LTF2014 UNIVERSITY OF Southampton

LANGUAGE TESTING FORUM 2014 21– 23 November

Engaging multi-disciplinary perspectives

HARTLEY SUITE, BUILDING 38, UNIVERSITY OF SOUTHAMPTON, SO17 1BJ

CONTENTS

TECHNICAL COMMUNICATION

Fast eduroam wireless Internet is available for delegates from organisations participating in the eduroam federation.

Eduroam users will be able to connect immediately using the username and password they already have from their home institution (eduroam users should have set-up their computer at their home organisation before coming to Southampton).

For delegates who cannot access eduroam, please contact the registration desk for temporary login information.

SPONSORED BY

CLLEAR Research Centre

PROGRAMME

Friday 21 November 2014		
17:00-18:00	**DELEGATE REGISTRATION** **RECEPTION**	Hartley Suite Building 38 Highfield Campus
18:10-18:30	**WELCOME AND INTRODUCTION** Clare Mar-Molinero (Associate Dean) Roumyana Slabakova (Director CLLEAR)	
18:30-19:30	**OPENING PLENARY TALK** Stakeholders and consequence in test development and validation	Barry O'Sullivan
20:00	**DINNER** Ceno Bar & Restaurant, 119 Highfield Lane, SO17 1AQ	

Saturday 22 November 2014		
09:00-09:30	**PAPER 1 (FEATURED TALK)** The potential uses and limitations of eye-tracking technology in research into language testing	Stephen Bax
09:30-10:00	**PAPER 2** Designing low-stakes English language tests by teachers: Using multi-disciplinary knowledge and skills in assessment designing process	Lin Fang
10:00-10:30	**PAPER 3** Reconsidering the development of pragmatics tests: The case of the discourse completion test	Afef Labben Moez Athimni
10:30-11:00	**REFRESHMENT BREAK**	
11:00-11:30	**PAPER 4** Computer technology and language testing	John H.A.L. De Jong

11:30-12:00	**PAPER 5** Introducing opportunities for learning-oriented assessment to large-scale speaking tests	Anthony Green Liz Hamp-Lyons
12:00-12:30	**PAPER 6** C-tests outperform Yes/No vocabulary size tests as predictors of receptive language skills	Claudia Harsch Johannes Hartig
12:30-14:00	**LUNCH AND POSTER PRESENTATIONS** (Full abstracts detailed below)	
14:00-14:30	**PAPER 7** Interfaces between corpus linguistics and language testing/assessment	Yu-hua Chen
14:30-15:00	**PAPER 8** Exploring corpus analyses to inform writing assessment: A pilot study	Franz Holzknecht
15:00-15:30	**PAPER 9** What can expert judgement teach us about using the CEFR for young learner test development?	Amy Malloy
15:30-16:00	**REFRESHMENT BREAK**	
16:00-16:30	**PAPER 10** Virtual face-to-face speaking tests using web-based video conferencing technology	Fumiyo Nakatsuhara Chihiro Inoue Vivien Berry Evelina Galaczi
16:30-17:00	**PAPER 11** The planning strategies of young teenagers in a speaking task: Cultural trends	Gwendydd Caudwell
17:00- 18.00	**PANEL DISCUSSION** Tony Green, Liz Hamp-Lyons, Jennifer Jenkins, Barry O'sullivan	Chaired by Richard Kiely
18:30	**DINNER** Banana Wharf, Ocean Village, SO14 3JF	

Sunday 23 November 2014		
09:00-09:30	**PAPER 12 (FEATURED TALK)** Feedback as first principle	Liz Hamp-Lyons
09:30-10:00	**PAPER 13** Investigating washback: A study of an English speaking component in the French Baccalauréat	Gemma Bellhouse
10:00-10:30	**PAPER 14** An investigation of students' writing performance in the Test of English for Academic Purposes (TEAP)	Mikako Nishikawa M. Honma K. Nakamura T. Matsudaira S. Shiozaki K. Yanase
10:30-11:15	**REFRESHMENT BREAK**	
11:15-11:45	**PAPER 15 (FEATURED TALK)** Looking into reading: The use of eye tracking to investigate test-takers' cognitive processing	Tineke Brunfaut Gareth McCray
11:45-12:15	**PAPER 16** SLA theories on developmental sequences and the assessment of L2 writing	Christian Krekeler
12:15-13:15	**CLOSING PLENARY TALK** English language teaching and testing at the crossroads	Lianzhen He
13.15-13:30	**LTF 2014 – FAREWELL** Followed by buffet lunch	

POSTER PRESENTATION TITLES Saturday 22 November 2014

- A communicative approach to Arabic language proficiency testing in the light of Diglossia
- The revitalisation of formative assessment for developing academic writing and enhanced practices
- Analysing academic listening needs from a cognitive perspective
- The relationship between teachers' L1 Arabic varieties and students' L2 English pronunciation • Effect of Language Learning Strategies (LLS) instruction on Saudi EFL students' strategy use and proficiency
- Investigating assessment literacy in Tunisia: The case of EFL advanced reading teachers
- Bridging the gap: The effectiveness of short-term intensive IELTS writing preparation in Japan and 'relearning' academic conventions
- Investigating the relationship of word frequency and learner proficiency in an English proficiency test
- How well do different task formats elicit L2 learners' pragmatic competence?
- International and local raters: Comparing ratings and rationales on a speaking test across international contexts
- Predictive validity of TOEFL iBT: Quantitative and qualitative perspectives
- Language testing and vocabulary research – a symbiotic relationship?
- Investigating the validity of discourse completion tests: Effects of rejoinders and prompt enhancement
- Investigating the constructs measured by EAP writing tests for use in Japanese university entrance examinations
- Comparing textual features of the IELTS and TOEFL iBT reading texts: An empirical approach
- Evaluating the effect of testing on teaching within the English as a life skill programme in Sri Lanka- A case study
- The effects of planning time on Taiwanese learners' language performance and strategy use in a TOEFL iBT integrated speaking task
- The comments are useful but you do not understand: Supporting lecturer-written feedback with effective and mutual understanding

ABSTRACTS OF PAPER PRESENTATIONS

Friday 21 November 2014

Plenary Speakers:

Stakeholders and consequence in test development and validation

Prof Barry O'Sullivan, British Council/ Roehampton University, London

Ever since the introduction of the concept of test consequence as an aspect of validity, the profession has argued about its importance and relevance. While it has been generally accepted that test consequence is important, the degree to which this is the case and the way in which it might impact on test development has been debated and, more recently, challenged. Like others, my position on the topic has changed over the past number of years, from one of scepticism (the concept of 'consequential validity' is itself an error), to more conciliatory (consequence is somehow important to all aspects of test development and validation), to my current view (presented here). It is now clear to me that stakeholders offer test developers a key to understanding the way in which test consequence can be operationalised in development and validation models. By taking stakeholder groups into account at the conceptualisation stage of development we are essentially building consequence into the test design. This, in turn, enables an accurate prediction of the impact of decisions made on the basis of test performance, and creates a clearly considered *a priori* and *a posteriori* role for consequence within the development and validation model. Consequence can therefore be seen as a critical and operationalisable source of validation evidence. Involving stakeholder groups in this way has other consequences, this time for the developer. By acknowledging the importance and relevance of stakeholder groups to test development, we must also recognise the importance of communicating our validation arguments appropriately to them.

Bio

Professor Barry O'Sullivan is currently working with the British Council in London as Head of Assessment Research & Development. His recent work includes the design, development and validation of a placement test to be used by the British in their centres across the world and the design, development and validation of a new business to business language test called Aptis. Barry is particularly interested in issues related to performance testing, test validation, test-data management and analysis and scaling and calibration. Barry's publications have appeared in a number of international journals and he has presented his work at international conferences around the world. He is currently working (with C. Weir) on a major project documenting a history of language testing within the British Council. In addition to his work in the area of language testing, Barry has taught in Ireland, England, Peru and Japan.

English language teaching and testing at the crossroads

Prof Lianzhen He, Zhejiang University

English has been given much emphasis since China’s opening up to the outside world in 1978, as is evidenced by the development of English language teaching and testing over the years. But there has also been some controversy over the emphasis given to the English language and some high-stakes English tests, such as the Matriculation English Test (MET), College English Test (CET) and Graduate School Entrance English Exam (GSEEE), have long been under the spotlight, facing severe criticism at some point. This talk, following an outline of the major developments in both fields, highlights the ongoing reform in the test of English language proficiency in the broader context of the reform of National College Entrance Exam, the different voices from different people, and discusses the possible impact of this reform on English language teaching and testing in China.

Bio

Prof He’s main research interests are language teaching, language testing and discourse analysis. She was a senior visiting scholar at University of California at Los Angeles (UCLA) in 2004 and Benjamin Meaker Visiting Professor at the University of Bristol in 2014. She is the Deputy Chair of the Advisory Board of Foreign Language Teaching and Learning in Higher Education Institutions in China and is a member of the Language Assessment Quarterly Editorial Advisory Board. She has directed more than 10 large-scale research projects on language tests, published widely in applied linguistics and language testing, including 28 English textbooks which are used nationwide in Chinese universities, a monograph on cognitive computer adaptive language tests (2004) and a number of journal articles on language teaching and assessment.

Saturday 22 November 2014

Paper 1- The potential uses and limitations of eye-tracking technology in research into language testing

Stephan Bax, University of Bedfordshire

Recent research in the field of language testing has successfully made use of innovative eye tracking technology, in conjunction with other tools, to investigate aspects of cognitive processing in reading tests (e.g. Bax 2013). However, given that the use of this technology is still in its infancy in our profession, it is timely now to explore possible ways in which it could potentially provide better insights into elements of language tests themselves, and of candidates' behaviour during those tests. Like any new technology, however, there are inevitable limitations and pitfalls which also need to be taken into account. In this context, the talk will show samples of eye tracking videos and graphics from recent research projects, as the basis for discussion of how we could potentially extend the use of this developing technology in order to research not only reading tests but also tests of other skills. It will address some of the problems of using eye tracking in general, and issues which arise when we seek to use eye tracking for researching different skills, as well as ways of overcoming them. Finally, it will set out a number of possible future avenues which might usefully be explored using eye tracking, in conjunction with other tools, to the potential benefit of test designers and also of language teachers and learners more generally.

Paper 2- Designing low-stakes English language tests by teachers: Using multi-disciplinary knowledge and skills in assessment designing process

Lin Fang, University of Bristol

Low-stakes summative assessments are used as an essential tool for monitoring students' learning progress, motivating students and helping teachers to make decisions. How teachers from different educational and professional background design the low-stakes summative assessments and what knowledge and skills applied in this process are under-researched. In the small number of existing research, teachers' test-design processes are often investigated with reference to the procedures for developing high-stakes assessments in the sense that teachers are normally required to follow the 'standard' test designing procedure as the high-stakes assessments. However, test-developers (i.e., teachers) for low-stakes assessment are not necessarily professionally trained as item writers and they may fail to follow the standard procedures as for high-stakes assessment, due to a number of reasons. In this proposed case study research, I explored the test development process of a mid-term English exam in a Chinese university where half of teachers were not Chinese. The coordinator, who was an experienced foreign teacher from America, was trying to innovate 'direct items' into the test paper to improve the test validity and reliability. In this case study, I recruited all teachers who were involved in the test designing process as participants. Test designing process was observed by equipping softwares on participants' computers, which captured images of the computer screen as the video. Apart from data on observation, there were other major data that included curriculum documents, semi-structured interviews with each participant after test design, participants' journal study and focus group among participants after test administration. This case study aimed to examine 1) how participants cooperated with each other to design each test item, produce and finalize the test; 2) how they use multi-disciplinary knowledge and skills, knowledge on language assessment, applied linguistics and education, to generate test item. Preliminary analysis on the data indicated that although teachers were provided with test specification, their interpretations on the document varied. While teachers designed test items, their language assessment literacy were intertwined with their pedagogical knowledge, knowledge on second language acquisition, understanding on test-takers, perspectives on local context, interpretations on curriculum, syllabus, teaching materials and perceptions on their identities.

Paper 3- Reconsidering the development of pragmatics tests: The case of the discourse completion test

Afef Labben, University of Tunis; Moez Athimni, University of Carthage

A survey of the field of Interlanguage Pragmatics (ILP) shows that the Discourse Completion Test (DCT), also referred to as 'production questionnaire', has been the most frequently used test to investigate second/foreign language learners' ability to perform speech acts in a target language despite the harsh criticism leveled against its low construct validity and its failure to represent the features of authentic discourse. Interestingly, focusing on the statement of objectives of a number of ILP studies using DCTs, one can notice that such studies rarely refer to the DCT as a test. In addition, an overview of the DCT design process as described in several ILP studies shows that ever since its first use as a measure of socio-cultural competence (Blum-kulka, 1982), there has been a tendency to use or adapt one of the existing DCT versions used in previous studies based on the argument of comparibility of results. While a number of ILP researchers tried to strengthen the design of the DCT by the inclusion of rejoinders or by enhancing the prompt material (e.g. Billmeyer and Varghese, 2000) little attempts have been made to reconsider the DCT development process. McNamara and Roever (2006: 253) urge for the need for "more research on testing of sociopragmatic knowledge and design of discourse completion tests for testing purposes." With reference to research in the fields of language testing and psychometrics, the present paper shows that, whether used for research or instructional purposes, the DCT shares several qualities with language tests. As such, it is argued that the DCT should be treated as a language test and not as a questionnaire and should, thus, undergo a rigorous developmental process. Based on recent models of language test construction, the paper provides an overview of the stages of DCT development with a special focus on the development of DCT specifications.

Paper 4- Computer technology and language testing

John H.A.L. De Jong, Pearson

Computer technology was introduced into testing in the 1930s for the purpose of automated scoring. The history of computer-based test administration began much later, in the early 1970s, mainly in the military and in clinical testing, triggering research on adaptive testing and innovative item types. By the mid-1980s the College Board introduced ACCUPLACER and in the 1990s high stakes educational achievement and language tests started to appear. At a conference in Athens in 1995 ETS expressed the expectation that by 1999 CBT testing would be available in 170 countries. The rapidly expanding use of Computer-Based Testing (CBT) since 2000 has brought with it an increasing interest in the potential of computer-based systems to provide improvements in areas other than test administration. One such area involves the development of innovative item types which incorporate features and functions that are not possible with conventional test administration methods. Technology now allows us to incorporate video, sound, increased interactivity and simulation into the items that are developed for examinations. The use of innovative items holds out the promise of allowing us to improve our measurement of the skills that our examinations are attempting to tap into. Innovative item types allow us to measure the same things better than we could previously, or allow us to measure something more, or different, than we were previously able to. This factor is becoming increasingly important as technological advancements lead to ever more complex working environments. Innovative item types hold out the promise of allowing us to, more directly measure cognitive and behavioural skills that are vitally important for the tasks we are asking people to carry out. In this contribution I will discuss the use of automated scoring of speaking and writing as well as the use of virtual peers to address performance skills.

Paper 5- Introducing opportunities for learning-oriented assessment to large-scale speaking tests

Anthony Green, Liz Hamp-Lyons, University of Bedfordshire

Grounded in recent developments in formative assessment and in broader concepts of learningoriented assessment [LOA] (Carless 2005, 2007; Carless, Joughin, Liu et al. 2006; Joughin 2009), we describe a model of learning-oriented language assessment (LOLA) which we have applied to the speaking component of a large-scale international examination. We describe our findings on how official teacher support materials for this speaking test may, or may not encourage the teachers to put into practice aspects of LOLA through formative or dynamic assessment practices during preparation for the test. We have identified some opportunities for learning oriented practices to be strengthened in the available resources for teachers, and for the strategies implied by the identification of these opportunities to also be promoted in test preparation. The study suggests that putting LOLA into practice brings the personas of assessor and language teacher closer together, giving assessors the potential to implement teacherly skills, but that for this potential to be fully realised, some changes would need to be made to the current speaking test structure, the preparation materials, and the constraints imposed on interlocutors. Our study has shown that if effective LOLA is to occur in preparation for large-scale speaking tests, further development of the support materials for teachers and learners planning to take this test is essential so that learners may take full advantage of opportunities for learning. With this in mind, training programmes are recommended for materials developers and teacher trainers. These should be combined with ongoing observations of teachers in action to identify opportunites to improve on current practice.

Paper 6- C-tests outperform Yes/No vocabulary size tests as predictors of receptive language skills

Claudia Harsch, University of Warwick; Johannes Hartig, German Institute for International Pedagogical Research, Frankfurt/Main, Germany

Placement and screening tests serve important functions, not only with regard to placing learners at appropriate levels of language courses but also with a view to maximizing the effectiveness of administrating test batteries. With the advent of computer-administered tests (CAT), adaptive testing becomes more and more attractive (e.g. Frey/Seitz, 2009). Here, reliable pre-tests become a prerequisite to optimize the CAT procedure. The purposes of placement, screening or pre-testing require a reliable format which is simple and quick in administration and scoring and poses little demand on the test takers while covering as many items as possible. We examined two widely reported formats suitable for these purposes, the discrete decontextualized Yes/No vocabulary test (Alderson/Huhta, 2005; Meara, 2005) and the embedded contextualized C-test format (e.g. Eckes/Grotjahn, 2006), in order to determine which format can explain more variance in measures of listening and reading comprehension. Our data stem from a large-scale assessment with over 3000 students in the German secondary educational context; the four measures relevant to our study were administered to a subsample of 559 students. Using regression analysis on observed scores and SEM on a latent level (Kunnan, 1998), we found that the C-test outperforms the Yes-No format in both methodological approaches. The contextualized nature of the C-test seems to be able to explain large amounts of variance in measures of receptive language skills. The C-test, being a reliable, economical and robust measure, appears to be an ideal candidate for placement and screening purposes. Moreover, positive washback is to be expected since solving C-test items draws on a range of processes and skills, all of which are needed to process language in context (e.g. Sigott, 2005; Qian, 2008). In a side-line of our study, we also explored different scoring approaches for the Yes-No format (Huibregtse/Admiraal/Meara, 2002). We found that using the hit rate and the false-alarm rate as two separate indicators yielded the most reliable results. As an additional benefit, these two indicators can be interpreted as measures for vocabulary breadth and as guessing factor respectively, thus giving substantive feedback to learners and teachers and they allow controlling for guessing.

Paper 7- Interfaces between corpus linguistics and language testing/assessment

Yu-hua Chen, University of Nottingham Ningbo

With the advance of computer technology, Corpus Linguistics (CL) has progressed significantly with a wide range of applications reported in various branches of Applied Linguistics, and yet how it can contribute to Language Testing and Assessment (LTA) has received comparatively limited attention in the literature. Starting from test design, writing individual test items, to test validation and aligning to a framework of reference, corpora and corpus approaches can be used to inform various stages of test development. At the initial stage of test design, for example, specialised or large general corpora can provide actual evidence of natural occurring language use in various discourses, which can form a solid foundation for test developers to identify the appropriate context for test users and test purposes. Native corpora such as the British National Corpus (BNC) or the Contemporary Corpus of American English (COCA) also provide reliable sources of authentic lexical or grammatical concordance data, which allows test writers to determine whether the vocabulary or structure tested in a given context is suitable or not. In terms of non-native corpora, when developing a rating scale, exam boards can make use of learner corpora as the empirical underpinnings to reveal what learners typically can do and cannot do as opposed to relying on practitioners' judgment alone. This paper will not only provide an overview of corpus-informed and corpus-based applications in LTA but also highlight the key areas which deserve more focused research at the intersection of CL and LTA.

Paper 8- Exploring corpus analyses to inform writing assessment: A pilot study

Franz Holzknecht, University of Innsbruck

As part of a pilot study the research in this paper exemplifies how the analysis of learner corpus data can inform writing assessment. The study investigates second language learners' "functional competence" as described in the CEFR (p. 125 ff.), the assessment of which poses certain challenges to language testers. Functional competence is "concerned with the use of spoken discourse and written texts in communication for particular functional purposes" (CEFR, p. 125). For the assessment of writing, "macrofunctions" are particularly important. These are defined as "categories for the functional use of […] written text consisting of a (sometimes extended) sequence of sentences" (CEFR, p. 126). However, the framework does not include an extended enumeration of macrofunctions, but only an unfinished list. In addition to the lack of specificity when it comes to translating functional competence into levels on the illustrative scales, this leaves language testers somewhat in the dark regarding the assessment of these features. To investigate if learner corpus data can be informative in this respect, 835 test takers' writing samples of two languages (Italian and English) and two CEFR levels (B1 and B2) are analysed. All writing samples are based on standardized writing tasks developed for a national high-stakes exam. The tasks specifically target the macrofunctions listed in the CEFR. Analytical software tools such as Antconc are used to answer the following research questions: Does the inclusion of macrofunctions in writing prompts mean that test takers actually perform these functions? Which macrofunctions are test takers capable of performing at different CEFR levels and in different languages? The results of this pilot study show that learner corpus analyses offer a different perspective for language testers and can inform test design and validation.

Paper 9- What can expert judgement teach us about using the CEFR for young learner test development?

Amy Malloy, Oxford University Press

The political and institutional rise of the Common European Framework of Reference (CEFR) has led to its widespread use within both language learning and testing (Lim & Khalifa, 2013). Whilst it has contributed greatly to international literacy around language learning, its suitability for language test development (in particular for young learners) has been questioned (Hasselgreen, 2005). The young learner construct lacks accurate representation, in particular the listening scales (Field, 2013), risking the fairness of score interpretations on tests aligned with it for this purpose (Bachman, 1990). Within test development, the role of the expert judge has been deprioritised in recent years in favour of statistically-based alignment to the CEFR scales. This mixed-methods study reprioritizes it within young learner test development in the form of the young learner teacher. It investigates the area of discrepancy between professional judgement of young learner listening item difficulty and benchmarking items to the CEFR, by quantitatively identifying consistency among young learner teachers, and qualitatively examining their decision-making processes using concurrent reporting. The presenter will show that a limited sample demonstrated professional judgement to be a more accurate measure of young learner listening item difficulty than referring solely to the CEFR descriptors, with a consideration of the child's cognitive development and conceptual understanding being a more reliable factor to refer to than a lexical or grammatical syllabus. Cultural background was also shown to be an influencing factor of professional judgement. The findings also suggest the CEFR to be in need of adaptation for use with young learners, including a need for sub-levels within the bands, an age-related pathway to reflect cognitive development, and age-appropriate linguistic markers for each level.

Paper 10- Virtual face-to-face speaking tests using web-based video conferencing technology

Fumiyo Nakatsuhara, Chihiro Inoue , University of Bedfordshire; Vivien Berry , British Council; Evelina Galaczi, Cambridge English Language Assessment

Applied Linguistics, SLA and language testing research have a reciprocal relationship. Issues proposed by Applied Linguists and SLA researchers have been investigated in testing contexts, and the results have fed back into Applied Linguistics and SLA research. One such area is the relationship between the nature of learner language and the conditions under which the language is elicited. This issue is especially salient in the assessment of speaking, since a change in exam conditions through the use of, for example, technology, could fundamentally impact on the underlying construct. This presentation reports on a preliminary exploration and comparison of test-taker and examiner language and behaviour across two different delivery conditions for the same L2 speaking test: a face-to-face test administration, and a test administration using web-based video conferencing technology. The study examined (i)testtakers' linguistic output and scores on the two modes and their perceptions, and (ii)examiners' test management and rating behaviours across the two modes, including their perceptions of the two conditions for delivering and rating the speaking test. 32 testtakers and four trained examiners participated in the study. The test-takers took two versions of the same speaking test under face-to-face and computer-delivery conditions. A convergent parallel mixed methods design was employed to allow for more in-depth and comprehensive findings from multiple perspectives. The study included feedback interviews with test-takers, their linguistic output during the tests (especially types of language functions) and scores awarded under the two conditions. Examiners provided written comments justifying their scores, completed a feedback questionnaire and participated in retrospective verbal report sessions to elaborate on their interlocuting and rating behaviour. All test sessions were observed by three researchers and field notes were taken. The findings suggested that while the two modes generated similar test scores, there were differences in test-takers' functional output and examiners' interviewing and rating behaviours. In particular, some interactional language functions were elicited differently in the two modes, and the examiners used different turn-taking techniques in the two conditions. This presentation will discuss the importance of test administration conditions to measure the construct defined in the tests, and offer recommendations for further research.

Paper11- The planning strategies of young teenagers in a speaking task: Cultural trends

Gwendydd Caudwell, British Council

In assessing the validity of tests of spoken English, distinctions are made between spontaneous or interactional speaking skills and more structured, extended turns that reflect aspects of spoken English use. As is the case with previous studies (e.g. Foster & Skehan 1996; Ortega, 1999; Wigglesworth, 1997) this study focuses on planning in what Weir (2005) refers to as an extended informational routine. The specific focus here is on planning for such a task in a large-scale, computer-based speaking test. In addition, since we are interested in exploring how young teenagers

(13-15 year olds) use the time allowed for planning in such a test, the task in question has been designed specifically with this agegroup in mind. Specifically, planning strategies used by candidates from at least two different cultural backgrounds, from regions such as South Asia, the Middle East or Europe are examined via data collected through a questionnaire administered during the test event. The questionnaire has been developed based on findings in the literature and on advice from individuals with expertise in the teaching of this age-group. This presentation reports on the results of the data collected and explores the relationship between planning strategies employed and whether these strategies might have had an impact on the performance being measured. It also examines whether there are tendencies towards particular strategies in the different culture groups included in the study. Implications, limitations and further research steps to be taken are also discussed.

Sunday 23 November 2014

Paper 12- Feedback as first principle

Liz Hamp-Lyons, University of Bedfordshire

Although the concept of ‘feedback’ has been known in second language teaching for many years, it also occurs in many disciplines. Recently, however, the concept of feedback has been expanded and elaborated substantially in second language acquisition, aligning with a similar trend in fields such as neuroscience, environmental science, human resource management, nursing, electrical engineering biology and other disciplines. In this presentation I discuss how ‘feedback’ has become a key area of research and practice in language teaching, learning and assessment. In particular I will link together the growing understanding of the role of feedback with two current issues in writing (and speaking) assessment: the use of rubrics, and the role of automated essay scoring/automated writing evaluation.

Paper 13- Investigating washback: A study of an English speaking component in the French Baccalauréat

Gemma Bellhouse, University of Oxford

The Baccaleuréat has long been considered more than just a test, but also a social institution of governmental legislation effecting enormous ramifications on teaching methods (Colin, 1900). Within the domain of speaking assessment, this present study explores the effects of a new English speaking test component in the Baccaleuréat and particularly its influence on language learner strategies (LLS) both within and outside of the French secondary school classroom context. This study also analyses consequences of the test alignment with the CEFR B2 level. The sampling frame for the principal participants includes 10 English teachers and 10 classes of 211 students from 3 secondary schools in southwest France. The research design is a mixed methods framework including instruments of semistructured and nonstructured interviews, questionnaires, follow-up questions via email, an online survey, and document analyses. The quantitative data is investigated with descriptive and exploratory statistics, reliability coefficients, and multiple regression. Qualitative data was analysed and coded using constant comparative analysis. The intentions of this test were to implement the communicative approach. Positive washback and consequential validity are observed in more encouragement of speaking by teachers and learners becoming more aware of their deficits in communicative abilities. The open-ended test design allows students of all proficiencies to interpret authentic English materials, another characteristic of communicative methods. However, as the test prompts are publicly known, students use strategies of memorisation, a sign of negative washback. This study provides evidence that the sought B2 level may not be accurate due to a lack of teacher training and a generous rating scale. 66% of the students believe in the importance of the test, but the weight of the speaking test is so low that high-proficiency students have relaxed their studying and low-proficiency students may give up to focus on other, more important subjects. An average 20% increase in strategy use demonstrates evidence of positive washback, but the group and individual student differences vary greatly. Extrinsic motivation, anxiety, and proficiency account for 16% of the high variance of the effects on the student LLS. The relationships of the variables are supported by the qualitative data.

Paper 14- An investigation of students' writing performance in the Test of English for Academic Purposes (TEAP)

Mikako Nishikawa, University of Bristol and Eiken Foundation of Japan; M. Honma, K. Nakamura, T. Matsudaira, S. Shiozaki, K. Yanase, Eiken Foundation of Japan

This study offers a preliminary review of a new test called the Test of English for Academic Purposes (TEAP), which was developed by the Eiken Foundation of Japan in a collaborative effort with Sophia University in Japan. TEAP evaluates the preparedness of Japanese high school students in academic English for college entrance. While the results for entrance exams in Japan are traditionally reported as pass or fail, TEAP gives a score report with meaningful feedback about the test takers' academic readiness in four language skills. Furthermore, TEAP follows the guidelines set by the Ministry of Education, Culture and Sports, Science and Technology (MEXT) to test a range of contents that are in line with the national course of study for Japanese high school students. The targeted English proficiency levels for TEAP are from A2 to C1 in the Common European Framework of Reference (CEFR). TEAP became available to the public in 2014 after impact studies were thoroughly conducted by the Center for Research in English Language Learning and Assessment (CRELLA) at the University of Bedfordshire in the United Kingdom. There are two sections to the writing component of the TEAP: Task A is a short summarization of a single text and Task B requires the abilities to synthesize information from multiple sources such as graphs and articles. In 2011, 112 high school seniors participated in the second pilot study for the writing test. According to the previous report by Weir (2014), student performance varied mostly in "coherence and cohesion" compared to other criteria such as "main idea", "lexical range and accuracy", and "grammatical range and accuracy." Majority of the test takers were at level A1 or A2 (11% at A1 level, 52% at A2 level) while only 37%of the test takers achieved B1 or above (1% at B1 level, 36% B2 level.) This presentation offers a detailed text analysis of the Task A essays graded as A2 and B1 to illustrate the strong correlation found (R2>0.76) between "coherence and cohesion" and "lexical range and accuracy" in determining their levels.

Paper 15- Looking into reading: The use of eye tracking to investigate test-takers' cognitive processing

Tineke Brunfaut, Gareth McCray, Lancaster University

In this paper, we will report on a study exploring the use of eye tracking to gain insights into test-takers' cognitive processing whilst completing reading tests. Only recently, language testing researchers have started experimenting with eye-tracking technology as a tool for item validation, with promising initial findings (see e.g., Bax, 2013; Bax & Weir, 2012; Gorin, 2006; McCray, 2013; McCray, Alderson & Brunfaut, 2012). However, to our knowledge, no guidelines exist on how to apply the technology to best fulfil this function. In methodological experimentation with eye-tracking technology to look into reading test items, McCray, Alderson & Brunfaut (2012) found that a combination of retrospective interviews with eye- tracking traces providing the stimulus proved to generate rich data on test-takers' cognitive processes. The study reported on in this presentation therefore aimed to further explore the utilisation of eye tracking, and also the synergy with stimulated recalls, to look into cognitive processing. To this end, 25 English second language speakers completed two versions of the Aptis reading component whilst their eye movements were being recorded. Each reading task was immediately followed by a stimulated recall. The innovative methodology – including a detailed analysis of eye movement metrics and of the stimulated recalls test-takers produced– proved to be particularly useful. Eye-tracking visualisations revealed several overall processing patterns and tendencies, which were triangulated by the stimulated recall findings. Although both the eye tracking and stimulated recall analyses led to data on both lower- and higher-level processing, the eye movement analyses allowed for more insights into lower- level reading processes, whereas the stimulated recall data were highly useful in revealing more higher-level reading processes. In addition, the findings resulting from the two methods mutually confirmed the other results, thus providing a solid basis on which to draw conclusions on test-takers' cognitive processes. This methodology, although quite labour intensive, was found to reveal vital information on what processes underlie correct item responses, and thus to be extremely valuable as part of test validation research. The specific methodological design and analyses used in the study will be elaborated upon in this talk.

Paper 16- SLA theories on developmental sequences and the assessment of L2 writing

Christian Krekeler, Konstanz University of Applied Sciences

Can SLA research on developmental levels assist in the assessment of L2 writing? It has been tentatively suggested that, in addition to grammatical accuracy, language testers should take account of interlanguage norms and developmental levels (Ellis 2001). It is hoped that an acquisitional perspective might increase the validity and the fairness of language assessment. However, a strong case has been made against the use of developmental levels as a basis for scoring assessment and especially as a basis for decision-making (Purpura 2005). It is argued that knowledge of developmental levels is still limited to a narrow range of morphosyntactical features and that it would be too complex to determine developmental levels with sufficient confidence. This presentation aims to contribute to the ongoing debate by applying this approach to L2 writing assessment. The presentation consists of three parts: (1) I will briefly describe SLA theories on developmental levels and mainly refer to Pienemann's processability theory (PT). PT is an SLA theory of language processing which is based on the assumption that the acquisition of language procedural skills follows predetermined sequences (Pienemann 2011). (2) Examples of grammatical accuracy and developmental levels in written texts will be given to illustrate possible scoring procedures. The examples also highlight that, if developmental levels were to be used for the scoring of L2 writing, it would be fraught with difficulties. (3) Possible uses of developmental levels for the assessment of L2 writing and their limitations will be discussed. It will be asked whether "developmental scores" can complement accuracy scores. I will suggest that an acquisitional perspective may be useful for specific types of assessment, such as diagnostic assessment, norm-referenced assessment and informal classroom assessment.

ABSTRACTS OF POSTERPRESENTATIONS

Saturday 22 November 2014

Poster 1- A communicative approach to Arabic language proficiency testing in the light of Diglossia

Rahaf Alabar, Goldsmiths, University of London

Despite the vast research by linguists on language proficiency testing, very little is written about Arabic proficiency tests. The overall account, which emerges from the literature, does not touch on how to assess Arabic learners' linguistic competence. Instead, it identifies, to some extent, the standards of proficiency as a basis for any potential construction or development of an Arabic proficiency test. Those standards are basically identified from the framework developed by the American Council on the Teaching of Foreign Languages (ACTFL). However, those standards are based on personal unshared principles of language experts, and can be debated in terms of how the issue of diglossia and its direct relation to the concept of "Arabic for communication" is perceived. My research has been inspired by Cummis statement that says, "many linguists […] have argued that language proficiency can be validly assessed only in naturally-occurring communicative contexts" (in press, 26-27). This statement squares with the idea of language for communication in daily life, which highlights the importance of addressing the bias towards teaching and testing the grammatical forms of language at the expense of the skills' knowledge in using language for natural purposes in realistic situations. In the Arabic context, this implies the need to give attention to teaching and testing the spoken form of Arabic, as it is the core aspect of communication in terms of the two language skills; listening and speaking. Therefore, the main purpose of my research is to determine the extent to which ACTFL can be efficient as a language scale for measuring non-native speakers' Arabic communicative proficiency taking into consideration the interactional competence as a norm of analysis. For this purpose, teachers of Arabic, assessors and learners will be interviewed to contribute to the analysis process. Thereafter, necessary changes and improvements will be made on the scale to facilitate any future attempts to accomplish a set of tests and scoring scales and procedures. The last step will be evaluating an existing proficiency test against the newly constructed scale in order to test its efficiency and practicability. My methodology relies on an analysis and a critique from different perspectives; considering all test users who can possibly be involved in constructing the test, validating or taking it. The process aims to create a critical account using an analytical method. The importance of my research lies in the contribution I hope it makes in the field of Arabic language proficiency testing especially that this domain in particular has not been highlighted in the Arab world. The outcome expected by my research is to pave the way for any attempts to construct standardised tests of Arabic that meet the growing demands for accurate information concerning Arabic language learners' proficiency levels.

Poster 2- The revitalisation of formative assessment for developing academic writing and enhanced practices

Ahmed Al Khateeb, University of Southampton

This presentation will investigate the effect of introducing formative assessment in a hybrid context for language learning and particularly for online-enhanced writing. That is because such assessment places the emphasis on monitoring writers' progress, identifying their strengths and weaknesses and how their shortcomings can be improved. Formative assessment during the writing process helps to change common negative attitudes to academic writing. It makes it a practice of ongoing development that involves a series of cyclical strategies in order to empower writers by allowing them to improve their writing skills, self-confidence and express themselves more effectively so that they have the ability to describe precisely what they feel. Accordingly, this presentation will reveal how the new 'unassessed' testing approach to writing construct, which is based on formative assessment, has positively contributed to the development of learners' written practices and performance in EFL writing and promoted this skill as being more than simply a method of determining achievement in a short point of time. Furthermore, with the recent development of social platforms such as wikis, learners have been provided with tools which aid them to increase their achievement and communicative competency, as a result of assistance from peers in an informal way. Based on the research findings, and in light of the conclusions which were drawn at the end of the course, the presenter will highlight the significance of formative assessment and how it is necessary that it be officially adopted by teachers and teacher trainees. He will also discuss the key implications of this assessment approach, which would provide it with an equal status to other assessment approaches such as interim or end of year assessment.

Poster 3- Analysing academic listening needs from a cognitive perspective

Sahar Alkhelaiwi, Lancaster University

Determining and describing language-learning needs is the first step in the construction of language curricula (Graves, 2000). Language for Specific Purpose courses (including teaching and assessment materials) should especially be derived from an analysis of the target language use situation so that L2 learners can be prepared for the situation in which they will eventually function (Hyland, 2009). Listening is a vital form of input for acquiring knowledge in an academic milieu (Jordan, 1997) – also so for English Language and Literature undergraduates at a Saudi Arabian university. This study concerns a needs analysis to investigate the academic listening 'target, present and learning' needs of these students, so that informed decisions can be made concerning the development of listening comprehension materials for use in an English for Specific Academic Purposes (ESAP) course for the target population. It is hoped that this will constitute an improved design for EAP needs analysis frameworks directed at developing listening materials for university level. To inform the development of appropriate academic listening materials for the selected population, a five-phase needs analysis approach has been adopted. The first phase – focussed on in this poster – involved 'a spoken target discourse analysis' of five Linguistics and Literature lectures. The cognitive listening processes likely to be used by an 'idealised' lecture attendee were explored by means of an overarching model posited by Aryadoust, Goh and Kim (2012). To this framework, other specific listening models were added to characterise the integrated manner in which academic listening functions: Field's (2013) listening cognitive processing, Young's (1994) lecture structure, and Flowerdew and Miller's (2005) contextualised-listening. Although many listening processes are so automatic that they are not observable as they occur, the target discourse analysis revealed several types of cognitive listening processes that students are likely to employ and need to comprehend lectures. Overall, those listening to the selected lectures were likely to have 'encountered' a total of 37 cognitive processes, which can be categorised into 1) cognitive listening processes, 2) lecture structure, and 3) relating input to other materials. The poster presentation will provide more details on the first phase's data collection and analysis methods, and findings.

Poster 4- The relationship between teachers' L1 Arabic varieties and students' L2 English pronunciation

Wafa Alotaibi, University of Southampton

The study investigated the factors that affect Saudi students' L2 English pronunciation, their preferences and dislikes towards non-native Arabic accented teachers of English. In the course of the investigation, a detailed comparative analysis was made of the Arabic and English phonetic systems, different theoretical frameworks were explored including transfer and markedness theories, and the Contrastive Analysis Hypothesis, and as well as identifying a range of factors that affect pronunciation, studies were also examined that identified the typical pronunciation difficulties faced by Arab learners of English. The purpose of the study was to establish where there is any significant phonological influence of Arabic teachers' language variety on students' L2 English pronunciation of segmental consonants. Data was collected by interviewing teachers, conducting a survey among students, doing classroom observation, and making recordings of teachers' and students' pronunciations. The typical demographic profile of the survey respondent was an 18 year old female fresher student having been learning English for 12 years at a governmental school in the Saudi education system. The most preferred accents were Saudi then Jordanian whereas Sudanese and Yemeni were least preferred, and the main reason in both cases related to the clarity of the accent.

Use of or exposure to English was largely restricted to either the school or work environment, or to watching movies. The initial result of the data analysis reveals that there is an influence from teachers' own Arabic dialects on their students' English pronunciation in terms of new vocabularies while the is no significant effect on familiar words.

Poster 5- Effect of Language Learning Strategies (LLS) instruction on Saudi EFL students' strategy use and proficiency

Ibrahim Alzahrani, University of Southampton

Scholars in the field of Language Learning Strategies (LLS) are constantly calling for more research on the effect of LLS. The present study plans to check the effect of LLS instruction on the proficiency and strategy use among Saudi EFL college students. Researchers in the area of LLS provided similar frameworks for strategy instruction. In the present study, strategy instruction is delivered through Styles- and Strategies- Based Instruction framework which was built by Weaver and Cohen in 2006. Integrated in regular language classrooms, LLS are going to be explicitly taught using this framework through its five steps: 1) Strategy Preparation: where teachers should find out how much their students know consciously about LLS and if they are able to use them; 2) Strategy Awareness-Raising: where students should be more aware of the process of learning, their Learning Style Preferences (LSP), the LLS they already use and others suggested by their teachers or classmates, and their responsibilities as language learners; 3) Strategy Instruction: where teachers should describe, model, and give examples of LLS; 4) Strategy Practice: where students should try LLS out; and 5) Personalization of Strategies: where students are advised to evaluate LLS they are using and if they can use them in other contexts. For the purpose of strategy instruction, four Meta-cognitive Strategies: Organize/Plan learning, Manage learning, Monitor learning, Evaluate learning, in addition to 16 Taskbased strategies underlie four categories: Using what you know, Using imagination, Using organizational skills, and Using a variety of resources are going to be explicitly taught to students in the experimental groups. Prior to strategy instruction, students' Learning Style Preferences (LSP), are going to be measured through a style survey where students should know their preferred LSP. Also, students' strategy use is going to be measured through a strategy use survey as a kind of raising their awareness. Moreover, pre- and post-proficiency tests are going to measure students' performance before and after strategy instruction. At the end of the intervention, some students and teachers are going to be interviewed to get their feedback on strategy instruction.

Poster 6- Investigating assessment literacy in Tunisia: The case of EFL advanced reading teachers

Moez Athimni, University of Carthage

In the last decades, the area of language testing has witnessed a movement from traditional testing which focuses on the assessment of learners' general knowledge of the language to what has been referred to as 'alternative' (e.g. Herman et al., 1992), 'authentic' (e.g. Bachman and Palmer, 1996; Newman et al., 1998) or 'performance' assessment (e.g. Solano-Flores and Shavelson, 1997). However, the extent to which such theoretical developments have informed testing practicum is still debatable. Several research studies (e. g. Zhang & Burry-Stock, 1997; Mertler, 2003; Galluzzo, 2005) have characterized the way teachers evaluate the performance of their students as "incongruent with the recommended best practices" (Volante & Fazio, 2007:2). Popham (2009:5) argued that "many of today's teachers know little about educational assessment." In response to such evidence, a number of researchers (e.g. Stoynoff & Chapelle, 2005; Inbar-Lourie, 2008; Malone, 2008) have coined the term 'assessment literacy' to refer to what teachers and instructors need to know about assessment (Taylor, 2009). Today, assessment literacy is widely recognized as an essential component of teacher training. The present paper explores Tunisian university EFL advanced reading teachers assessment literacy by investigating the way they construct their reading tests. Data were collected by means of a structured interview administered to third year EFL reading teachers from different higher education institutions in Tunisia. Results show that teachers did not receive any training in language testing neither as part of their university courses nor during their professional development programs. Results also show that most reading teachers had no guidelines or specifications for developing or selecting test items or tasks. In most cases, they relied on their intuitive skills or used old test versions as templates to design and construct their own tests. Such findings urge for the need to promote assessment literacy in the pre-service and in-service EFL teacher training programs.

Poster 7- Bridging the gap: The effectiveness of short-term intensive IELTS writing preparation in Japan and 'relearning' academic conventions

Tony Clark, University of Bristol

Intensive IELTS preparation courses are becoming more and more popular, as prospective students - or workers - aim to reach a required band score within a specified (and often short) time frame. Having taught intensive IELTS preparation courses for some time, it has become clear to me that certain aspects of the preparation process require investigation; none more so than the transition from culture-specific academic norms to fit with British or international expectations. Stakes are high, and Japanese learners will have to quickly absorb writing conventions required in the IELTS exam (e.g. structuring paragraphs, organising an argument, taking a stance and using topic sentences). Without these, they stand little chance of achieving the score they need. As a result of such difficulties, writing instruction is a significant part of the course. The main aim of this study is to increase our understanding of the practicalities regarding written aspects of exam preparation, especially over a short-term period.

Poster 8- Investigating the relationship of word frequency and learner proficiency in an English proficiency test

Yu-hua Chen, University of Nottingham Ningbo; Shaida Mohammadi, Pearson

Over the past few decades, it has been increasingly recognised that corpora can contribute to the development and validation of language testing and assessment. However, little has been reported in the literature regarding the extent to which the word frequency information from native corpora interacts with candidates' performance in relation to their proficiency. This study aims to address this issue by comparing candidate performance in one English proficiency test and the frequency bands of words tested in the items. The frequency lists come from a new General Service List (Brown, 2014) and the 5,000 most frequently used words in American English (Davies and Gardner, 2010), both of which were compiled from native corpora, the former the Cambridge English Corpus (CLC) and the latter the Contemporary Corpus of American English (COCA). A number of item types are selected for investigation as they are expected to assess different abilities such as integrated or productive vocabulary skills. The results will shed light on the relationship between L1 word frequency and L2 proficiency, and the findings can also help test developers or item writers, for example, to determine 'whether an item should be tested, given its frequency or authenticity' for a particular level or task type (Barker, 2006).

Poster 9- How well do different task formats elicit L2 learners' pragmatic competence?

Edit Ficzere, University of Bedfordshire

The importance of testing L2 learners' pragmatic knowledge is becoming evident as a result of increasing research. Bachman (1990) identifies it as one of the major components of language competence, which combines social and linguistic knowledge. Pragmatic competence should, thus, be included when evaluating L2 learners' communicative competence. Current pragmatic tests are mainly based on the Cross-Cultural Speech Act Realization Project (Blum-Kulka et. al., 1989), which examined speech act realization from a cross-cultural perspective. Using Speech Act Theory as a theoretical framework, however, has been criticized lately for overlooking the importance of the discursive side of pragmatics. The main objective of this research is to investigate an effective approach to assessing L2 learners' pragmatic competence in speaking. It focuses on examining this knowledge in extended discourse at CEFR B2-C2 levels for the reason that increasing language ability might free up more cognitive capacity allowing learners to attend to pragmatic features. The study aims to identify criterial features defining the level of L2 pragmatic competence and examines the extent to which different speaking task formats allow test takers to display their pragmatic competence. It also investigates how the identified criterial features can be operationalised in rating scales, while considering features which are salient to raters when awarding scores for pragmatic competence. This poster reports on a preliminary study for the first phase of the bigger research project. The first phase focuses on identifying criterial features and on how well different task types may differentiate learners at B2-C2 levels in terms of pragmatic competence. Ten university students took part in the pre-pilot study, which included a written DCT task (adapted from Roever, 2006) consisting of ten items and a selfdevised monologue task consisting of four items. The results indicate that the written DCT task did not distinguish different level learners as clearly as the monologue task. The various responses to the monologue tasks showed clear difference between levels in terms of the length of responses, higher level learners including more elaboration in formal situations, the range of linguistic devices used and the occurrence of routine formulas.

Poster 10- International and local raters: Comparing ratings and rationales on a speaking test across international contexts

Luke Harding, Lancaster University; Mark Griffiths, Trinity College London

In large-scale, international tests of oral proficiency some examination boards and testing organisations have opted to train and employ local raters whose first language is not English in international contexts, rather than recruiting English-speaking expatriates or deploying examiners from a central team based in an English-speaking country. This shift towards a local rater model raises questions of rater consistency, an issue which has begun to receive attention in the research literature (see Kim, 2009; Xi & Mollaun, 2011; Zhang & Elder, 2011; 2014). It also raises questions concerning the constructs which are being operationalized through raters' interpretations of criteria across contexts where World Englishes norms may apply, and where cultures of English language education and teacher training may differ. In such contexts, it is possible that there will be different interpretations of correct or incorrect forms in spoken English, different levels of "error" sensitivity, and different perspectives on what features of spoken communication are valued. It is therefore important to take account of what raters know about acceptable local norms and variations in order to inform revisions to speaking performance descriptors. Against this background, the current study posed two research questions: (1) Do raters from the UK, India and China assign the same grades to a set of spoken performances? (2) Do raters from the UK, India and China differ in their orientations towards aspects of speaking performance in assigning grades? If so, what is the nature of these differences? In order to address these questions, a mixed-methods study was designed. 24 trained raters (eight British; eight Indian; eight Chinese) rated the same 30 speaking performances in an online video rating task. These performances included 10 European candidates, 10 Indian candidates and 10 Chinese candidates drawn equally from A2 and B1 level exam performances. The online task required examiners to rate each candidate according to a four-point marking scale, and then to provide a retrospective written justification of their marking decision. This poster will provide details of the online rating task, and present some preliminary quantitative and qualitative analyses.

Poster11- Predictive validity of TOEFL iBT: Quantitative and qualitative perspectives

Claudia Harsch, Ema Ushioda, Christophe Ladroue, University of Warwick

The gate-keeping role of English language tests in the process of admitting international students to universities in the UK is undisputed. What is, however, not yet fully understood is to what amount these tests can predict preparedness for academic studies and academic success. The work-inprogress we report on is focusing on one such test, the TOEFL iBT®; the research is funded by the ETS TOEFL ® COE Research Programme (RFP 2012-21; duration 2013 – 2016). We are exploring both from a quantitative and qualitative perspective the relationship between students' linguistic preparedness, as expressed by students and their tutors in interviews, their exploitation of pre- and in-sessional language support classes, their TOEFL iBT test scores, and their academic success, as expressed in final academic grades. The outcomes aim at informing stakeholders such as university admissions of appropriate entrance levels with regards to TOEFL iBT score profiles and support placement decisions for pre- and in-sessional language support with regards to TOEFL iBT score profiles. We will present preliminary analyses and findings both for the qualitative interview data (students =25; tutors=31; academic year 2013/14) and the quantitative data sets on test scores and academic grades (n=400; academic years 2011/12 and 2012/13). For the latter, we are examining via correlational, regression and graphical analyses the explanatory power of the TOEFL iBT test scores on the dependent variables academic grades, pre-sessional test scores and in-sessional attendance. With regard to the qualitative data, we are exploring the relationship between the perceived linguistic preparedness, the exploitation of language support, and the TOEFL iBT test scores. Since we are still waiting for the final academic grades for this cohort of students (due by the end of 2014), we will use students' assignment grades during the academic year 2013/14 as preliminary indicators for academic success. Our study aims at filling a gap in investigating the predictive validity of TOEFL iBT with an under-researched qualitative perspective.

Poster 12- Language testing and vocabulary research – a symbiotic relationship?

Benjamin Kremmel, University of Nottingham

Vocabulary research has proven a prolific area of study over the last decades. Applied linguists interested in the learning and teaching of FL vocabulary have therefore developed measurement instruments in all shapes and sizes, which have then become employed by SLA researchers for a variety of research purposes. The problem, however, is that these assessment tools were predominantly developed by vocabulary researchers, people with relatively little concern for testing or general psychometric principles. Unfortunately, this shows in the scarcity of validation evidence available for vocabulary tests, even for those that have been in use for decades now. While an interdisciplinary collaboration between vocabulary researchers and language testers is desirable in principle, the potential for synergetic complementation between the two fields seems not yet fully realized and the integration of language testing expertise in vocabulary research appears overdue. This study presents an attempt to bridge the gap between the two research fields in the development of a new vocabulary test. In challenging existing assumptions and starting from empirically grounded rationales, it explores the usefulness of different item formats for tests of vocabulary size. As different test formats will provide different kinds of information for the test score users, an account for what any particular format and the scores it yields can and cannot tell about the lexical abilities of the test-takers seems necessary to determine the best-performing format(s) for use in a new vocabulary test. This paper will therefore report on a study that compares the informativeness of two recognition and two recall formats. Native and non-native speakers of English were given 36 vocabulary items in the different formats in a Latin square design. The word knowledge of the participants was afterwards verified in face-to-face interviews and written meaning recall measures, probing whether their scores derived from guessing, partial knowledge or mastery of different word knowledge aspects of the target items. The results provide insights into the workings of different test formats and can serve as the basis for constructing new tests or improving existing vocabulary tests to arrive at more useful measurement instruments for both FL teachers and SLA researchers.

Paper 13- Investigating the validity of discourse completion tests: Effects of rejoinders and prompt enhancement

Afef Labben, University of Tunis

Despite the criticism leveled against elicited data and the call to use 'authentic' corpora, recent research in Interlanguage Pragmatics (ILP) still relies on the use of the Discourse Completion Test (DCT) to make inferences about learners' pragmatic ability or investigate the linguistic and cultural specificities of pragmatic behaviors. A survey of a number of ILP studies shows that different versions of this instrument exist and that researchers often opt for comparing data collected via different DCT types. Billmyer and Varghese (2000) designed an unstructured DCT with enhanced situational prompts to investigate the effects of prompt enhancement on subjects' production of requests and concluded that enhanced DCT items produce "more robust external modification and elaboration than do the archetypal content-poor prompts which most DCT studies to date have used (p.543)." Johnston et al. (1998) investigated the effects of different types of rejoinders on the production of complaints, requests, and apologies and concluded that speech act strategy selection is sensitive to rejoinder type. Such findings raise doubts as to the construct validity of different DCT types and call into question the comparability of the data they yield. The present study seeks to validate different DCT versions in a non-Western context. Specifically, it investigates the effects of prompt enhancement and rejoinder type on Tunisian EFL students' production of requests and apologies in English. Six different DCT versions varying according to absence or inclusion of different types of rejoinders (positive and negative), and absence or inclusion of contextual details in situational prompts were administered to six equivalent and homogeneous groups of respondents (total number = 240). Each version included twelve apology and request situations set in different contexts. After completing the DCT, a structured interview was conducted to investigate informants' evaluation of the DCT items to which they responded. Results show [section in progress] that context enriched DCT prompts produced longer utterances and affected respondents' request strategy selection and modification differentially […]. The study has implications for pragmatics research, instruction, and testing.

Poster 14- Investigating the constructs measured by EAP writing tests for use in Japanese university entrance examinations

Yumiko Moore, University of Bedfordshire

Drawing on the socio-cognitive frameworks for validating academic literacy tests (Shaw and Weir, 2007 and Chan 2013), this study aims to establish whether writing tasks in English for Academic Purposes (EAP) in two new tests are valid for Japanese university entrance purposes. This study focuses only on writing tests, as writing is the key EAP skill based on which students are frequently assessed in their degree courses. The findings will contribute to the appropriate development and use of EAP tests for university admission purposes in Japan. Data will be gathered from multiple sources in Japanese universities. The mixed methods approach has been chosen for this study because a combination of qualitative and quantitative approaches provides a better understanding of research matters than either approach alone (Johnson and Onwuegbuzie, 2004). First, the initial sample for a survey of the current academic writing tasks will consist of c.100 teachers from a number of Japanese universities who teach courses in English, with follow-up semi-structured focus group meetings (five groups of four people). They will be also asked for comments on the current writing tasks used for entry at their institutions and provide copies of the tasks, which we will look at. Secondly, (a) course syllabi and course assignment samples, and (b) English test papers used in the entrance examinations during the 2013 and 2014 academic years will be analysed by expert judges. 200 students will take the TEAP writing test and an IELTS sample test together with a cognitive processing questionnaire, which will then be followed by a post-test questionnaire. In addition they will fill out the same questionnaire in respect of an academic writing task in English carried out on their courses. Statistical analysis will examine the overlap between the constructs measured in each (eg through factor analysis).

Poster 15- Comparing textual features of the IELTS and TOEFL iBT reading texts: An empirical approach

Nathaniel Owen, University of Leicester

Traditional subjective approaches have long been used to develop typologies of textual features to be included in language tests (Bachman et al, 1995; Enright et al, 2000; Khalifa and Weir, 2009). More recently, researchers have started to critically examine automated text analytic software such as CohMetrix (Graesser et al., 2004) to determine whether such tools offer an opportunity to standardise textual features in particular tests (Weir et al, 2009) and comparison across tests at different levels of the CEFR (Green et al, 2013). IELTS and TOEFL iBT are two tests of English designed to assess readiness for a course of higher education and are targeted at the same stage of test taker language proficiency. The utility of analytic software for comparing reading texts of IELTS and the TOEFL iBT remains unexplored. Exploratory research was conducted with 48 texts (24 IELTS and 24 TOEFL). These texts represented 16 complete reading tests (8 IELTS and 8 TOEFL) each from different, official test preparation materials produced by University of Cambridge ESOL examinations and Educational Testing Service (ETS) respectively. Texts were analysed using Coh Metrix v3.0 (McNamara et al, 2013). Paragraph structure and spellings remained consistent with the source material. Coh Metrix v3.0 includes eight separate measures of latent semantic analysis (LSA), seven measures of syntactic complexity and new measures of text easibility (McNamara et al, op. cit.). Texts were assigned group membership (IELTS and TOEFL) and descriptive statistics calculated for each group. A series of t-tests were conducted, adopting a stricter p-value (p < .01) to minimise the propensity for type I error (Brown, 2001). Fourteen metrics recorded statistically significant differences. These metrics represented lexical (three metrics), syntactic (five) and text-level (six) complexity. Discriminant function analysis revealed that these fourteen metrics significantly differentiated between texts for the two tests, accounting for 74% of between group variability. Closer scrutiny revealed that word count (DESWC .702) and latent semantic measures of cohesion (LSAGN -.411) were the best predictors, while metrics of word familiarity and text easibility metrics were poor predictors. Cross-validated classification showed that 85.4% of texts were correctly classified as IELTS or TOEFL using this regression model.

Poster 16- Evaluating the effect of testing on teaching within the English as a life skill programme in Sri Lanka- A case study

Umashankar Singanayagam, Anthony Green (Director of studies), Lynda Taylor (2nd supervisor), University of Bedfordshire

It has long been considered that tests play a powerful role in the Sri Lankan education system, and exert a significant washback on language teaching. In Sri Lanka, for the first time, School Based Assessment of Speaking (SBA) and National Test of Speaking (NTS) have, recently, been introduced within the English as a Life Skill Programme. The new system of assessment is intended to encourage more teaching and (hence learning) of spoken English in the classroom. Recent researches, however, suggest that although the relationship between testing and teaching is commonly made it is not clear whether it exists and, if it does, what the nature of its effect might be. The nature and extent of its effect can only be established through empirical investigation. This case study investigates the washback effect of the new change from the teachers' perspective. The research question, therefore, at its general level is: Does speaking receive more attention from teachers when a new system of assessment of this skill is introduced and does the assessment of speaking have washback on the educational processes, and the participants in teaching and learning? For this purpose, data were collected from participants at three levels- decision making, intervening and implementing- through questionnaires administered to 48 teachers and 480 students, who are the participants at the implementing level, 36 classes were observed with follow up interviews, taught by 12 teachers. There were interviews with 12 officials representing decision making and intervening level, and related documents were also analysed. The results of the study seem to indicate that the introduction of the new system of assessment has exerted some influence on teaching. However, many factors prevented the assessment from having the intended positive washback and the nature of the washback seemed to vary from teacher to teacher and between the SBA and the NTS. This study also shows how crucial the stakes and proximity of a test in determining types and intensity of washback and the extent teachers can therefore become agents for promoting intended positive washback.

Poster 17- The effects of planning time on Taiwanese learners' language performance and strategy use in a TOEFL iBT integrated speaking task

Helen Tan, University of Bristol

Integrated speaking tasks, which require test-takers to respond to a speaking question by synthetizing information obtained from reading and listening prompts are, as yet, underresearched in terms of planning time and learner strategy use, particularly in light of their cognitive complexity. This study investigates planning time effects on the speaking performance of Taiwanese college students (n = 67) and their strategy use when performing an integrated speaking task from TOEFL iBT. A mixed-methods research design was employed, involving first video recording students' performance on the task under two pre-task planning time conditions (30 seconds and 120 seconds) followed by a stimulated recall session in which they were probed about their strategy use during the task. The speech samples were subsequently rated by three experienced EFL teachers, using separate rating measures derived from TOEFL iBT speaking rubrics on language use, delivery, and topic development. In addition, the learners' lexical usage (Vocabprofile) and fluency performance (de Jong & Wempe, 2009) were also computed and compared across both groups. The stimulated recall sessions were transcribed and coded, by adapting Swain et al.'s (2009) coding scheme on TOEFL iBT strategy use in a second (as opposed to a foreign) language setting. Five strategy categories (Approach, Affective, Cognitive, Communication, and Metacognitive) were coded and the coded strategies were later enumerated to compare frequencies across planning time groups. The results reveal no significant differences between task performance and learners' strategy use across planning time conditions, suggesting that the length of planning time might not be a sensitive factor on influencing their language performance and strategy use behaviours. In addition, moderate positive correlations between task performance and the frequency of cognitive strategy use were found in the 30 second planning time group while a similar pattern of correlations was detected in the 120 second planning time group on the frequency of communication strategy use. The findings of this study shed light on learner strategy use while dealing with cognitively complex speaking tasks, with learners' variable task familiarity in the research context potentially offsetting planning time effects.

Poster 18- The comments are useful but you do not understand: Supporting lecturer-written feedback with effective and mutual understanding

Lulu Zhang, University of Southampton

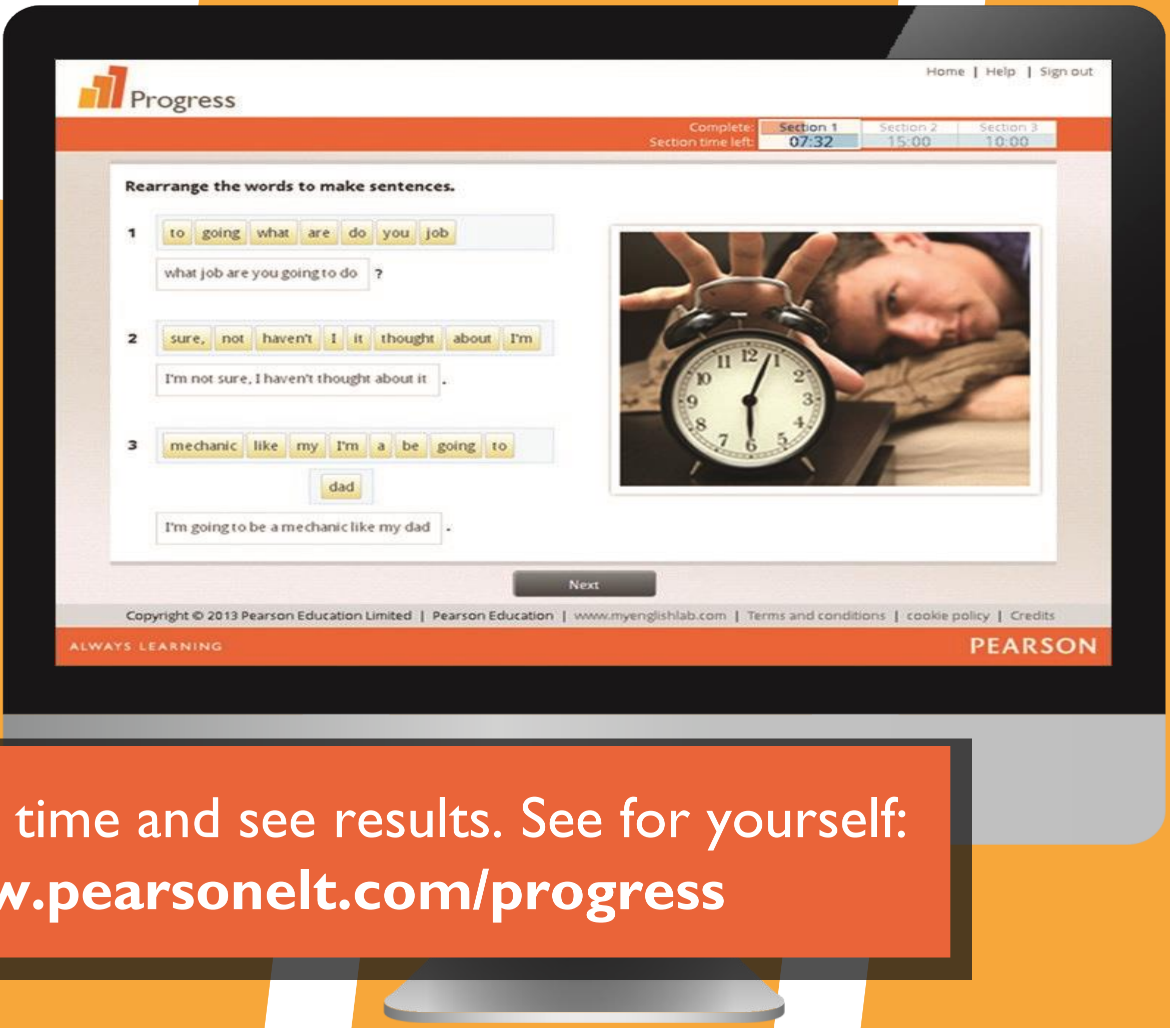
Providing feedback on students' writing is a central pedagogical practice in higher education. Feedback is an essential component in the learning circles. It is an ongoing instructional process that allows students to track their performance and makes adjustments in their efforts, directions and strategies. Growing attentions have been paid to the usefulness and effectiveness of feedback. However, the usefulness of feedback is put into question by the students' responses. A large number of students took feedback for granted and considered feedback useless. This qualitative study employing case study captured students' perceptions toward lecturer-written feedback and lecturers' responses to feedback. Those five student participants were full-time Master students in a UK university. They were either native English speakers or non-native English speakers. Two lecturers, one experienced and one new in the university, participated in the study. The purpose of the study was to make feedback more effective to students and better mutual understanding between students and lecturers. It aimed to raise students' awareness of being writers of their own writing and activate student self managed learning to make choice rather than being passive for learning and writing. It suggested that lecturers should take students' needs into consideration. The research categorized lecturers' written feedback from the perspectives of functions of speech, i.e., referential feedback, expressive feedback, and directive feedback. Students' interviews and lecturers' interviews were conducted based on the analysis of feedback practices to explore how students perceived the feedback and how lecturers gave feedback on students' essay writing. The study found that there were some gaps between students' and lecturers' perceptions toward written feedback. The gaps focused on the global issues and error corrective feedback, praise comments and criticism comments, direct comments and indirect comments and also the interpretations on the purpose and use of feedback. Students with different backgrounds had different preferences to those comments. Those differences could be attributed to gender, nationality, previous writing experience, teaching experience and students various expectations on their achievements.

Progress

Progress!

The first ever online, fully automated and institutionally administered progress test.

- Objectively and accurately measure students' progress throughout their co
- Fully automated speaking and writing sections
- Measure small increments in progress on the Global Scale of English



Progress

Home | Help | Sign out

Complete: Section 1 Section 2 Section 3
Section time left: 07:32 15:00 10:00

Rearrange the words to make sentences.

1 to going what are do you job
what job are you going to do ?

2 sure, not haven't I it thought about I'm
I'm not sure, I haven't thought about it .

3 mechanic like my I'm a be going to dad
I'm going to be a mechanic like my dad .

Next

Copyright © 2013 Pearson Education Limited | Pearson Education | www.myenglishlab.com | Terms and conditions | cookie policy | Credits

ALWAYS LEARNING PEARSON

Save time and see results. See for yourself:
www.pearsonelt.com/progress

ALWAYS LEARNING

PEARSON