

Dina Tsagari / Ildikó Csépes (eds.)

Collaboration in Language Testing and Assessment



PETER LANG

Internationaler Verlag der Wissenschaften

**Bibliographic Information published by the Deutsche
Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the
Deutsche Nationalbibliografie; detailed bibliographic data is
available in the internet at <http://dnb.d-nb.de>.

Cover Design:
Olaf Glöckler, Atelier Platen, Friedberg

ISSN 1612-815X
ISBN 978-3-631-63529-2

© Peter Lang GmbH
Internationaler Verlag der Wissenschaften
Frankfurt am Main 2012
All rights reserved.

All parts of this publication are protected by copyright. Any
utilisation outside the strict limits of the copyright law, without
the permission of the publisher, is forbidden and liable to
prosecution. This applies in particular to reproductions,
translations, microfilming, and storage and processing in
electronic retrieval systems.

www.peterlang.de

Table of Contents

Foreword <i>John H.A.L. de Jong</i>	9
Chapter 1. Issues and Challenges in Combining SLA and Language Testing Research <i>Riikka Alanen, Ari Huhta, Scott Jarvis, Maisa Martin and Mirja Tarnanen</i>	15
Chapter 2. Replicating Results from a CEFR Test Comparison Project Across Continents <i>Jamie Dunlea and Neus Figueras</i>	31
Chapter 3. Adaptations of CEFR Descriptors to Local Contexts <i>Karin Vogt</i>	47
Chapter 4. Co-operation in Europe: ‘Local’ Practices and ‘Global’ Requirements <i>Carole Sedgwick</i>	63
Chapter 5. Working with the CEFR: The Bilkent Linking Project <i>Carole Thomas</i>	79
Chapter 6. A Collaborative Approach to National Test Development <i>Gudrun Erickson and Lisbeth Åberg-Bengtsson</i>	93
Chapter 7. Comparing the Hungarian School-Leaving Examination with International Language Examinations <i>Györgyi Együd, Zoltán Kiszely and Gábor Szabó</i>	109
Chapter 8. Local Institution, Global Examination: Working Together for a ‘Co-certification’ <i>David Newbold</i>	127
Chapter 9. Matura’s Rocky Road to Success: Coping with Test Validity Issues <i>Slobodanka Dimova</i>	143
Chapter 10. Synergies and Tensions in Computerised Language Testing <i>Anders Johannessen Fikke and Hildegunn Lahlum Helness</i>	159

Chapter 11. Translation Assessment Methodology and the Prospects of European Collaboration <i>June Eyckmans, Winibert Segers and Philippe Anckaert</i>	171
Chapter 12. Preliminary Collaborative Steps in Establishing CEFR Sign Language Levels <i>Laura Sadlier†, Beppie van den Bogaerde and Joni Oyserman</i>	185
Chapter 13. Collaboration in Understanding Results – Self- assessment of EFL Writing <i>Anne Dragemark Oscarson</i>	199
Chapter 14. Effects of CLIL Teaching on Language Outcomes <i>Lisbeth M. Brevik and Eli Moe</i>	213

Local Institution, Global Examination: Working Together for a ‘Co-certification’

David Newbold¹

University of Ca' Foscari, Venice

The gap between major testing organizations whose products are intended for a world market, and the institutions which use them across the globe can seem potentially unbridgeable from a local perspective. For the testing organization, the challenge is to produce language tests which are objective, culturally unbiased, politically correct, and universally valid (and one could add some more qualities to this list). But for the test taker, and for the teachers and institutions who have to make choices about which tests to use, the resulting tests may be perceived at best as anodyne, and at worst inappropriate. This paper charts the progress and pitfalls of an experiment in *co-certification* – a collaborative process by which a local institution worked with an international assessment agency to adapt an existing suite of tests, with the aim of making it more suited to local needs. Six years on, the co-certification appears to be a ‘niche’ product requiring a considerable investment of time and energy by both partners. However, we believe that it is increasingly in the interests of global testers to be sensitive to local needs and contexts; that the project we have described shows that collaboration is possible, and can lead to better tests; and that the model outlined could be adapted to other, quite different, contexts.

Key words: co-certification, external assessment, CEFR

1. Collaboration – but between whom?

1.1 *The need for collaboration*

The growing importance of the need for collaboration in language testing has been evidenced in a number of ways over the past two decades: it has been built into the codes of practice of testing organizations such as ILTA and EALTA, it has become the focus of seminars and conferences, and it is beginning to generate models to promote language testing reform (e.g. Andrade & Green, 2010). Winding up his 2008 lecture celebrating forty years of progress in language testing Charles Stansfield claims that

we have been doing the right things since the early 1990s’ and goes on to spell out what these right things are: ‘We have collaborated with each other, and we have developed new kinds of tests, expanding our field and its reach within our countries. We must continue to do this, responding to opportunities to use our skills to contribute to a fair and just society. (Stansfield, 2008, p. 323)

¹ newbold@unive.it

Collaboration, in short, has become the premise for developing good tests. But collaboration between whom, and about what? Apart from the obvious example of collaboration between researchers and teachers, which is the lynchpin of educational research, one thinks of contexts such as academics working together on international projects (made possible with the development of the Internet), or to assessment agencies working together to develop a test framework, or new tests (as in a recent Cambridge – University of Michigan agreement). These are examples of collaboration between equals, involving sharing research findings, working across languages within a framework, and trying to develop better tests. But what of the relationship between test developer and test user, which is clearly not one of ‘equals’. What sort of collaboration is possible – and desirable – here?

1.2 At the test developer – test user interface

Kunnan (2000) begins his introduction to the notion of fairness by referring to the joint responsibilities of test developers and test users; whereas the test developer has the duty to produce material which does not discriminate between test takers, the test user has a monitoring function. This is a timely reminder to institutions who make use of external certification that to do so does not grant a license to abdicate responsibility for assessment. The temptation to do so, though, can be great. Testing agencies have become multi-million dollar businesses, operating on a global scale, browbeating users with claims about their tests, and submerging them with glossy documents about validation processes to prove their point. To the local institution which uses the tests, the gulf between the knowledge and resources of the assessment agency, and their own knowledge and experience, may seem unbridgeable, and best left as it is. However, this is never the case. There is no such thing as a perfect test; assessment agencies have everything to gain when a local institution approaches them to make suggestions about improvements. But how often will they listen? In the case study reported on in this chapter, a local institution approached an internationally known assessment agency with a view to adapting an existing test. The result was a lasting professional relationship across the test developer – test user interface, and the development of a ‘co-certification.’

2. Co-certification envisioned

2.1 The context: university reform

The co-certification grew out of the 2000 Italian university reform, itself a result of the 1998 Bologna process intended to streamline European university courses, making them more comparable and, at the same time, more competitive. In Italy, this meant reducing the first degree (‘laurea’) from four years to three, and introducing a second level two year ‘laurea specialistica’.

The reform coincided with the appearance of the Common European Framework of Reference (CEFR), which presented a window of opportunity for language faculties or departments to update their programmes, linking them (at least nominally) to the levels described in the Framework, and introducing, perhaps for the first time, communicative objectives. Typically, for a first degree course in modern European languages, the target levels by year were identified as B2 (year 1), B2+ (year 2) and C1 (year 3), at least for students with prior knowledge of the language. With English, a compulsory curriculum subject from the first grade in primary school, this was always going to be the case, reflected in the minimum (B1) level requirement in English for *all* incoming students, which has since been implemented by most Italian universities.

At the University of Venice Ca' Foscari Department of European and Postcolonial Studies a team of language teachers began work on a project leading to a document ('Vertical Integration for the Reform') which outlined a new syllabus, after listing the main problems which they felt needed to be addressed. For example, problems with the final year language course were felt to be:

- the course was too exam dependent;
- there was a need for more skills-based teaching;
- although writing was a focus of attention, it was consistently done badly;
- students seem inhibited about speaking, possibly (the teachers reported) through self consciousness; the result of too much attention to error (One teacher put this very bluntly: 'They can't write and won't speak').

This was confirmed in a survey of final year students from the old regime (reported in Newbold, 2004). 100 per cent of respondents identified speaking as an essential component of any degree course in languages, the figure dropping slightly for writing (92%), followed by listening (84%) and reading (76%) – maybe because students thought reading skills could be developed through personal study. Knowledge of grammar and sound systems were rated less important than direct acquisition of skills. Astonishingly, however, speaking skills had neither been taught nor assessed throughout the old four year degree course in English – partly due to the constraints of the university context (large classes and limited resources). In the new three year syllabus, among the framework-related objectives, speaking finally arrives on the scene.

2.2 External certification in the Italian education system

The new syllabus was drawn up in December 2003. Around the same time, on the crest of the wave that the CEFR had begun to move, and a protocol signed by the major assessment agencies working in Italy and the Ministry of

Education ('Progetto Lingue 2000') state schools began to operate as test centres for external certification, offering preparatory courses for well-known tests such as Cambridge PET or Trinity GESOL; the cost of the tests was subsidized, or, in some cases, borne completely, by local authorities. Part of the thinking behind the move was that external certification, linked to the framework, would be a source of motivation for students and help teachers focus on framework-related objectives. Ten years on, with such tests now a regular feature of Italian school life, (although there is less funding available), the choice seems to have been a long-sighted one.

The universities followed suit in their own agreement with the assessment agencies ('Campus One') which recognized the part that external certification could play in the assessment process. To start with, this meant that certification obtained at school could be used to exonerate students from B1 entrance tests, which have now become a compulsory university admission requirement throughout the country – giving schools further incentive to continue their commitment to external certification, and considerably lightening the assessment load for the universities. For example, at the Faculty of Languages in Venice the number of incoming students with appropriate certification currently (2011) stands at around 20%, a figure which is likely to be similar in other Italian universities.

Certification is also recognized as an alternative to in-house language tests in many degree courses. The advantages for the universities are obvious: with an increasing number of courses requiring an exit level stated in terms of the CEFR, and limited resources to carry out mass testing, certification (provided by one of the agencies stipulated in the Campus One agreement) eases the strain. The Campus One agreement has now expired, but the use of certification continues to grow, with individual universities or faculties free to choose which assessment agencies they recognize.

2.3 Language testing in a language faculty

In modern language faculties, however, certification was not initially embraced with the same enthusiasm. For one thing, the specialized nature of language courses in language faculties, with their emphasis on literature, linguistics, and translation, meant that certification could be used to substitute only the most generic component of a degree course. Secondly, the language faculties were less willing than science faculties to relinquish their traditional role in the assessment of students. After all, language teaching and assessment is at the heart of a languages faculty, and the faculties should have the necessary competences, and will, to assess their own students. Why enlist outside agencies to do the task? Nonetheless, there was a growing awareness of the discrepancy between a traditional approach to testing rooted, to quote Spolsky (1978), in the 'pre-scientific' period, and the powerful description of language

ability which the framework provided. Besides, in the wake of the Bologna agreement, and increased student mobility, English was beginning to emerge as an everyday lingua franca in universities across Europe, needed for carrying out research, for listening to visiting academics, or for interacting with foreign students. Surely language assessment in a languages faculty should reflect these real needs?

2.4 Co-certification conceived

At the University of Venice, the new syllabus described in 2.1 above set out attainment targets for each year which we felt met these real needs, as well as reflecting the more academic skills (such as translation and academic writing), which have always had a place in a traditional modern languages faculty. The targets consisted mainly of *can do* statements, taken bodily from the Framework, or slightly modified (modification in italics), e.g.:

Sustained monologue: describing experience

Can give straightforward descriptions on a variety of familiar subjects within his/her field of interest, *after reasonable preparation*.

or specially written, e.g.:

Written production: Summaries

Can write very brief summaries of a variety of text types (news items, descriptions, short narrative texts) getting the main point across although with limited accuracy.

How to assess real life skills (such as spoken interaction) generated considerable debate, and inevitably we began to look at how it was done elsewhere, in the new suites of CEFR-related assessments which were emerging in international language testing. Some of them, such as the Cambridge ESOL and Pitman City and Guilds suites, were adapted from existing tests and calibrated to the Framework; others, such as the Trinity ISE suite were a direct result of the Framework. They provided insights into the form our own assessment might take, revealing a range of testing techniques and formats that some of us, in the staid environment of a traditional language faculty, had not realized existed. For example, speaking was elicited by two examiners through peer interaction (Cambridge), one examiner in a one-to-one structured conversation format (Trinity), a facilitator who recorded the interaction but took no part in the assessment (City and Guilds), and so on.

It was clear that we had a lot to learn from the expertise of the assessment agencies. At the same time, it was also clear that no single certification (all of which had been designed as free standing proficiency tests) could substitute, in the content it tested, our own yearly exams. Thus, the idea began to grow of

adapting an existing certification to reflect the needs and profile of university students specializing in English. This certification, if it could be developed, would harness the expertise and experience of a professional language testing organization to the pedagogical aims, and syllabus constraints, of a local institution. We realized we were looking for a container – the test format which best suited our purpose, and the content of which could most easily be modified.

From the outset we were particularly interested in the Trinity ISE (Integrated Skills in English) suite because of its focus on the productive skills – speaking and writing - which seemed to respond to the needs identified by students. It also included portfolio as part of the assessment of writing, which required students to produce three types of text (correspondence, factual writing, and creative writing) in their own time. This process approach to writing was also attractive. At this point two routes seemed to be available to use: we could produce our own test, loosely inspired by the features which we liked most in the ISE certification; or we could ask Trinity to work with us to adapt its tests, and offer them in the university. Since the protocols referred to above made this possible, we chose the second route. The Dean of the Faculty approached Trinity College; Trinity College replied that they were interested; and the idea of a *co-certification* was born.

3. Co-certification in practice

3.1 The basis of a collaboration

The co-certification was not, however, intended to replace the in-house exam completely, since it would have a cost for students and to oblige students to pay for assessment within the state system would not be acceptable. Rather, it would stand alongside an in-house exam as an alternative, for those students interested in its dual function, as an equivalent to the exam, and as internationally recognized certification, which students could use (for example) when applying for a job, or for higher education courses in other countries, including the UK. We felt that this would appeal to students, who would see the co-certification as a worthwhile investment in their future.

An initial meeting in Venice between the Dean of the Faculty, the Trinity Director of Language Examinations, the Trinity National Coordinator for Italy, and the teachers responsible for the project, cleared the ground, assigned roles, and led to the signing of a three page contract setting out the nature of the agreement. Premised on the ‘common interest of both parties to organize English language exams for students of the University’, and that ‘the organization of such exams is compatible with the institutional aims of both parties’ it allocates the responsibilities of each party as follows (translation from Italian):

Trinity College (...) agrees to make available its specific competence in the field of language testing, administering English language exams for students of the University through its own specially selected experts. (...)

The University of Venice (...) agrees to make available its specific educational and cultural competence in the preparation of the exams (...)²

This is the basis of the collaboration which we will discuss in the rest of this chapter. The agreement extends to use the logo of both institutions on the certificate issued to successful candidates, and a financial side: all expenses will be borne through the enrolment fees paid by students (which are equivalent to the fees for the standard ISE suite).

3.2 Agreeing roles

The first problem was to decide who was to relinquish what. From the start it was apparent that the University could have no role in the assessment process. It could provide content – in the form of questions for the portfolio and, possibly, the controlled written exam – but it would not then assess candidates' responses. To do so would mean that the exam would collapse into two separate assessments, the Trinity component, and a smaller, independent add-on component, a sort of optional extra, defeating the idea of a 'co-certification'. Besides, Trinity could rely on a body of professional raters, for whom assessment of writing – once the underlying constructs were clarified – would not be problematic.

In the end, we agreed to provide writing tasks for the portfolio, but not the controlled written exam (which tested similar competences to those of the portfolio, in exam conditions). Of the three short texts that students had to write for the portfolio (correspondence, factual writing, and creative writing), we felt that *creative writing* was least suited to our purposes - which is not to say that it has no place in an EFL programme at university level. In contrast, *correspondence* and *factual writing* looked like real life tasks, high in face validity, whereas the *creative writing* component seemed to have been born at a lower level than C1, with a younger audience in mind. To take the first example at C1 level from the 2011 main ISE suite:

C1: Write a short story for a writing competition ending with the words "She couldn't believe the audience's reaction. Applause was ringing in her ears. It had been a struggle, but looking back she knew it had been worth it".

This did not seem to fit with a writing programme geared to preparing students to write their final dissertation (on a literary or cultural topic) in English. The obvious writing task we needed was 'critical writing' (whatever that meant),

² Agreement dated 16.09.2004

which would be able to contain the experiences and competences of a student in a humanities faculty, specializing in foreign languages and literatures.

Thus, the university contribution to the co-certification would boil down to one task in the portfolio, which itself carried less weight than the controlled written exam. On paper, this is not very much. However, the portfolio returns in the oral, when the examiner (who rates it) also asks questions about it. In addition, students have to give a presentation on a topic of their choice during the oral, and there was no reason why this could not reflect the work and interests developed by the student during the course. In short, we felt that the university dimension would be present throughout the adapted version of the certification. (For more about this ‘university dimension’, see Newbold 2009.)

3.3 Agreeing constructs

The next problem was to agree what the concept of critical writing, in a series of meetings between Trinity and the University. Initially, we drew up a list of underlying constructs which we felt were part of the ability to write ‘critically’, e.g.:

- evaluating
- exemplifying
- contrasting and conceding
- effective organization
- making comparisons
- using persuasion
- using a formal register, etc.

These were easy to agree. More problematic were the identification of topic areas and how to relate critical writing to the CEFR. It is no secret that the CEFR has proved a difficult document to use, for syllabus designers, teachers and testers alike, and much has been written about its limitations (e.g. Morrow, 2005 and, specifically from the testing perspective, Weir, 2005).

Relativistic language, and the *can do* statements are part of the problem. Modifiers such as *short*, *simple*, *complex* or *subtle* presuppose an intuitive understanding by users of the Framework, while *can do* statements seem exclusive, rather than inclusive, when they attempt to exemplify problems at a given level, for example:

Formal discussion and meetings

B1: Can put over a point of view clearly, but has difficulty in engaging in debate.

In the real world, observers of interaction at this level might be more familiar with an ‘opposite’ profile, e.g.:

Can engage in debate, but has difficulty in putting over a point of view clearly.

Weir (2005) is concerned primarily with content validity (‘the scales are premised on an incomplete and unevenly applied range of contextual variables/performance conditions’) and theory-based validity (‘little account is taken of the nature of cognitive processing at different levels of ability’). In drawing up our own descriptor, we also had to take into account the specific purposes for which we were developing the test, requiring clarification of those areas of background knowledge, as well as strategic competences, which were required. Douglas (2000) provides salutary advice here, and reminded us that getting the level of detail right in the definition of the construct may be problematic. This turned out to be the case. We began by looking at the CEFR descriptor for ‘overall written production’:

Can write clear, well-structured texts of complex subjects, underlining the relevant salient issues, expanding and supporting points of view at some length with subsidiary points, reasons and relevant examples, and rounding off with an appropriate conclusion. (Common European Framework of Reference for Languages, p. 61)

Our descriptor took shape by focusing on target content, as well as the constructs referred to above, and it grew in the writing as we tried to cast the net wider to include a range of candidate profiles, the additions in italics being added at a later stage:

Can write a critical appraisal of a work of art, such as a novel, a film, or a collection of poetry, *or present a critical overview of a cultural phenomenon, such as an institution or a lifestyle, or of an economic, historical or linguistic issue*, isolating and developing the main thrust of the argument with some assurance, identifying supporting themes *or typical features*, and evaluating the work appropriately against the background to which it belongs.

In some respects unwieldy, it was intended as a working document for candidates and item writers, and could be modified further if need be. However, seven years into the project, it has consistently generated questions which students consider to be appropriate and challenging. In the most recent (2011) version of the co-certification, the critical writing portfolio section consists of the following questions:

1. Some political commentators believe the European Union (EU) has failed in its wider purpose of promoting social and political unity. Write an essay discussing the role of the EU today and highlighting its achievements and limitations.

2. *'Does the idea of utopia still have relevance in today's society?'* Write an essay exploring this question and saying whether or not you believe it is necessary to hope that the future will be better than the present.
3. *'Is it right for society as a whole to bear the costs of an economic crisis generated by the speculations of a minority?'* Write an essay exploring this question and discussing some of the moral issues involved.
4. Write a critical review in which you compare and contrast the work of a contemporary 'popular' writer with that of a contemporary 'literary' writer. Predict whether they will still be seen as such, one hundred years from now.
5. *'Do schools and universities spend too much time testing students?'* Write an essay examining the issues involved in testing students and proposing alternatives to the current system of assessment.

Two years after the introduction of the co-certification at C1 level, it was extended down to B2 level (the lowest level that we felt it was possible to satisfactorily test 'critical writing'), and offered to students as an alternative to the 1st year in house exam. The descriptor for this level was identified as:

Can write a clear and detailed description and evaluation of a work of art (such as a film or a novel) or a cultural phenomenon (especially with regard to current lifestyles in the society in which one lives), by synthesizing information and comparing and contrasting differing viewpoints, using appropriate exemplification and showing evidence of effective structuring.

3.4 Working for washback

The co-certification quickly established itself as an interesting, and motivating, alternative to the in-house exam, with around a third of final year students choosing to take the higher version. From the beginning, students were quite clear as to why they had chosen this option: in a questionnaire asking them to identify reasons, 21 out of 39 chose 'It is an opportunity to get an internationally recognized certificate in English'; 6 chose 'It will be recognized as a valid alternative to the university exam.' The remainder showed less brazenly utilitarian reasoning, such as 'The focus is on skills, not grammar' (4) or 'Preparing for the co-certification will help me with my English generally' (3).

The attractiveness of the co-certification was confirmed by a pass rate of around 90% which was higher than the pass rate of the in-house exam. This needs to be seen in context: until the 2000 reform of the universities, intended to harmonize degree programmes across Europe, a dropout rate of between 60% and 70% meant that failure was an experience shared by most university students. The reform forced the universities to face up to the need for

responsible testing and realistic objectives. In Venice, the introduction of the new framework-related syllabus referred to above was gradual, and went hand-in-hand with the development of the co-certification. Thus, although the idea of the co-certification grew out of the need for syllabus reform, it is also true to say that it has helped to shape the teaching and the in-house tests. A contextualized approach to writing, and the introduction of speaking tests, have become features of the teaching programme and faculty exams, as a direct result of the co-certification. This looks like good washback, in the 1993 definition of Alderson and Wall as “the way that tests are perceived to influence classroom practices, and syllabus and curriculum planning” (p 117). In a very real way, students have been impelled to do things (speaking!) ‘that they would not otherwise necessarily do’.

At the time of writing the gap between the results (of the co-certification and in house tests) has narrowed, but still exists. This can be explained by a number of factors:

- Although the tests purport to measure the same things, they are not identical.
- Co-certification has a cost so students treat it seriously. In-house exams can be taken as often as the student likes, so they may try their luck without preparing for it properly.
- Students wishing to do the co-certification usually discuss the possibility with their teachers beforehand, and are advised about their suitability.
- Students find preparing for the co-certification, and the exam itself, motivating.

4. Collaboration strained and regained

4.1 Routine collaboration

For five years the collaboration between the university and Trinity College functioned smoothly, with both parties investing time and energy into the promotion of an exam which clearly had high face validity for students. At the university, we put on short courses to help students prepare for the oral, giving them a chance to make presentations, and introducing them to the highly structured phases of this part of the exam. Trinity, for their part, consistently managed to give us a quick turnaround with results, and provided us with numerical scores (in addition to the usual grades), which could be converted to a scale used for university exams. In this way, students were able to choose whether or not to use their co-certification score, or to do the in-house exam a few weeks later. (Most chose the former option).

Particularly useful for teachers were the annual reports on the co-certification produced by Trinity, and the feedback sessions for teachers given

by the oral examiners at the end of their visits. Since the oral examiners also rated the portfolio, this meant feedback was both on candidates' writing and speaking skills. The interest generated by the C1 level exam led to the introduction (after much discussion) of a B2 version of the co-certification, which Trinity thought might be marketable in secondary schools. Trinity also used our students to pilot a C2 version of the main ISE suite, not for use at the university, but which owed a lot (in its third portfolio task) to the critical writing component of the co-certification.

4.2 Crisis

Collaboration in any field, if it is to work, requires not only shared objectives, but also mutual respect and trust. If this is the basis of the collaboration, it will be easier to overcome the problems which can crop up in any professional relationship. For the co-certification a moment of crisis came in 2009 when Trinity rejected as inappropriate most of the portfolio titles which had been suggested by the University.

Until this year, we had used a team of two item writers, offering ten critical writing tasks for each level, of which Trinity then chose five. These were then published alongside the five tasks for the two other sections of the portfolio (correspondence and factual writing); students have to choose one task for each section. In 2009, six of the suggestions for ISE 2 and five of the ISE 3 titles were considered inappropriate, and doubts were expressed about the wording of three of the surviving titles. The reasons given were:

- the wording of the items did not follow the house style
- some topics were too similar to the previous year
- some titles did not appear to elicit the required level for the language
- some topics chosen by the university were not appropriate

Although we felt the first point could be easily addressed, the others were more worrying. We thought that major topics (such as university reforms, or cinema) which came up in successive years were not problematic, if they focused on different aspects. The idea that the titles should themselves elicit target language seemed trickier, and required clarification. Finally, the idea that Trinity, not the University, was the final arbiter in deciding which topics were appropriate, seemed to call into question the roles which had been assigned at the outset of the co-certification agreement. Surely, the local institution knew better than the external organization which topics were most suitable for its students? For example, one of the titles rejected for the ISE 3 exam required students to reflect on the way in which Italy has changed from a country of emigrants to one with a large immigrant community within the space of a generation.

Here it seemed that the topic of immigration had been avoided on principle, perhaps because ‘immigration’ was seen as a controversial topic, or perhaps because of sensitivities developed in other testing contexts (in the UK Trinity College is a major provider of English language certification for UK citizenship seekers). Assessment agencies rightly have policies about taboo or offensive topics, which can be summed up ‘better to err on the safe side’. However, the perspective changes in a local context. For young Italian adults, we felt that the question was intellectually stimulating, could allow them to refer to their own experience, and was in no way controversial.

Trinity responded to our request for a crisis meeting, which was attended by the newly appointed CEO as well as all management involved in the development of the test. We felt this to be an impressive example of sensitivity to local needs. The meeting led to the reaffirmation of roles (i.e. the university provided the questions for the critical writing element, which presupposed the topics chosen were appropriate) and to a number of resolves, e.g. we would pay closer attention to the items, to bring them more into line with the Trinity house style – for which closer attention to Trinity guidelines for item writers would be needed – and we would increase the number of item writers (from two to four). We also began to discuss other forms of collaboration between the two institutions. In short, we came away with the impression that relationship between us – and the co-certification - had been strengthened.

4.3 Consolidation

At the time of writing, the co-certification remains a popular option for students. The higher level version is the most popular (around 70 candidates per year, compared to 30 – 40 for the B2 level), although costing more. The preference is understandable; for students who are going to major in English, it seems better to wait and do the higher level, rather than do two levels of the same certification. The take up for the B2 version comes mainly from students doing other languages, or from different faculties.

Inevitably, to keep the project going as an *alternative* to an existing exam requires a considerable investment of time and energy on the part of the hosting institution, while Trinity have to bear costs (such as the specially printed exam papers as well as the human resources made available), which are probably difficult to justify in economic terms. But we feel the ‘add-on’ value of a co-certification has made the project worthwhile, since

- it has been instrumental in reforming the teaching syllabus
- it is perceived by the faculty to be a guarantee of CEFR level
- it is perceived by students to be more relevant to their needs than other external exams
- it is consistently reported by students to be a fair test, and satisfying to do

On this last point, year after year, writing the portfolio texts, and the experience of the oral exam with an outside examiner, are identified as the most positive and enjoyable features of the co-certification; unsurprisingly, the controlled written exam, the most traditional part of the exam, and the one most reminiscent of the existing in-house exam, is the part which elicits least enthusiasm.

5. Conclusion: the future of co-certification

5.1 Is co-certification a niche product?

100 test takers a year is a very small number in the world of language testing. At the beginning of the project we hypothesized that other Italian universities might be interested in our version of the certification, and that this could generate revenue, for example, to provide extra preparatory courses for our own candidates. This has not turned out to be the case. Or rather, teachers from other universities have shown interest, but their institutions have not adopted the certification. Why, they reason, should we adopt an exam which issues a certification bearing the logo of a university other than our own? Perhaps we should have anticipated this objection.

In contrast, the B2 level was developed with good secondary school students in mind, as well as our own first year students. Here the university logo on the certificate is seen as an added bonus, not a threat (The faculty of languages at Ca' Foscari is the largest in Italy, both in the number of students and in the number of languages taught); but the timing of the exam, which is available only once a year, does not fit in well with the school calendar. Schools are more likely to choose to enter for the main ISE suite, which is available every month.

So is the co-certification destined to remain a niche product, with little or no commercial potential, and therefore of little interest to large assessment agencies? A quick Internet search for 'co-certification' gives 900,000 hits, but on closer inspection most of these turn out to be for CO certification – of interest to automobile manufacturers who have to run carbon monoxide checks on new cars. Apart from the project reported here, there appear to be no other examples of co-certification in language testing.

5.2 Possible future scenarios

Yet there are plenty of contexts in which a local component in a global exam might seem possible and desirable. An obvious one is where the local component is an ESP requirement, to be added to a general English test, and for which the Ca' Foscari Trinity co-certification could provide a model. In this case, the local institution might be a professional or business organization. Perhaps one of the most challenging areas in which testers will have to work in the future concerns the increasing emergence of English as a Lingua Franca,

and the extent to which non-standard varieties of English will need to be incorporated into models of assessment. In spite of the reservations, which many professional testers have towards ELF (e.g. Taylor 2006), this issue is unlikely to go away. Standing the organizational model described in this chapter on its head, co-certification in an ELF context might mean a single global exam assessed locally by non-native speaker examiners able to assess candidate's performance against locally defined performance criteria.

Whatever the scenario, co-certification offers a chance for collaboration, which brings its own rewards. Any partnership which thrives on a mutual understanding of the tester's objectives and the candidates' profiles is likely to yield fairer, more valid tests. Today, in a world in which instant international communication has become routine, international partnerships are becoming easier to set up and sustain. If they are driven by a spirit of collaboration and shared objectives – and not just by financial gain – language testing can only benefit from them.

References

- Alderson, J. C. & Wall, D. (1993). Does washback exist? *Applied Linguistics* 14, 115-129.
- Andrade, M. S. & Green, B. A (2010). Guiding principles for language assessment reform: a model for collaboration, *Journal of English for Academic Purposes* 9, 322-334.
- Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.
- Douglas, D. (2000). *Assessing Languages for Specific Purposes*, Cambridge: Cambridge University Press.
- Kunnan, A. (Ed.) (2000). Fairness and validity in language assessment. Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida. *Studies in Language Testing* 9, Cambridge: Cambridge University Press.
- Morrow, K. (Ed.), (2005). *Insights from the Common European Framework*. Oxford: Oxford University Press.
- Newbold, D. (2004). Which English for a Modern Languages Faculty? *Didattica delle lingue straniere: Testing e Multimediale in Quaderni di Ricerca del CLI*, Venice, Università di Ca' Foscari Venezia.
- Newbold, D. (2009). Co-certification: a new direction for language testing? *English Language Teaching Journal* 63, 51-59.
- Spolsky, B. (1978). Linguists and language testers, in B. Spolsky (ed), *Papers in applied linguistics: Advances in language testing, series 2*, (pp. 6-11), Arlington, VA: The Centre for Applied Linguistics.

- Stansfield, C.W. (2008). Where we have been and where we should go? *Language Testing* 25, 311-326.
- Taylor, L. (2006). The changing landscape of English: implications for language assessment, *English Language Teaching Journal* 60, 51-60.
- Weir, C. (2005). Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing* 22, 281-300.