Weaving CEFR Scales into Rubrics for EFL Writing Assessment in Indonesia

As there is a substantial research void in the area of high-quality and effective writing EFL assessment in Indonesia, the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001) might be a suitable standardised option. However, it has often been underlines that the CEFR scales are inappropriate for marking since they were developed for the description of language proficiency levels in general and only contain broad “can do descriptions”. In fact, it was not designed to be used as a rating scale or for test development without adaptions and modifications to render it suitable for those purposes (North, 2000). In this perspective, the descriptions can serve as a guideline when assessing language skills as stated by Davidson and Fulcher (2007), which encourage test developers to see the framework as a “series of guidelines from which tests and teaching materials can be built to suit local contextualized needs”. These circumstances has inspired the current research project, whose aims it is to design feasible and practicable scoring rubrics for writing assessment in tertiary Indonesian EFL context. The major issue addressed in this research is “How is the design of writing assessment rubrics developed form CEFR levels that improve Indonesia lecturers’ rating performance? The rubrics will originate from CEFR levels and are meant to make writing assessment more reliable and objective. This study will employ both qualitative and quantitative methodologies using scale-revision and rater-training approach adopted from Harsch and Martin (2012). Eight EFL Indonesian lecturers will in phases experience several rater trainings on how to apply the rubrics. Raters’ responses and recommendations from each rater training will be analyzed to revise the rubrics while the intraclass correlation coefficient will be used to determine the level of consistency among raters’ scores.

Assessing productive skills at universities: a comparative study between CLES and CertAcles

This paper introduces a research effort undertaken to further cooperation and recognition amongst two Higher Education certification systems within the NULTE network, CLES in
France and CertAcles in Spain. Our research focuses on productive skills as this is the main goal of university certification systems. Our first step was to undertake the comparative analysis of C1 writing assessment in both certification systems, from the point of view of specifications, rating scales and linking to the CEFR. Our goal is to find a common path towards the certification of language competence in a university context, based on the descriptors of the CEFR - and thus comparable and internationally transferable-, but sustained on different context-based assessment models.

Models of communicative competence (Hymes, 1972; Canale & Swain, 1980; Bachman, 1990; Celce-Murcia, Dornyei & Thurnrel, 1995; Bachman & Palmer, 1996) have tilted the balance towards the validity of assessment and thus towards using direct testing methods while implementing mechanisms that ensured reliability of results (design of specifications and scales, rater training, double marking, etc.). The most critical aspect of validity in a writing test is that of construct validity, that is, whether the test designed is successful in measuring the element required for successful writing in English (Kroll, 1998). In the case of university certification systems, the element required is the ability to go beyond the knowledge-telling model characteristic of immature writers towards the knowledge-transforming model of a mature writer (Bereiter & Scardamalia, 2003).

COFFEE BREAK (11.05-11.25)

11.25 – 11.55 Ewa Behling, Anna Soltyska (Zentrum für Fremdsprachenausbildung, Ruhr-Universität Bochum) and Mirja Englert (Zentrum für Sprachenlehre, Universität Paderborn)

Writing assessment calibration project: aligning B2 writing tasks and assessment criteria within the UNIcert® certification programme

The presentation outlines a calibration project which has been conducted over twelve months at the university language centres of the Ruhr University in Bochum and the University of Paderborn. The project aimed at ascertaining that, in line with the theory of constructive alignment (Biggs & Tang 2007), course learning outcomes and assessment tasks in both institutions equally match the CEFR level descriptors and the same minimum standards need to be met to obtain UNIcert® II certificates for the skill “writing” at both institutions. The project addressed level B2 exams and entailed designing, implementing and assessing written tasks for EAP courses. The procedure, to which four teachers/raters from both institutions contributed, involved rating students’ texts using institutions’ own rubrics in the first stage of the project (winter term 2018/19) and developing an improved rubric to be used in the second stage (summer term 2019). The assessment results obtained in each project phase were analysed both quantitatively (through the comparison of raw scores and analysis of inter-institutional reliability) and qualitatively (based on a detailed analysis of individual scores and criteria and subsequent discussion of assessment decisions made by the raters). The ultimate overview of assessment results is yet to be finalised (in autumn 2019). It is hoped that calibration projects irrespective of their scope prove to be an efficient tool for setting standards for internal and external contexts (Reich & Timukova 2015), at the same time helping to revise current teaching and assessment practices, and increase the assessment literacy in participating institutions.

11.55 – 12.15 Question collection for pm discussion
12.15 – 1.15 LUNCH
Integrated Tasks, Elements of Task-Based Language Teaching and Their Value for the Assessment of Writing Skills

Task-based language teaching (TBLT) has become a significant topic in the field of second language learning. TBLT provides opportunities to practice the use of language for meaningful purposes (Norris, 2009). Therefore, it supports the integration of skills through completing practical and functional use of L2. Despite increased familiarity in implementing TBLT as a way to integrate language skills in current second language classroom, the implications for assessment of integrated writing skills is not well recognized due to complicated issues related to measurement, test design and examinee performance (Cumming, 2013). This presentation discusses possible ways of assessing writing tasks in the task-based language teaching context. TBLT criteria for designing or identifying suitable tasks by Willis & Willis (2007) and a classroom activity that matches these criteria will be presented first. Research-based evidence supporting the need to integrate language skills in academic environment will be discussed next. The presentation will conclude with ways to assess integrated writing skills in the classroom through relevant tasks.

Keywords: task-based language teaching, integrated task, integrated writing assessment, second language learning

A Phenomenological Investigation of Pre-sessional Students’ Perceptions and Practices of Formative Assessment in English Writing

Formative assessment (FA), also known as assessment for learning, is aimed at helping the students’ learning while the learning is still occurring and checking the students’ progress during the instructional process (Bennett, 2011; Abrams, 2007). FA has become an important aspect of English teaching and has been attracting considerable attention from language assessment researchers, and it is often associated with idea of teachers’ feedback, peer-assessment, or self-assessment (Carless, 2012; Crossouard & Pryor, 2012; Birjandi & Tamjid, 2012). This phenomenological research aims to examine ESL learners’ perceptions and practices of FA when used as tools to promote learning:

1) What are participants’ attitudes to FA in their English writing?
2) To what extent do students engage with FA tools – feedback, self-, and peer-?
3) What are participants’ perceptions of FA in terms of utility of learning?

Purposive sampling method, classroom observation, and semi-structured, one-to-one interview are used in this study (Palys, 2008; Lavrakas, 2008; Denscombe, 2017; Creswell & Creswell, 2018). A coding method is used as the focus of the content – all data are transcribed by creating a coded Word document and a clean read transcription (Elliot, 2005). A categorical content analysis that focuses on separate parts of the texts to find themes or contexts is adopted to analyze the findings (Vostal & Hughes & Ruhl & Benedek-Wook & Dexter, 2008). Lastly, all data are organized and interpreted according to the research questions.