

A30

Hildén Raili & Martikainen Marja K.

Rater feedback on speaking assessment

The background of this study is a 3-year research project financed by the University of Helsinki, Finland, with a focus on the assessment of oral proficiency in foreign languages. The project aim is to build a validation argument regarding relevance, utility, and sufficiency for the illustrative scales of oral proficiency included in the new National Core Curricula (NCC). This scale with subscales for overall task completion, fluency, pronunciation, range and accuracy is empirically aligned to relevant scales of the CEFR.

Research problems address issues of inter-rater and intra-rater reliability of expert members of the Swedish, English, German and French juries, each 5-10 in number. Inter-rater reliability is scrutinized in relation to overall and task specific ratings on the subscales. Rater comments and interaction patterns emerging at the sessions are also taken into account. The findings are compared across members of each jury and also across languages. Intra-rater reliability is deduced from the correspondence between thematic commentaries of a rater and his/her ratings attached to congruent categories.

The data consist of level ratings assigned by the raters, feedback sheets and video recordings from rating sessions. A multi-method approach is adopted implying statistical analysis of numeric matrices and content analysis for qualitative data. Some preliminary results from the German jury already support appropriateness of the selected techniques for the purpose. By June 2009, the analyses will be replicated for the three other languages. Envisaged results have impact on design of an oral course to be introduced at upper secondary level in 2010.