EALTA Special Interest Group Meeting

Assessing Speaking

Thursday, 5 May, 2016, 14:00 – 17:30

(4P Language Centre Building)

13ᵗʰ EALTA Conference, Universitat Politècnica de Valencia, Valencia, Spain

PROGRAMME

14.00-14.10	**Welcome**	
14.10-14.40	**Kathrin Eberharter, Nivja de Jong, Benjamin Kremmel, Carol Spöttl** (Assessing Speaking SIG)	Speaking assessment practices across Europe and beyond – preliminary survey results
14.40-15.00	**Raili Hildén** (University of Helsinki), **Mikko Kurimo** (Aalto University), **Martti Vainio** (University of Helsinki)	Digital test of oral proficiency for large-scale use – starting from pronunciation
15.00-15.20	**Nahal Khabbazbashi & Evelina Galaczi** (Cambridge English Language Assessment)	Technology and the 'what' of auto-marked speaking tests
15.20-15.50	**COFFEE BREAK**	
15.50-16.10	**Armin Berger** (University of Vienna)	Assessing academic speaking skills – rating scales as the de facto test construct
16.10-16.30	**Margit Hesselager** (University of Southern Denmark)	Teacher Talk: Assessing Speaking in Tertiary Education
16.30-16.50	**Anita Hegedűs** (University of Pécs)	The Construct of Speaking Ability in a Bilingual English for Medical Purposes Exam
16.50-17.00	**Jayanti Banerjee** (Assessing Speaking SIG)	Discussant
17.00-17.30	**Group discussion**	

Paper 1: Speaking assessment practices across Europe and beyond – preliminary survey results

Kathrin Eberharter (University of Innsbruck), Nivja de Jong (Utrecht University), Benjamin Kremmel and Carol Spöttl (University of Innsbruck)

Developing and delivering speaking examinations takes up considerable resources for everyone involved in the process and poses particular challenges to those who have to assess speaking skills. The EALTA Assessing Speaking SIG was founded in an attempt to support members in overcoming some of these challenges and help establish sound assessment practices throughout Europe.

At the inaugural meeting of the EALTA Assessing Speaking SIG in Warwick 2014, the idea was born to create a survey to systematically collect data on the practices of speaking assessment across different European contexts. This would help identify the lacks, needs and wants of the community to plan further actions in promoting best practice in speaking assessment. A draft collection of survey items was developed during the second SIG meeting in Copenhagen 2015. This first batch of items was then expanded, ordered and digitalized and an online survey was launched in early 2016 on the EALTA members discussion list.

The aim of the survey is to investigate the state of speaking assessment practices in Europe and beyond. The survey intends to capture a wide spectrum of assessment practices across different teaching and learning contexts, languages and proficiency levels. This includes the parameters of the test design and delivery, such as, for instance, the types of tasks, length of examination and delivery mode, but also investigates the assessment criteria employed.

This presentation will provide a summary of the results from the first data collection period (January-March 2016) and provide a basis for the discussion of the needs, in terms of resources and training, of language teachers and assessors within and outside of the EALTA community.

Paper 2: Digital test of oral proficiency for large-scale use – starting from pronunciation

Raili Hildén (University of Helsinki), Mikko Kurimo (Aalto University), Martti Vainio (University of Helsinki)

Major challenges of testing speaking in large-scale contexts adhere to practicality. Modern digital technologies, finally, enable storing samples for computerized rating. At the same time, measurable dimensions of the speaking construct, like pronunciation, grow in importance.

We introduced a digital speaking test for Finnish matriculation exam. The pilot test of Swedish consists of a read-aloud task, short turns and a series of extended answers in a video supported test environment, and a problem-solving task in pairs. The performances were assessed by human raters for pronunciation and fluency, and analyzed for prosodic features by a speech recognition tool. Following research questions were addressed:

1. How well does a speech recognition system trained for student speech perform in these tasks?
2. How accurately do student speech sounds match to native speakers?
3. How well do the human and computer ratings for pronunciation coincide?

We sampled school-based Swedish from ca 500 students to train the recognition tool. The word error rates (WER) of non-native and native samples were compared. Ten Swedish teachers familiar with the CEFR scales rated a random sample of non-native students' Swedish for fluency and pronunciation. Tentative results suggest that the speech recognizer trained for student speech and adapted to the speaker is useful in monitoring what the students speak and aligning the read speech with the corresponding text. This automatic processing allows the comparison of student speech sounds to the native speech sounds, computerized rating of the goodness of pronunciation, and the comparison of human and computer ratings.

Paper 3: Technology and the 'what' of auto-marked speaking tests

Nahal Khabbazbashi & Evelina Galaczi (Cambridge English Language Assessment)

Technology has played an important role in speaking tests in shaping the way the speaking construct is conceptualised, elicited, marked and marketed, leaving exam providers with challenging decisions in the design, development and scoring of their speaking tests. In our presentation, we focus on a recently developed online placement test of speaking. The test consists of five parts including read aloud items, questions on familiar and abstract topics and photograph description. It was originally intended for the test to be marked by human examiners. However, the market demand for a cost-effective and quick testing solution required a consideration of alternative automated approaches to marking speaking. In our talk we discuss the challenges faced in selecting a marking model in light of the (often conflicting) demands of validity, reliability and practicality. We then present the results of a pilot study that was specifically designed to inform this decision; 240 speaking tests from our exam covering a range of proficiency levels were marked by four raters and an automarker. Candidates' CEFR classifications using the different marking models were subsequently compared. While the findings relate to the specific assessment context of our placement test of speaking, they can be of relevance to different assessment contexts and can also contribute to wider discussions on the 'what' that is being assessed in technology-mediated speaking tests and the different scoring models which technology can open up in assessment.

Paper 4: Assessing academic speaking skills – rating scales as the de facto test construct

Armin Berger (University of Vienna)

This presentation focuses on the construct of a speaking test administered at the end of the Practical Phonetics and Oral Communication Skills module as part of the English Language Competence programme in the English Department at the University of Vienna. After describing the basic design of the test, I would like to elaborate on the rating scales as the de facto test construct, covering academic presentation and interaction skills. The construct has been developed further by the English Language Teaching and Testing initiative – a working group consisting of applied linguists and language teaching experts from the Universities of Graz, Klagenfurt, Salzburg, and Vienna. Extending and adapting the CEFR descriptors to suit the local needs, subdividing levels C1 and C2 into narrower yet meaningful and practical level descriptions, and counterbalancing the intuitive approach to describing the progression between the scale levels were some of the major challenges faced by the test developers. I would like to report on how these challenges were addressed in a series of scale development workshops and a multi-method validation study involving participants from all five Austrian English departments. I will share some of the practical lessons learned from this project that might be useful in similar contexts, and conclude by pointing out how this local test development and validation project can enrich our understanding of academic speaking, adding concrete details to the two advanced levels of the CEFR.

Paper 5: Teacher Talk: Assessing Speaking in Tertiary Education

Margit Hesselager (University of Southern Denmark)

In 2013 a new policy was introduced at the University of Southern Denmark according to which professors who teach in English but have no formal or native English background need to have their English proficiency assessed. The test designed for that purpose requires for assessors to sit in on a 45 minute lecture hosted by the given test taker after which the assessors write a report based on their observations. The report refers to an assessment template designed for the purpose and anchored in the CEFR framework. Challenges resulting from the design of the template or arising during the assessment procedures are discussed regularly by stakeholders. Some challenges adhere to the situated nature of the test. Another challenge arises when very concrete constructs intended to qualify the assessment are unable to encompass certain unpredictable qualities of language produced at the advanced levels. Finally, there can be a tension between the need for precision among assessors and the wish to ensure that the report and test are understandable to test takers. Exchanges spurred by issues such as the above have led to a recent revision of several details in the template as well as to minor adjustments in the assessment approach. The presentation will introduce the template and explain some core alterations.

Paper 6: The Construct of Speaking Ability in a Bilingual English for Medical Purposes Exam

Anita Hegedűs (University of Pécs)

PROFEX is a bilingual (English - Hungarian) EMP (English for Medical Purposes) examination. It has three profiles: medicine, pharmacy and general health care. The speaking paper of PROFEX consists of an introductory conversation on the candidate's professional background, two simulated conversations (one between a health care professional and a patient and another one between two health care professionals) and a presentation. Speaking performance is evaluated along three traits: communicative competence, adequate use of medical terminology and grammatical accuracy. They receive equal weight in the case of the simulated conversations, while in the assessment of the other two tasks communicative competence is more important than the other two traits.

Although the role of background knowledge is a pivotal issue in testing languages for specific purposes, topical knowledge is not evaluated in this exam but is taken for granted and medical professionals are not involved in any stage of the evaluation. The examiners are language teachers who must have at least three years of experience in teaching EMP. PROFEX being a bilingual examination, the prompts for the tasks are provided in Hungarian. Therefore, the speaking paper inevitably involves an element of foreign language mediation. However, not providing sufficient input would involve testing the candidate's background knowledge, which is not what the exam is designed for. To overcome the above-mentioned problems, examiners receive training each year where these issues are dealt with. Only examiners participating in the trainings are employed by the examination centre.