

Dina Tsagari / Ildikó Csépes (eds.)

## Collaboration in Language Testing and Assessment



**PETER LANG**

Internationaler Verlag der Wissenschaften

## Foreword

**John H.A.L. de Jong<sup>1</sup>**

*Pearson / VU University Amsterdam*

The selection of articles in this book, dedicated to the memory of Felianka Kaftandjieva, has been gathered by Dina Tsagari and Ildikó Csépes around the theme of collaboration. Felly has collaborated with many of the authors in this book and has most certainly influenced them in their attention to detail, their respect for measurement rigour and their enjoyment in research discoveries. She taught us to savour the reward of finding meaningful results achieved by painstakingly searching for truth in data. The editors and the authors of this book by collaborating on its publication show their wish to honour her memory by bringing the lessons she taught us into practice.

The articles were originally presented at conferences of the European Association for Language Testing and Assessment (EALTA) from 2008 to 2010 in Athens (Greece), Turku (Finland) and The Hague (The Netherlands). The collection illustrates the international and inclusive character of EALTA, involving researchers in large scale testing as well as in classroom assessment.

As scientific disciplines develop they tend to diversify into a growing number of branches that specialise in ever smaller areas in order to study these areas in more detail. General linguistics as it developed around the beginning of last century subdivided into branches like psycholinguistics, sociolinguistics and applied linguistics around the middle of last century. The latter then again split up into the studies such as child language, second language acquisition and language testing. Language testing into listening, speaking, reading and writing. Studying at increasing levels of granularity is a necessity brought about by the expanding human knowledge, but occasionally we have to zoom out again to understand how our fragmentary understanding relates to the real phenomenological world. This reverse direction has also occurred within language testing where we have seen how initial discrete point testing was replaced by communicative four skills testing, which in its turn is currently being exchanged for integrated skills testing to get at the language how it is used in real life. Specialists in listening comprehension therefore find themselves having to collaborate with specialists in the testing of speaking, reading experts with writing experts. Similarly second language acquisition scholars seek collaboration with language testers.

With the ease of communication created by electronics and the relatively decreasing cost of international travel we also see increasing collaboration between scholars from research centres worldwide.

---

<sup>1</sup> John.DeJong@Pearson.com

Indeed the first chapter in this volume by Riikka Alanen, Ari Huhta, Scott Jarvis, Maisa Martin and Mirja Tarnanen exemplifies both these forms of collaboration. Four authors from the University of Jyväskylä, Finland collaborated with one researcher from Ohio University in the USA. They point at a lack of co-operation between language learning and language assessment research and contend that both fields come together in defining the construct of L2 proficiency. To illustrate their point they report on the CEFLING project, a study on the relationship between linguistic features of writing performance and functionally defined language proficiency levels. The functional description is taken from the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001). Their study aims to detect which linguistic developmental phenomena go in parallel with the set of levels of increasing proficiency defined in the CEF. Their study shows how our knowledge and understanding can be advanced by combining the depth and the attention to detail of SLA with the rigour and the larger numbers typically used in language testing.

In the next chapter Jamie Dunlea, based in Japan, and Neus Figueras, born and raised in Spain collaborate on finding out whether the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001) applies as well outside of its original European context as it does inside. For their study the authors have chosen to evaluate the relationship between the EIKEN tests developed in Japan and the CEFR. The EIKEN tests are a seven-level set of tests made and administered by the Society for Testing English Proficiency (STEP), a non-profit foundation established in Japan in 1963. In order to properly introduce their study they provide a thorough discussion of standard setting methodologies. Preliminary studies based on research in Japan had established a hypothesized relation between particular scores on the EIKEN tests and the CEFR. The present study used judgements from teachers based in Catalonia and found slightly higher cut-offs than the original cut-offs found in Japan, but within their predefined interval of acceptable difference. The authors conclude therefore that the difference is small and that the interpretation of the CEFR across the two cultures is closer than some critics of the usage of the CEFR outside the European context would have expected.

In the third chapter Karin Vogt starts out by discussing the quality of the set of performance and ability descriptors that came with the publication of the CEFR. She points at several inconsistencies and inadequacies and proposes a method for extending and improving the original set of descriptors to apply to the vocational sector. Therewith Karin introduces yet another form of collaboration, that is, collaboration between professional language testing and language testing in the professions. The chapter reports on developing new descriptors in three specific professional contexts and presents a thorough study using both qualitative and quantitative methods to validate the newly developed

descriptors. The chapter is an example of how the CEFR can be applied as intended: a framework of reference allowing development beyond the illustrative set of descriptors provided in the 2001 publication.

Carole Sedgwick reports in Chapter 4 how she lost her initial enthusiasm for the ideal of the Bologna Declaration (1999) which is to enhance transnational mobility of European citizens by creating comparability of degrees in higher education. Based on her research she fears the richness of local diversity might be threatened by the centralizing tendencies in the Bologna Declaration. She based her study on thesis writing as it is practised and valued in two European countries: Hungary and Italy and found that local academic traditions tend to be undervalued in an approach that attempts to homogenize the approach to thesis writing. Carole's study exemplifies the European dilemma between unity and diversity. Wishing to recognize the equality of languages, the European Union tries to withstand the natural development of English becoming a *lingua franca*. On the other hand, the Union's ideal of human mobility is assumed to require a homogenization of tertiary degrees across the European Union, but, as Carole points out, in thesis writing this leads to standardization on an Anglophone model.

In Chapter 5, Carole Thomas presents an example of post-hoc linking of an existing exam to the CEFR. The object of study is the Certificate of Proficiency in English (COPE) developed and administered by Bilkent University in Turkey. Engaging in this activity revealed initial uncertainty among teachers about the levels and required gaining greater knowledge and understanding of the CEFR. An important side-effect was that by involving teachers in the process of standardization they were trained on the CEFR. This collaboration has resulted in a growing number of teachers with experience and understanding of the CEFR.

A similar process of collaboration in the Swedish context is described by Gudrun Erickson and Lisbeth Åberg-Bengtsson in Chapter 6. In Sweden there is a long tradition of involving teachers in standardized assessment. In fact the teachers are entirely responsible for grading their students when exiting secondary education, but a national standardized examination is provided to assist teachers in this task. This examination is developed in close collaboration between teachers and experts. Students too are involved in piloting the exam tasks and in providing feedback on their experience with them.

Györgyi Együd, Zoltán Kiszely and Gábor Szabó in Chapter 7 present a critical discussion of the national school-leaving exam in Hungary. They compare this exam to two other exams accredited by the Hungarian state: the exam from the European Consortium for the Certificate of Attainment in Modern Languages (ECL) and Cambridge First Certificate in English (FCE). The results from administering the three tests to pilot groups suggest that the reading tasks on all three exams are quite similar in difficulty but that the

writing tasks on the school-leaving exam are significantly easier. These results are remarkable as the Hungarian Accreditation Board for Foreign Language Examinations (HABFLE) sets quite strict requirements on the alignment of international exams to the CEFR. Apparently the national school-leaving exam is not held to meet these requirements.

In Chapter 8, David Newbold describes how the University of Venice has managed to find a way to satisfy both the need for external recognition for English language exam results and relevance of the examination with respect to a local curriculum. As a result of the Bologna Declaration (1999) Italian Universities experienced the need to revise their curricula and for foreign languages they choose to model these on the CEFR. The revision of the curriculum for English required reviewing the examinations. They choose one of the internationally operating language test providers (Trinity) and collaborated with them to produce a localized version that is recognized both by the university and by the external test provider. Winner in this effort of co-certification is obviously the test taker, who taking a local exam, obtains an internationally recognized certificate at the same time.

Slobodanka Dimova in Chapter 9 presents an interesting record of the introduction of a new examination system in the Republic of Macedonia. As part of secondary education reforms the national “Bureau for Development of Education” decided to introduce a Matura exam to replace the traditional school-based examination. Although positive washback on instruction and enhanced credibility of the secondary school diploma were expressed as main goals, the introduction of the Matura also had a mixed set of further objectives including the control of teaching curricula based on educational standards. Objections against the Matura came from the general public and from students fearing a reduced probability of passing the exam. Dimova suggests that a major cause of the upheaval about the new exam was that authorities failed to involve stakeholders in the design and implementation of the innovation. From her study based on surveys conducted with students and teachers it would appear that these stakeholders felt there was a discrepancy between the curriculum as taught and the exam. She concludes that collaboration with stakeholders in the design phase could have been beneficial for a better understanding: the developers knowing more about the expectations of the students and the students realizing the intentions of the developers.

Chapter 10 provides a report by Anders Johannessen Fikke and Hildegunn Lahlum Helness from the University of Bergen, Norway on three parties involved in moving a computer-based test to a new platform. The authors discuss sources of tensions in this collaborative effort involving the contracting authority, the test developers and the software engineers. The chapter shows how ultimately success can be achieved by surmounting technological hurdles, but most of all by a cooperative attitude of parties involved.

June Eyckmans, Winibert Segers and Philippe Anckaert in Chapter 11 advocate the collaboration between translation trainers and test developers because they see the need to further professionalize translation assessment practices. With the advent of the communicative approach to language learning, teaching and assessment, translation went out of the window with grammar-translation method. Thirty years later we see a renewed interest in translation because of the needs of our international society. The authors describe the current situation in translation assessment and point at the shortcomings from a measurement perspective of the Translation Services-service requirements issued by the European Committee for Standardization. It is therefore that they call for standardization in measurement principles and methodology. In order to achieve this they suggest developing a Common Framework of Reference for Translation Competence and the adaptation of the EALTA Guidelines for Good Practice to also cover Translation testing.

Yet another new field is introduced in Chapter 12 by Laura Sadlier and Beppie van den Bogaerde who report on Irish and Dutch scholars in sign language (SL) collaborating on investigating the necessary adaptation of the CEFR to make it appropriate for learning, teaching and assessment of SL's. Clearly no adaptation is required in the context of the written skills, but for the spoken skills the use of the CEFR would require replacing the auditory-verbal concepts by visual-gestural ones. The authors describe current developments in creating assessment instruments for testing SL skills and present examples from tests developed in Ireland and in the Netherlands.

Chapter 13, authored by Anne Dragemark Oscarson, revisits collaboration between teachers and learners. The author reports on a study conducted in Sweden where students were required to self-assess their writing skills based on writing exercises and self-assessment instruments. Teachers and students appreciated the self-assessment, but correlations between teacher grades and learners self-assessment were moderate for overall writing (explaining one-third or less of the variance) and low for specific writing tasks. Nevertheless the self-assessment exercise is seen as positive as learners acquire greater insight in the learning task. In focus interviews students did express their appreciation and mentioned they became more aware of the learning process and its requirements. In addition, the students indicated that the self-assessment equipped them with life-long learning skills.

In the last chapter, Chapter 14, Lisbeth M. Brevik and Eli Moe from Norway describe how researchers benefitted from their collaboration with teachers to investigate the effect of Content and Language Integrated Learning (CLIL). The researchers allowed teachers and students to choose whether they wished to participate in a CLIL or in a control group. Parallel tests of listening and reading were developed and administered one at the beginning of the experiment and the second at the end of the same school year. Both groups