QMS).

The aim of the QMS is to guarantee minimum quality standards for ALTE exams used on the ALTE Code of Practice. In the current phase of the project, the working group is focusing on the different organisational factors within the ALTE Membership and the speakers will draw attention to the diversity that exists. It is not the aim to make all ALTE members conform to the same models of assessment for all 24 languages represented, and it important to recognise the varied linguistic, educational and cultural contexts within which the examinations are being developed and used. It is argued that an appropriate balance is required between the need to guarantee the professional standards to users and the need to take into account the differing organisational features of the ALTE institutions and the contexts in which their exams are used.
Three Instruments for Measuring Reading Components and Oral Reading Fluency.
Isabella Barbier, Jared Bernstein, Michael Morrow, and Ognjen Todic, Ordinate Corporation

Some reading theorists (e.g. Perfetti 1985, Gough & Juel 1992) have focused on the componential nature of literacy and reading comprehension. These proposed models of cognitive components in reading development have inspired a wide range of fruitful research in early reading. In order to understand the nature of reading difficulties in adult and child populations of low-performing readers, we implemented a series of instruments that measure aspects of memory, word and non-word decoding, language skill, and oral reading fluency. One such instrument (telephone mediated) has collected many samples of oral reading and other speech-borne tasks from 700 adult “Level 1” readers. Another of these instruments (running on a notebook computer) has collected two thousand samples of 4th-grade students’ oral reading, which are being analyzed in parallel by human experts and by automatic fluency analysis. A third computer-based instrument has collected performance data on components of oral reading fluency from a sample of 500 adults age 16 and older for whom a wide range of data on demographic background is available. The instruments are different: One instrument merely supports a human administrator; the second implements a mixed initiative administration; and the third presents a fully automated procedure that a human administrator merely monitors. The poster will describe and compare the three instruments. If the data and analyses are publicly available at the time of LTRC, we may also briefly compare the evidence for a componential analysis of low-level reading performance between the adults and children in these samples.
POSTERS

Testing the Academic Language Proficiency of English Language Learners in U.S. Schools. Micheline Chalhoub-Devile and Craig Deville, University of Iowa

In the last decade accountability has been driving educational reform in elementary and secondary schools in the United States. School reform powered by accountability emphasizes the importance of testing—testing to ascertain the effectiveness of schools, testing to compare districts and states, testing to document the rigor of school curricula and teacher effectiveness, testing to decide whether students are eligible to graduate or advance to the next level, etc. The accountability push demands the testing of all students to ensure that no students are excluded from efforts to raise educational standards.

Teachers and administrations are increasingly being asked to assess students who are non-native speakers of English as well (LaCelle-Peterson and Rivera, 1994; August and Hakuta, 1997). As Abedi, Lord, Hofstetter, and Baker (2000) write, the enactment of the Improving America’s Schools Act of 1994 dictates that “all children should be given educational experiences to assist them in achieving high standards. It follows that children previously excluded from large-scale assessments because of...limited proficiency in English must have the opportunity to participate” (p. 16). Administering standardized assessments to English language learner (ELL) students is a complex matter because the administration of these assessments is conducted in a language the ELL students are still learning. ELL students’ performances on standardized tests are sometimes confounded with their English language ability, which can negatively impact the interpretation of their knowledge and skills in the various subject areas of the tests.

Considerable research has been done on the effectiveness of various types of accommodations to help circumvent the linguistic barrier that ELL students face. Accommodations have included the translation of tests, extended time, small-group administration, use of dictionaries, read aloud of directions, giving oral responses, among others. While it is important to continue the research to better understand the types of accommodations needed that would result in meaningful scores from ELL students taking academic-based standardized tests, it is also important to explore other language-based approaches. As Hurley and Tinajero (2001) argue, “[t]he question is whether [we] are attempting to assess content knowledge or language fluency” (p. 106). This question drives the purpose of the present project.

The fundamental research question is how to determine whether ELL students have the requisite language skills needed to take standardized achievement tests that result in accurate test scores reflecting their skills and abilities in the content areas rather than their language proficiency. The present proposal focuses on the initial phase of a project, the design and development of the ELL test. To design such a battery of ELL tests that spans K-12 is a challenge because ELL students vary considerably in their cognitive development, first language literacy, length of time in the U.S., migrant status, among other things. In addition, such tests must take into account the increasingly complex and abstract language students learn and need as they grow older. These factors necessitate the construction of several ELL tests, each appropriate for a particular grade span, such as grades 3-5, and 6-8. We will report on the ongoing test development process, including initial test specifications.
Tertiary English Language Placement Assessment for the United Arab Emirates.
Christine Coombe, Higher Colleges of Technology, Peter Davidson, Zayed University, Glenn Fulcher, University of Surrey, Elizabeth Howell, Higher Colleges of Technology, and Paul Jaquith, United Arab Emirates University

This poster documents a project to develop and implement a national tertiary English language placement assessment in the United Arab Emirates (UAE).

The poster situates the initiative in the context of the UAE and the three national tertiary institutions, namely the United Arab Emirates University, the Higher Colleges of Technology, and Zayed University. The rationale for devising and implementing such an assessment is also described.

The poster describes the phases of the initiative from pre-planning, formation of an exam writing task group, item writing, piloting, validation to administration. Culturally sensitive theoretical issues including the tensions between secondary and tertiary educational systems, between the function of the assessment for placement or admission, and the role of the National Assessment and Placement Organization (NAPO) are highlighted. Logistical issues pertaining to administration and marking are addressed, including the need for training of administrators, the need for transparency versus security, time and resources for item writers, statisticians, and item banking, and the need for co-operation between the three national tertiary institutions.

The poster presents an overview of the results of the first administration and summarises the implications of the assessment, such as the impact that implementing such a placement exam will have upon learning in the high school system.
Testing Oral Language: Which English? Rebecca L. Damron, Oklahoma State University

While the TSE is administered widely, large International Teaching Assistant programs do not rely solely on the TSE for making final decisions (Halleck and Moder, 1995). In fact, passing TSE scores for South Asian speakers of English do not indicate readiness for classroom teaching. ITA performance tests require testees to produce longer segments of comprehensible spoken discourse than the TSE, which reveals prosodic weaknesses in the development of discourse topics.

The study reported in this poster presentation involved the prosodic analysis of the discourse produced by 20 subjects in the ITA performance test administered at Oklahoma State University. While all subjects took and passed the TSE, one half of the did not pass the ITA test. Each subject's ITA test discourse samples were transcribed and segmented into intonation units (as per Chafe, 1994; DuBois et.al, 1992).

Results indicate one of the important differences was in the prosodic structure of intonation units. Subjects who failed the ITA test produced intonation units that were longer and had fewer accented syllables than American English intonation units. The poster will include samples of transcribed discourse from the ITA performance test administrations at Oklahoma State University.

The implications of this study for the inclusion of prosodic criteria in test evaluation and in training for further testing of individual testees will be addressed in the poster.
Development of a Test of Professional English Communication – Engineering. Dan Douglas, Iowa State University, Thomas Orr, University of Aizu, Randy Thrasher, International Christian University, and Katsumi Yuasa, International Language Centre

In order to meet a need for the evaluation of the English ability of professionals in fields such as engineering, finance, medicine, law, and mass media, an international consortium is developing a new series of proficiency tests: Tests of Professional English Communication (TOPEC). The first test in the series is the Fundamental Engineering English Test, designed by a team of specialists in engineering English and specific purposes assessment in Asia and North America for initial implementation in Japan. Intended for practicing engineers as well as students of engineering, the test is based on a currently well-known model of specific purpose language ability which includes language knowledge, strategic competence, and specific purpose background knowledge. The test assesses knowledge of the fundamental English that engineers normally use in their professional activities and attempts to include assessment of strategic competence and background knowledge as well. After several trial tests in 2001, the official examination is scheduled for introduction in April, 2002.

This presentation will include a discussion of the theoretical background of the new test, a report on the trial testing and analysis, a description of the current version of the test, and a brief description of integrated specific purpose English training materials to accompany the test.
Interactive v. Monologic Discourse. Gene B. Halleck, Oklahoma State University

The administrative burden of doing "live" oral proficiency interviews has necessitated the development of so-called semi-direct tests in which examinee responses are elicited by pre-recorded questions on audiotape or videotape (Stansfield & Kenyon, 1992; Shohamy, 1994; Cascallar, 1997; O’Loughlin, 2001). This poster reports results of a study comparing responses elicited by two instruments designed to assess oral proficiency: a face-to-face Oral Proficiency Interview (OPI) and a tape-mediated Video Oral Communication Instrument (VOCI) for ESL/EFL.

Native language backgrounds of the examinees (N=30) who took both the OPI and the VOCI include Amharic, Arabic, Armenian, French, Gujerati, Hindi, Kannada, Mandarin, Marathi, Russian, Spanish, Swazi, Tamil, and Thai.

In addition to the holistic scores for each instrument, this poster illustrates two other aspects of test-taker language output: syntactic maturity and lexical density. These aspects of learner language output are analyzed in responses to questions requiring three different tasks: talking about the here and now, narration, and supporting an opinion.

Tables shed some light on the relationship between test format and task type. With respect to text type, responses to questions without an interlocutor (on the VOCI) were significantly longer than responses to comparable tasks in the OPI. With respect to lexical density, however, results are not so clear. For some tasks lexical density was higher in monologic responses to tape-mediated questions than in the more interactive responses to similar questions in the OPI. But for other tasks, there was no significant difference between the lexical density of responses in the two test formats.
The present study investigates reliability and validity of new type of translation test with multiple-choice format as a reading comprehension measure. In the present study translation tests means the test measuring the participants reading comprehension ability through examining how well they perform in translating English prose into Japanese language. As a reaction to the new reform in English language tests, the institution I am currently working for has tried to change some parts of English language tests for university entrance examinations. Though translation test (from English to Japanese) may be reliable and valid test for measuring reading comprehension ability (Buck, 1992), it could be a tough work for English teachers to spend some time for scoring. In addition, some English language teachers feel that open-ended translation test often can be difficult task for examinees, we have changed the format of translation test into a multiple-choice one. However, this change is completely done on the speculation, so this study has been conducted. One narrative prose was selected. (1) Cloze test (30 items), (2) short-answer question test (6 items), (3) translation test (6 items), and (4) multiple-choice translation test (6 items) were developed. The participants were Japanese high school students (N=72). In order to examine the effect of test method proficiency level, two-way ANOVA was performed on the scores of the tests with four different test methods. Moreover, various correlation studies were done in order to compare the examinees’ test performance in translation tests with two different methods. The present also discuss the possibility of use of MC translation test for high-stakes testing such as one of the components of university entrance examination tests in Japan.

Akihiro Ito (PhD in applied linguistics, Hiroshima University, Japan) is currently an Associate Professor of applied linguistics in Aichi Gakuin University, Japan. His research interests include language testing and second language acquisition. His recent publication is Japanese EFL learners’ acquisition on English relative clauses: A cross-sectional approach (Tokyo: Libor Press, 2001).
Setting Multiple Standards on Performance Assessment in Foreign Language. Yoko Kozaki, Mukogawa Women’s University

The purpose of this study was to demonstrate the viability of a new standard-setting method for performance assessment, which may be particularly suitable for occupational language testing situations where both language and subject expert judges participate in the decision-making process. The method was constructed such that an established standard can reflect the real world requirements, which are informed by expert judges with differing qualifications and professional experiences. It was characterized by (a) judges’ independent involvement in the activity, (b) the use of multiple standard-setting exercises, (c) the introduction of partial inter-judge agreement in setting a standard, and (d) the handling of the error in decision by means of Multifaceted Rasch Analysis. The procedure geared to revealing and accommodating judge variability permitted the detection of sources of error, with the importance attached more to judges’ self-consistency than to inter-judge consensus. Sound application of this method was made possible with the aid of the computer program FACETS (Linacre and Write, 1993) for Multifaceted Rasch Analysis, which can deal with the possible misclassification of examinees, especially attributed to judges’ lack of self-consistency. FACETS also served to quantify a standard as a measure that transcends particularities of judging sessions, that is, examinee ability and judge severity for particular sessions. The data presented are from two sessions in 1998 and 2002, both held upon completion of a professional training program in medical translation from Japanese into English. The two data sets combined are for a total of 15 translations, of which 3 are anchor performances, and eight judges consisting of Japanese and English native translators, an English native rewriter, Japanese and English native medical doctors, and a Japanese native physiologist.
Assessing Transfer of Cognate and Morphological Awareness in Bilingual Children.
Valerie Malabonga, Mohammed Louguit, Silvia Caglarcan, Dorry Kenyon, Diane August, Center for Applied Linguistics, and Maria Carlo, University of Miami

Researchers will share the development and results of two language tests constructed for a national biliteracy study funded by the U.S. government. The two tests represent cognate and morphological awareness, two components believed to be important for native Spanish-speaking children becoming successful English readers. The tests were developed specifically to look into Spanish-English bilingual children’s transfer of their cognate and morphological knowledge from Spanish to English.

In the Cognate Awareness Test, half of the English words were Spanish cognates while half were noncognates. The cognates and noncognates were matched in terms of frequency. Researchers hypothesized that the native Spanish-speaking children would perform better on cognates than on noncognates. The test was piloted on 110 native Spanish-speakers in 4th and 5th grade. Children were given a 61-item, paper-and-pencil test where they were asked to mark one of four words or phrases that was most closely related to the target word.

Rasch analysis of the pilot data showed that there were 13 misfitting items, and that the test was too long, difficult and stressful for the children. Thus, researchers removed all misfitting items, added a few easy items and made the format more “user-friendly.” Rasch analysis yielded a reliability coefficient of .75 for the test with the 13 misfitting items already removed. Native Spanish-speakers found noncognates more difficult than cognates \( (t = 10.75 (23), p < .001) \). Fifth graders also scored better than fourth graders \( (t = -2.50 (79), p < .05) \).

The revised version of the test was then used with 169 4th grade children. The 37 native English-speakers found cognates more difficult than noncognates whereas the 132 native Spanish-speakers found noncognates more difficult than cognates \( (t = -3.35 (167), p < .01) \). Noncognates: \( t = 3.43 (167), p < .01 \). Likewise, of the 124 Latino children, the 55 Spanish-instructed children had higher scores on cognates than the 69 English-instructed children \( (t = -3.06 (122), p < .01) \), but there were no significant differences between the performance of the Latino children from the two languages of instruction when it came to noncognates. The revised test was also an improvement over the pilot version because the Rasch map indicated that there were enough words for the various abilities of the children. However, the test was still a little difficult because the mean difficulty of words was slightly higher than the children’s mean ability.

In the Extract the Base Test, children were asked to change the form of a word so that it fit in the sentence; e.g. farmer, My uncle works on a _____ (farm). Four categories of word transformations were presented: orthographic change (emptiness>empty), phonological change (cleainess>clean), both changes (recognition>recognize) and no changes (dangerous>danger). Researchers hypothesized that negative transfer will occur with the native Spanish-speakers making more errors.

The assessment was piloted on 162 native Spanish- and native English-speakers in 3rd, 4th and 5th grade. Children were given a 32-item, paper-and-pencil test where they wrote the new form of the word. They were given partial credit if they correctly represented but misspelled the word (e.g., publik for public). Rasch analysis yielded a reliability coefficient of .98 for the test. Rasch analysis also showed that the children had the most difficulty with words that changed both orthographically and phonologically, followed by words that changed only orthographically. Fifth graders scored significantly better than third graders \( (F = 4.38 (2,123), p < .05; \text{Mean Difference} = 9.15, p < .05) \). A frequency analysis of children’s errors showed that the most common error was the logical “chopping off” of word endings. Similar percentages of Spanish- and English-speakers were making these logical “chopping” errors.

In summary, the Cognates Awareness Test seemed sensitive enough to assess positive transfer of children’s knowledge of cognates from Spanish to English. However, the Extract the Base test did not appear able to assess transfer from Spanish to English as native Spanish- and native English—speakers were making similar types and percentages of errors.
Balancing Continuity and Innovation in EFL Examinations - Revising CPE. Nick Saville, University of Cambridge Local Examination Syndicate, and Cyril Weir, University of Surrey

Foucault (1977) has emphasised the need to write the history of examinations as they are an important force in society. This talk takes an historical perspective and addresses the development of the Certificate of Proficiency in English (CPE) - 1913 to 2013.

The University of Cambridge Local Examinations Syndicate (UCLES) provides examinations for a wide range of purposes and educational contexts in over 100 countries world-wide. It began its role in 1858 as one of the public examining bodies taking responsibility for school-leaving examinations in Britain, and in 1913 it began testing English when the Certificate of Proficiency in English (CPE) was instituted (Roach 1971, Spolsky, 1995).

Using historical documents and archive data, the speakers describe how the CPE has evolved through successive revisions in 1930, 1966, 1975, 1984 and most recently in the 1990s. The current revision completes a cycle of developments carried out by UCLES EFL over a decade and the revised CPE now forms part of a system of criterion-related examinations linked to the Common European Framework of Reference which has been opted by the Association of Language Testers in Europe - ALTE. The revised CPE (introduced in December 2002) demonstrates how much language teaching and testing has changed during the last century. Nevertheless it is important to recognise that the innovations which have been introduced are grounded in educational traditions and approaches to language teaching and assessment which can be traced back to the early days of the 20th century.

As well as reviewing these developments, the talk will address the extent to which the examination now matches a framework of test fairness qualities and, in conclusion, the speakers will examine the ways in which EFL/ESOL examinations might develop in the next decade.

References
Standardization of Marking and the Reliability of Oral Tests. Shen Mingbo, Guo Qian, Zhang Wenxia, and Xing Ru, Tsinghua University

Oral tests, which evaluate the speaking skill (a productive skill) of learners, have a comparatively high validity. However, the reliability of oral tests is low. Since the questions in oral tests are subjective ones, the marking of oral tests usually lacks in objectivity and consistency. This arises not only from the inter-item consistency of exam questions, but from the inter-rater and intra-rater reliability. It has been a major concern for many researchers to improve the reliability of oral tests. This article addresses itself to this problem by describing the mark scheme and rater training of the oral tests of Tsinghua Proficiency Test (I), which was started in 1996 and recognized by the Ministry of Education in 1999. The research focuses on the detailed design of rating scale and rater training. Research methods include questionnaires and videos. The research reveals that validity and reliability of oral tests are closely related, and test reliability and rater reliability are also closely related. This research is valuable in that it provides some understanding of the relationship between validity and reliability, which has not been studied in depth in nationwide oral tests as well as small-scale oral tests in China.

Republication of a test that has been out of print is a surprisingly involved process that in many ways recapitulates the multiple phases of development of new tests. In the beginning, there is a lot of thinking and planning regarding the need for the test, its intended purpose, and how well its design is suited to the anticipated uses of the test. In the case of tests that were created decades ago, it is especially important to establish relevance and validity in the context of current language teaching and testing practices. There are also ethical and legal considerations (Reed and Stansfield, 2002), not to mention contract negotiations with the test authors, owners, and/or former publishers. Once all that is done, the test materials have to be evaluated for quality, relevance, and reproducibility. A test is seldom usable “as is” or “as was,” and so extensive modification is typically necessary. Once updated and assembled, the test has to be tried out, and then revised again based on test taker and test administrator feedback. It is also necessary to communicate with current users and potential customers. And then there is advertising, taking orders, filling orders, and providing customer support. Throughout it all, there are concerns about validity and appropriate uses, all balanced with considerations of the commercial viability of the product. This poster will illustrate the many steps involved in bringing the MLAT and the PLAB back into operational use and in sustaining their availability. An evaluation of the whole endeavor will be made by taking into account some of the major costs and benefits.
Demonstration of Three Computerized English Certification Tests Developed for Novice to Intermediate Level Examinees. Diane Strong-Krause, Brigham Young University

This poster session will provide an overview of a series of three English certification tests developed at Brigham Young University. The computerized tests are aimed at examinees with lower language proficiency levels ranging from Novice High to Intermediate Mid. The exams are task-based and cover the four skill areas—listening, speaking, reading, and writing. The tasks are independent but are related by a common theme. For example, an examinee must invite a friend to a movie, indicating the movie, the day and time (by leaving a phone mail message). Then the examinee must postpone the outing because of illness (by writing an email to the friend). Later, the examinee must explain to the doctor symptoms of the illness and then listen to the instructions given by the doctor (followed by a set of multiple-choice items). At the doctor’s office, the examinee reads an article in a magazine (followed by a series of questions about the article). One of the goals of the test developers was to design tests that might be interesting and relevant for the examinees. The poster will include a review of the tests’ purposes and provide examples of tasks.
POSTERS

Investigating High-stakes Test Impact at the Classroom Level. Carolyn E. Turner, McGill University

The awareness of the potential impact of high-stakes tests in L2 educational settings has been heightened in the recent literature (e.g., special edition Language Testing 1996:13(3), "washback effect"; symposium presented at LTRC 2000, Context & Method in Washback studies). This is generally referred to as "washback" in L2 education. Evidence from ashback studies in many parts of the world demonstrates the global concern for this phenomenon. We have learned that washback is complex in nature, and is contextually bound terms of location, purpose and methodology of the study. Much of the discussion focuses on the negative effects of tests (e.g., political uses and potential abuses of tests). It is too simple to conclude, however, that introducing new exams will affect all teachers and students the same way (Alderson & Hamp-Lyons 1996, Cheng 1998, Wall 1999, Watanabe 1996).

This poster will present research conducted over a 2-year period, which investigated the washback effects from the development stage through the implementation stage of the eaking section of a high-stakes provincial ESL exam at the secondary level in the province Quebec, Canada. The purpose of the study was to examine the nature of impact at the classroom level as experienced by different teachers and students in the same educational stem (i.e., evidence of change in behavior/practice and perceptions/beliefs). Recent studies ve found that washback is unpredictable at the classroom level (e.g., Andrews et al., 2002). No levels of participants were investigated: the macro level (Ministry of Education people teachers from across the province) and the micro level (5 case studies, 5 different teachers and their students). Macro-level data was collected across the 2-year study. Micro-level data collected in the 5 case study classrooms during 3 distinct observation periods over 6 months: baseline data period, period right after speaking exam and rating scale were officially reduced into the educational system, period right before administration of the speaking exam. Instruments developed and employed were: questionnaires (Likert-scale data and free responses), interviews, classroom observation guide (nature, number and length of speaking activities), post-observation "chats" guide. In addition, the speaking exam characteristics were identified. In order to analyze and triangulate the different data sources, an in-depth analysis of data organization took place. SPSS and the qualitative software, NUDIST, were used.

The macro-level results reveal a general direction of positive washback as teachers see that the speaking exam represents the curriculum and their teaching strategies. Among other things, they use the new speaking scale and similar tasks as instructional tools. The micro-level results (i.e., the classroom case studies) have proven more complex. The evidence demonstrates that certain types of washback concerning defined test characteristics teaching content/method may vary across teachers but may be predictable for individual teachers when data from teaching context, teacher beliefs and practice are triangulated and tied together. The research methodology enables this insight. A general framework of phenomenon is put forth.
What Can Discourse Analysis Tell Us About OPT Performance? Yang Lu, University of Reading

In an multi-task oral proficiency test we would often find that (1) some test-takers tend to elaborate, expand and enhance their replies and opinions; (2) some seems more likely to initiate in discourse when involved in two- or three-way communication; (3) some would not give up what they are saying and always try to get back where they are interrupted; and (4) some like to develop their partners’ ideas by supporting or contradicting with elaborated justifications.

What would be your reaction as the examiner? Are you going to give more scores to the ones who tend to do the above? Would you regard them as higher-level test-takers than the ones who do not behave the same in discourse? Do you think these discoursal tendencies could also show that these test-takers are also linguistically more competent?

This poster session will reveal the answers to the questions and report the hypotheses and methodology of a study which focuses on the growing interest in the validation of the notion of discourse competence first proposed by Canale and Swain (1980) and later reformulated in Bachman’s (1990) model of Communicative Language Ability. Results of previous studies have been mixed or contradictory: some suggest that grammatical and discourse competence are closely associated with each other (Bachman and Palmer 1982, Milanovic 1988), while others have suggested that they are distinctive from each other (Allen et al. 1983). Other investigators have found that the results seem to depend on the statistical procedures used (e.g. Swain, 1985; Harley et al. 1990).

The research predicts that test-takers’ spoken discourse competence resulted from discourse analysis positively correlates with examiners’ holistic and analytical judgement of the candidates’ discoursal performance, and discourse competence is an independent component in test-takers’ spoken communicative competence that positively correlates with their overall oral proficiency and linguistic performance. Thirty UCLES First Certificate English (FCE) Speaking Test events were analysed to test the hypotheses using a framework based on a modification of the Birmingham School Approach to spoken discourse analysis (DA) and examining the correlations between DA results and FCE overall, discoursal, global, and linguistic scores.
The UCLA Web-based Language Assessment System (WebLAS). Priyavada Abeywickrama, Lyle Bachman, Nathan Carr, Brent Green, Greg Kamei, Mkyung Kim, Soojong Kim, Michael Pan, Christine Salvador, Yasuyo Sawaki, Sang-kuen Shin, Sung-Ock Sohn, Viphavee Vongpumivich, and Xiaoming Xi, University of California at Los Angeles

Every year at UCLA, over 2,000 students are who enrolled in or are seeking to enroll in language classes in ESL, Japanese and Korean take language tests. The results of these tests are used for the purposes of placement, diagnosis, progress, and grading. Currently, all of these tests are administered in paper and pencil format, which presents a number of problems. The paper and pencil format limits the types of tasks and responses that can be included; the preparation of new assessment tasks and the scoring of student responses to these tasks are very labor-intensive; and infrequent scheduling is inconvenient for many students.

In order to address these problems, UCLA is developing a web-based language assessment system (WebLAS). Some of the advantages of a web-based system are greater administrative efficiency, more flexible and frequent scheduling, more authentic, interactive and valid measures of language ability, and improved response-time in scoring and notification of results, via web posting or email. At the same time, issues such as test security, working with multiple fonts, and how to deliver multi-media rich tasks and how to score open-ended responses need to be addressed.

The WebLAS consists of three dynamically-database driven components: 1) a Task Creation System, which is user-friendly for content specialists (classroom teachers) and easily integrated into a task bank, 2) a Task Delivery System, which consists of attractively designed, engaging web-based assessment tasks, linked to a database and a potentially automated assembly algorithm, and 3) a Data Management System, which provides for automated collection, scoring, processing, analysis, reporting, storage and retrieval of assessment results.

This poster will demonstrate the key components of the WebLAS. The Task Creation System enables test developers to create and edit assessment tasks on-line and to store these in a data base. It also requires test developers to specify the scoring rubric for the machine scoring of tasks with open-ended responses. A number of prototype web-based assessment tasks from the three languages (ESL, Japanese, Korean) that can be delivered via the Task Delivery System will be demonstrated. Finally, the prototype of Data Management System, with the response capture and automated scoring algorithms and will be demonstrated.
RESEARCH NETWORK

Researching and Supporting Classroom-based Assessment Systems. Geoff Brindley and Helen Slatyer, *Macquarie University*

This research network session will provide an opportunity to discuss the challenges faced in implementing and supporting assessment systems that rely on teacher-developed tasks as a basis for assessing and reporting outcomes in language learning programs. The session will focus on an ongoing research study into task-based assessments of language competencies that are used in the Australian Adult Migrant English Program (AMEP). The study aims to identify factors affecting the difficulty of teacher-developed assessments and to incorporate the research findings into a bank of assessment tasks that teachers can access online. In collaboration with a national working group of teachers, a set of professional development resources has also been created to assist teachers to develop their own assessment tasks. The online task bank and professional development resources are currently being piloted.

A number of ongoing issues and challenges that have arisen from this project will form the basis for discussion. These include:

**Technical issues**
- How can tasks that assess the same learning outcome be made parallel?
- How can minimum standards of achievement be set for individual tasks?
- What criteria for validity and reliability should be applied to teacher-conducted assessments?

**Professional development issues**
- How can guidelines be created to guide teachers in the task design phase?
- What are the minimum skills and knowledge necessary for teachers to be able to construct valid and reliable assessment tasks? What level of expertise can reasonably be achieved?
- What models of professional development are most effective in assisting teachers to acquire skills in assessment task design? How can internet technology best be harnessed to achieve this?

**Practical/political issues**
- What type and level of institutional support is necessary to support and maintain task development and quality control?
- What can be done to facilitate maximum teacher involvement and collaboration in professional development activities?
RESEARCH NETWORK

Researching and Supporting Classroom-based Assessment Systems. Geoff Brindley and Helen Slatyer, Macquarie University

This research network session will provide an opportunity to discuss the challenges faced in implementing and supporting assessment systems that rely on teacher-developed tasks as a basis for assessing and reporting outcomes in language learning programs. The session will focus on an ongoing research study into task-based assessments of language competencies that are used in the Australian Adult Migrant English Program (AMEP). The study aims to identify factors affecting the difficulty of teacher-developed assessments and to incorporate the research findings into a bank of assessment tasks that teachers can access online. In collaboration with a national working group of teachers, a set of professional development resources has also been created to assist teachers to develop their own assessment tasks. The online task bank and professional development resources are currently being piloted.

A number of ongoing issues and challenges that have arisen from this project will form the basis for discussion. These include:

Technical issues
- How can tasks that assess the same learning outcome be made parallel?
- How can minimum standards of achievement be set for individual tasks?
- What criteria for validity and reliability should be applied to teacher-conducted assessments?

Professional development issues
- How can guidelines be created to guide teachers in the task design phase?
- What are the minimum skills and knowledge necessary for teachers to be able to construct valid and reliable assessment tasks? What level of expertise can reasonably be achieved?
- What models of professional development are most effective in assisting teachers to acquire skills in assessment task design? How can internet technology best be harnessed to achieve this?

Practical/political issues
- What type and level of institutional support is necessary to support and maintain task development and quality control?
- What can be done to facilitate maximum teacher involvement and collaboration in professional development activities?
RESEARCH NETWORK

A Qualitative and Quantitative Analysis of EAP Reading Test Data. Nathan T. Carr, University of California at Los Angeles

Reading is a central activity in academia, and tests of second language academic reading, particularly tests of English for academic purposes (EAP), are used to make high-stakes decisions about hundreds of thousands of students and prospective students every year. It is not clear, however, that these tests reflect current reading research on the reader and task variables involved in reading. In fact, some notable exceptions aside (see Alderson, 2000), there seems to be, in general, a mismatch between theory and practice in many large-scale high-stakes EAP reading assessments. Furthermore, current measurement models may not adequately model reader and task variables that are either implicitly or explicitly included in the designs of large-scale high-stakes EAP reading tests. Therefore, this study investigates the extent to which widely used EAP reading tests are based on current research about the nature of reading, and the adequacy of currently available measurement models for modeling design features of these tests. The author’s dissertation investigates the matter of the reader and task variables involved in reading by addressing the following research questions:

1. What kinds of tasks are used in high-stakes EFL tests of reading, and how well do they reflect current thinking about the task variables involved in reading?
2. What do these tests appear to be measuring, in terms of the skills, abilities, processes, or other reader variables that they involve, and how well does this reflect current reading research about such variables?
3. What interactions between task and reader variables appear likely in these tests, and how well do these anticipated interactions reflect current research about such variables?

Subsequent research will involve expanding the study to include the analysis of additional tests, and will also investigate the extent to which currently available measurement models are able to model the reader and task variables and interactions that are design features of large-scale EAP reading tests. Specifically, it will address these questions:

4. To what extent do IRT, multifaceted Rasch measurement, and the Marcoulides-Drezner latent variable approach to multivariate G-theory give similar information regarding test items, i.e., to what extent do they scale items similarly?
5. To what extent do these three methodologies provide similar estimates of ability?
Assessing Multimodal Texts in Multilingual Classrooms. Harriet Davis, Assumption Convent School, and Yvonne Reed, University of the Witwatersrand

Since the mid-1990s, the New London Group (and others) have argued that the idea and scope of literacy pedagogy needs to be extended 'to account for the context of our culturally and linguistically diverse and increasingly globalised societies; to account for the multifarious cultures that interrelate and the plurality of texts that circulate' (NLG, 2000:9) Kress, Jewitt, Ogborn and Tsatsarelis suggest that assessment needs to be rethought in the context of multimodality and recommend that 'teachers be given the means to become highly reflexive of their practice in this respect, particularly of the interrelation between modes and shapes of knowledge, and between mode and receptiveness by students' (2000: 176-77). In similar vein, Stein and Newfield (forthcoming) suggest that 'Assessment of multimodal texts in the English classroom is a complex, multilayered task which makes new demands on teachers' competencies and has implications for teacher education.'

Having been introduced to new ways of thinking about literacy pedagogy while studying for her MA degree at the University of the Witwatersrand, Harriet Davis decided to change some of her classroom practices. As an alternative to a literary essay, she requested 40 grade 11 students in two multilingual, multicultural classes in a Johannesburg high school to respond to themes, issues, characters, or 'key moments in the plot' in Ngugi wa Thiongo’s novel A Grain of Wheat by developing (i) film story boards and (ii) posters or three dimensional artefacts. The learners participated in self and peer assessment of their work. Through a combination of classroom observation, questionnaires and interviews, Davis collected a wealth of data on learners’ responses to both the tasks and the assessment processes. Analysis of this data, undertaken in collaboration with Yvonne Reed, has raised a number of questions. For some of these we have tentative answers and for others none. Thus our plan for future research is (i) to do further empirical work with grade 11 high school students and (ii) to begin a research project with pre-service and in-service language and literacy teacher education students.

These are some of our questions:
1. Is it possible to frame criteria that both scaffold a multimodal task and also encourage innovative responses?
2. If an educational institution requires that marks or symbols be awarded for multimodal responses to texts, should ‘excellence in execution’ contribute to the mark or symbol (as has been traditional practice in the assessment of essays)? One of our tentative answers to this question, is that whether marks are awarded to the artefact, may depend on whether the focus is on demonstrating learning from/about the text or on learning to design, though the two are not easy to separate.
3. What can students learn from participating in self and peer assessment of multimodal responses to texts? How can teachers facilitate such learning?
4. In multicultural classrooms is the language of instruction likely to have a constraining effect on the quality or ‘accuracy’ of learners’ assessments and on their ‘take-up’ of teacher assessments? If there is such a constraining effect, can it be mitigated and if so, how?

At LTRC we are hoping to find out where and how any similar research is underway. We would like critical comment on our questions and also suggestions in regard to methods for taking our research forward.
RESEARCH NETWORK

Investigating the Dependability and Consistency of Speaking Test Ratings. Barbara K. Dobson, Jeff S. Johnson, and Amy D. Yamashiro, University of Michigan

This research network presentation is part of a larger validation study examining a large-scale, high-stakes, intermediate-level foreign language test. The first part of the study examines the dependability of the speaking test rating scale using Generalized Analysis of Variance (GENOVA). The second part of the study uses multi-faceted Rasch to investigate how the four-point rating scale and raters are performing.

For the 2001 test administration, six speaking prompts were produced. Under standard testing conditions a candidate responds to a prompt selected by the speaking test examiner. If the candidate receives a clear pass, that sole rating is used. If the candidate receives a borderline pass or a failing mark, one or more additional raters will evaluate an audiotape of the performance.

The participants included examinees at the target level of the examination battery and others who had ability levels higher and lower than the targeted test population. They were recruited to take the entire examination battery under standard testing conditions, and to take a second speaking test, administered on a separate day by a different speaking test examiner. In other words, each of the 53 participants for this study provided two speaking test samples which were rated independently by four raters. The first rating was done during the live test and the other three were done later using videotape.

These data were analyzed using a fully-crossed (P x T x R) design in GENOVA, where for each person (P) there were two tasks (T) and for each task there were three raters (R). Then, the steps on the rating scale and the consistency across the raters were analyzed using multi-faceted Rasch modeling. The researchers hope the presentation will lead to discussion about the methodologies used in this study as well as broader issues surrounding the validation of speaking assessments.
Item Complexity and Latent Trait Ability. Parisa Daftari Fard, Iran University of Science and Technology

The concept of unidimensionality of reading comprehension (Mathews, 1990; Weir et al., 1990; Hillocks and Ludlow, 1984) leads scholars to believe that there might be a one-to-one correspondence between item difficulty and the level of cognition it measures (Cf. Alderson, 1990). Other researchers, however, question the relationship between 'item difficulty and the item level' (Alderson and Lukmani, 1989 as cited in Alderson, 1990). In addition, there is research that suggests that reading comprehension might not be unidimensional (Daftari, 2002).

The present research examines the hierarchical characteristics of the item types and the correspondent latent trait ability the items measure. An in-depth investigation of the theoretical models of reading comprehension (Daftari, 2002) was used to account for the nature of reading comprehension (RC). Then two tests with varying item formats were developed to measure 24 abilities of RC, each with at least two items for each of the ability.

The tests were administered to 479 heterogeneous students (247 students took the SBRTa and 232 took the SBRTb). The Rasch model, along with confirmatory factor analysis, was utilized to answer the following question: is there a one-to-one correspondence between item complexity and the cognitive level of the items. The results show that there is no relationship between item difficulty and the latent trait ability; complex items might measure the so-called low ability, too. It is concluded that test developers might focus their attention on two aspects of RC ability: level of difficulty and cognition.
RESEARCH NETWORK

Building a Computer-adaptive Face-to-Face Oral Interview: The BEST Plus. Dorry M. Kenyon, Stephanie Stauffer, Mohammed Lougait, and Carol van Duzer, Center for Applied Linguistics

In this session, participants will see a computer-adaptive face-to-face oral interview built using the many-faceted Rasch model to provide adaptivity. As background to the next test, the original Basic English Skills Test (BEST) was developed in the early 1980s as an easily administered assessment of the speaking abilities of non-English-speaking (and often re-literate) adult refugees and immigrants to the United States. Most often used for placement and progress testing, the oral interview section consists of about 50 scripted questions. The test administrator scores each question using a simple rubric immediately after hearing the examinee’s response, before asking the next question. In general, the original EST oral interview takes about 15 minutes to administer.

Responding to needs for a shorter administration time, as well as greater accuracy of measurement for accountability assessment, researchers at the Center for Applied Linguistics we developed a computer-adaptive version of the BEST oral interview. With this instrument, the interviewer reads the test questions from the computer screen and enters the scores into the computer after the examinee responds. As in any computer-adaptive test, the computer program estimates the examinee’s ability and then selects, based on the examinee’s current ability estimate and within content and other constraints to be discussed in this session, the next question for the interviewer to read from the screen. The assessment procedure stops once one of three conditions is met: a lower measure bound is met, the standard error drops below a pre-determined level, or a pre-determined number of questions have been administered.

The design of the BEST Plus represents a complex compromise of several competing values. First, there is a need to keep scoring as simple as possible, since many test administrators in adult education may have minimal background in the field. Second, there is a need to cover certain content domains and keep a ‘conversation’ going in an adaptive test. Third, there is a desire to have high measurement accuracy but keep the assessment time as short as possible.

Data for calibrating the item pool for the BEST Plus came from over 2300 adult ESL students. In this session, results from an initial reliability study, in which two test administrators each tested the same group of 32 students with the adaptive instrument, will be presented. Feedback on the approaches we took to meet the competing ends of the adaptive testing situation will be greatly appreciated. Have we reached the um design of such an adaptive test, balancing stakeholders, content, and measurement?
Test of Interlanguage Pragmatic Knowledge of Chinese EFL Learners. Liu Jianda, City University of Hong Kong

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 have been based on the prototype developed by Hudson et al. (1992; 1995), with the exception of Röver (2001), who developed a web-based test to measure the interlanguage pragmalinguistic knowledge of ESL/EFL learners. This study will employ and validate three test formats developed by Hudson et al. – Written Discourse Completion Test (WDCT), Multiple-choice Discourse Completion Test (MDCT), and Discourse Self-assessment Test (DSAT) – to test the interlanguage pragmatic knowledge of Chinese EFL learners. Two speech acts (requests and apologies) with three variables (imposition, power, and familiarity) are included in this study. Much effort is devoted to the development of the three test papers in this study. The development procedure follows 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; and pilot testing (both paper-and-pencil and think-aloud), to validate preliminarily the test papers. Altogether 213 subjects (both native Chinese speakers and native English speakers) participated in the development of the tests. An additional 310 Chinese university EFL learners will take part in the actual data collection. The validation will be based mainly on Messick (1989), which includes a range of statistical procedures and analyses of test-taking process by means of verbal protocols.

References
An Exploration of the Validity of the Paired Candidate Interaction. Lynette May,
Nanyang Technological University

One commonly used task for eliciting assessable language is the Language
Proficiency Interview (LPI) or, in the American context, the Oral Proficiency Interview (OPI).
However, research on the validity of LPIS have raised concerns including the innate
artificiality of the task (Van Lier, 1989; Lazaraton, 1992; Johnson, 2001), the limited range of
discourse that can be elicited in a typical interview (Shohamy, 1983), the power differential
inherent in the interview (van Lier, 1989) which can only be exacerbated when the
interviewer is performing the dual role of assessor, the uniformity of interviewer (Brown &
Hill, 1998) and rater behaviour (Brown, 2000), and the essentially co-constructed nature of
the interview, which is perceived in the rating process as being the product of the interviewee
only (McNamara, 1997). Thus test developers are beginning to explore more authentic and
genuinely communicative performance assessments which can be used to elicit rateable
language samples from test takers (McNamara, 1996).

A form of paired candidate interaction is used in a number of high stakes
examinations, including several UCLES tests of spoken English (UCLES Research Notes,
2000). It is perceived to be more symmetrical and genuinely interactive in nature, as there is
the potential for the sharing of “power”, including topic nomination and abandonment. It was
found to be preferred by students (UCLES Research Notes, 2000). Perhaps more importantly,
the assessor is not performing a dual role as interlocutor.

However, the pair interaction has been criticized as lacking in reliability (Foot, 1999),
in that a test taker could be disadvantaged if paired with an interlocutor of differing oral
proficiency or communicative ability. It is possible that many of the areas of concern related
to LPIS will also be problematic in the paired learner interaction. As a high stakes mode of
speaking assessment, it warrants further investigation. Swain (2001) calls for consideration
of dialogue that is jointly constructed by language learners.

Through a pilot study involving twelve test takers and four assessors, I intend to
explore aspects of the validity of the paired candidate interaction. The orientation of the
methodology is primarily qualitative, with discourse analysis, and retrospective verbal
protocol analysis being used to gain different perspectives on the test taker’s performance and
the process through which assessors reach their final rating.
RESEARCH NETWORK

Assessing Teacher Language Proficiency: Turning a Stick Into a Carrot. Tim McNamara, The University of Melbourne, Jin Kyungae, Korea Institute of Curriculum and Evaluation, and Annie Brown, The University of Melbourne

The global spread of English is making increasing demands on educational systems worldwide. Since 1996, Korea has introduced English as a compulsory component of the primary school curriculum, but financial constraints since 1997 have limited the number of native speakers of English available. The burden of introducing the new curriculum has thus fallen to home room teachers, many of whom have limited English proficiency. A response in many systems to issues of teacher competence has been to introduce proficiency tests, which are often seen as threatening to teachers. Language testing researchers are often invited to become involved in the development of such tests. Is there an alternative?

The paper presents a report on one such case involving a group of Australian language testing researchers who were invited to develop a teacher proficiency test in a joint research project with the Korean education system, known as the Classroom Communicative Competence project. A way was sought to use assessment principles to support and encourage the efforts of teachers to improve their competence, rather than to place additional pressures on them. Assessment typically involves the freezing of a performance for careful consideration against relevant formal criteria, with the goal of assessing strengths and weaknesses. In diagnostic assessment, this analysis leads to the formulation of specific goals as a focus for change. In this project, a group of 25 lead teachers from provinces throughout Korea attended a week-long workshop in Seoul conducted in Korean by Korean language teacher educators. The workshop introduced a diagnostic assessment instrument in which teachers were trained in the use of an analytic framework to raise their awareness of the patterns of language and interaction used by both teachers and students in elementary English classes. They were then further trained in the formulation of specific goals for improvement in each area, and then applied the framework to videotapes of their own classrooms.

The paper describes (1) the training course and the analytic framework used, (2) the accompanying instruments for the analysis of lessons and (3) the results of the use of a measure of change in teacher consciousness over the course of the workshop. An analysis of the evaluative responses of participants to the workshop is also reported. The project is offered as a case study of the way in which assessment principles can be used creatively to support, rather than merely to police, the work of teachers facing the language demands of globalization.
RESEARCH NETWORK

A Unit-Based Learning Progress Report: Can it Apply to Traditional University Students as an Alternative Method of Progress Assessment? Niu Jian, The Hong Kong Polytechnic University

'A Unit-Based Learning Progress Report' was developed by the researcher as an alternative method of progress assessment for the distance learners at China Radio and Television Universities (RTVUs), the Open University of China. The instrument aims to assess students' pre-tutorial learning performance and aims to enhance students' independent learning. It consists of 3 sections in design: 'Evaluating Yourself', 'Your Problems and Tutor Feedback' and 'Talk to Your Tutor'.

In the Research Network session, the researcher will show a sample of the progress report and will discuss with participants whether the instrument can be applied to traditional university students as an alternative method of progress assessment to monitor students' pre-class learning performance. Two questions will be specifically discussed:

1) Is it necessary and also feasible to introduce the 'Unit-Based Learning Progress Report' in a classroom teaching context? Why or why not?

2) What might be the positive and/or negative impact upon learning and teaching, if the instrument is introduced in a classroom teaching context? Will the impact be different in any way from that in a distance education context?

It is expected that possible collaborative research might be initiated during the discussion to answer the two questions empirically.
RESEARCH NETWORK

Investigating the Role of Vocabulary Knowledge in Academic Reading Ability. John Read, *Victoria University of Wellington*

The strong correlation between tests of vocabulary knowledge and reading comprehension is well established. However, although there are various explanations for the nature and strength of the relationship, there has been comparatively little research with second language learners to investigate it thoroughly. One question that arises with such research is how to operationalize the construct of vocabulary knowledge. Is it sufficient to estimate the number of words that the learners have some acquaintance with (*breadth* of knowledge), or is it necessary to measure the quality (or *depth*) of their knowledge as well? Studies by Schoonen and Verhalleen (1998) in the Netherlands and Qian (1999, 2000) in Canada have produced some evidence that vocabulary depth – as measured by Read’s word associates test – accounts for a significant amount of the variance in reading test scores in addition to that predicted by a vocabulary size test.

To explore these findings further, the ongoing research project reported here focuses on the acquisition by international students at an English-medium university of the subtechnical vocabulary contained in Coxhead’s Academic Word List. Three measures have been developed and piloted: a) a breadth test, with a word-definition matching format; b) a depth test, using the word associates format; and c) a context test, with the vocabulary items embedded in a reading passage. The presentation will include the results of a study to develop these vocabulary tests, to be administered together with a measure of reading comprehension ability. Participants will be invited to discuss issues such as: how the pattern of relationships among the variables should be interpreted; how adequately the vocabulary tests operationalize various components of the construct; and the extent to which other vocabulary measures will be required to investigate the lexical dimension of learner performance in reading and in other skill areas, such as writing or listening.
The Bilingual Screening of Interpreter Applicants. Hanne Skaaden, Tatjana Felberg, Randi Havnen, and Mona Myran, University of Oslo

The testing of interpreter aptitude is not only an expensive endeavor, it is also a complicated one – both due to the diversity of skills required for the activity of interpreting and due to the complexity of the skills to be tested (Moser-Mercer 1994). A component present in most test batteries is the testing of bilingual skills. This is perhaps the most basic skill, but also the most complex of the skills required for the interpreting activity.

When the Norwegian Directorate of Immigration took over the hearings of asylum seekers from the police in July 2000, the directorate was faced with a need for the services of interpreters in a diversity of languages. Although Norway got an accreditation exam for interpreters in 1997 (cf. Mortensen 1998), the examination has not yet produced enough accredited interpreters to fill the need of the directorate. In order to strike a balance between the paragon and the possible, a bilingual lexical knowledge test was set up to eliminate candidates who obviously lacked the bilingual vocabulary necessary in order to fulfill the interpreter task.

The aim of this paper is to present results from the lexical knowledge test that was administered as a partial recruitment tool into the interpreter function. Along with an evaluation of the preliminary results from the test, the paper discusses the need for quality measurements in public service interpreting where the possibility of monitoring the interpreter’s performance is restricted due to professional secrecy. It is concluded that there is an evident need for testing, although the limitations of such tests should be stressed.

References

An Impact Study of Hong Kong’s Exit English Test for University Graduates:
Investigating Whether the Status of a Test Affects Students’ Test-preparation Activities.
Bernadette Stoneman, The Hong Kong Polytechnic University

The English language proficiency of Hong Kong students has been perceived as declining for some years, and the government sees an exit English test for Hong Kong’s university graduates as one way of motivating students to improve their English. One local university, the Hong Kong Polytechnic University (PolyU), proposed and subsequently received government funding to investigate and to develop a Graduating Students’ Language Proficiency Assessment (GSLPA). This test was officially launched and implemented by PolyU in 2000 for its own graduating students. The test was obligatory to all language-major students but voluntary to students from other disciplines. Over 2000 PolyU students took the test each year.

Three years into the operation of the GSLPA, in July 2002, the University Grants Committee (UGC) announced its decision to implement a Common English Proficiency Assessment (CEPA) for all students graduating from all the eight UGC-funded universities in Hong Kong. There are approximately 14,500 such students per year. UGC’s choice of test as the CEPA is the International English Language Testing System (IELTS). The taking of this CEPA-IELTS test is voluntary at this stage, but UGC is strongly encouraging all graduates to take it, to the point of offering to reimburse students the full cost of sitting the test. All PolyU students have received notice from the university’s Academic Secretariat about UGC’s decision regarding CEPA-IELTS.

Within the context of an exit English test for Hong Kong, the stated function of an exit test as a tool for assessing students’ English proficiency is the same for both the GSLPA and the IELTS. Similarly, the stated purpose of an exit test to motivate students to improve their English is the same for both the GSLPA and the IELTS. However, as regards the nature of the two ‘versions’ of Hong Kong’s exit English test, the GSLPA and IELTS differ. GSLPA is a locally developed test that is implemented at the single-institutional level; IELTS is an internationally recognised test implemented at across-institutional level and with overt government sanction. Regarding tests and their washback effects, Alderson & Hamp-Lyons (1996) put forth a hypothesis worthy of investigation: that the amount and type of washback will vary according to, amongst other things, the status of the test. By status Alderson & Hamp-Lyons were referring to the level of stakes that the test carries. However, it seems reasonable to also refer to the status of a test as 1) the degree of government endorsement that the test receives, 2) the level of test implementation, and 3) whether the test is recognised internationally. The present case of Hong Kong moving from one ‘version’ of its exit English test to another ‘version’ that has a different status offers a unique opportunity to test Alderson & Hamp-Lyons’ (1996) hypothesis.

The research study investigates whether and to what extent the status of a test as defined above affects students’ test-preparation activities. The study is conducted in PolyU where different ‘versions’ of the exit test are used at different times. The period of investigation straddles the last year of operation of the GSLPA and the first implementation of the IELTS. The focus of the proposed research network session at LTRC 2002 will be to invite critical comments regarding the research design and research question framework. The researcher also wishes to share experiences with fellow researchers on problems concerning designing and using questionnaires, interviewing students, and observing language classrooms.
RESEARCH NETWORK

The Language Proficiency Assessment for Teachers of English in Hong Kong. Alan Urmston, Hong Kong Examinations Authority

The Language Proficiency Assessment for Teachers (LPAT) is a criterion-referenced test of the language ability of teachers of English (and Putonghua or Mandarin). Teachers (who are not already exempted due to being subject trained and qualified) in primary and secondary schools in Hong Kong are required to reach a “benchmark” level 3 on a 5-point scale of proficiency in five skill areas of reading, writing, listening, speaking and classroom language in order to be able to teach in schools after 2005.

The LPAT was developed as a result of recommendations made by the Hong Kong Education Commission in 1995, in response to research findings from the University of Hong Kong, which showed that only 14% of local English teachers were subject trained. In March 1997, a pilot scheme was carried out to determine benchmark language proficiency levels for language teachers and the first administration of the Assessment was carried out in March 2001, amidst protests from teachers who objected to having to take the test.

This paper will be divided into two parts. The first part will briefly describe the development of the LPAT, through the pilot scheme to determine the benchmark levels leading to the first and second administrations of the Assessment. The presenter will describe the changes that the Assessment has gone through in order for it to be seen as a fair test of the language proficiency of school teachers. In the second part of the presentation, the washback effects, in terms of both the socio-political and educational impact of the Assessment on English language teaching in Hong Kong will be examined.

It is anticipated that colleagues will find the presentation an interesting and informative account of how an educational system, in which the major second or foreign language is taught mainly by non-native, unqualified people, attempts to raise its standards of language provision.
Validity Evidence of Performance-based Tasks Measuring Text Structure Knowledge.
Viphavee Vongpumivitch, University of California at Los Angeles

Reading researchers have hypothesized that the knowledge of text structure, i.e., how ideas in the reading passages are organized and presented, is an important component of academic reading ability. However, while this knowledge has been primarily measured in reading tests through multiple-choice items, there is, at present, no performance-based task that has shown sufficient evidence of validity in measuring the knowledge of text structure.

The goal of the present study is to compare validation evidence provided by three performance-based tasks: (1) a recall task in which the examinees replicate the hierarchy of information in a reading passage, (2) a diagram task in which the examinees fill in all main ideas and supporting details in a passage in a blanked diagram, and (3) an incomplete outline task in which the examinees fill in the appropriate main ideas or supporting details in a partially-blanked outline. Ultimately, the task that shows the strongest evidence of validity will be included as a part of the reading section in a web-based ESL placement examination.

To answer the question of the extent to which each of the three tasks measure the knowledge of text structure, two sets of data have been gathered from a group of university ESL students. First, the examinees’ responses to the three tasks have been collected for quantitative analysis using confirmatory factor analysis. Second, a subset of examinees’ provided verbal protocol data for qualitative analysis of the processes in which they were engaged as they performed the three tasks.

This research network section will focus on the analysis of the verbal protocol data. Interpretation of the results as validation evidence of inferences about examinees’ text structure knowledge will be discussed. Comments and suggestions in terms of further analysis and interpretation of the data are welcome. The presenter also encourages interactive exchange with researchers having experience with verbal protocol data.
The two nation-wide tests of CET (College English Test for non-language majors) and TEM (Test for English Majors) in China have been in official operation for more than a decade. Their scores are widely used and frequently interpreted by associated parties or score users for academic or occupational purposes. Both CET and TEM have two levels of testing; one is designed for sophomores at the foundation stage and the other for the advanced stage. As currently there is the tendency that the CET test for the advanced stage, named CET-6, is also taken by sophomores across China, a quest may logically arise for comparability between the two tests, CET-6 and TEM-4.

Nevertheless, there has been little research on the comparability of the two influential tests in China. To bridge this gap, this study attempts not only to search for relationship, if any, in terms of test construct but also to explore the extent to which the two sets of scores can be appropriately compared or equated, with a focus on the common influences that shape the learning experiences and achievements of all second-year students in higher education in China.

To this end, about 200 subjects (including both English majors and non-English majors) from relevant academic institutions (both a bit higher than norm data) were selected to take the two different tests within a short time. Analyses were conducted to look into the comparability between the two tests. Presently, statistics and analyses show that one set of scores achieved from the CET-6 may be more or less the same as that of the TEM-4, but the language proficiency they measure is different in essence.

Needless to say that this discovery may well reflect the actual different patterns of English learning resulted from differences in teaching syllabuses for either language majors or non-language majors. Can we now say that the advanced stage of English proficiency for non English majors is more or less equal to the foundation stage of English majors in China? This question could be of critical relevance when new designs of these tests are under discussion.
The list below is intended to enable contact among LTRC 2002 participants. Please note that this list does not comprise all participants. Only symposia presenters and authors of the abstracts compiled in this book are listed. For the abstracts with 2 or more authors, only the presenting author is included.

<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
<th>Institution</th>
<th>City</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arkoudis, Sophie</td>
<td><a href="mailto:sophiaa@unimelb.edu.au">sophiaa@unimelb.edu.au</a></td>
<td>University of Melbourne</td>
<td></td>
<td>AUSTRALIA</td>
</tr>
<tr>
<td>Bachman, Lyle</td>
<td><a href="mailto:bachman@ucla.edu">bachman@ucla.edu</a></td>
<td>University of California at Los Angeles</td>
<td></td>
<td>USA</td>
</tr>
<tr>
<td>Barbier, Isabella</td>
<td><a href="mailto:isabella@ordinate.com">isabella@ordinate.com</a></td>
<td>Ordinate Corporation</td>
<td></td>
<td>USA</td>
</tr>
<tr>
<td>Brindley, Geoff</td>
<td><a href="mailto:gbrindley@ling.mq.edu.au">gbrindley@ling.mq.edu.au</a></td>
<td>Macquarie University</td>
<td></td>
<td>AUSTRALIA</td>
</tr>
<tr>
<td>Brown, Annie</td>
<td><a href="mailto:a.brown@unimelb.edu.au">a.brown@unimelb.edu.au</a></td>
<td>University of Melbourne</td>
<td></td>
<td>AUSTRALIA</td>
</tr>
<tr>
<td>Butler, Frances</td>
<td><a href="mailto:butler@cse.ucla.edu">butler@cse.ucla.edu</a></td>
<td>University of California at Los Angeles</td>
<td></td>
<td>USA</td>
</tr>
<tr>
<td>Carr, Nathan</td>
<td><a href="mailto:carr@ucla.edu">carr@ucla.edu</a></td>
<td>University of California at Los Angeles</td>
<td></td>
<td>USA</td>
</tr>
<tr>
<td>Chalhoub-Deville, Micheline</td>
<td><a href="mailto:m-chalhoub-deville@uiowa.edu">m-chalhoub-deville@uiowa.edu</a></td>
<td>University of Iowa</td>
<td></td>
<td>USA</td>
</tr>
<tr>
<td>Chang, Jaehak</td>
<td><a href="mailto:je434@columbia.edu">je434@columbia.edu</a></td>
<td>Columbia University</td>
<td></td>
<td>USA</td>
</tr>
<tr>
<td>Daftari Fard, Parisa</td>
<td><a href="mailto:pdaftaryfard@yahoo.com">pdaftaryfard@yahoo.com</a></td>
<td>Iran University of Science and Technology</td>
<td></td>
<td>IRAN</td>
</tr>
<tr>
<td>Damron, Rebecca</td>
<td><a href="mailto:rld@okstate.edu">rld@okstate.edu</a></td>
<td>Oklahoma State University</td>
<td></td>
<td>USA</td>
</tr>
<tr>
<td>Davidson, Fred</td>
<td><a href="mailto:fgd@uiuc.edu">fgd@uiuc.edu</a></td>
<td>University Of Illinois at Urbana-Champaign</td>
<td></td>
<td>USA</td>
</tr>
<tr>
<td>Davies, Alan</td>
<td><a href="mailto:a.davies@ed.ac.uk">a.davies@ed.ac.uk</a></td>
<td>University of Edinburgh</td>
<td></td>
<td>UK</td>
</tr>
<tr>
<td>Davison, Chris</td>
<td><a href="mailto:cdavison@hkucc.hku.hk">cdavison@hkucc.hku.hk</a></td>
<td>University of Hong Kong</td>
<td>HONG KONG</td>
<td></td>
</tr>
<tr>
<td>Douglas, Dan</td>
<td><a href="mailto:dandoug@iastate.edu">dandoug@iastate.edu</a></td>
<td>Iowa State University</td>
<td></td>
<td>USA</td>
</tr>
<tr>
<td>Elder, Catherine</td>
<td><a href="mailto:c.elder@auckland.ac.nz">c.elder@auckland.ac.nz</a></td>
<td>University of Auckland</td>
<td></td>
<td>NEW ZEALAND</td>
</tr>
<tr>
<td>Fox, Janna</td>
<td><a href="mailto:jfox@ccs.carleton.ca">jfox@ccs.carleton.ca</a></td>
<td>Carleton University</td>
<td></td>
<td>CANADA</td>
</tr>
<tr>
<td>Geranpayeh, Ardeshir</td>
<td><a href="mailto:Geranpayeh.A@ucl.ac.uk">Geranpayeh.A@ucl.ac.uk</a></td>
<td>University of Cambridge</td>
<td></td>
<td>UK</td>
</tr>
<tr>
<td>Gui, Shichun</td>
<td><a href="mailto:itscgui@gdvnet.com">itscgui@gdvnet.com</a></td>
<td>Guangdong University of Foreign Studies</td>
<td></td>
<td>CHINA</td>
</tr>
<tr>
<td>Halleck, Gene</td>
<td><a href="mailto:halleck@okstate.edu">halleck@okstate.edu</a></td>
<td>Oklahoma State University</td>
<td></td>
<td>USA</td>
</tr>
<tr>
<td>Howell, Elizabeth</td>
<td><a href="mailto:Elizabeth.Howell@hct.ac.uk">Elizabeth.Howell@hct.ac.uk</a></td>
<td>Higher Colleges of Technology</td>
<td></td>
<td>UAE</td>
</tr>
</tbody>
</table>
Hudson, Thom
tdh@hawaii.edu
University Of Hawaii
USA

Jones, Neil
Jones.N@ucl.ac.uk
University of Cambridge Local Examinations Syndicate
UK

Kozaki, Yoko
akahr500@tcn.zag.ne.jp
Mukogawa Women’s University
JAPAN

Lim, Youn-Hee
ylim@ets.org
Educational Testing Service
USA

Masters, Geoff N.
Masters@acer.edu.au
Australian Council for Educational Research
AUSTRALIA

Niu, Jian
jian.niu@polyu.edu.hk
Hong Kong Polytechnic University
HONG KONG

Rea-Dickins, Pauline
P.Rea-Dickins@bristol.ac.uk
University of Bristol
UK

Saville, Nick
saville.n@ucl.ac.uk
University of Cambridge Local Examinations Syndicate
UK

Shen, Mingbo
shenmingbo@yahoo.com
Tsinghua University
CHINA

Ito, Akihiro
akito@dpc.aichi-gakuin.ac.jp
Aichi Gakuin University
JAPAN

Kenyon, Dorry
dorry@cal.org
Center for Applied Linguistics
USA

Johnson, Jeff S.
elisj@umich.edu
University of Michigan
USA

Kobayashi, Miyoko
m.kobay@kan.kuis.ac.jp
University of Warwick
UK

Lee, Yong-Won
ylee@ets.org
Educational Testing Service
USA

Liu, Jianda
enjlhu@cityu.edu.hk
City University of Hong Kong
HONG KONG

May, Lynette
lamay@nie.edu.sg
Nanyang Technological University
SINGAPORE

McNamara, Tim
ttimcna@unimelb.edu.au
University of Melbourne
AUSTRALIA

Pan, Michael
mipan@humnet.ucla.edu
University of California at Los Angeles
USA

Reed, Yvonne
yvonne_reed@languages.wits.ac.za; University of the Witwatersrand
SOUTH AFRICA

Shin, Sang-Keun
sshin@humnet.ucla.edu
University of California at Los Angeles
USA

Read, John
John.Read@vuw.ac.nz
Victoria University Of Wellington
NEW ZEALAND

Ross, Steven
s95001@ksc.kwansei.ac.jp
Kwansei Gakuin University
JAPAN

Shaw, Stuart
Shaw.S@ucl.ac.uk
University of Cambridge Local Examinations Syndicate
UK

Shohamy, Elana
elana@post.tau.ac.il
Tel Aviv University
ISRAEL
Skaaden, Hanne
hanne.skaaden@if.uio.no
University of Oslo
NORWAY

Storey, Peter
pstorey@ied.edu.hk
Hong Kong Institute of Education
HONG KONG

Turner, Carolyn
carolyn.turner@mcgill.ca
McGill University
CANADA

Wang, Lin
lwang@ets.org
Educational Testing Service
USA

Yamada, Minako
yamada@tkk.at.tne.jp
The University of the Air
JAPAN

Zhou, Yuemei
yuemei@mail.shufe.edu.cn
Shanghai University of Finance and Economics
CHINA

Stansfield, Charles W.
cstansfield@zlti.com
Second Language Testing Inc.
USA

Strong-Krause, Diane
ds23@email.byu.edu
Brigham Young University
USA

Urmston, Alan
aurmston@hkea.edu.hk
Hong Kong Examinations Authority
HONG KONG

Xi, Xiaoming
xxm@ucla.edu
University of California at Los Angeles
USA

Yamashiro, Amy D
amydyama@umich.edu
University of Michigan
USA

Stoneman, Bernadette
egstonem@polyu.edu.hk
Hong Kong Polytechnic University
HONG KONG

Taylor, Lynda
taylor.l@ucles.org.uk
University of Cambridge
Local Examinations Syndicate
UK

Vongpumivitch, Viphavee
viphavee@hotmail.com
University of California at Los Angeles
USA

Xi, Zhongen
zhongenxi_wenli@hotmail.com
Shaoxing University
CHINA

Yang, Lu
lrooyl@rdg.ac.uk
University of Reading
UK
LIST OF PARTICIPANTS

This list comprises participants whose registration took place before 27 November, 2002.

Abeywickrama, Priyan
ailey, Alison
erry, Vivien
rooks, Lindsay
rown, Kathleen
ck, Gary
rpetner, Helen
ang, Jinping
au, Michael
vali, Nalini
eng, Venice
ara, Tetsuro
ng, Tim
is, Harriet
eechter, Roelof
ille, Craig
zhengdong
itt, William
othy
i, Yo
ah, Aishah
ungae
unyoung
ami, Satoko
sha, Shinji
ven, Vincent
y, Antony
ristina
ung Shik
athy
icz, Jo
-in-min
-ar-fun
ena
i
hi, Mineo
ye
eyyed Abbas

priyan@ucla.edu
aimbens@ucla.edu
vberry@hkcc.hku.hk
lbrooks@oise.utoronto.ca
kabcontact@hotmail.com
gbuck@ballard-tighe.com
helen@cal.org
cindych33@hotmail.com
Michael.Engl@polyu.edu.hk
rchavali@hkeaa.edu.hk
eyvenice@polyu.edu.hk
Chihara@wilmina.ac.jp
eltchung@cityu.edu.hk
h.davis@webmail.co.za
R.van.Deemter@let.leidenuniv.nl
craig-deville@uiowa.edu
eegzedon@hotmail.com
bhewitt@ncsc.dni.us
02900034e@polyu.edu.hk
voinnami@hotmail.com
egaishah@polyu.edu.hk
kajin@kice.re.kr
youngk77@hotmail.com
skawakami@toiec.or.jp
cats@ares.eonet.ne.jp
vincent.kortleven@elt.kuleuven.ac.be
akunnan@calstatela.edu
clee@hkeaa.edu.hk
vshlee@ucla.edu
gkathy@polyu.edu.hk
jolewkow@hkusua.hku.hk
sammi_liu79@hotmail.com
eygfliu@polyu.edu.hk
weiminwk@online.SH.CN
lorenas@ucla.edu
ev.lui@cityu.edu.hk
research@toiec.or.jp
eyjanem@polyu.edu.hk
sa_mousavi@yahoo.com
<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Okabe, Junko</td>
<td><a href="mailto:okabe@lit.aichi-pu.ac.jp">okabe@lit.aichi-pu.ac.jp</a></td>
</tr>
<tr>
<td>O'Loughlin, Kieran</td>
<td>kieran.o'<a href="mailto:loughlin@rmit.edu.au">loughlin@rmit.edu.au</a></td>
</tr>
<tr>
<td>Opanon-amata, Prakaikaew</td>
<td><a href="mailto:prakaikaew.O@chula.ac.th">prakaikaew.O@chula.ac.th</a></td>
</tr>
<tr>
<td>Pang, Francine Soi Meng</td>
<td><a href="mailto:francine.pang@cuhk.edu.hk">francine.pang@cuhk.edu.hk</a></td>
</tr>
<tr>
<td>Park, Siwon</td>
<td><a href="mailto:siwon@hawaii.edu">siwon@hawaii.edu</a></td>
</tr>
<tr>
<td>Pellatt, Valerie</td>
<td><a href="mailto:mslvip@bath.ac.uk">mslvip@bath.ac.uk</a></td>
</tr>
<tr>
<td>Poehner, Matt</td>
<td><a href="mailto:mep158@psu.edu">mep158@psu.edu</a></td>
</tr>
<tr>
<td>Prapphal, Kanchana</td>
<td><a href="mailto:kanchana.p@chula.ac.th">kanchana.p@chula.ac.th</a></td>
</tr>
<tr>
<td>Purpura, James</td>
<td><a href="mailto:jp248@columbia.edu">jp248@columbia.edu</a></td>
</tr>
<tr>
<td>Qi, Luxia</td>
<td><a href="mailto:engiluxi@cityu.edu.hk">engiluxi@cityu.edu.hk</a></td>
</tr>
<tr>
<td>Quan, Zhang</td>
<td><a href="mailto:gzhou@hku.hk">gzhou@hku.hk</a></td>
</tr>
<tr>
<td>Reed, Dan</td>
<td><a href="mailto:djreed@indiana.edu">djreed@indiana.edu</a></td>
</tr>
<tr>
<td>Ronning, Sidsel</td>
<td><a href="mailto:sir@udi.no">sir@udi.no</a></td>
</tr>
<tr>
<td>Shen, Sabrina</td>
<td><a href="mailto:Sabrina.a.shen@inet.polyu.edu.hk">Sabrina.a.shen@inet.polyu.edu.hk</a></td>
</tr>
<tr>
<td>Shimada, Megumi</td>
<td><a href="mailto:shimada@u-gakugei.ac.jp">shimada@u-gakugei.ac.jp</a></td>
</tr>
<tr>
<td>Shimizu, Yoko</td>
<td><a href="mailto:yukos@ec.ritsume.ac.jp">yukos@ec.ritsume.ac.jp</a></td>
</tr>
<tr>
<td>Smart, Cameron</td>
<td><a href="mailto:cjsmart@hkeaa.edu.hk">cjsmart@hkeaa.edu.hk</a></td>
</tr>
<tr>
<td>Spratt, Mary</td>
<td><a href="mailto:ecspratt@polvu.edu.hk">ecspratt@polvu.edu.hk</a></td>
</tr>
<tr>
<td>Tang, Rosina</td>
<td><a href="mailto:rwyang@hkcn.com">rwyang@hkcn.com</a></td>
</tr>
<tr>
<td>Tchigaeva, Svetlana</td>
<td><a href="mailto:egsveta@polyu.edu.hk">egsveta@polyu.edu.hk</a></td>
</tr>
<tr>
<td>Thrasher, Randy</td>
<td><a href="mailto:thrasher@ocj.ca.hk">thrasher@ocj.ca.hk</a></td>
</tr>
<tr>
<td>Tsai Feng Yin, Sarah</td>
<td><a href="mailto:sarahft.yin@cettw.com">sarahft.yin@cettw.com</a></td>
</tr>
<tr>
<td>Walton, Owen</td>
<td><a href="mailto:owalt@fluencygroup.com">owalt@fluencygroup.com</a></td>
</tr>
<tr>
<td>Wilson, James</td>
<td><a href="mailto:ecjames@inet.polyu.edu.hk">ecjames@inet.polyu.edu.hk</a></td>
</tr>
<tr>
<td>Winke, Paula</td>
<td><a href="mailto:paula@exal.org">paula@exal.org</a></td>
</tr>
<tr>
<td>Wu, Weiping</td>
<td><a href="mailto:wwp@cuhk.edu.hk">wwp@cuhk.edu.hk</a></td>
</tr>
<tr>
<td>Yanli, Zhang</td>
<td><a href="mailto:Johnwei@online.sh.cn">Johnwei@online.sh.cn</a></td>
</tr>
<tr>
<td>Yongqiang, Zeng</td>
<td><a href="mailto:zengyq@gdufs.edu.cn">zengyq@gdufs.edu.cn</a></td>
</tr>
<tr>
<td>Yuasa, Katsumi</td>
<td><a href="mailto:k.yuasa@ile-japan.com">k.yuasa@ile-japan.com</a></td>
</tr>
<tr>
<td>Yuen, Weith</td>
<td><a href="mailto:elweith@cityu.edu.hk">elweith@cityu.edu.hk</a></td>
</tr>
</tbody>
</table>
... in the field of child language... there has been no up-to-date, easily accessible source to provide guidelines, stimulate ideas and relate theory to practice.

The Times Higher Education Supplement on Child Language Teaching & Therapy

‘A substantial literature has developed on the topic... it makes sense to concentrate more of this work into a single journal... Language Testing does this admirably’

The Times Higher Education Supplement on Language Testing

‘More than any other journal in the field, Second Language Research reflects the growing interest in relating second language acquisition research to general theoretical concerns in linguistics and psychology... highly recommended.

Choice, American Library Association on Second Language Research

Try our journals for FREE!
- contact us today at ejournal@hodder.co.uk or www.arnoldpublishers.com

25th Annual Language Testing Research Colloquium

July 22nd to 25th 2003

The University of Reading

Learner Interfaces with Language Testing & Assessment

Proposal Submission Deadline: January 31st 2003

Contact: Dr Barry O'Sullivan
School of Linguistics & Applied Language Studies
The University of Reading
Whiteknights
Reading RG6 6WB
Berkshire United Kingdom

www.rdg.ac.uk/AcaDepts/IL/teru/ltrc2003/
teru@reading.ac.uk

The University of Reading