Sauli Takala, Bernd Voss

**Guest editors’ introduction**

Erstveröffentlichung in / First published in:


DOI: [https://doi.org/10.1515/cercles-2014-0001](https://doi.org/10.1515/cercles-2014-0001)

Diese Version ist verfügbar / This version is available on:

[https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-712806](https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-712806)

„Dieser Beitrag ist mit Zustimmung des Rechteinhabers aufgrund einer (DFGgeförderten) Allianz- bzw. Nationallizenz frei zugänglich.“

This publication is openly accessible with the permission of the copyright owner. The permission is granted within a nationwide license, supported by the German Research Foundation (abbr. in German DFG).

[www.nationallizenzen.de/](http://www.nationallizenzen.de/)
This special issue of *Language Learning in Higher Education* is devoted to the field of language testing and assessment, an area often underrated in higher education, where other concerns tend to be more in the focus of attention. Our call for papers made clear that our aim was “to cover a wide range of interrelated themes, in theory and practice, such as assessment and self-assessment, formative and summative assessment, performance standards and standard setting, use and impact of tests, tailoring and developing tests for special purposes, backwash effects (desirable/undesirable), quality issues and ethical concerns. Also considered would be contributions dealing with programme assessment and evaluation . . .” In other words, we were inviting contributions from a wider range of perspectives than is often associated with this field. As a result, the 12 articles selected and presented here cover a rather wide variety of issues often more concerned with the users of language tests, i.e. with those who have to apply them, to develop them within their own institutional constraints, and to interpret and defend the results, than with full-time researchers talking to full-time researchers.

The articles fall quite naturally into four broad categories. We begin with four articles dealing with tests, their design and their implications. This is followed by three articles focusing in various ways on the issue of comparability across languages (and institutions) when using in-house tests. The third group, again comprising three articles, deals with current issues such as language assessment literacy or the introduction of a quality management system to support the development and use of in-house tests. And finally there are two articles by test providers allowing insights into the deliberations and decision-making processes that provide the background to their test designs.

In more detail: MaríA Pilar Alberola reports on a study of ESP assessment in speaking, in a study programme in tourism management. The aim of the programme is a maximally close integration of instruction and assessment. This kind of continuous, formative assessment (assessment for learning) is less commonly reported, and the article is a welcome contribution to redress the balance. The author focuses in particular on the orientation of the programme and the

Sauli Takala and Bernd Voss

Guest editors’ introduction

DOI 10.1515/cercles-2014-0001

MaríA Pilar Alberola: University of Jyväskylä (retired). E-mail: sjtakala@hotmail.com

Bernd Voss: Technische Universität Dresden. E-mail: bernd.voss@mailbox.tu-dresden.de
approach adopted in developing the assessment system (test formats, rating criteria and scales) while less attention is devoted to the empirical testing and validation of the system. Accordingly, the article is a promising pilot study that invites replication on a larger scale.

Satu Tuomainen’s article addresses the situation in higher education in Finland in which all students have to pass an LSP test in whichever official language of the country (Swedish or Finnish) is not their first language and a test in a foreign language (usually English). As universities face serious resource problems, LSP teaching and testing are under strong pressure to economise. Tuomainen reports on the option of students gaining exemption from LSP study by passing a pre-instruction test – presumably there is also pressure in other countries to use tests to avoid unnecessary teaching. While the study is small-scale, it shows how demanding the development of such an exemption system is, and it provides a promising model for the design of this kind of testing option. It could also serve as a useful basis for a larger-scale validation study.

Antonios Tsopanoglou, S. George Ypsilandis and Anna Mouti report on a pilot study which explores the awarding of partial credit to plausible options (partially correct distractors) in multiple-choice language test items. Topics addressed deal mainly with reliability and the potential consequences (e.g. pass/fail) for individual test takers. The outcomes using the traditional dichotomous scoring method (correct option = 1, incorrect option = 0), an experimental scoring method (correct option = 2, highly plausible option = 1, plausible option = 0, clearly incorrect option = −1), and a negative scoring method (correct option = 1, incorrect option = −1) were compared. On the whole, the correlation between the methods was very high (.959–.999), but a more detailed analysis at the individual level suggested some differential impact. For this reason, a larger-scale follow-up on the use of polychotomous scoring would be of considerable interest to the language testing and assessment community, and to language teaching in general.

Finally in this group of articles, Yuyang Cai offers a strongly methodological investigation in which an approach based on a sophisticated bi-factor multidimensional item response theory is used to compare two theories of grammatical knowledge. One theory (Bachman 1990) is claimed to conceptualise grammatical knowledge as vocabulary and syntax, while the other (Purpura 2004) is considered to combine grammatical form and meaning. In an empirical study involving 1,491 second-year nursing students in China, a 15-item multiple-choice grammar test (interpreted as corresponding to A2–B1 in the CEFR reference levels) was administered. Although the number of items was limited, it was judged to be sufficient for the purpose. Several models were explored and the empirical evidence is interpreted as supporting the appropriateness of subdividing grammatical knowledge into form and meaning. The author concludes, however, that further
studies are needed in order to deepen our understanding of the issue and to provide a better foundation for developing grammar instruction and assessment.

In the second group of articles, Karl-Heinz Eggensperger reports on an ongoing initiative and resulting projects that are developing C-tests for placement purposes, i.e. to assign students to the appropriate proficiency level in UNICert®-based language teaching programmes across different universities. Using French as a foreign language as his example, he explains how it is possible to arrive at common cut-off points between the levels and how C-testlets can be selected and calibrated to serve as anchor items for more reliable and standardised placement decisions. In this way, learners said to be at the same level of comparable programmes in different institutions can be certain that they are so (and shown to be so). Amongst other things Eggensperger’s article demonstrates, as one of the reviewers observed, how in a collaborative effort item banking is feasible also when sophisticated psychometric methodology is not available (or deliberately avoided).

A higher degree of comparability across languages and institutions is also the driving force behind the project reported by Astrid Reich and Mirka Mainzer. They describe an attempt to use the same writing task in different languages and to develop comparable grading systems in a succession of collaborative sessions devoted to task design and the development of suitable evaluation grids. The article provides informative insights into a process based on collaborative consensus formation between professional language teachers when designing, implementing and evaluating an examination task which is appropriate for the relevant language level and has been and can be used in different languages and institutions. As Reich and Mainzer show, there are no shortcuts here: working towards higher degrees of comparability will remain an ongoing process, but the process itself can also be seen as a significant move in this whole undertaking.

Comparability across languages is also a key issue in the article by Anna Nunan. She makes a strong case (and plea) for a task-based approach to language learning flanked and supported by equally task-based assessment procedures, not only as a generally desirable approach to language learning and teaching in higher education but also as a way of overcoming the difficulty of finding comparable frameworks of course design and assessment formats for all the languages taught in an institution. While there is usually no problem in treating European languages in comparable ways, this is much more difficult when it comes to languages with non-alphabetic writing systems such as Japanese or Chinese, where reading and writing tasks make special and different demands on the learner. Nunan outlines in some detail the implications of testing task performance and language use (rather than language knowledge) and the adjustments that need to be made across languages, and sets out the conditions under which task-based
assessments could be standardised even across languages of different degrees of closeness to the L1 of the learners. Given that “student assessment should improve performance, not just audit it”, this seems to be a path worth pursuing even if the present report on work in progress is more at the stage of identifying needs than providing all the answers.

The third group of articles focuses on framework conditions for the use of language tests. Beverly A. Baker, Rika Tsushima and Shujiao Wang explore how confident people other than language assessment specialists are in using and interpreting tests and test scores. Language assessment literacy, as it is called, is an interesting and topical issue, e.g. for university admissions officers. At a time of increasing internationalisation, applicants from other countries present universities (the example here is from Canada) with a great variety of tests and test scores (cl)aiming to show that they have sufficient command of the language of instruction to be able to follow their chosen university programme. Universities need to take transparent and defensible admission decisions. The article reports on a survey of the information and awareness levels of Canadian decision-makers, as part of a larger project aiming to provide appropriate guidance, workshop materials and training. The authors draw attention to the need for the language testing community to help non-specialists to interpret and thus make proper use of tests and test results. The initiative they report is certainly worthy of support, not only in its native country but also as a model for universities worldwide.

Olga Kvasova and Tamara Kavytska address the assessment competence of university foreign language teachers in the Ukraine. The impetus for the study derived from the need felt in the universities to develop assessment practices in line with what the Bologna process was judged to require. In general terms, this was considered to presuppose that assessment literacy needs were to be ascertained and training to be made more professional. In order to have a better foundation for development work, the authors replicated a European survey of language testing and assessment needs and supplemented it with a more tailor-made survey. While some obvious problems were identified in the provision of training, the results are interpreted to indicate that the current level of assessment literacy is relatively satisfactory. This is a pioneering study in the Ukraine and can undoubtedly make a contribution to developing syllabi for pre-service and in-service teacher training in language testing and assessment. The study illustrates how assessment literacy can be approached in a situation in which there is no previous local development work to draw on.

The topic of Giuliana Grego Bolli’s article is the use of a quality management system (QMS) to improve test development. The empirical work she reports focuses on the Certificates of Italian Language (CELI), originally created in the early 1990s at the Università per Stranieri di Perugia. Since 1990, the development
of the test has been linked to the cooperation in language testing and assessment co-ordinated by ALTE (Association of Language Testers in Europe). This is described in some detail, and the QMS procedures followed are considered to have made a substantial contribution to enhancing the quality of the test. The article provides a useful case study of how an existing test can be further developed by a systematic process in which quality concerns are highlighted. Results of the validation of the test are reported in an appendix by providing analysis data in the form of several tables, including rating procedures.

Finally, there are two articles by test providers that yield interesting insights into the reasoning and decision-making processes that lie behind their test designs. In the first, Rudi Camerer is concerned with the thorny problem of testing “intercultural competence”. Starting from a detailed (and very critical) analysis of the personality-based tests that are frequently used but basically unsuitable for this purpose, he suggests as an alternative that a practicable construct of intercultural communicative competence can be found using a performance-oriented approach instead, taking the notion of politeness in international encounters as a key issue and focusing on English as a lingua franca rather than as the language of English-speaking countries. As his outline of a criterion-based test of intercultural competence in English demands and shows, “politeness comes before grammar” (although language accuracy is not neglected), and it is worth noting the admission that not everything that is worth knowing might be testable.

In the second of these last two articles, Sibylle Plassmann and Beate Zeidler take the reader step by step and in considerable detail through the questions, the reasoning and the eventual decisions that shaped a new test of German for academic purposes aimed at foreign applicants for admission to German universities and developed by the institution the two authors work for. The article is more a workshop report than a research paper in the strict sense of the term. It does not, for example, attempt to engage in a discussion of comparable efforts in other countries where the need to ascertain prospective students’ command of the language of instruction presents a similar challenge, and where the same problems hold. But it provides an instructive overview of and insight into the problems and challenges of test design, and it is for this information value to expert and non-expert alike that we offer it to our readers here.

Looking at the articles collected and presented here in retrospect, it is clear that they cannot claim to be a comprehensive (let alone representative) sample of what is currently going on in the field of language testing and assessment in higher education in general, and in language centres in particular. Rather, they provide a snapshot impression of current discussions, issues and concerns. Still, a few general notes and comments may be in order.
We note that in a fair number of cases we have to do here with pilot studies, and we would like to offer as one explanation for this the fact that language teaching staff in higher education are typically employed full-time to teach language and are therefore measured by the results of their language teaching today, rather than by the impact their research might have on language teaching and testing tomorrow. As a result, research is often of secondary concern (or even left entirely to an individual’s private initiative and time), and there may be little room for large-scale research projects that go beyond piloting.

It is worth observing in this context that interesting pilot studies are not often followed up in large-scale replication projects. It seems that in the academic world in general (except perhaps in medicine) replication studies, i.e. studies testing interesting pilots on a larger scale to see whether promising preliminary results can be replicated and more confidently generalised, do not have a sufficiently high prestige value to attract much interest and support. This seems wasteful and a pity, and pilot studies such as the ones presented here would deserve a closer look by institutions and organisations capable of carrying out larger projects.

Secondly, a common interest and concern is raised in a number of articles showing that language teaching programmes and the in-house tests attached to them try to reach out beyond their special niche to achieve higher degrees of comparability across languages and institutions. Often, the crucial step forward is to organise collaborative consensus formation, i.e. to implement networking and iterative procedures as bottom-up modes of action, rather than working top-down, e.g. by using tests offered by external providers. This seems a healthy sign as it helps to ensure a close link between teaching programmes and test content and construct on the one hand, but now with ranges of acceptability that go beyond the individual language, programme or institution, on the other.

A third very topical concern raised in these articles is that language tests and their results are used not only by language teachers and their learners, but also in more abstract and sometimes quite serious ways by decision-makers not necessarily familiar with this field. This should serve as a useful reminder that it is not enough for experts to understand language tests: efforts can and must be made by the language testing community to provide non-experts too with the necessary background and information, and thus to help users of tests and test results to make informed decisions.

Finally, what may also be special about this special issue is an article from Ukraine, a country from which we do not usually hear very much outside its own borders, and where all the changes the Bologna process has set in motion seem to be happening at the same time.

We would like to thank the authors of the articles, our reviewers, and the editors-in-chief for their efforts and support in producing this special issue. It has
been a pleasure reading many interesting manuscripts and seeing that assessment issues are not the esoteric concern of a handful of experts best left to their own devices, but affect everyone in the field of language learning and teaching in higher education, and beyond. We hope that in this special issue on language testing and assessment readers will find many points of interest, information, advice, and perhaps guidance, but in particular inspiration and encouragement for their own work.

References


Bionotes

Until his retirement Sauli Takala was professor of Applied Language Studies at the University of Jyväskylä, Finland. His work has covered language testing and assessment, curriculum development, and language policy and planning. He was co-ordinator of the DIALANG project during its first stage, a member of the advisory group for the development of the CEFR, and a member of the team that developed the Council of Europe’s manual for linking tests and examinations to the CEFR. He is a founding member of EALTA and served as its second president. Currently his work focuses mainly on projects involved in linking examinations to the CEFR.

Until his retirement Bernd Voss was professor of English Language, Literature and (Foreign) Language Teaching at Technische Universität Dresden, Fakultät für Sprach-, Literatur- und Kulturwissenschaften. His work has focused on language teacher training and education, language learning and teaching (in particular in higher education), language processing, language testing and assessment, classroom language and communication, and media in language teaching. He has served as board member and/or chairman of various organisations, including the European Confederation of Language Centres in Higher Education (CercleS), the Deutsche Gesellschaft für Fremdsprachenforschung (DGFF), and the Arbeitskreis der Sprachenzentren (AKS). He is a founding member and honorary president of UNIcert®.