EALTA Special Interest Groups for

The *Assessment of Writing* and *Assessment for Academic Purposes*

Thursday, 23rd May (13.30-17.30)

PROGRAMME

13.30-13.35 Welcome and Introduction

Presentations

13.35-14.00

Helen Heaney (University of Vienna)

Extending C1/C2 descriptors to assess academic writing in Departments of English at Austrian universities

Austrian education is currently undergoing long overdue standardization procedures, with the introduction of nationwide school-leaving exams scheduled for the academic year 2014/2015, for example. Similar issues of standardization were first addressed at Graz, Klagenfurt, Salzburg and Vienna Universities in 2006, when applied linguists and language teachers initiated a series of workshops to professionalize assessment practices for exit-level English-language exams in their BA programmes by developing analytic rating scales, first for writing and then for speaking. This paper charts how anchor descriptors were extracted from the CEFR for a bare pass (C1) and the top grade (C2) in Grammar, Vocabulary, Textual Competence and Pragmatic Competence and then fleshed out to help the English departments assess their students' academic writing skills in line with the CEFR.

Once anchor statements had been taken from the "criterion levels" C1 and C2 in the relevant subsections of linguistic, sociolinguistic and pragmatic competence, exam performances on typical writing prompts set at the four departments were rated by the workshop participants with reference to those anchor statements and including brief justifications. Subsequently, extended descriptors were formulated for C1/C2 based on the teams' justifications, a mid-way level (C1.1) was inserted, and Pragmatic Competence was replaced by more workable definitions of Task Achievement. Finally, benchmark performances were identified for as many levels as possible for each dimension and the scales were modified slightly during that procedure. Work on the writing profile should now continue, both in relation to rater training and validation of the assessment instrument itself.

14.00-14.25

Nükte Durhan (Middle East Technical University Northern Cyprus Campus)

Assessing Integrated Writing Tasks for Academic Purposes

A recent trend in assessing English for academic purposes is to integrate reading and writing rather than testing them as discrete skills. The justifications for integrated reading/writing assessment tasks are that they enhance authenticity and improve validity as these tasks elicit reading and writing processes that are essential for academic writing, have the potential to promote positive washback, encourage the focus on discourse synthesis skills and evaluate language abilities in a way consistent with construction-integration models of literacy. Now, there is a strong argument in literature for including the appropriate use of source material as part of the construct of writing ability. In addition to these promises, however, the literature points out several limitations. The most relevant ones to our context are mixing the measurement of writing ability with ability to comprehend source material, involving genres that are not well-defined and so are difficult to score and misrepresentation of candidates' language proficiency because of heavy reliance on the language of the source text. This study is aiming at showing how assessment tools that are tailored towards the specific aims of a particular course and certain instructional applications can overcome limitations and increase beneficial effects. In the case of METU NCC freshmen reading into writing EAP courses, an important part of assessment consists of thematically related essay writing tasks. Instructional practices include focusing on key thematic vocabulary, integrating extensive reading to course syllabus but carefully limiting the texts to course theme and modeling and scaffolding integrated tasks leading to the use of discourse synthesis processes.

14.25-14.50

Ying Zheng (Pearson)

An investigation into the writing construct(s) measured in Pearson Test of English Academic

This study examined the underlying writing constructs as assessed in Pearson Test of English Academic (PTE Academic) and then further explored item type difficulty and item type effectiveness. The statistical analysis of PTE Academic writing items indicated that the six writing item types fell into two categories. Although these item types all assess the writing construct, the results showed two different sub-constructs, namely, the *Analytical/Local Writing* construct and the *Synthetic/Global Writing* construct. Two gap-filling item types and the dictation item type primarily assess the *Analytical/Local Writing* construct. These three item types predict this construct comparatively well, with dictation having the highest standardized coefficient. Two summary item types with the essay item type primarily assess the *Synthetic/Global Writing* construct. The reading-to-write item type predicts this construct best, suggesting an item type that can measure test-takers' global writing ability well.

The results indicated that using a variety of tasks in assessing writing abilities of test takers allows test takers to use various strategies and deploy their different cognitive and linguistic resources in different ways. The results support earlier literature that listeners use different processing strategies and that different listening tasks measure different aspects of the construct. The findings also provided support for using integrated listening with writing and reading tasks with writing in the assessment of academic writing. Another important finding of this study is the importance of dictation and the role it can play in the assessment of writing skill(s).

14.50-15.15

Yeliz Akel & Hale Kızılcık (Middle East Technical University, Ankara)

<u>The Role of Text and Question Types in Predicting Reading Comprehension Difficulty of Reading Tests: A Study on Low-stakes Exams</u>

Assessing text difficulty, which is of crucial importance to write good reading tests is a challenging task. Most research studies regarding variables affecting text difficulty are conducted in relation to standardized high-stakes exams. However, it is a pedagogical and an ethical responsibility to study the impact of the texts and questions on students' success in low-stakes exams.

This study explores the difficulty of the reading parts in the midterm and final examinations administered in ENG 101 given by the Department of Modern Languages at Middle East Technical University. Based on the findings, test specifications will be revised to ensure that the tests do not show much variance in terms of the factors that are found to have a significant impact on the difficulty of the tests.

The study is carried out at two stages. First, 5 ELT experts are asked to rank 4 texts in term of their difficulty based on specific assessment criteria by SourceRater (vocabulary level, syntactic complexity level, degree of academic orientation and paragraph structure). Then, their rankings are correlated with each other's rankings and the rankings of SourceRater. Second, the questions in these exams are coded applying the question type classification used by Ozuru et. al. (2008) in order to identify what kind of processing skills students are required to use to answer the questions. These findings are expected to help testers to write tests at a standardized difficulty level.

****

15.15-15.45   Coffee Break

****

15.45-16.10

İrem Gedil, Bekir Ateş, and Hardy Griffin (Istanbul Şehir University)

In February, the testing office at Şehir University conducted a survey of faculty members and students in the faculties who had gone through the preparatory program (SEPP). Those survey results indicated that approximately 94% of classes utilize lectures supported by Powerpoint visuals. All of the students surveyed enjoyed the use of visuals to concretize the abstract concepts from the lectures.

Up to that point, our formative and summative assessments used listenings with no visual element; however, we created a video with still visuals supporting the lecture section of the Pre-Faculty module exam. After the exam, we surveyed the students to further understand the impact of the visual element in listening assessment.

In the quantitative part of the survey, 24 out of 51 students said the visuals helped them, only 1 student said they disturbed her/him, and 26 said the visuals "had no effect". When asked if they would like to see visuals on future tests, 25 said they would, 10 said they would not, and 16 said they were not sure.

The qualitative answers to the survey were the most interesting in that they demonstrated that students were able to remember the connection between the pictures and the ideas presented in the lecture. By the time of the ELEA SIG meeting on May 23, we will have conducted one further survey on listening without visuals and can present that data as well.

16.10-16.35

Carolyn Westbrook (Southampton Solent University)

Investigating the relationship between input task characteristics and performance on an EAP written output task

Nowadays, particularly in the field of EAP, integrated performance testing (see TOEFL iBT, PTE Academic, TEEP) is gaining importance since integrated tasks "reflect authenticity of task and response" (Pearson Language Tests, 2010) and "real-life performance is increasingly seen as the criterion of choice against which test tasks are judged" (Shaw and Weir, 2007: 17).

A number of researchers have investigated the effects of independent vs integrated tests (Gebril, 2009; Esmaeili, 2006) while others have looked into the effects of input characteristics on performance in listening tests (Wagner, 2010; Coniam, 2000; Ginther, 2002; Ockey, 2007). However, there seems to be a lack of research on the impact of listening task characteristics on performance on a written output task.

This presentation will discuss the researcher's PhD work-in-progress which looks at performance on independent IELTS listening and writing tasks compared with an integrated EAP listening-into-writing task.

The presentation will illustrate the findings of the pilot study and will discuss changes to be made in the main study.

Initial findings suggest that performance on the integrated task is generally better than the IELTS writing task and longer length essays are being produced. Another finding is that one form of input allows students to produce more from the source text than others; however, there may be a number of reasons for this, and these are not clear from the pilot study.

16.35-17.00

Dr Annie Brown and Louise Courtney (Australian Council for Educational Research)

Assessing writing skills in the Middle East – a case study from the UAE

The Australian Council for Educational Research is developing a system-wide testing regime for students in the UAE at grades 3, 5, 7& 9. The assessment includes testing of Arabic and English (EFL) writing at each level. Over the last three years English-speaking test developers have worked intensively with Arabic education experts on site in the UAE. What started as a dual task test with separate rubrics and scales has now developed into a single-scale assessment tool for writing competency in Arabic. Complex written prompts have been replaced with visual prompts, and a single rubric has been refined after taking into account first round assessment results and feedback from the expert panel. L2 (EFL) is assessed using a range of tasks from single word through to free writing. The English writing test was also developed and modified with expert educators in the UAE.

Other test developers and researchers may be interested in trans-cultural issues around differences in approach to the assessment of writing. Can all grades be effectively scored on a single scale in writing? Does this work for both L1 and L2? How has future capacity been developed for UAE test developers and classroom teachers? What are the implications for marker training and washback effect in the UAE? Psychometric data and prompts for writing assessment are presented and issues for the future of assessing writing in the Middle East are considered.

17.00-17.25

Jonathan Rees (University of Birmingham)

Assessing the academic writing skills of 'home' undergraduate Science students

This presentation draws on data from the second year of an ongoing collaborative research project at the University of Birmingham between the School of Biosciences and the English for International Students Unit. The project aims to produce a tool for the 'quick' assessment of academic writing skills for home undergraduate Science students.

It briefly reviews the background to and the scope of the project (as reported in detail at last year's SIG). It then proceeds to present both score and 'self-assessment' questionnaire data* from this year's undergraduate cohort (n= 220).

The data from this year's cohort is then compared with that from last year's cohort, revealing a surprising level of consistency both in terms of student performance and student self-assessment.

The next section of the presentation provides a detailed analysis of student performance on the final section of the test, the most 'holistic' in its assessment type. This section requires students to rewrite an excerpt from an authentic student essay, with particular reference to correcting any perceived problems of coherence and cohesion. Drawing on student responses to this task, the extremely wide range of 'academic discourse' competence represented within the cohort is illustrated and examples of basic errors in register, punctuation, tense selection and use of discourse markers are provided.

Finally, possible explanations for this diversity in discourse competence are proposed.

17.25-17.30 Closing Remarks and Feedback

POSTER PRESENTATIONS

Jayne Pearson (King's College London)

Participation and Control in EAP Writing Assessment

Awareness is growing, both inside and outside the testing community, of the impacts of testing on individuals, as well as the wider social and institutional consequences of tests (Shohamy 2001, McNamara, 2001). Testing of writing in EAP contexts should be focussed on the very principles these courses espouse: of preparation for future learning with a view to enabling students to participate in their academic communities. This is not always the case when one- shot, product-oriented summative testing is the model. As a result, students who should be finding their voice in a new academic culture are silenced, identifying their own learning progress with that of a numerical score. Hence, the student is the object of the assessment as opposed to the subject /agent of it (Boud, 2000).

This paper will outline planned research to take place at King's College London ELC as part of a PhD project on how EAP test takers can negotiate power with those who are making judgements of their ability though *assessments for learning* (Black and William, 2002). The study aims to examine whether the implementation of *processfolio* writing assessment could a) increase students' sense of control over their development as writers and b) increase their participation in the assessment process itself.

The nature of and rationale for the design of the study will be described, in line with the view that traditional psychometric definitions of validity and reliability are not the only criteria needed to widen our understanding of assessment in the EAP context.

Sonja Zimmermann (TestDaF Institute)

Fair assessment? – Modifications for disabled candidates in the TestDaF

Free access to language testing is understood as a central criterion of quality assurance as it is defined in various codes and guidelines of associations that aim to establish high quality in language testing (e.g. ALTE 1994, EALTA 2006). The ethical context hereby is striving for fairness. Consequentially, institutions that develop, administer and assess language exams endeavour to offer exams that are modified according to the needs of examinees with disabilities.

The paper presents current accommodation practices for test takers with special needs in use with the Test for German as a Foreign Language (TestDaF), focusing on special modifications offered for blind test takers. Specifically these participants can choose between two test versions: one in Braille, the tactile writing system used by blind people, and one electronic version that is processed via PC, using a screen reader. Both versions are described and discussed according to the following questions: In what ways are test design and test administration accommodated? What impact does the use of technology have? What challenges do the changes made pose for appropriate, standardised test development and administration?