

## María Luisa Carrió-Pastor, ed.: New Challenges for Language Testing: Towards Mutual Recognition of Qualifications

Tove Larsson

To cite this article: Tove Larsson (2017) María Luisa Carrió-Pastor, ed.: New Challenges for Language Testing: Towards Mutual Recognition of Qualifications, *Studia Neophilologica*, 89:2, 312-317, DOI: [10.1080/00393274.2017.1354719](https://doi.org/10.1080/00393274.2017.1354719)

To link to this article: <https://doi.org/10.1080/00393274.2017.1354719>



Published online: 25 Aug 2017.



Submit your article to this journal [↗](#)



Article views: 9



View related articles [↗](#)



View Crossmark data [↗](#)



## REVIEW

María Luisa Carrió-Pastor, ed.: *New Challenges for Language Testing: Towards Mutual Recognition of Qualifications*. Newcastle upon Tyne: Cambridge Scholars Publishing, 2016. ISBN: 1443802069. 180 pp.

This edited volume provides new insight into test development and accreditation in foreign-language assessment. Aspects such as testing strategies and student motivation are also discussed in relation to the main theme. The book includes ten chapters that all address different aspects of language testing. The book is divided up into two parts, each comprising five chapters, the first focusing on test development and the second on accreditation in foreign language teaching (FLT). The book also includes an introduction written by the editor, María Luisa Carrió-Pastor, where the chapters are introduced. In the present review, a brief summary of the chapters will be given first, along with a short evaluation of each chapter; this will be followed by an evaluation of the volume as a whole.

In the first chapter, “Development of bilingual and monolingual English-for-medical-purposes exams”, Anita Hegedus compares two oral sub-tests in English for Medical Purposes that were carried out at the University of Pécs in Hungary: one bilingual, for which the instructions are in Hungarian and the test itself is in English, and one monolingual, which is entirely in English. The study aims to investigate the role of the first part of the sub-tests (which is an introductory conversation), the source language of the input and the assessment of each of the tests. The material is made up of a random sample of 100 mark sheets at level B2 (see the *Common European Framework of Reference for Language*, Council of Europe, 2001) from the bilingual test. However, Hegedus explains that “[a]ssessment sheets from the monolingual exam were not included in the study because only test exams have been carried out, and thus a sufficient sample of mark sheets was not available” (p. 7).

The results show that the mean score for the introductory conversation for the bilingual test was higher than for the two remaining parts of the test, and that the correlation between this part of the test and the total score was somewhat weaker than for the other parts. The differences between the scores for all the different parts of the test were furthermore statistically significant. Based on these results, the author concludes that “this task [the introductory conversation] is not a valid indicator of measuring speaking skills in English for Medical Purposes” and that the sub-test “impacts negatively on reliability” (p. 10).

However, the results and conclusions drawn would perhaps have benefitted from a more in-depth discussion, as it might not be immediately clear to readers whether the above-mentioned (relatively strong) claims are supported by the statistical tests presented in the chapter. Furthermore, while Hegedus carefully investigated the role of the introductory conversation in relation to the rest of the sub-test for the bilingual test, no results were presented for the monolingual test, due to scarcity of data. The two tests could therefore not be compared. It might thus have been preferable to have changed the title and aims of the study to reflect this, thereby enabling more attention to be paid to the bilingual test throughout the study, as this was an investigation that yielded promising results.

The second chapter, by Marta Conejero López, also addresses a test of oral proficiency in an English for Specific Purposes (ESP) context. This chapter is entitled “Speaking skills testing for business administration undergraduates: How to assess persuasive speeches in B1 Business English courses”. In this work-in-progress report, Conejero López gives an account

of one way in which Business English students' speaking skills (in particular with regard to persuasiveness) can be tested and assessed during a ten-hour course module. The materials were developed for students at Universitat Politècnica de València in Spain.

The course module described is intended to help the students improve their persuasiveness and covers preparatory work, a presentation, self-assessment and tutorials. In preparation for the test, the students study relevant vocabulary and phraseology in class; they also watch a short video, where central persuasive strategies are introduced. To prepare the students for the self-assessment, some marking rubrics are presented and discussed.

The test itself involves students preparing and giving a three-minute "persuasive speech" (p. 18). The speeches are video recorded, and the recording is to be submitted along with the script for assessment. The students subsequently assess their own performances. As the main objective of the course module is for the students to "gain confidence with speech content, speech quality and persuasive strategies choices" (p. 20), aspects like overall fluency and grammar only receive limited focus. Towards the end of the module, tutorials are carried out, where the students are asked to discuss the results with their teacher. According to Conejero López, expected benefits of this course module include improvement of students' persuasive speech production and increased student motivation.

In this chapter, the author not only provides an inspirational account of the module design but also shares her materials and links, which will no doubt be useful for Business English teachers around the world. However, since the chapter aims to present on and share teaching materials for a course module rather than to present the results of a study, a slightly different structure would perhaps have suited the paper better; the current IMRD (Introduction, Method, Results, Discussion) structure leads the reader to expect presentation and discussion of actual (rather than expected) results).

Chapter 3 is called "Measuring linguistic competences through Erasmus+ Online Linguistic Support (OLS): Benefits and drawbacks" and is written by María Boquera Matarredona. As explained by the author, OLS is an initiative taken by the European Commission that enables exchange students in the Erasmus+ program to assess their language skills before and after their stay in the host country. The test is compulsory for all Erasmus+ participants and is made up of 65 multiple-choice or gap-filling questions. Having described the test, the chapter reports on advantages and disadvantages of it.

Several advantages are reported. For example, the test "stimulates and encourages learning languages before and during mobility" (p. 43). Moreover, the test increases the participants' self-awareness, as they get feedback on where their linguistic strengths and weaknesses lie. At an organizational level, the author states that the test provides data on European students' linguistic performance to governments and Erasmus agencies that can be used for research. The author draws two conclusions based on the data from the test: (i) English is still the main language studied in Europe, and (ii) the exchange students who have taken the test "have already achieved quite a reasonably good level" (p. 46). With regard to disadvantages, she points out that the test is not timed or controlled, which means that students can ask someone else for help (43). Nonetheless, the author concludes that OLS "greatly contributes to the fulfilment of the [Erasmus+] objectives" (p. 46).

With its well-structured format, this chapter is both reader-friendly and informative. It provides useful background information about the test with illustrative examples of what the test questions look like. However, whereas some results were presented from the test, the discussion of its advantages and disadvantages does not seem to have a clear empirical basis. More explicit referral to the test results would perhaps have served to further strengthen the discussion.

The fourth chapter, "Assessing writing for higher education: Time to transform?" is written by Elaine Boyd. The chapter reports on a UK fellowship scheme that "uses fiction writers to support students in their academic writing" (p. 47). The main aims of this model

are to put more emphasis on the development of coherence and to enable students' "voice" to come through more clearly by using storytelling techniques.

The chapter begins with a description of guidelines and assessment criteria that are currently used in the United States and the United Kingdom, and the author concludes that these are not only mechanical, but also not in line with what subject tutors typically require. For example, these criteria do not encourage students to develop their voice, or an academic identity.

As an alternative, the author proposes revised assessment criteria through which storytelling techniques are applied to academic writing (the rationale for suggesting criteria being that "teachers teach to the mark scheme" (p. 55)). Among other things, these criteria would allow for more focus on progression, where information about how far students have come is provided; this way, the author states, the students are not "being judged against an end model, which seems long distant at the start" (p. 55), thereby also serving to increasing students' confidence. The criteria would also enable students to develop their authorial voice.

Boyd has written a well-structured chapter addressing a topical theme. In doing so, she questions the sometimes-rigid norms and practices applied around the world, thereby providing interesting new perspectives on how academic writing could be taught more effectively. As the approach described is likely to be of interest to many practitioners and curriculum designers, it would, however, have been useful if the model had been described in more concrete terms to add some clarifications. For example, does the author suggest that *only* storytelling techniques should be taught in writing classes, or is the proposed model a complement to more traditional techniques and assessment criteria?

In Chapter 5, María Luisa Carrió-Pastor discusses peer assessment and motivation in her study entitled "Should peer assessment be included in foreign language testing? The role of motivation in testing". Peer assessment is here defined as involving "the grading of the work of other students" (p. 61). The study aims to explore (i) a new way of assessing students' English proficiency, (ii) the "interrelationship" between peer assessment and motivation and (iii) whether peer assessment increases students' motivation.

To investigate this, 30 students (out of a total of 60 students) were selected as peer assessors, based on their "language skills and motivation" (p. 66); their job was to assess 60 oral presentations together with two instructors in an English for Specific Purposes (ESP) course at the Universitat Politècnica de València in Spain. The grading criteria were made available to all students before the oral presentations and covered delivery, content, organization and language. Following Panadero et al. (2013), the students' assessments were subsequently compared to that of the two instructors. After the presentation, the students were asked to fill out a questionnaire where questions about their motivation were asked.

The results show that the student assessors gave the presentations higher scores on average than the instructors. With regard to motivation, a majority of the students marked that they agreed or strongly agreed that peer assessment increased their motivation. These results lead the author to conclude that a combination of peer assessment and instructor assessment should be an integral part of foreign language assessment.

In this chapter, Carrió-Pastor provides a clear and well-described account of how peer assessment can be used in an ESP context. She thereby adds to the growing body of research advocating student involvement in the assessment process. As the results show great promise, it would be most interesting to see if future, more large-scale investigations could confirm the results.

The sixth chapter opens the second part of the volume, where accreditation requirements and needs are addressed. It is written by Gillian Mansfield and is entitled "The feeling's mutual? Reflecting on 'mutual' as key word in (the) context of fostering language centre collaboration and intercultural competence". Here, the author explores "the ways in which

the European Confederation of Language Centres in Higher Education (CercleS) works in mutual agreement and recognition of its members' work" (p. 77); she then proceeds to give suggestions for how CercleS can "extend mutual recognition further in the concept of the other" (p. 78).

After having explored the semantics of the word "mutual", Mansfield discusses CercleS in relation to the Council of Europe and to European language policy. She then brings up English as a Lingua Franca in business contexts (BELF) as a possible alternative model for how to view non-native speakers of English. Concepts such as "intercultural communicative competence" and "intercultural competence" are also discussed.

Based on this discussion, Mansfield emphasizes the need for increased awareness of "the other", in the sense that participants in intercultural encounters should focus on acceptance of cultural differences, rather than imposing "one's own as the expected norm" (p. 97). Based on this, she suggests that a new CercleS focus group should be implemented with the aim of better integrating intercultural competence in language classes. Such a group would, among other things, "further a mutual understanding of the other" (p. 99).

In this chapter, Mansfield argues convincingly for the need for increased focus on intercultural competence in EFL teaching. In an inspirational manner, she provides a helpful overview of CercleS and its mission and discusses concepts of great relevance to language teachers.

The seventh chapter, "Local and global accreditation needs: Quality, sustainability, and the role of the CEFR", is written by Neus Figueras. It discusses the Common European Framework of Reference (CEFR) and the impact it has had on local assessment systems. The chapter also addresses potential challenges for such systems with regard to sustaining quality over time.

The chapter begins with a general overview and background of the CEFR, where it is, among other things, stated that this framework aims to reconcile "two apparently divergent ends in Europe", namely diversity and standardization (p. 109). The author subsequently groups different kinds of exams offered in Europe into categories based on the purpose of these exams (for receiving a degree vs. a language certificate, etc.).

As pointed out by the author, there are, however, threats to the longevity of any such exam systems. The first threat mentioned is financial in character; since it is a challenge to keep a project going after the initial enthusiasm has faded, permanent staff and budgets are necessary. The second threat is political, as policies issued can change the original direction of any projects.

Through this chapter, the reader obtains a helpful overview of the CEFR and how it is used at institutions around Europe. In addressing possible issues pertaining to sustaining exam systems, the author also adds valuable suggestions as to how to avoid these issues from threatening the systems' continued existence, which will most likely be of help to practitioners and administrators alike.

The eighth chapter, written by Oksana Polyakova and Julia Zabala, is entitled "Comparative analysis of the state testing system in the Russian language for foreigners and the language accreditation model for the Spanish association of language centres in higher education: Towards mutual recognition". This chapter compares two language examination models: the *State Testing System in the Russian Language for Foreigners* (TORFL) and the *Language Accreditation Model for the Spanish Association of Language Centres in Higher Education* (CertAcles). In doing so, the authors aim to "introduce a proposal for mutual recognition" to overcome "barriers for academic mobility" (p. 120).

The chapter starts with a description of the two tests, and goes on to discuss differences and similarities between them. The authors note, for example, that both tests are proficiency exams whose results are officially recognized in the respective countries. However, certain differences are addressed, too. For example, whereas TORFL is used to test non-native

speakers' Russian proficiency, CertAcles can be used for several different languages. Nonetheless, the authors conclude that "the similarities between both models are more significant than the differences" (p. 137) and propose that mutual recognition of these tests by both countries would benefit "much needed exchange between students and researchers from Spanish and Russian academic institutions" (p. 137).

While the chapter provides an interesting comparison between these two tests, and the authors rightly conclude that mutual recognition seems advantageous for both countries, the chapter would perhaps have benefitted from more discussion of why these particular tests (and countries) were chosen for evaluation. Such justification would not only have served to strengthen their claims, but it would also most likely have broadened the applicability of the results.

The ninth chapter is entitled "Current trends in e-testing: The case of the eLADE – the University of Granada B1/B2 online Spanish accreditation exam". It is co-authored by Aurora Biedma Torrecillas, Lola Chamorro Guerrero, Alfonso Martínez Baztán, Adolfo Sánchez Cuadrado, Sonia Sánchez Molero, Steven Sylvester, César Amador Castellón, Jesús Puertas Melero and José Rodríguez Vázquez. The chapter describes an online test, the eLADE, that is described to be "completely reliable" (p. 141). The test is aligned with the CEFR and is recognized by both ACLES (the Spanish Association of Higher Education Language Centres) and CercleS.

The test assesses reading and listening comprehension, as well as written and spoken production and interaction at B1 and B2 level. The chapter begins with an overview of the scales and descriptors used for the test, followed by a description of the test. The test has to be taken at an institution where the test-taker's identity can be confirmed. The chapter concludes with a brief account of the grading protocol.

High reliability results are reported for the test, meaning that the test measures what it is supposed to measure; relatively high discrimination scores are also reported, thereby suggesting that the test can be used to successfully "differentiate between candidates of higher and lower language proficiency" (p. 150).

In this chapter, the authors provide a detailed account of the eLADE that will most likely be helpful for test developers and policymakers alike. However, while a description of what *kinds* of questions and tasks are included in the test, no actual example questions are provided (in fact, administrators are asked to sign a confidentiality agreement, thereby agreeing not to disclose any questions, p. 148), which makes it slightly difficult to evaluate the test's potential usefulness for other settings and CEFR levels. The chapter could perhaps also have benefitted from more detailed discussion of what sets this test apart from other, similar tests.

The tenth, and final, chapter is written by Cristina Pérez-Guillot and Asunción Jaime Pastor and has the title "Analysis of B2 listening tasks in UPV CertAcles certification exams". The chapter describes and discusses the listening comprehension part of the CertAcles test (the certification developed by the Spanish Association of Higher Education Language Centres, ACLES).

The authors discuss previously used language descriptors and taxonomies for listening skills, and go on to provide an overview and an analysis of test scores from the B2 listening comprehension test that they have developed. This test is, like the CertAcles exam as a whole, described as being "based on the CEFR descriptors" (p. 163). This comprehension test includes multiple-choice questions, "multi-matching activities" (where the students are asked to select the right option from a longer list) and sentence completion exercises.

Based on the results of the analysis, the authors note, for example, that aspects such as task layout and format can affect the test-taker's results, which lead them to conclude that the instructions "should be formulated as clearly as possible", preferably using lexico-grammatical features that are typically attained at a lower level than the level

evaluated (p. 171). The order in which the tasks are presented is also shown to potentially have an impact on the students' test results.

The authors situate the study well vis-à-vis previous research, and the reader is provided with detailed information about the test. However, the analysis part of the chapter could perhaps have been devoted more space to allow for a more thorough description of the method(s) used; the authors draw several interesting conclusions, but they seem to be based merely on the distribution of the test scores investigated, and the reader is therefore left with many questions with regard to how these can show, for example, that the task layout has an impact on the students' test results.

As the individual chapters have already been evaluated briefly in the previous section, this section will be devoted to a brief evaluation of the volume as a whole, starting with two critical comments. Some of the many strengths of the book will subsequently be addressed.

First, the majority of contributions primarily discuss language testing and accreditation in a Spanish context. Since one can expect the needs and challenges of language testing to vary somewhat across L1s and countries, it would perhaps have been preferable if the Spanish focus were reflected in the title or more explicitly addressed in the introduction and description of the volume. Second, a different ordering of the chapters would have helped the reader attain a better overview earlier on. One could, for example, have chosen to start with the more general chapters currently placed in the second half of the book, as many of the first chapters refer to associations and frameworks introduced and discussed in these chapters.

However, these minor weaknesses do not diminish the value of the volume; some of its many strengths will now be discussed. First of all, the editor has managed to bring together authors working on a wide variety of projects, using many different methods and metrics; the volume thereby paints a multifaceted picture of language testing and what challenges lie ahead. Second, the volume holds together in that many themes are echoed in several chapters of the book. One such theme is the importance of increasing student confidence in testing situations (discussed in Chapters 2 and 4), which is an often-neglected aspect of assessment.

All in all, the volume makes a strong contribution to the important discussion of how best to assess language proficiency, which is of great interest to the field. The authors have definitely succeeded in fulfilling the aim of exploring "new ways of testing and implementing assessment" (p. vii). The volume covers descriptions of an impressive number of new and innovative tests and methods, along with more general discussions of testing and accreditation. It will no doubt be of interest to universities, policymakers and individual researchers. With its predominately empirical basis, the book furthermore has practical uses, for example for foreign-language teaching practitioners.

This review appeared originally in the LINGUIST List at <http://linguistlist.org>.

Tove Larsson

*Uppsala University, Sweden*

 [tove.larsson@engelska.uu.se](mailto:tove.larsson@engelska.uu.se)  <http://orcid.org/0000-0002-0489-2697>

© 2017 Tove Larsson

<https://doi.org/10.1080/00393274.2017.1354719>



## References

- Council of Europe. 2001. *Common European framework of reference for languages: Learning, teaching, assessment (CEFR)*. Cambridge: Cambridge University Press.
- Panadero, E., M. Romero & J.-W. Strijbos. 2013. The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation* 39, 195–203.