

Poster presentations – Abstracts

Inga Arffman *Identifying translation-related sources of differential item functioning in international reading literacy assessments*

International reading literacy assessments are a highly collaborative effort. This is not least true of the quality control procedures implemented in these assessments to ensure equivalence and validity. An important part of these procedures are judgemental item analyses, where items functioning differentially across countries are reviewed for e.g. translation problems and errors and those interpreted as biased are removed from the final analyses. However, research suggests that the analyses have often not been able to consistently and accurately identify sources of differential item functioning (DIF), including those related to translation.

The purpose of this study was to examine whether it would be possible to improve the ability of judgemental item analyses conducted in international reading literacy assessments to identify translation-related sources of DIF. A judgemental analysis was conducted of three Finnish items which in the PISA 2000 reading test were identified as DIF but for which no (clear) translation-related explanations were suggested. The items were compared linguistically to the corresponding items in the English and French source versions, at the same time taking into account the cognitive processes required to answer them. Clear translation-related explanations were found for all three items, suggesting that it is possible to improve the ability of judgemental item analyses to identify translation-related sources of DIF.

This, however, necessitates closer collaboration between content specialists and linguists, national teams and international test developers, testers and tessees. Suggestions are given as to how to develop not only judgemental item analyses but also other quality control procedures.

Jayanti Banerjee e.a. *Testing listening: exploring a “third way”*

A recurring debate in the design of listening tests is whether the listening input should be played once or twice. Persuasive construct arguments can be made on both sides of the question (Buck, 2009) and research to date has not helped to break the theoretical impasse (e.g. Fortune, 2004; Pranculiene, 2004). Stakeholders generally prefer to hear listening input twice and many language testing organizations replay the listening stimuli. However, playing a listening text twice is not necessarily theoretically sound. In particular, it fails to address a key point of contention, which is that when asked to repeat what they have said, speakers virtually always reformulate. Even when the same words are repeated, the intonation is invariably different. In fact, listeners are unlikely to hear exactly the same text twice.

In this project, researchers and test developers have collaborated to craft input texts that come in two complementary parts so that when the second input is played, it is a reformulation of the [content of the] first text. The poster will show how we developed a beta version of the listening items and will present the issues arising for the development of such test items. It will also illustrate how our work operationalises the listening construct and will initiate a discussion about the construct validation of such an approach to testing listening and suggest further opportunities for collaboration, particularly with stakeholders. Finally, we will provide access to a beta version of the listening items for colleagues to experience.

Eszter Benke *Workplace language use and a test task-based needs analysis*

The content of teaching and testing of languages for specific purposes is increasingly described in terms of the tasks and language activities in which language users engage in their work contexts. In a rapidly changing business environment language teaching and testing have to be responsive to the newly emerging challenges language users have to take on. After seven years of operation, the Foreign Language Examination Centre of the Budapest Business School perceived the need to revisit the issue of language needs in a test task-based needs analysis. The purpose of the survey was to conduct a needs-based validation of the test tasks of an operational LSP examination system to investigate the ‘fit’ between foreign language target tasks frequently performed in the specific work context and the LSP examination tasks. Workplace language use informed the survey carried out with more than 600 LSP language users in business, catering and tourism related jobs. In the questionnaire respondents were expected to indicate on a five-point Likert scale the frequency and relevance of the language activities in their work which the test tasks simulated. Additional open-ended questions inquired about common text types that these language users had to employ to perform their organizational responsibilities using a foreign language.

One of the most significant results is the apparent emergence of new discourse types and modes with new structures; a phenomenon which should be adequately addressed by the examination.

Sanja Fulgosi e.a. *Structural Analysis of the Croatian Language Domains: Examples from the Croatian National Assessment*

The Croatian educational system is facing the process of introducing student assessment system in order to determine information on the long term progress in the academic performance and factors that affect students' accomplishment. The cooperation between CITO and Croatian National Centre for External Evaluation of Education (NCEEE) represents the important progress in assessment within the international project Development of Instruments in Croatian National Assessment. The main objective of that project was the education and capacity building of NCEEE experts. The most important activity within the project was the assessment of educational achievements of the 2nd grade gymnasium programme students. Along with broadly specified goal, the aim of the project was to analyze sample data in order to determine some of the factors that affect language domain constructs and possible external factors influencing students' performance. The current research analyses the data on Croatian language performance in four tested domains: grammar, theory of literature, reading of literary and non-literary text. The results justify used division of tested domains. In addition, some demographic variables influence the students' achievement. One of the specific influences is the one of gender on the reading performance results, which could be derived from the thematic of used texts. The results of presented analysis could be used to improve the validity of instruments and quality of reading framework for the long term assessment as well as to provide considerable information on target student population.

Sébastien Georges e.a. *TCF a Test made by Crosslinking between FFL experts and psychometricians*

A new item for examinees of the "Test de Connaissance du Français" (TCF) gets already a long story for test makers. 1) Experts of French as a Foreign Language (FFL) request to item writers a set of items following specifications both of the CEFR and TCF. Back and forth between item writers and experts are done to reach the conformity expected. Then, items are stocked. 2) Psychometricians select 11 anchors from the item bank and set them in defined locations. Experts of FFL have to complete the 69 remained locations with the new item-stock to get a pre-test. 3) Pre-test is administered to a sample of at least 500 pseudo-examinees similar to real ones. Data are analysed through the Classical Test Theory and the one parameter Rasch model (IRT). 4) All items that do not fit the criteria are rejected ($\pm 20\%$), others are either banked with an anchoring procedure ($\pm 70\%$) or discussed with expert of FFL when they are borderline with the criteria ($\pm 10\%$). 5) Psychometricians controlling item exposure select a mix of new and older items from the bank and propose to the experts of FFL a combination of 80 items. Experts examine the linguistic shape of this combination and ask to psychometricians to change items where it is necessary. When the combination satisfies both experts and psychometricians respectively on the side of linguistic and the side of measure parameters the new version of the TCF is administered to examinees.

Erna Gille *The European Survey on Language Competences (ESLC) in the Netherlands*

The European Commission outlined The European Indicator of language Competence in 2005 and proposed a detailed strategic approach for the creation of a European Survey on Language Competences. The survey aims to provide member states with comparable data on foreign language competence and knowledge about good practice as well as progress towards the objective of improving foreign language learning. The survey will test the two most widely taught official European languages (English, French, German, Italian and Spanish) in each participating country from a representative sample of pupils in their final year of lower secondary education.

The survey will assess students' proficiency in listening, reading and writing. The results will be related to the Common European Framework of Reference (CEFR).

The survey will be carried out in 2011. The final report will be delivered to the European Commission in 2012.

The Field Trial took place in February – March 2010. In this poster I will outline the survey in some detail and present what the schools have reported to us about the survey.

Robert Marcz *ECL – Comparable Language Testing and the Use of Less Frequently Taught Languages*

The purpose of this poster is to introduce the European Consortium for the Certificate of Attainment in Modern Languages (ECL), the collaboration it coordinates in the field of language testing and how these activities contribute to the use and development of the less frequently taught (LFT) languages of Central Europe.

The Consortium, consisting of 13 partner institutions, and the cooperation between them are introduced. Examples of concrete collaboration in the field of language testing and test development are presented. The tests used for the ECL language exams have been developed according to the CEFR and with the help of the Manual. The tests are developed by native speakers living in the target language countries. This wide cooperation and collaboration in the field of test development as well as the common test specifications produce comparable language exams and certificates.

10 of the 13 languages of the ECL exam are Central European, LFT languages. It is an important aim of the Consortium to motivate the studying of these languages, which, as by-product of this activity, also contributes to the better understanding of nations living next to each other. Data concerning the number of candidates taking the ECL exam in the LFT languages is presented.

As a washback effect, communicative language books are also being developed to teach these languages.

The development of the ECL language exam system also requires the collaboration of policy makers of the Central European countries. Examples for such cooperation are presented.

Juan Carlos Rodriguez Pozas *Enhancing the Effectiveness of CAELTIC Training Programme through Collaborative Online Oral Assessment*

This poster will illustrate how the effectiveness of CAELTIC training programme is enhanced through the collaborative online assessment of oral foreign language proficiency in WebCEF.

Since the formation of Cuban PhD holders is currently a key priority in the VLIR-UCLV collaboration programme, language proficiency in English has become a prerequisite for academics and student mobility throughout Europe and a requirement to get involved in PhD programmes in Flanders. The CAELTIC (Center for Advanced English Language Training for International Collaboration) programme aims at building the capacity of the local members of the VLIR-UCLV programme to communicate in English for academic purposes. Accordingly, the Common European Framework of Reference for Language (CEFR) has been recognized by the local language teachers as a useful instrument to ensure a globally accepted proficiency in English of CAELTIC graduates and their optimal performance in the intercultural European setting. However, the daily practice of oral assessment within the CEFR but outside Europe raises an interesting question as to the reliability of the implementation.

WebCEF (www.webcef.eu) is a web-based online environment for cross-national and cross-border collaborative assessment of oral language proficiency within CEFR. The Caeltic group will describe their experiences in the use of this collaborative assessment tool, focusing on its contribution to the implementation of the CEFR beyond Europe, particularly in CAELTIC training programme and to the validation of the daily practice of oral assessment on the basis of collaboration with colleagues from European national communities linked into an international community of practice

Ingrid Radtke *Collaboration in developing a basic skills reading and writing test for immigrants*

The poster aims to present the development of a test in basic reading in writing skills for immigrants where the collaboration of different proficiencies is vital in order to achieve good validity in the test.

Vox, the Norwegian Institute for Adult Learning works to improve basic skills in the adult population in the areas of literacy, numeracy and the use of ICT.

As part of this, we have developed a Framework for Basic Skills for adults, ie. description of levels of competence for each of the basic skills, and on top of this, develop online tests that serve both as diagnostic and summative tests.

As it concerns the reading and writing skills, we see that we need a specified test that helps to discover the lack of these skills in the immigrant population. The challenge in developing such a test is that one has to divide between general reading and writing difficulties of immigrants and problems to understand the norwegian language.

Therefore we need to include both, experts who teach norwegian as a foreign language and experts who are knowledgeable when it comes to reading and writing difficulties. Both expert groups are involved in the item production by checking whether the content (on the basis of the developed Framework for Basic Skills) is appropriate for a certain language knowledge (on the basis of commonly used textbooks and teaching materials), or help to detect basic decoding problems in reading and writing skills. The poster will present how we organise this process.

Cathy Wendler e.a. *Providing Additional Meaning to TOEFL® iBT Reading Scores*

The TOEFL® iBT measures the ability to use and combine listening, reading, speaking, and writing skills to communicate in an academic environment (Chapelle, Enright, & Jamieson, 2008). The Reading measure is one of four sections on the test; examinees receive section scores as well as a total test score. Examinees also receive performance feedback that describes tasks that examinees within a particular score range are able to do. Beginning in fall 2009, information on the Lexile® measure was provided to examinees as part of their TOEFL reading performance feedback. The Lexile measure allows examinees to better understand their reading level.

The Lexile measure represents two dimensions: (1) a reader-level measure that indicates a person's reading ability on the Lexile scale and (2) a text-level measure that represents the difficulty of the text that is read. The same scale is used for both reader ability and test difficulty. When used together, they provide an estimate of the person's reading level and help determine the appropriate difficulty level of reading material. The higher the Lexile measure, the higher the examinee's level of reading ability.

A study was conducted this past year to link the Lexile scale to the TOEFL Reading scale. This poster will present the results of the study. It will also present information on what the Lexile scale represents, discuss how Lexile information is provided to TOEFL examinees, and explain how this information can be used by students and teachers to monitor reading progress.