

EALTA SIG Seminar Assessing Writing/ EAP Assessment
University of Warwick, Centre for Applied Linguistics
6/7th February 2015

Friday, 6 th February 2015				
Room A0.14 Social Science	Time	Presentation	Presenter	Affiliation
	14:00-14:30	Welcome		
	14:30-15:20	Assessing writing through editing – A step back or forward?	Sonja Zimmermann	TestDaF Institute, Ruhr-University Bochum, Germany
	15:20-15:50	Coffee Break		
	15:50-16:40	Assessing writing in monolingual and bilingual EMP tests	Vilmos Warta	University lecturer, director of PROFEX Testing Centre, head examiner of sTANDEM Testing System, Hungary
	16:40-17:30	<i>Processfolio</i>	Jayne Pearson	King's Collage London, UK
Saturday, 7 th February 2015				
Room A1.11 Social Science	Time	Presentation	Presenter	Affiliation
	9:30-10:20	Scoring writing in the Swedish national test	Eric Borgström, Anne Palmér, Maria Westman	Department of Scandinavian Languages, Uppsala university, Sweden
	10:20-11:10	Setting Educational Standards for Academic Writing Benchmarked Against the CEFR	Ebtesam Abdulhaleem	PhD students, University of Warwick, UK
	11:10-11:40	Coffee Break		
	11:40-12:30	Assessing the Scoring Validity of an EAP Reading Comprehension Test	Anita Hegedus	Department of Languages for Specific Purposes, Medical School, University of Pécs, Hungary
	12:30-13:20	Using the assessed 'Feedback Viva' to increase student engagement in academic writing and to prompt reflection on written feedback	Martin Seviour and Walter Nowlan	Nottingham Trent University
	13:20-14:50	Lunch Break		

14:50-15:40	A presentation of the Norwegian digital test at B2 level	Hanne Lauvik	Norway
15:40-16:30	Tests, constructs and the world of academia	Bart Deyger	KULeuven, Belgium
16:30-17:00	Coffee Break		
17:00-17:50	Adapting a writing rubric	Berna Torlakli	EALTA Member
Seminar Close			

Abstracts

Sonja Zimmermann: *Assessing writing through editing – A step back or forward?*

Over the last two decades performance-based assessment has played an increasingly dominant role. In fact, constructed-response tasks like the time impromptu essay have come to be the most widely accepted format to assess students L2 writing ability (Cumming 1997; Weigle 2002). Besides the serious critique that general conclusions about the writing ability of examinees are difficult to draw based on just one writing product, in one genre, on a new topic, and in a low-authentic, timed context (Bouwer et al. 2014; Weigle 2007), free-response writing in a timed assessment context is also usually limited to first draft writing. The writing process – like proposed in existing models of writing (e.g. Hayes and Flower 1980) – though presupposes also revision skills which are not explicitly assessed through such writing tasks and are not reflected in the corresponding rating scales. The paper presents a draft of an automated editing task and wants to discuss to what extent this task can contribute to a wider representation of the writing construct by not only measuring core text production processes like drafting but also revision skills.

Vilmos Warta: *Assessing writing in monolingual and bilingual EMP*

The focus of teaching English has remarkably changed in the past decades. Teaching English for Specific Purposes (TESP) has emerged from the shadow of general English and has become more and more demanded and widespread all around the academic world. Inevitably, English language testing systems should have responded to this change. Now, there are a great number of newly established English for Specific Purposes (ESP) testing systems and projects to develop them both in the academic setting and in the real market. This altered approach is especially present in medicine. There is a vast need to test English for Medical Purposes (EMP) competences giving a reliable and valid feedback on test-takers oral and written EMP skills both for candidates and for stake-holders. Here, I intend to present and compare the methods of writing skills assessment in two EMP testing systems harmonised to the principles of the Common European Framework of Reference for Languages, PROFEX and sTANDEM. The former, PROFessional EXamination, is a bilingual EMP test accredited by the Hungarian Language Testing Accreditation Board in 2000. The latter, STANdardised English for Medical Purposes, is the result of a project supported by the European Union, in which ten partners, the Jagiellonian University, the Innsbruck Medical University, the Elanguest English Language School, the University of Pecs, the University of Medicine and Pharmacy of Târgu-Mures, the University of Edinburgh, the University of

Burgundy, Dijon School of Medicine, the Tokyo Medical University, the European Association of Science Editors and the European Association in Health Care Communication worked together. The presentation shows that despite the different assessment methods, both EMP assessment systems give a valid and reliable picture about the test-takers' writing skills.

Jayne Pearson: *Processfolio*

This presentation summarises an ongoing project at King's College London on the impacts of an alternative writing assessment in EAP, termed processfolio. The rationale for the project will be outlined through a brief overview of the literature and initial exploratory data, namely the questionable impact of current testing procedures on pre-sessional EAP courses on a) students' sense of agency in assessment practices and b) their writing development due to the focus on the writing product over process. The current project, using an action research methodology, explores the extent to which processfolio encourages more active participation on the part of EAP students, both with the writing process and practices of assessment for learning. To this end, I will outline the implementation of the processfolio assessment, present initial research findings for the first cycle of the project in 2014 and implications for the second cycle in 2015 and beyond. This is an example of research conducted within the practitioner's own local context, but is intended to expand on practitioner reflexivity to encompass critical emancipatory action research principles for wider engagement with colleagues and stakeholders in EAP writing assessment.

Eric Borgström, Anne Palmér, Maria Westman: *Scoring writing in the Swedish national test*

This presentation reports on a recently initiated (design stage) research project, aiming at evaluating the effects on test validity of using different scoring methods on tests in the Swedish national writing assessment.

In 2011, a new scoring method was introduced in Swedish national writing assessments. In the new test construction, teachers are to score student texts analytically (grades A-F), in three or four dimensions (i.e. content, structure, language resources and correctness) using matrix and benchmark-texts. These scores are then to be summarized to a single grade according to a regulating principle. In prior test constructions, teachers were to score essays more holistically, guided by their professional judgment and supported by tools such as matrixes and benchmark-texts. The new scoring method was assumed to add clearness to the scoring procedure and to enhance the reliability of the test scores. However, no empirical evidence supporting these assumptions were collected.

The idea behind this study is to investigate the consequences of using different scoring methods, such as this analytic method compared to a holistic and a semi-holistic method. Analysis will be made on inter rater-reliability, co-scoring sessions and teacher questionnaire.

Ebtesam Abdulhaleem: *Setting Educational Standards for Academic Writing Benchmarked Against the CEFR*

The Preparatory Year (PY) programme at a University in Saudi-Arabia has recently started using CEFR as a reference tool to assist in curriculum development, textbook selection and assessment. However,

CEFR usage is only limited to curriculum and assessment and it is not used by teachers and students. The use of CEFR at the PY is based on general perceptions of the programme's requirements. Each unit is working independently to adopt the framework to their needs. Curriculum unit, for example, uses the CEFR to specify the outcomes and choose textbooks. Assessment unit refers to it when designing exams. Nevertheless, the interdependence the CEFR brings between curriculum, assessment and pedagogy (Little, 2011) is not noticed at PY. For this reason, this study aims at adapting and using the CEFR writing descriptors to conduct a needs analysis study for the purpose of identify the educational minimal and ideal standards that are needed and expected from the local context-medical students at PY. The study will be targeting different stakeholder groups to triangulate and gather more cohesive and comprehensive information regarding those educational standards. In conclusion, the study aims to contribute to the literature of using CEFR to link curriculum, assessment and pedagogy. It is also hoped that this initiative will be further researched and made use of by all parties at PY and similar programmes.

Anita Hegedus: Assessing the Scoring Validity of an EAP Reading Comprehension Test

This paper undertakes to investigate the scoring validity of an EAP reading comprehension test at level B2 based on item analysis. The test was an English for Medical Purposes test administered at the PROFEX examination at the University of Pécs, Hungary in April, 2014. First, pre-testing was carried out on a sample of 30 and statistical analyses using SPSS (frequencies and descriptives) were done. Based on the results of the analyses, 5 items were modified. The same descriptives and frequencies statistics were carried out after the live administration of the test on a sample of 171 to see whether the test was appropriate for the test population and whether the decisions made after the pretest were justified. Following the statistics performed using SPSS, IRT analysis with the program of Winsteps was done to confirm the decisions made after pretesting. The results of the analyses showed that the decisions done following the pretest increased the reliability of the test and the test was proved to be appropriate, although slightly easy, for the test population.

Martin Seviour, Walter Nowlan: Using the assessed 'Feedback Viva' to increase student engagement in academic writing and to prompt reflection on written feedback

Teaching academic writing forms a significant part of the input on the Pre-sessional EAP courses at NTU. Over the last two years we have been looking at ways to shift the focus of our teaching and assessment away from the 'end product' and towards the academic writing process itself.

As part of this shift we have introduced a summatively assessed 'feedback viva' in which the student and the teacher who has marked the essay draft discuss, in a semi-structured format, the student's essay and the written feedback on the draft. The key aims of the feedback viva are to provide an opportunity for the student to demonstrate engagement with the assignment and to reflect on and respond to the feedback.

The presentation will outline the context in which the feedback viva is set as well as the practicalities of the process including standardisation and moderation of the assessment and the development of assessment criteria. We will evaluate the advantages and drawbacks of the feedback viva in the light of student and teacher responses to a survey as well as consideration of published reports on similar initiatives (e.g. Franks and Hascomb, 2012).

Hanne Lauvik: *A presentation of the Norwegian digital test at B2 level*

Vox, Norwegian Agency for Lifelong Learning, has developed a new digital test of Norwegian as a second language from level A1-B1. The writing test measures at two levels: the candidates can take a A1-A2 or a A2-B1 test. From December 2015 the B2 level will be included in a B1-B2 test.

The CEFR level to enter higher education in Norway is set at B2 level, and some employers (especially within the health and social care sector) require B2 level for employees. The target group for the test is therefore not just academia, and the test will have to balance the needs of both higher education and the job market.

With developing a digital test there are many opportunities, but also some challenges. In this presentation I will discuss the main challenges we faced in developing a new digital B2 test from scratch, as well as the solutions we chose to these challenges.

The presentation will focus on the challenge of measuring at two levels in one test, task formats, number of tasks, text length, and criteria.

Bart Deyger: *Tests, constructs and the world of academia*

Discussing three complementary studies, this presentation reflects on the context and criterion validity of the two main B2 language tests that grant access to Flemish higher education.

The first qualitative study examines which skills are required of students on the first day of university education in Flanders and how these skills relate to the university entrance tests. The results highlight issues of fairness, construct-irrelevant variance and context validity. Following up on this, the second study is a longitudinal qualitative analysis of 24 L2 students' experiences during their first year at a Flemish university. The data reveal factors uniting the very heterogeneous group of L2 students and indicate that the fiercest obstacles this group faces are not necessarily language-related. Thirdly, a concurrent validity study shows how the operationalizations of both B2 tests compare, to what extent both tests measure the same skills in the same way and to what extent both tests assign the same level of ability to the same candidates.

These studies are part of an on-going PhD project and not all analyses are finalized. Consequently, the presenter would greatly welcome any constructive feedback and critical comments.

Berna Torlakli: *Adapting a writing rubric*

This study aimed to adapt the writing rubric that is used to evaluate essays in Preparatory School at Toros University. A six-week workshop was conducted to achieve the aim of the study. During the workshops, ideas were exchanged and experience was shared. After the methods to evaluate writing papers were reviewed, the methods that are used at Toros University and in the international exams such as TOEFL and IELTS were examined. In this concern, the rubric that is in use in Preparatory School at Toros University was examined in detail and it was decided to make some changes in some parts to make it more clear and understandable. In this study, qualitative data was used. The participants of this study included the researcher of the study and a writing instructor who were at the same time the Testing Coordinators at Toros University Preparatory School.