SAIL 12

Rethinking English Language Certification

New Approaches to the Assessment of English as an Academic Lingua Franca

David Newbold

Edizioni
Ca'Foscari

Rethinking English Language Certification

SAIL
Studi sull'apprendimento
e l'insegnamento linguistico

Collana diretta da | A series edited by
Paolo E. Balboni

12

Edizioni
Ca'Foscari

SAIL
Studi sull'apprendimento e l'insegnamento linguistico

URL http://edizionicafoscari.unive.it/it/edizioni/collane/sail/

Rethinking English Language Certification

New Approaches to the Assessment of English as an Academic Lingua Franca

David Newbold

Venezia
Edizioni Ca' Foscari – Digital Publishing
2017

Rethinking English Language Certification: New Approaches to the Assessment of English as an Academic Lingua Franca
David Newbold

The author wishes to point out that any errors of fact which may be present in this volume, and any other shortcomings, for which he apologizes, are entirely his responsibility.

Edizioni Ca' Foscari – Digital Publishing
Università Ca' Foscari Venezia
Dorsoduro 3246, 30123 Venezia
http://edizionicafoscari.unive.it | ecf@unive.it

1a edizione novembre 2017 | 1st edition November 2017
ISBN 978-88-6969-195-9 [ebook]
ISBN 978-88-6969-166-9 [print]

Certificazione scientifica delle Opere pubblicate da Edizioni Ca' Foscari – Digital Publishing: tutti i saggi pubblicati hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima sotto la responsabilità del Comitato scientifico della collana. La valutazione è stata condotta in aderenza ai criteri scientifici ed editoriali di Edizioni Ca' Foscari.
Scientific certification of the works published by Edizioni Ca' Foscari – Digital Publishing: all essays published in this volume have received a favourable opinion by subject-matter experts, through an anonymous peer review process under the responsibility of the Scientific Committee of the series. The evaluations were conducted in adherence to the scientific and editorial criteria established by Edizioni Ca' Foscari.

Rethinking English Language Certification: New Approaches to the Assessment of English as an Academic Lingua Franca / David Newbold — 1. ed. — Venezia: Edizioni Ca' Foscari – Digital Publishing, 2017. — 130 p.; 16 cm. — (Sail; 12). — ISBN 978-88-6969-166-9.

URL http://edizionicafoscari.unive.it/it/edizioni/libri/978-88-6969-166-9/
DOI 10.14277/978-88-6969-195-9/SAIL-12

Rethinking English Language Certification
New Approaches to the Assessment of English as an Academic Lingua Franca

David Newbold

Abstract

The premise for this volume is that the time has come to rethink English language certification to reflect the needs and profiles of users of English as a lingua franca, in which the dynamics of interaction are rather different from that of communication with native speakers. After an analysis of existing certifications, their scope and limitations, we describe an experiment in co-certification in which an international examining board and a higher education institution joined forces to produce a local version of an international exam, within a framework of English as a lingua franca.

Keywords Testing. Certification. English as a lingua franca. Higher education.

Rethinking English Language Certification
New Approaches to the Assessment of English as an Academic Lingua Franca
David Newbold

Table of Contents

Foreword 9

1 What Is Certification? 13

2 Certifying English to Access Higher Education 29

3 A Critique of the Sample Material on the TOEFL, IELTS and PTE Websites 43

4 An Experiment in 'Co-Certification' 59

5 The Spread of English as an Academic Lingua Franca in Europe 73

6 Co-Certification Revisited 91

7 The Shape of Certification to Come 103

References 121

Foreword

This volume takes a close look at English language certification in higher education, which has developed phenomenally over the last decade in response to student and teacher mobility, the growth in English medium instruction, and the need to communicate in English lingua franca. Today, IELTS has overtaken TOEFL as the main provider of academic certification, with more than two million tests each year, and the market continues to attract new products, such as the Pearson Test of Academic English.

But just how suited are these certifications to the needs of students and higher education institutions in Europe? Testers need standards; as a result, all the 'global' exams on the market are modeled on native speaker standards, and aimed at test takers who intend to pursue their studies in an English speaking country. Although the test providers promote an image of global mobility ("Be anything and study anywhere" the TOEFL homepage extravagantly announces), it is one-way travel that is on offer – towards the linguistic standards (and inevitably, the cultural models) of the UK and USA.

The premise for this volume is that the time has come to rethink English language certification to reflect the needs and profiles of users of English as a lingua franca, in which the dynamics of interaction are rather different from that of communication with native speakers. After an analysis of existing certifications, their scope and limitations, we describe an experiment in "co-certification" in which an international examining board (Trinity College London) and a higher education institution (Ca' Foscari University of Venice) joined forces to produce a local version of an international exam, within a framework of English lingua franca.

The volume concludes by looking at possible future directions which English language certification may take, as well as the possible pitfalls for test developers. There are likely to be many of these, ranging from the elaboration of test constructs, to rethinking the notion of 'error', developing an assessment framework which can account for extensive variety, and ensuring fairness within the fluid norms of ELF. These are the challenges now facing international examination boards, and they are demanding; but to ignore them altogether will mean losing touch with the reality of how English is evolving in Europe and the world.

Premessa

Questo volume studia il fenomeno delle certificazioni di lingua inglese all'università, un fenomeno in crescita esponenziale da più di un decennio, alimentato dalla mobilità internazionale di studenti e docenti, dalla diffusione dei corsi in lingua inglese negli atenei europei e dal bisogno, ormai generalizzato, di poter comunicare ovunque in inglese come lingua franca. Attualmente la certificazione più richiesta è IELTS, con più di due milioni di somministrazioni annue, seguita da TOEFL e da prodotti più recenti come PTE (Pearson Test of English Academic).

È utile, pertanto, chiedersi se le certificazioni linguistiche siano prodotti che rispondono ai bisogni degli studenti e delle istituzioni universitarie europee. Un ente certificatore ha bisogno di standard; di conseguenza, tutti gli esami per il mercato globale fanno riferimento a un modello di *native speaker english* e mirano ad avere come clienti studenti che intendono frequentare l'università in un paese di lingua inglese. Nonostante gli enti certificatori promuovano un'immagine di mobilità globale (*Be anything and study anywhere* proclama il sito web di TOEFL), si tratta ancora di un viaggio a senso unico, verso gli standard linguistici – e anche culturali – del Regno Unito e/o degli Stati Uniti d'America.

Il volume parte dalla convinzione che sia giunto il momento di ripensare la certificazione di lingua inglese in modo tale che i test rispecchino i veri bisogni e i reali profili di chi usa l'inglese come lingua franca, in contesti dinamici molto diversi da quelli in cui l'interazione è con parlanti nativi. Dopo un'analisi delle certificazioni attuali, dei loro scopi e dei loro limiti, sarà preso in esame un progetto di 'co-certificazione' per il quale un ente certificatore internazionale (Trinity College London) creò, in collaborazione con un'istituzione 'locale' (Università Ca' Foscari Venezia), una versione di un esame internazionale per un contesto specifico dell' inglese lingua franca.

Il volume si conclude con una panoramica dei possibili futuri scenari della certificazione di lingua inglese e delle possibili insidie che i *test designers* incontreranno. Tra queste, l'elaborazione dei costrutti, il bisogno di ripensare il concetto di 'errore', la messa a punto di una griglia di valutazione che permetta una grande varietà di risposte e, contestualmente, la garanzia di un trattamento equo nonostante la fluidità delle norme linguistiche dell' inglese come lingua franca. Questi scenari rappresentano sfide impegnative per gli enti certificatori, ma chi li ignorasse potrebbe rischiare di perdere il contatto con la realtà, con il modo stesso in cui la lingua inglese si sta evolvendo in Europa e nel mondo.

1 What Is Certification?

Abstract The first chapter looks at how English language certification has developed over the past decade, in the light of the massive growth of English as the language of choice for international communication, and the related needs for language assessment. It shows how certification received a boost from the publication of the CEFR at the beginning of the new millennium, offering the opportunity for all boards to validate their exams in line with a common, functional based, description of language competences. Although the major international boards use quite different approaches to implement these assessments (as we shall see in later chapters), we conclude by suggesting that they nonetheless share five common objectives, by attempting to produce tests which are authentic, valid, fair, secure, and which have a positive impact.

1.1 The Scope and Limitations of Language Certification in Assessing English

Language testing fulfils a variety of functions, and it can take many forms. At school, the purpose of a test might typically be to check what a student has learnt (or not learnt) at the end of a teaching unit (in a 'progress' test, set to monitor a list of objectives, or a 'diagnostic' test if it is intended to identify problems); in a language school or in higher education it might be used to decide which class or group the test taker should attend (an 'entrance' or 'placement' test). Prospective employers might need to appoint staff, or universities select students, on the basis of a test which provides evidence of an overall level of competence in the language (a 'proficiency' test). In an increasingly mobile, globalized society, governments may require immigrants to pass a 'citizenship' test, which will include an element of language competence, and which is intended to give some sort of indication of how the test taker has integrated (or will integrate) into their adopted country.

It should be clear from this incomplete list that tests come in all shapes and sizes, that they serve a vast range of purposes, and that they may be more or less important to the test taker. Of course, they are all intended to provide accurate information about the test taker. As a rule, though, as students progress through the educational system and into higher education, or the world of work, or international mobility, the stakes become higher; they have more to lose if they fail the test. Similarly, it is more crucial for the organization which requires the information the test is

intended to provide to be accurate and reliable. It is in this area of 'high stakes' tests that language certification thrives. 'Certification' usually refers to an independent assessment (independent, that is, of the test user, the organization requiring the assessment), which is valid, fair and reliable – three key concepts in testing which Messick (1998) puts together under the overarching umbrella term 'validity'. Crucially, the certifier is a professional organization which specializes in language assessment, and which typically may have developed from an educational institution (such as a university), or a government agency, (such as a cultural institution). Equally crucially, the certification has been through a process of validation, to guarantee the claims made by the certifier about the language competences the test is supposed to measure.

In recent years language certification, particularly for English, has developed enormously. This is largely due to the undisputed role which English now enjoys as the world's lingua franca. Whereas certification for other languages may be understandably bound to a (more or less overt) promotion of the culture, and cultural values, of the native speakers of that language, English certification has gradually moved away from a 'one language, one culture' approach to a policy of promoting the use of English for international communication, and a lifestyle which sees English occupying the free time, as well as the workplace, of today's globally mobile citizens. This is evident from the slogans used on the homepages of certifying agencies, or examining boards (as we shall refer to them in this study). Cambridge English declares that their certification can help test takers "achieve their goals for study, work and life";[1] IELTS is "the high-stakes test for study, migration or work";[2] while TOEFL entices candidates with the invitation to "pursue your dreams and go anywhere with the TOEFL test".[3]

The wealth of materials offered to potential candidates by the boards, through well-maintained websites, and traditional paper-based publicity, are an indication that certification also means big business. TOEFL and IELTS, the principal international tests for access to higher education, both number around 2 million test takers per year; IELTS is the current leader, having reached 2.7 million in 2015,[4] while for the same year TOEFL does not appear to have issued candidate numbers but instead claims that, since its inception in 1964, it has administered "thirty million [tests]

1 Cambridge http://www.cambridgeenglish.org/exams/ (2017-02-01).

2 https://www.ielts.org/ (2017-02-01).

3 http://www.toeflgoanywhere.org/toefl-practice (2017-02-01).

4 https://www.ielts.org/ielts-for-organisations/why-accept-ielts-scores (2017-02-01).

and counting".[5] The certification which has most candidates, however is, TOEIC (Test of English for International Communication), a test of English for the international workplace, administered, like TOEFL, by ETS (Educational Testing Services). TOEIC numbers more than five million tests annually,[6] 1.5 million of which are taken in Japan.

Certification comes with a cost, which varies depending on the type of test, the level, and where it is taken. The higher the level, the more expensive the test. Typically, at the time of writing, in Europe, a 'complete' test (i.e. one which assesses speaking, listening, reading and writing) at a higher intermediate level (B2) will cost around 150 euros. This, too, is a reminder of the 'high stakes' which certification usually implies: most candidates are in their teens or early twenties, the cost of the certification is a not inconsiderable sum, and it is viewed as a form of investment for the future.

The rapid growth in the demand for English language certification in Europe over the past decade has been fuelled, at least in part, by the consequences of the Bologna Process, which began with the 1999 Bologna Declaration and which aimed, among other things, to make European universities more competitive in the world market of higher education, by streamlining courses and ensuring that qualifications were mutually recognized by member states (Reinalda, Kulesza-Mietkowski 2005), but also that courses were accessible to an international student body, which in essence meant offering courses in English (Coleman 2006). This in turn led to universities setting minimal levels of proficiency in English for prospective students, to be certified by recognized examining boards. The implications of English Medium Instruction (EMI) for certification will be discussed fully in chapter 5. However, the need for certification in European universities extends far beyond EMI courses. Many universities now require a minimum certified level of English (B1 or B2) for *all* incoming students, whatever their course of study. This is because it is assumed that they will need English to carry out research, and possibly also to interact with students and staff on mobility programmes, for example with Erasmus.

But what does certification certify? And how do tests vary from one examining board to another, if they are all supposed to be assessing the same skills, as exemplified in the Common European Framework of Reference (CEFR)? The Framework, which by declared intent, provides a reference for the learning, teaching, and assessment of European languages, has rapidly established itself as an unavoidable standard setter for language policy makers, curriculum planners, and examining boards throughout

5 https://www.ets.org/toefl/institutions (2017-02-01).

6 https://www.ets.org/Media/Tests/TOEIC/pdf/TOEIC_sw_sample_tests.pdf (2017-02-01).

Europe and beyond. All certifications in Europe today are linked, in some way, to the Framework. They may be set at a stated level (for example, Cambridge First Certificate and Trinity College ISE 2 are both set at B2): to pass the exam, and obtain the certificate, means demonstrating the language skills which are a feature of that level; or they may not be set at any specific level (TOEFL and IELTS are examples) but the range of scores they produce can be interpreted in terms of the Framework. Thus an overall band of (say) seven at IELTS indicates a low C1 level, while in the TOEFL Internet-based test the same C1 level is indicated by a score of 95 or higher. Those boards which have developed a suite of exams directly from the Framework, such as Trinity College Integrated Skills in English, or calibrated an existing suite to the levels of the Framework, such as the Cambridge ESOL exams, are at pains to indicate the basis on which they make claims about levels.

Recognizing this need, the compilers of the Framework issued the manual *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (Figueras et al. 2005). The manual went through a lengthy piloting process, to which the major exam boards as well as the European testing organizations ALTE and EALTA contributed (Figueras & Noijons 2009), and which was intended, along with the more general aim of "competence building in the area of linking assessments to the CEFR", to increase the transparency of examination providers. For 'examination providers', read providers of certification. For descriptions of validation processes, see Papageorgiou (2007) for the Trinity ISE suite, and Khalifa and French (2008) for Cambridge ESOL. There can be no doubt that the major examining boards have emerged from the validation process with stronger products, certification which is the fruit of research, academically acceptable, and easier to market. At the same time, 'certification' which has not been through a rigid and transparent process of validation against the CEFR risks not being recognized as such, at least in Europe.

The Framework itself reflects current orthodoxy on what it means to know a language. It appeared on the crest of a communicative wave at the beginning of the new millennium, and as a result offers a description of language (any language, or at least, any European language) in functional communicative terms. Knowing a language, for the compilers of the Framework, is about doing things with language, a notion grounded in Austin (1975) and reinvented by the theorists of the communicative approach, such as Widdowson (1978), and Canale and Swain (1980). The Framework lists these functions (or rather, lists examples of functions) as *can do* statements, which are categorized into macro-functional areas of *reception*, *production* and *interaction*. Unsurprisingly, in a communicative approach to language use, *spoken interaction* turns out to provide the longest list of examples of language use.

Language certification has harnessed itself to the Framework in its approach to language competences, and, to some extent, in the use of terminology to describe those competences ('written production', 'spoken interaction', etc.). Thus speaking is no longer seen as a monolithic skill, such as the prepared monologue required in a pre-scientific age of testing (Spolksy 1976), but ranges across a range of competences; a good test needs to elicit samples of language which reflect this range of competences. How examining boards actually do this, however, varies greatly from one board to another. Thus assessment of speaking might be carried out through a range of formats, such as:

- paired speaker format, in which candidates speak to each other, are prompted by a facilitator and scored by a non-participating (but physically present) examiner (Cambridge ESOL).
- one-to-one speaking, in which candidates converse with a physically present examiner (Trinity GESE and ISE).
- one-to-one speaking, in which candidates speak with an interlocutor/facilitator; the exam is recorded and scored later (City and Guilds ESOL).
- responses to taped material delivered over the Internet (TOEFL, Pearson PTE Academic).

It seems reasonable to assume that individual candidates may be more comfortable with one format, and less so with others. Some students may prefer responding to prompts on the internet to interacting with a live examiner, and vice versa; in a paired interaction, some candidates will feel more comfortable talking to peers than to an examiner, while others may fear their own score may be compromised by the performance of by their partners, and so on. The paired interaction format has been the subject of extensive research (O'Sullivan 2002, Norton 2005, Brooks 2009), suggesting both advantages and problems, making it possible for boards to flag paired interaction as more 'authentic' (i.e. closer to real life language use) than individual interaction with the examiner, or, conversely, to promote their preference for individual interaction as more 'controlled', as in the following rationale for the traditional one-to-one format of the City and Guilds ESOL test of speaking: "Candidates are examined individually and converse only with the examiner and not another candidate, resulting in a controlled environment in which candidates can perform at their best".[7]

This type of variability, one might assume, will impinge on the candidate's performance, and compounded with other factors (such as test content, scorer reliability, conditions of administration), will lead to very

7 http://www.cityandguilds.gr/en/ESOLqualifications/oraltestsISESOL/Pages/isesol.aspx (2017-06-27).

different results for the same candidate, depending on the certification chosen. However, generic framework-related certifications are allegedly all testing the same things - competences described and exemplified in the CEFR - at the same level. Unsurprisingly, although individual boards have carried out a lot of research into their own tests, there is a dearth of comparative research, in which different tests and test results are compared. Such a study would be difficult and costly to organize, and is unlikely to be in the interests of the boards (who would see their exams branded as 'easier' or 'more difficult' as a result, and would need to realign their marketing strategies as a result). But for potential test-takers, it is important to realize that there are differences between tests, which are immediately perceptible in the test structures, and to choose carefully the one which is best suited to them.

To a large extent, these structural differences reflect an approach to assessment which the examining board may have nurtured over a long period of time, and which has become part of a house philosophy. The first Cambridge exam (Cambridge Proficiency) was delivered in 1913; all three candidates failed. This might have been due to the fact that the exam lasted for twelve hours, and required knowledge of French grammar (for a translation task), as well as phonetics and English literature. Later versions of the exam moved away from grammar translation to a more structural view of language (with a strong focus on sentence level syntax), and then, from the late seventies onward, to a "gradual shift […] away from structural approaches to language teaching towards approaches which involved using language as a means of communication" (Weir et al. 2013). Trinity College London began life assessing the performing arts, offering qualifications in music (from 1878) and then drama. It held its first exams in English as a foreign language in 1938, and since then has continued to promote a performance-based approach to assessment, with a the main focus on production, rather than testing knowledge of rules. The provider of City and Guilds ESOL certification began its business in the same year, 1878, and has a history of issuing vocationally and technically orientated certification, ranging from "Beauty Therapy to Business, Construction to Conservation and Digital Technology to Tourism".[8] General English language certification (the International ESOL suite) is only a part of this operation, a complement to the vocational and technical certification, and thus projecting an image of relevance to the world of work.

In contrast, the provider of TOEIC and TOEFL, Educational Testing Service (ETS), is an American institution founded in 1947, at the height of the structural linguistics era associated in the US with Leonard Bloomfield. This was reflected from the start in the tests, with the overarching concern

8 http://www.cityandguilds.com/qualifications-and-apprenticeships#fil=uk (2017-03-03).

for accuracy of measurement, and use of new technology, leading to the computer-based test (1998) which was then rapidly superseded by the Internet-based test (2005). TOEFL refers to itself as "the most researched" language test in the world, quoting more than 150 test reports of "rigorous research".[9]

More recent tests include Pearson Test of English (PTE Academic) and the Ireland-based Test of Interactive English (TIE). To distinguish itself from existing certification, the PTE draws attention, as the first feature of the test on the home page, and as the first point to be made in an introductory video, to the rapid reporting of test results: test takers will usually have their results in five days. This reflects not just the growth in need for certification, but the ever increasing need for university applicants, job seekers, and other test takers to provide evidence of their level in English at short notice. The Test of Interactive English requires candidates to carry out three pre-test preparatory tasks (reading a novel, following a news story, and carrying out an 'investigation'), which are then discussed with the examiner. These are supported by a 'logbook' which the candidate brings to the exam, a feature which owes something to the European Language Portfolio, developed in tandem with the CEFR, and whose aims are to motivate learners and to provide a record of linguistic and cultural skills acquired (Stoicheva, Hughes and Speitz 2009).

This brief overview of the approaches taken by six different boards should give an initial glimpse into the kind of variability, and the range of task types, a potential candidate may be faced with. At the same time, all boards share at least five common concerns, which are reflected in the frequent updates to tests, based on the development of new technologies, and research into language testing and assessment, and the nature of language itself. Updating tests are also of course a marketing strategy in an increasingly important global market. The five shared concerns are that tests should

1. assess 'real' or 'authentic' English
2. be recognized as valid
3. be fair and inclusive
4. be secure
5. have a positive impact

9 https://www.ets.org/c/mrm/ets00068/ (2017-03-03).

1.2 The Primary Shared Concerns of Examination Boards

1.2.1 Authenticity

Examining boards typically claim that their tests certify real, realistic, or real-life English. Cambridge English informs would-be test takers that "the Speaking test is taken face to face, with two candidates and two examiners. T]]]]his creates a **more realistic** and reliable measure of your ability to use English to communicate".[10] Trinity College London, in the foreword to the teacher's handbook for ISE, explains that "this integrated approach reflects how skills are used together in **real-life** situations".[11] IELTS, on a webpage for students, claims that its "content reflects **real-life** situations around the world".[12]

This emphasis on the authentic nature of tasks and language has a double purpose: it reassures teachers and students that the underlying approach is a communicative one, and it also sends a message to potential recognizing institutions that successful test takers will be able to use the language appropriately in the educational or work environment in which they may find themselves as a result of obtaining the certification. In the language assessment literature 'authenticity' refers to the degree to which a task reflects features of language use in real life, in what has come to be known as the target language use (TLU) domain (Bachman, Palmer 2010, 33). For Green (2014, 228) there are at least two types of authenticity: situational authenticity, "the fidelity with which real life tasks are reproduced in an assessment task", and interactional authenticity, "the extent to which an assessee engages the same mental processes in an assessment task as in target language use in the world beyond assessment". Bachman and Palmer (1996) consider authenticity to be a fundamental quality in a good test, along with *usefulness*, *reliability*, *construct validity* and *interactiveness*.

But authenticity and real life are not the same thing. Fulcher (2015) warns against the dangers of circular reasoning when making claims about test content, such that

> The test has *authentic* content. So: The test is valid because it measures *real-life* language use. (8)

10 http://www.cambridgeenglish.org/exams/first/exam-format (2017-06-27) (emphasis added).

11 From the forward to the Teachers Guide to ISE. URL http://www.trinitycollege.com/site/?id=3196 (2017-03-03) (emphasis added).

12 https://www.ielts.org/about-the-test/how-we-develop-the-test (2017-03-04) (emphasis added).

Take for example examiner-candidate interaction in speaking tasks, in which the candidate is required to react to an input, perhaps by showing surprise, giving advice, or expressing sympathy. This role-playing function is a feature of many speaking tests, and it is designed to sample a range of everyday communicative functions. However, depending on their cultural background, candidates may feel more or less inhibited about taking the initiative when interacting with an examiner (Fulcher, Reiter 2003) than they would in real life, when interacting with peers. But objections on the grounds of inauthenticity could be made throughout virtually any test, from multiple choice tests of receptive skills, to the content of reading and listening texts, to test administration, such as second hearings of listening texts, and time limits for written production. It would be difficult not to agree with Spolsky (1985) when he comments:

> Setting authenticity as a criterion raises important pragmatic and ethical questions in language testing. Lack of authenticity in the material or method used in a test weakens the generalizability of results. Any language test is by its very nature inauthentic, abnormal language behaviour, for the task is not to give so much as to display knowledge. With examinees who do not know or who are unwilling to play by the rules of the game the results of formal tests will not be an accurate and valid account of their knowledge. (31)

What is true of any language test is potentially more so of certification, especially generic certification which, intended for an international market, has no one clearly defined TLU domain. In-house tests in schools or universities are developed with a clear test taker profile in mind; progress tests produced by publishers for their course books have an explicit syllabus to check. But free-standing international certification needs to look elsewhere for its certainties, and it does so by investing in test validity and reliability, through research and development, through pilot studies and test taker feedback, to guarantee responsible management of high-stakes testing.

1.2.2 Validity and Validation

Traditionally, a test has validity if it measures what it is supposed to test. Thus a test of grammar does not necessarily give any information about (say) a learner's ability to speak, even though exam results may suggest a correlation between the two traits (Fulcher 2003, 203). Similarly, a test of listening which requires test takers to read the questions is more than just a test of listening. The notion of validity used to be viewed (Cronbach, Meehl 1955) from various perspectives: content validity, concurrent valid-

ity, construct validity, and predictive validity. To this we should add the popular notion of 'face validity', which concerns test takers' (and test users') perceptions that a test feels right, because it is set at the right level, or appears to test the right things. Content validity reflects the degree to which test samples from the stated target language use domain of the test, while concurrent validity is achieved when a second, alternative measurement from an independent source (such as another test, or an expert opinion) confirms the test result. Predictive validity refers to the ability of the test to predict a test taker performance in some future scenario; for example, TOEFL and IELTS both claim predictive validity in that they are intended to demonstrate to potential higher education institutions how well test takers will be able to operate successfully in English in that institution. Interestingly, their certificates are date stamped; the predictive validity is guaranteed for two years only. The central, underlying validity, however, is construct validity, a 'construct' being the underlying skill, or skills, which a test intends to measure. This may be an abstract skill such as 'reading', or a hypothetical trait or sub-skill such as 'reading for gist', which is operationalized through the test tasks.

Today, largely due to the work of Messick (1975, 1989), a researcher for Educational Testing Services (the organization responsible for TOEFL and TOEIC), construct validity has come to be seen as the core validity of a test, a super-ordinate notion or underlying framework which embraces all other features of validity and extends to notions of fairness and reliability. Messick's much quoted definition of validity has become the consensus view on validity (see D'Este 2012, 65 and Fulcher 2015, 107):

> an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assesment. (Messick 1989, 13)

Validity and validation are of course closely connected: validation is defined as "the collection of evidence which supports the validity of the inferences that may be made on the basis of assessment results" (Green 2014, 242); in other words, the confirmation of the meaning of test scores. Thus, when examining boards make claims about their tests, such as the level of the CEFR at which they are set, or the underlying skills which they are supposed to measure, then these claims need to be validated. Most examining boards devote considerable human and financial resources to these procedures.

1.2.3 Fairness

All tests are meant to discriminate. For example, a test set at B2 of the CEFR needs to separate those test takers who can perform at the level of B2 or above from those who can't. But the test must not discriminate for the wrong reasons, by adopting (consciously or unconsciously) criteria unrelated to the language skills ostensibly being assessed. If it did, it would be unfair. Most investigations of test fairness see it as an aspect of test validity (Messick 1989, as noted above, but see also Kunnan 2000, and Xi 2010), although it clearly has implications for test reliability, too: a test which is unfair may be consistent in the (unfair) measurements it yields, but it will not provide a reliable assessment of the skill(s) it claims to measure.

In *Language testing: the social dimension* (2006) McNamara and Roever approach the issue of fairness from Messick's distinction between construct under-representation – when a test fails to assess the test taker as completely as it should – and construct-irrelevant variance, when a difference in test scores between candidates is not due to a difference in the skill(s) being measured, but to some other factor(s). Construct-irrelevant variance is thus one aspect of unfairness; when it is systematically built into a test, it becomes test bias, and systematically harms one group of test takers when compared with another. This might happen when a test appeals (say) to the socio-cultural knowledge of one group, making it easier for that group to pass the test, when compared to another group; when in fact this knowledge is not part of the test construct.

As the title of the volume suggests, McNamara and Roever take a wide view of fairness issues, and are particularly concerned with the use made of test results:

> biased tests harm all stakeholders because students might get exempted from language programs although they would benefit from them, others do not get admitted to a program in which they would excel, universities or employers reject perfectly qualified applicants and accept less qualified ones, and society is deprived of potentially excellent doctors, lawyers, language teachers, or electricians and must make do with mediocre ones. (McNamara, Roever 2006, 82)

This wider view is of paramount importance to examining boards, who have to tread very carefully to avoid producing tests which might have cultural bias, and it is enshrined in the principles of good practice of testing associations, such as ALTE (Association of Language Testers in Europe) and EALTA (European Association for Language Testing and Assessment), and individual examining boards. Fairness is the first quality described in the 2016 Cam-

bridge English *Principles of Good Practice* document.[13]

The notion of fairness extends to that of inclusion of candidates with disabilities, such as the visually impaired or hard of hearing, or with learning difficulties such as dyslexia. All examining boards issue policy statements about how they will attempt to accommodate special needs candidates; not to do so would expose them to possible legal action (in the UK, for example, on the basis of the 2010 Equality Act). Typically, candidates with reading or speech difficulties will be allowed extra time, blind candidates may be assigned readers, and the hard of hearing may be able to lipread texts which are used in listening tests, and which are read to them by a 'live' reader. These accommodations require the deployment of extra resources, and so need to be arranged well in advance of the exam, and disabilities will need to be confirmed by medical records.

Conversely, examining boards are at pains to stress that accommodations do not lead to an unfair positive discrimination, by treating some candidates more leniently than others. As the Cambridge exams website puts it, "Once special arrangements have been made, candidates with hearing difficulties or speech difficulties are assessed in exactly the same way as other candidates; they are not marked 'more leniently' because they have difficulty hearing or speaking".[14]

Another aspect of fairness concerns transparency, which essentially refers to the amount of information the examination board releases to test takers about the way in which they will be (or have been) assessed. This can take the form of specimen papers, sample responses, and rationales behind scores, which boards post on their websites, and which we will take a close look at in the next chapter. More general help and advice to candidates, in the form of worksheets and videos, is also usually available. However, it is unlikely that any examining board is completely transparent about the way in which it trains its raters, sets standards, weeds out underperforming items, or adjusts scores when it realizes that the level was not exactly the declared target level, i.e. when a task or an item, or a whole test, turns out to be easier or more difficult than anticipated.

For the candidate, however, it is probably true to say that the most useful information that needs to be provided by the examining board concerns the structure of the test itself; understanding how the test is structured, and how much time the candidate will have for each section, is essential, given the elaborate structure of most tests. To be unaware of how the exam works will lead to valuable time being lost as candidates try to figure out what they are supposed to be doing.

13 http://www.cambridgeenglish.org/images/22695-principles-of-good-practice.pdf (2017-06-27).

14 Giudance Notes for Special Requirements Speaking Tests. URL http://www.cambridgreenglish.org/help/special-requirements/ (2017-02-10).

1.2.4 Security

In 2014 a documentary for the BBC TV programme *Panorama* revealed large scale cheating at an ETS test centre in London. Candidates had paid three times the real test fee to be guaranteed a successful result in the TOEIC exam, which at the time was one of the Home Office recommended exams for non EU students wishing to extend their visas and remain in the UK. The scam was comparatively simple: all candidates had to do was pay the inflated fee, turn up at the test centre, have their photograph taken to prove that they were there on the day of the exam, and then let a proxy sit the speaking and writing parts of the exam for them. They then had to return to the test centre a week later to sit the reading and listening exam. This was handled differently; candidates were told which answers to select (all items were multiple choice questions) by an invigilator who simply read out the answers to all the people in the room.

ETS unsurprisingly claimed they were unaware of the scam, as did the school whose premises were being used for the test, administered by a bogus 'education agency'. The scandal had major repercussions. The Home Office deported 48,000 students who they believed (on slim evidence) might also have cheated in previous versions of the exam; and ETS were removed from the list of test providers for citizenship and visa extensions.

With high stakes tests, such as those relating to citizenship or job applications, there will always be a security risk, and the scam uncovered by Panorama was not the first example of organized cheating. The size of the operation, however, has forced examining boards to review their arrangements to guarantee test security, the ways in which test centres are appointed and managed, how invigilators are selected, and how candidates' identities are checked. In the current document giving information about the TOEFL iBT (like TOEIC, an ETS test), we read:

> You must present valid and acceptable primary ID. [...] Verification of identity at the test centre may also include
> – thumbprinting
> – photographing/videotaping
> – signature comparison
> – electronic detection scanning devices such as hand-held metal detectors/wands
> – biometric voice identification
> – other forms of electronic confirmation

If you refuse to present ID or to have your ID verified, you will not be permitted to take the test and your test fee will not be refunded.[15]

TOEFL also publish[16] a list of clothes items which candidates may expect to be examined before the beginning of the test, ranging from hats, scarves and headbands, to jackets, cuff links and jewellery.

In 2015 the UK government introduced the SELT (Secure English Language Test), not a new test, but rather a list of approved test centres where tests for citizenship and leave to remain had to be taken; the only approved tests being Trinity College ISE (Integrated Skills in English) and GESE (Graded Exams in Spoken English) and IELTS. Unlike a test delivered entirely on line, such as TOEFL, the Trinity exams and IELTS both include face to face exams, with an oral examiner appointed directly by the examining board, which probably makes them intrinsically safer than an online test administered by local agents.

Nonetheless, with the rapid need for reliable high stakes tests, examining boards have become more security conscious and this has sometimes led to structural changes to exam. In 2015 Trinity College updated its existing four skills exam (ISE), introducing, among other things, a multitext reading to writing activity, and a free-standing listening administered during the oral interview. It also got rid of a portfolio component, which for ten years had been a characteristic feature of the assessment; candidates produced three short written texts in their own time, and discussed them in the oral interview. The portfolios were scored and contributed to the overall assessment of writing. Justifying the disappearance of the portfolio (which had proved popular with test takers no doubt because it be written outside the stressful context of an exam) Trinity College suggested that it "could be more usefully harnessed as a teaching and classroom support tool",[17] and consequently developed a 'portfolio toolkit' resource for teachers. It seems clear, however, that security conditions also contributed to the decision. Of course, an exam with a take-home writing component could not have been chosen as a UK Secure English Language Test.

1.2.5 Impact

'Impact' is usually taken to refer to the effect (whether positive or negative) of assessments on educational systems and society as a whole. It is an extension of the notion of "washback", the effect of assessments on

15 2016-17 TOEFL ibT Test Registration Bulletin, 12. URL `https://www.ets.org/toefl/ibt/about/bulletin` (2017-02-10).

16 *What to Expect*. URL `https://www.ets.org/toefl/ibt/test_day/expect/` (2017-02-12).

17 ISE FAQs page. URL `http://www.trinitycollege.com/site/?id=3323` (2017-10-02).

teaching and learning. Beneficial washback (or "backwash"[18]) began to be recognized in the 1980s as a fundamental test quality, alongside validity, reliability and practicality, and perhaps the most important quality from a teaching perspective (Hughes 1989). In the 1990s, especially in the wake of Alderson and Wall (1993), empirical research into washback developed rapidly, with some researchers (e.g. Messick 1996) seeing washback as fitting a unified theory of test validity.

In an educational context it is easy to find instances of 'harmful' or negative washback: for example, when 'teaching to the test' implies leaving aside useful skills which might be part of a curriculum document (such as speaking) but which are not included in the test, perhaps for organisational reasons. Similarly, a test which is not perceived to be fair by test takers is likely to impact negatively on their attitude to learning the language. Tests need to be motivating for learners, and it is in the interests of teachers to produce motivating institutional tests, since they will have to live with the consequences.

This is not the case for external certification and examining boards who have no direct knowledge of individual test takers. When an individual fails to achieve a necessary grade on a high stakes test, and has to retake the test, there is usually not much comfort to be gleaned on the boards' websites. Take for example the following advice on the IELTS website,[19] such as the following note, which, after explaining that it is possible to retake the test "as soon as you feel ready to do so" continues

> Before applying for another test, take a moment to consider your options. Your score is unlikely to increase unless you make a significant effort to improve your English.

This claim does not take into account variability associated with conditions of administration, and, crucially, the test takers themselves, who may have had to travel a long distance to get to the test centre, and for whom there may be a high discomfort factor which could affect performance. Unsurprisingly, it seems that boards do not publish statistics about test retakes, but it may be that, in spite of the advice quoted above, retakes sometimes give quite different results for the same candidate.

When it comes to the wider picture of impact, examining boards are careful to project an image which suggests, firstly, that they are aware of the ethical dimension of possible social uses of language assessments, and secondly, that their tests can promote international mobility and make a

18 Hughes uses the term "backwash" because, he says, he can find the term in a dictionary; whereas "washback" is not present in most dictionaries.

19 `https://www.ielts.org/book-a-test/resitting-the-test` (2017-02-11).

contribution to successful international communication. But such claims need to be evidence-based. Writing for Cambridge ESOL, Taylor (2005) puts it thus:

> Today tests are increasingly used for 'high-stakes' gate-keeping and policy-making purposes, as well as to provide targets for and indicators of change; it is therefore even more important for test producers to provide appropriate evidence that the social consequences of using their tests in such ways is beneficial rather than detrimental. Until fairly recently, claims and assertions about the nature and extent of a test's impact were largely based upon impression and assumption. It was relatively simple for producers to claim positive washback for their own tests, or for users to criticise tests on the grounds of negative washback; but it was also too easy for both sets of assertions to go unchallenged. Impact research – such as that conducted by our own organisation – reflects the growing importance of evidence-based approaches to education and assessment which enable policy and practice to be justified in terms of sound evidence about their likely effects.

In recent years major boards have published a number of impact studies, such as Merrifield (2016), which looks at the use made of IELTS exam results by professional organizations, or Khalifa and Vidakovic (2014) which presents impact studies for Cambridge exams in mostly educational settings. Evidence of positive impact can be used for promotional purposes, is made available on websites, and filters through to slogans which highlight the global acceptability of certifications.

A test which can demonstrate a positive impact, in an unstable and increasingly mobile world, is a test which is attentive to change and the opportunities afforded by the development of English as the world's lingua franca. Some tests need to be highly specialized, to capture the skills needed for communication in specific environments or for specific professions, ranging from air traffic control, to maritime communication, to legal professions, in which wrong use of a single word may have devastating consequences. This book is not concerned with this kind of ESP test. Rather, it will take a close look at the biggest and most well-established tests of English for academic purposes especially in the context of Europe, and how they may need to evolve to maintain a high degree of validity and a positive impact on the development of English language as a means of international communication.

2 Certifying English to Access Higher Education

Abstract The need for English language certification is nowhere more apparent than in higher education in Europe today. This chapter provides an overview of the three best known tests for academic purposes: TOEFL, IELTS, and the more recently developed Pearson Academic. It examines the structure and scope of the tests, and includes an analysis of the image which the boards project of them, as promoting mobility and guaranteeing success in the workplace. However, the tests are designed primarily to predict the abilities of test takers to interact in a native-speaker environment, in particular in the US, the UK and Australia. Today, we argue, there is an urgent need for examining boards to engage with the reality of non-native interaction, to reflect the real language needs which have emerged in academic contexts in Europe and beyond.

2.1 Overview

In this chapter we shall look more closely at three certifications which are used as tests of academic English, two of which (TOEFL and IELTS) are well established in Europe, the third (PTA, Pearson Academic) being a more recent addition. In particular, we shall compare their structures and task types, and see how they reflect the current testing orthodoxies of the period in which they were first developed. We shall also compare scoring systems, and (not least importantly for potential users) the self image promoted by the boards. In the next chapter we shall look more closely at the tests themselves by examining some of the sample material each board makes available on its website.

All of these tests might be used for a range of career-significant purposes, including professional and vocational purposes – IELTS has a 'general training' version which has different reading and writing components from the academic version – but they are particularly chosen by students wishing to enrol for university courses, either in English speaking countries, or, increasingly, for higher educational institutions elsewhere in the world delivering courses through the medium of English (EMI).

Although, as we noted in chapter 1, these tests are not set at a specific level of the CEFR, but report numerical results (TOEFL, PTA) or bands (IELTS), the examining boards publish equivalence tables which suggest a relationship between the results on the test and a level on the CEFR; and, as a consequence, these tests will also be accepted by a wide range of institutions in Europe as evidence of a level on the CEFR, in place of other certifications which are set at specific levels.

The quite different formats and approaches of the three certifications reflect the historical circumstances of the exam boards themselves, and the current testing orthodoxies in which the certifications were created, developed, and (for TOEFL and IELTS) revised. The oldest exam is TOEFL, owned by Educational Testing Services, the US based not-for-profit organization which was set up in 1947 with the aim of fostering research into the measurement of educational achievement. The TOEFL was first administered in 1964, at the height of the psychometric period in language testing (Carroll 1983, Fulcher 2015). Since then, it has gone through major transformations, from paper based,[1] to computer based, to Internet based, but the original concern for an objective, reliable test can still be seen, for example in the use of machine marking for writing, alongside human raters; not to mention the extensive research that ETS has itself published on the exam, all of which is downloadable from over 9,000 reports which can be accessed through the ETS website.[2]

IELTS first appeared in 1980, an offshoot of UCLES (University of Cambridge Local Exams Syndicate), on the crest of a communicative wave. The 'communicative revolution' in language teaching had begun in Britain in the late 1970s (Widdowson 1978, Richards and Rodgers 1986), and Keith Morrow's provocatively entitled "Communicative Language Testing: Revolution or Evolution?" (1979), on the difficulties of testing second language communicative competence(s) had just been published. From the beginning, IELTS took more of a communicative, skills-based, 'whole text' approach, which is still reflected in the current version of the test, for example in the use of a live, face-to-face examiner for the speaking part, and an overall structure which is less fragmented than TOEFL (with its numerous short listening texts) or PTE, which takes a more task-based approach.

The PTE, we said, is the most recent academic test on the market, owned by the leading educational publisher Pearson. First administered in 2009, it was designed from the outset as a computer-based test. Like TOEFL, the speaking part involves responding to recorded prompts, and includes tasks such as reading aloud and repeating sentences, as well as describing graphic information and summarizing short texts. Unlike TOEFL, there are no long reading texts, but short texts which provide the input for single tasks, or indeed single questions. Whereas in a traditional test of reading the questions exploit the text, here it seems to be more a case of the text being chosen to fit the task type. Indeed, whereas the structure of TOEFL and IELTS, as presented by the boards themselves (see tables 1 and 2 below) focus more on texts and candidate behaviour, such as "listening to lectures"

1 According to the TOEFL website, 3% of TOEFL test takers currently use the paper based version. URL `https://www.ets.org/toefl/pbt/about` (2017-10-24).

2 `https://www.ets.org/` (2017-02-22).

or "discussion of ideas and issues related to presentation", the PTE (tab. 3) lists only task types, such as "repeat sentence" or "fill in the blanks".

2.2 Test Structure and Test Taking Procedure

2.2.1 TOEFL iBT Test Structure and Procedure

Table 1. Test of English as a Foreign Language (TOEFL): Structure

TOEFL
Part 1: Reading (60-80 minutes)
The Reading section includes 3 or 4 reading passages. There are 12 to 14 questions per passage.
Part 2: Listening (60-90 minutes or longer)
4 to 6 lectures, each 3 to 5 minutes long, 6 questions per lecture.
2 to 3 conversations, each 3 minutes long, 5 questions per conversation.
Part 3: Speaking (20 minutes)
6 tasks.
2 independent speaking tasks.
4 integrated speaking tasks which develop ideas from reading and/or listening inputs.
Part 4: Writing (50 minutes)
Task 1 requires synthesizing material from two separate written texts and identifying opinions or arguments (20 minutes).
Task 2 is a free standing writing task (30 minutes).

As the oldest of the three tests being considered, the TOEFL iBT has undergone the biggest changes, most notably in its method of administration. Although the paper-based test survives for use in places where internet access may be limited, it is a substantially different test, with no speaking, but with a grammar section and a very different scoring system. It is not available in Europe. The computer-based test, in contrast, was short lived, having been superseded in 2005 (Alderson 2009) by the web-based version. This change simply reflected the rapid development and increased availability of the Internet, which brought with it flexibility of administration – what Roever (2001) calls the "asynchrony principle" – along with lower costs, although with high stakes such as TOEFL test security is a major issue, impinging on data storage and conditions of administration, as well as candidate recognition.

The iBT has four parts, and has an administration time of around four hours. The test is administered at a single sitting, with a mandatory ten minute break after the reading and listening sections. The timing is variable because the test sometimes includes items which are being piloted for use in possible future tests. This means that ETS will analyse the scores from these items to determine their facility index and reliability, but will

not use the scores in the test result. In other words, candidates are being used as guinea pigs for future tests, and paying for the privilege. Candidates are not told which items are being piloted, and so they end up doing more tasks than they actually need to.

There seems to be an issue of fairness here, both in terms of transparency – test takers do not know what they are actually being tested on – and in terms of possible test bias: if test takers perform better when the test is shorter, then this might be due to systematic construct irrelevant variation (see chapter 1).

The first part of the test measures the productive skills, and relies heavily on multiple choice items. A notable feature of the second part of the test, which measures production, is that both speaking and writing, with the exception of the initial speaking tasks and the free standing final essay, integrate skills, so that speaking follows an initial reading or listening input, and the first writing activity is also a reading comprehension activity, requiring the test taker to synthesize material from more than one written source.

2.2.2 IELTS Test Structure and Procedure

Table 2. International English Language Testing System (IELTS): Structure

IELTS
Part 1: Listening (30 minutes)
Recording 1 – a conversation between two people set in an everyday social context.
Recording 2 – a monologue set in an everyday social context.
Recording 3 – a conversation between up to four people set in an educational or training context, e.g. a university tutor and a student discussing an assignment.
Recording 4 – a monologue on an academic subject, e.g. a university lecture.
Part 2: Reading (60 minutes)
Three long texts for reading comprehension.
40 questions to test reading for gist, main ideas, detail, skimming, understanding logical argument, recognizing writers' opinions, attitudes, purpose.
Part 3: Academic Writing (60 minutes)
Task 1 – Describe, summarize or explain data presented graphically, in tables, diagrams, etc.
Task 2 – Write an essay in response to a point of view, argument or problem.
Part 4: Speaking (11-14 minutes)
Part 1 – Conversation focusing on personal background and interests (4-5 minutes).
Part 2 – Presentation on a topic given on a prompt card. (1 minute preparation, 2 minutes for presentations, followed by brief discussion.
Part 3 – Discussion of ideas and issues related to the presentation (4-5 minutes).

The IELTS test is shorter (2 hours 45 minutes), is paper-based, and has a slightly more flexible administration: the speaking component can be taken separately, up to a week before or after the other parts of the test. From 2016, a computer-based version of the test has been made available in some countries (UK and China), but not in Europe; the structure is the same as the paper-based test. Candidates enrolling for IELTS have to make the choice which version of the test they want to do: "academic" or "general training", the latter having been developed for professional, vocational and migration purposes. The two versions maintain identical listening and speaking sections, and vary only in the reading and writing sections.

The first part of the test, listening, presents a conversation and a monologue "set in everyday social context", while the second part offers (for both the academic and general training version) a more academic setting for a dialogue and an extract from a lecture. Both reading and listening sections use a range of question types, such as matching, labelling and sentence completion.

Unlike TOEFL, the writing tasks are not integrated with other skills. There is a clear cut distinction between the shorter, first task (describing a process or phenomenon by interpreting graphically presented data) and the longer task of critical writing. The biggest difference, however, from both TOEFL and PTE, lies in the speaking task, which is a one-to-one conversation with a live examiner. This format has been maintained in the computer-based version.

Table 3. Pearson Test of English - Academic (PTE): Structure

PTE Academic
Part 1: Speaking and Writing (77 – 93 minutes)
Personal Introduction.
Read aloud.
Repeat sentence.
Describe image.
Re-tell lecture.
Answer short question.
Summarize written text.
Essay (20 mins).
Part 2: Reading (32 – 41 minutes)
Fill in the blanks.
Multiple choice questions.
Re-order paragraphs.
Fill in the blanks.
Multiple choice questions .
A ten minute break is optional.

Part 3: Listening (45 – 57 minutes)
Summarize spoken text.
Multiple choice questions.
Fill in the blanks.
Highlight the correct summary.
Multiple choice questions.
Select missing word.
Highlight incorrect words.
Write from dictation.

2.2.3 PTE Test Structure and Procedure

The PTE is a three hour computer-delivered test. Unlike TOEFL and IELTS, and most other well known tests such as Cambridge exams and Trinity ISE, which typically begin with listening and reading (TOEFL, IELTS) or with reading and writing (Cambridge, Trinity), the PTE starts with the test of speaking. The first task, arguably the most authentic, is not scored. The test taker has thirty seconds to "give your selected institution some information about yourself". In other words, the candidate can record a prepared personal statement which will be sent by Pearson to any institution which requests a test report, and (if it so wishes) can take this statement into account when deciding whether or not to offer the candidate a place.

This is an interesting additional element to the test. After twenty five seconds for preparation, candidates have half a minute to record their presentation. It is not clear whether they can simply read a prepared text which they take with them (presumably not), but the most obvious strategy for such an important opportunity would be to prepare a short presentation and memorize it. In either case, there would be no need for twenty five seconds preparation time. In contrast, an unprepared, improvised, presentation is likely to lead to a rejection.

The rest of the speaking part consists of a string of less authentic tasks, such as repeating a sentence, or simply reading aloud a sentence. The short question requires candidates to identify a word from its definition, while the "re-tell a lecture" item involves a more demanding listening activity.

For the final activity in this section, the free-standing argumentative essay, the candidate has just twenty minutes to write 200-300 words – exactly half of the time allowed for the same length essay in the IELTS. It is difficult to understand why the time allowed for a comparable writing test should be so different across the two tests.

The PTE is the only test which was designed from the outset as a computer based test, and this seems apparent in the consistency of procedure throughout the test, such as the use of the progress bar to indicate how much time is left to complete a given task. This does, however, mean that

the test taker needs to be thoroughly familiar with the procedure, to avoid being impeded by a computer method effect (Chapelle and Douglas 2006, 40 ff.). One warning issued at the beginning to candidates is likely to be particularly worrying, especially if they have not had much practice for the exam: "If you remain silent for longer than 3 seconds, the recording will stop" and the candidate will not be able to re-record. In real life, three second pauses can be quite natural, as much a part of the flow of speech as the sounds of the language.

2.3 Scoring

As noted, none of the three tests we are presenting in this chapter are based on a given level on a scale of proficiency, such as the CEFR. There is thus no 'pass' or 'fail' result, but scores are given on a continuum, whether the broad 'bands' of IELTS (1-9), the "Global Scale of English" based on a 10-90 scale used by Pearson, or the mark out of 120 for TOEFL.

These are not so different as they might seem at first glance. The TOEFL exam, as we saw, makes extensive use of multiple choice questions, for which the test taker who has no knowledge of the language will have a 25% chance of getting the right answer; as a result a realistic 'low' score on TOEFL starts a long way up the scale; and the 10-90 global scale of PTE seems to equate to the 1-9 of IELTS.

All three boards are at pains to explain how to interpret the results of their tests, but this is not an easy task. Traditionally there are two kinds of assessment grid: a holistic grid, which identifies overall levels of performance, and an analytic grid which looks at different components of a test, and the different criteria needed to assess them. Thus an analytic grid for a test of speaking might list very different criteria from a grid used to assess a receptive skill such as listening. For example, assessment of speaking might take into account factors which only belong to speaking, such as pronunciation and fluency, while assessment of listening might consider hypothetical sub-skills or enabling skills, such as inferring meaning or understanding main points.

Behind the overall grade or score for TOEFL, IELTS and PTE lurk some rather different grids. TOEFL allocates 30 points to each of the four skills. For speaking, it identifies four broad bands of performance, and claims that a score of 26-30 is 'good', 18-25 'fair', 10-17 'limited' and anything below this is 'weak', making the scale rather top heavy (as we surmised above). IELTS bands are given a label (ranging from 1 = 'non user' to 9 = 'expert user') and a brief, one sentence description of an overall level of competence. Thus Band 7, a crucial level which many universities will set as a minimum required entrance level, is labelled 'good user' and the short description reads:

> The test taker has operational command of the language, though with occasional inaccuracies, inappropriate usage and misunderstandings in some situations. They generally handle complex language well and understand detailed reasoning.

The PTE uses a "Global Scale of English" which it maps against the TOEFL and IELTS scores, and offers an easy-to-use conversion table on its website, so we learn, for example, that IELTS 7 is equivalent to PTE 65-72, which in turn spans the range 96-105 on the iBT. But this is holistic scoring, and it is far from giving the whole picture. For a start, large swathes of the scales are underused, although, as Bachman and Palmer point out, "test developers will generally need more scale levels than there are decision levels" since ratings are never completely consistent (Bachman and Palmer 2010, 343). All the boards provide quite detailed information about how these overall scores are reached, by converting information from analytic grids. If we look at how speaking is assessed in the TOEFL exam, behind the overall score between 1 and 30 we find that up to four points are awarded for each of the six assessed speaking tasks; these points are then converted to the score out of thirty. For each task, a grid is used which requires raters to identify a "general description", and assess "delivery", "language use" and "topic development". So it is a hybrid scale, since the general description, as well as being based on task fulfilment, takes into account delivery (speed, pronunciation, etc.), language use (vocabulary range and grammatical accuracy) and topic development, which concerns effective organization.

As with TOEFL, the separate components of IELTS (reading, writing, listening and speaking) are equally weighted. The final band score is the average for each component, rounded to the nearest half band. Speaking is assessed according to four criteria: fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation, each of which are also weighted equally. A two page document (labelled "Speaking Band Descriptors: Public Version"[3]) describes levels of performance, maintaining the nine bands of the overall score, for each of the four criteria.

The PTE scoring system is arguably the most complex, since the final score combines scores for 'communicative skills' – i.e. listening, speaking, reading and writing – with those awarded for 'enabling skills' – grammar, oral fluency, pronunciation, spelling, vocabulary and written discourse. A downloadable 72 page booklet explains the rationale behind the system. Since PTA is a machine-marked computer-based test, much of the booklet is devoted to the rationale of machine marking, and the claimed high reli-

3 https://takeielts.britishcouncil.org/sites/default/files/IELTS_Speaking_band_descriptors.pdf (2017-11-02).

ability which results. For each of the five speaking tasks, a list of assessed sub skills is provided, and this might include features such as "speaking under timed conditions" or "speaking at a natural rate", which are presumably less problematic for the software engineers to develop than the more content-based criteria such as "supporting an opinion with details, examples and explanations" or "using words and phrases appropriate to the context".

Understanding test structure plays a vital part in exam success, which is why the examining boards need to give a lot of information about how the test is administered and scored. Since many potential candidates may be able to make a choice between two or more certifications, this information is likely to be just as important as the price, in the decision about which test to take. For example, some candidates may feel more at home than others in responding to recorded prompts (TOEFL and PTE), while others will prefer the one-to-one interaction of the IELTS interview. Timing is another crucial factor, and here too we noted considerable differences in time allotted to writing tasks.

On each website, the information is strategically layered; an initial, simple and clear-cut presentation of the overall score grids leads, via links, to attachments which contain more complex analytic information. But although there is a lot of information, which is aimed at teachers and recognizing institutions as much as to candidates themselves, a lot more information about procedures is left unsaid. As noted above, IELTS label their published band descriptors for speaking and writing "*public* version" (italics added), implying that there is an in-house version for IELTS examiners. This is understandable, since some of the procedural information is likely to be of use only to examiners, and more extended descriptors (in the in-house version) might be used to help examiners make decisions in borderline assessments. However, no board is completely transparent about the way in which it reaches its decisions. Some of the unanswered questions an inquiring candidate might want to ask include:

- How do machine and human raters interact (in TOEFL)?
- What arrangements are in place to ensure inter-rater reliability (in IELTS)?
- How are more complex responses machine marked (in the PTE)?
- What happens when one version of a test produces anomalous results?

To take the last question: although extensive pre-testing is carried out for all three certifications, it is possible, and perhaps inevitable, that bad questions (or bad examiners) will slip through the net. What happens, then, when the test administrators realize that something has gone wrong in one version of their test? For example, if it produces results which are considerably above (or below) the average? Do they adjust the scores *post hoc* to reflect this, concluding that the error lay in the supposed level of the test, or do they let the scores stay as they are, implying that the cohort

which took the test are considerably more (or less) competent than the norm? Questions such as these are left unanswered in the publicly available materials provided by the examining boards.

2.4 Reporting Results

Test procedures and scoring techniques inevitably impinge on reporting, which refers to the way in which the examining board communicates test results to the candidate or the institution which requires the results. With the continued growth in the number of students entering higher education, or going on to pursue postgraduate study, or applying for mobility programmes, and the rapid development of the higher education market, with the appearance of ever more courses intended to attract international students, often at quite short notice, the speed with which examining boards can offer their results has become a significant factor in choosing one test rather than another.

All three examining boards, ETS, IELTS and Pearson, make claims about reporting times. Of these, the fastest are claimed by Pearson, within five working days. This claim is made on the home page of the PTE website as one of the major attractions of the test (along with "flexible test dates" and "accepted for visa applications"). TOEFL and IELTS indicate times of "approximately 10 days" and "13 days" respectively. Results are posted on line, and can be sent by the candidate to the institution they are applying to, or directly by the examining board. TOEFL issue a disclaimer that scores are valid only if provided directly by ETS. The report forms are one or two page printable documents giving information about the candidate and a breakdown of the scores across the skills; the TOEFL report includes a description of what users at the certified level can typically do, for each skill assessed, but this should not be taken as feedback on an individual performance.

Unlike generic certifications such as Cambridge English graded exams, and Trinity College London Integrated Skills, tests of English for academic purposes do not come with any special mentions, such as 'merit' or 'distinction', partly because, as we noted, there is no pass mark around which to position them. What they do share, however, is a validity date. All three tests are valid for two years; this means, or should mean, that the results should not be used by educational institutions as evidence of the applicant's level in English after a two year period has elapsed. Why should this be so? The IELTS website provides a reason: attrition.

> While it is up to each organisation to set a validity period that works for their purposes, the IELTS partners recommend a 2-year validity period for IELTS test results based upon the well-documented phenomenon of second language loss or 'attrition'.[4]

This brief note is followed by a link labelled "Read research relating to language attrition". Clicking on the link gives access to an in-house search box in which the user is invited to "find IELTS research". But writing in the search word 'attrition' leads to the response "Your research returned 0 result(s)".

Elsewhere, however, Taylor (2004) justifies aligning IELTS with the TOEFL two year validity period on the grounds of language attrition. Quoting early research by Weltens (1989) and Weltens and Cohen (1989) she refers to the rapid language loss experienced by low level learners who stop using the language, whereas higher level learners apparently experience a "plateau" for "a few years" before their skills begin to decline, suggesting the existence of a critical period for language retention. She goes on to conclude (14) that a "two year period has been selected as a reasonable 'safe period' for the validity of certification".

The IELTS website, quoted above, only 'recommends' a two year validity period, and of course any institution is free to continue to recognize certification as meaningful beyond the two year period if it so wishes, perhaps if it is validated by a letter from a teacher or an interview with the candidate; but this may not happen very often. After the two year period has elapsed, Pearson simply remove the evidence of the pass from their website, so potential test users are no longer able to access the introduction recorded by the candidate (see 2.2.3 above). For the would-be candidate the two year period imposed by all three boards might look suspiciously like a conspiracy, or a cartel, and it may mean that a student requiring evidence of a level in English more than two years after doing a certification will have to do the same exam again, even though he or she has been using English on a regular basis over that period. ETS, however, does suggest that students "who have successfully pursued academic work at schools where English was the language of instruction in an English-speaking country for a specified period (2 years)" may not need to do the TOEFL when applying for a university place.

The notion of 'English-speaking country' – i.e. those countries where English is the first language – is crucial here, since it excludes students who have been taught through the medium of English (EMI) or followed programmes delivered entirely in English (ETPs) in all other countries. It is a notion which does not take into account the current status of English as a global language, outside the 'English speaking countries'. Attrition

4 https://www.ielts.org/about-the-test/how-ielts-is-scored, (2017-03-03).

sets in when users are not exposed to a language over a period of time; but English is a difficult language to escape from. Weltens looked at secondary school learners of French in Holland in the 1980s; if he were to look at secondary school learners of English in Holland today, it might be more difficult for him to find evidence of attrition for competent users of English, since young Dutch people, like their contemporaries across Europe, are likely to be exposed to English as a lingua franca, or even active users of the language, on an almost daily basis.

2.5 How Examining Boards See Themselves

The scope of the tests, and their international nature, can be gauged not only by the test descriptions and the practice materials, but also by the promotional stances of the examining boards, and the extravagant claims they make about them. Global mobility seems to be key to all three tests, heralded on the home pages by similar sounding slogans:

"Be anything and study anywhere" (TOEFL)

"The test that opens doors" (IELTS)

"The test that takes you places" (PTE Academic)

As the would-be test taker moves further into the websites, the emphasis changes slightly. TOEFL and IELTS (as market leaders) focus on recognition of their certification, with TOEFL resorting to superlatives:

> The TOEFL Test Gives You an Advantage: Most Widely Accepted, Most Popular and Most Convenient Choice.
>
> The TOEFL test is the most widely respected English-language test in the world, recognized by more than 9,000 colleges, universities and agencies in more than 130 countries, including Australia, Canada, the U.K. and the United States. Wherever you want to study, the TOEFL test can help you get there.

IELTS has a more sober approach:

> An IELTS certificate is recognised as evidence of proficiency in English by more than 9000 education and training providers worldwide. Some universities in non-English speaking countries require an IELTS score, where courses are taught in English.
>
> The IELTS Academic test is suitable for entry to study at undergraduate or postgraduate levels, and also for professional registration purposes. It assesses whether you are ready to begin studying or training in an environment where English language is used, and reflects some of the features of language used in academic study.

Pearson, being a relative newcomer, can only refer vaguely to "thousands" of institutions which recognize the PTA, but relies instead on speed of reporting and flexibility of administration as its main selling points:

Fast

PTE Academic typically delivers results in five business days, so you don't need to worry about waiting for results.

Flexible Test Dates

We run test sessions 363 days of the year, at one of over 200 locations worldwide, so you can choose a time and place that suits you.

Approved

Approved by the Australian Government for visa applications and accepted by thousands of institutions in the UK, Australia, USA, Canada, New Zealand, and Ireland, including Harvard, Yale and INSEAD.

These messages are intended primarily for test takers, and accordingly they address them directly ("You don't need to worry", "Wherever you want to study", etc.). But the websites are aimed at three categories of users: potential test takers, test users (recognizing institutions), and teachers. The three-way focus is signposted most clearly on the TOEFL home page with its three menus labelled "For test-takers", "For institutions" and "For teachers and advisors". IELTS distinguishes between "Organisations" and "Teaching and Research", while Pearson has a drop-down menu for "Organisations" which is further divided into "universities" "researchers", "agents" and "teachers", a reminder of the diversity of interested parties seeking information about certifications.

One of the most noticeable features of self-promotion is to be seen in the quantity of research articles which each board refers to, provides links for, or makes available for download from the site. The clear intention is to imply that a particular certification has a strong theoretical basis, and that this has been demonstrated by serious research. Much of the research has been commissioned by the boards themselves, or produced in house. ETS claims to have published "more than 240 peer-reviewed research reports, books, journal articles and book chapters"[5] in support of test design and validity, making TOEFL the most widely researched certification. Perhaps unsurprisingly the main research focus varies according to the most salient features of each test; thus TOEFL is particularly interested in research findings which substantiate claims made for automated scores, IELTS in validity arguments for 'live' examiners, and Pearson in demonstrating that the PTE is a valid alternative to other high stakes tests.

5 https://www.ets.org/toefl/institutions/about/research_design/ (2017-10-07).

A closer look at the research, much of which is undoubtedly of a high standard, shows that the test under consideration does not always emerge in a completely favourable light. For example, in one of the more recent research articles which is forefronted on the TOEFL website, Bridgeman et al (2012, 91) find that, although the software programme *SpeechRater* used by TOEFL to score speaking does evaluate some aspects of communicative competence, it "fails to measure aspects of the construct that human raters can evaluate". Ultimately, however, it is to the examining board's credit to draw attention to articles which call into question test features. There is no such thing as a perfect test, but there is always a need for new research to reflect new developments in testing. It is in this light, it seems, that both IELTS and Pearson[6] invite researchers to apply for funding for projects relating to their tests, and more generally (IELTS) to the field of language testing.

In contrast, there is a dearth of unsolicited (by the examining boards) independent research into certifications. This is surprising, given the high stakes nature of the tests. One exception is Uysal's (2010) critique of IELTS. Uysal focuses on the writing test, and argues (among other things) that a test which purports to be "international" needs to look closely at the assessment criteria used, and rater training, to promote "rhetorical pluralism" rather than culture-bound, inward-looking, Western academic conventions. In short, she seems to be taking IELTS to task for its self-promotion as an 'international' test, if the language construct is modelled on a native speaker variety. This may be a quibble about an ambiguity; is it the test (or the "testing system"[7]) which is 'international', or the 'English'? But it is an interesting allusion to the role of English as the world's academic lingua franca, and the fact that, so far, no exam board has properly tackled the reality of English lingua franca in its constructs, test design, and assessment criteria. If IELTS is being used to place students in university programmes in non-native English speaking countries, then there is a washback issue with the IELTS construct, since this would become a threat to writing styles in world Englishes (a problem raised by Yamuna Kachru 1997). This is not a marginal point, but it needs to be seen within the wider context of a TLU domain, which is no longer native speaker English but English lingua franca. We shall return to this in chapter 5. In the next chapter, we will examine the sample materials published by the boards, to get a more detailed idea of the scope, but also the limitations, of existing certifications.

6 The Pearson 'call for papers' `http://pearsonpte.com/wp-content/uploads/2016/06/26959953_research_call_2016.pdf`, (2017-03-14).

7 IELTS stands for International English Language Testing System

3 A Critique of the Sample Material on the TOEFL, IELTS and PTE Websites

Abstract This chapter provides an in-depth look at the free sample material posted on their websites by the examining boards of their main academic exams, and highlights a number of problems. Firstly, much of the material presented is fragmented, (whereas, typically, complete practice tests are only available to be purchased). Secondly, some of the material appears to be outdated or problematic for other reasons, with a number of poor items or inadvertent language errors, which we note. This is surprising, given the desire to project a positive image of tests which we noted in chapter 2, and the financial resources available to the major boards. Overall, however, the sample material examined can give potential test takers useful insights into the scope, but also the limitations, of each certification.

3.1 The Shortcomings of Sample Materials

All boards necessarily publish information and advice about how to prepare for their certification. As well as detailed descriptions of the test structure, and information about scoring and reporting, there will usually be links to free materials, and publicity for practice tests and other materials which can be purchased. In this chapter we shall take a critical look at some of the free practice materials made available for TOEFL, IELTS, and PTE, on the assumption that would-be test takers – and teachers preparing students for the tests – will first look at these materials to get an idea of test structure, test items, and difficulty, before deciding to enrol for the test and invest in practice tests which are for sale.[1]

However, it is not easy to get a complete overview of any single test. The free material tends to be fragmented, taking the form of downloadable PDF files, listening files, and partially interactive pages which focus on different parts of the exam. None of the websites administer a complete, timed, practice test in the mode (Internet-based for TOEFL, paper-based for IELTS, computer-based for PTE) used in the actual test. This is perhaps not surprising, especially if the boards want to sell practice material or complete, fully interactive (in the case of the PTE) sample tests.

1 All the sample material analysed in this chapter was accessed in March 2017. It may of course have been removed, changed, or updated since then.

More surprising is the fact that a lot of the material is outdated, or problematic for other reasons, as we shall see. All three examining boards are guilty of posting problematic items, or texts which are plainly out of date, or which contain unintentional language errors. Of course, there is no such thing as a 'perfect test'. A 'test' is an abstraction, or rather a set of specifications, which Alderson et al. (1995, 294) refer to as "blueprints" which generate forms, the items and test tasks, which themselves, following the advice of Hughes (2003, 54) need to be sampled "widely and unpredictably" from the target language use domain. Items which do not work do not therefore necessarily indicate a 'bad test', but a glitch in the process of test production and validation.

However, it has to be of concern that all three websites offer problematic samples, some of which have been on offer for a long time. The boards do not give much information about where their sample material has been taken from, or whether it has been used in actual tests. If they are taken from previous tests, it is curious that some at least of the problematic items commented on in this chapter were not identified as such and weeded out after the test. What selection criteria were used, one might wonder, to choose sample items, if not to give an accurate idea of texts, tasks, and level of difficulty? Given the self image that boards are keen to promote of themselves (chapter 2), and the funds which they presumably have available for this purpose, it seems strange that they have not been more careful about which material to use in these key sections of their websites.

Nonetheless, for the potential user browsing these websites with a view to comparing the three tests and choosing one of them, the sample materials offer ample evidence of the ways in which the tests are different, although ostensibly testing the same skills and having the same function.

3.2 The TOEFL Sample Material

3.2.1 Critique of Sample Material for TOEFL iBT: Overview

The TOEFL website takes potential test takers to at least three different sources of free test material: a 32 page PDF document, an "interactive sampler" which can be downloaded and which reproduces the test interface, and a "Quick Prep" resource which offers further samples from all four sections of the test, in a PDF which includes embedded listening texts. Thus similar material is offered, in a paper version, in an interactive form, and in a semi-interactive paper version which connects to listening texts. This sample material is offered alongside other free resources, such as a "Test Prep Planner", and "Tips", all of which probably make it hard for potential test takers to understand where to go first to get an idea of

the test, or indeed which practice materials is likely to be the most useful, or the most recent. In addition, a separate link leads to a "TOEFL video library" with eighteen short video clips offering mini-tutorials on different parts of the test. In this critique we will consider in detail the paper and the interactive versions of the free test materials.

Neither resource offers a complete version of the test; the description of the interactive sampler as offering "free unlimited access to past TOEFL questions from all sections of the test" is misleading; the "unlimited access" means that one can return to the same sample and redo the questions. Other, more complete resources, are offered for sale on the same webpage.

Nonetheless, the sample material gives a good overview of what to expect in the exam. Both the PDF material, and the interactive sampler, which uses different questions, we are told, comprise actual questions from previous tests. Presumably the interactive sampler will give potential test takers a closer idea of what it is like to take the test, while the paper version is offered to help students get an idea of the test even if they do not have access to a PC or the required Windows 7+ operating system.

3.2.2 The PDF Sample

The PDF document has one reading text, with questions and answers; the tapescripts for two listening tasks, with questions and answers, questions, texts and tapescripts for all six speaking tasks, and questions for both writing tasks, with two sample answers for each, and a rationale behind the scores awarded.

The single reading text offers a possible explanation for the extinction of the dinosaurs. The text (for this reader at least) is well written, and interesting. Questions follow the order in which the information in the text is presented. Five are about understanding words in context, while other functions range from extracting main ideas, to inferring information, paragraph completion, and understanding writer's purpose.

The first 13 of the 14 items are MCQs with 4 options. To respond to the last question, number 14, which is worth two points, candidates have to select three correct pieces of information from six statements, and transfer them onto the answer sheet.

There appear to be a number of problematic items, especially those testing lexis. Q8 and Q9 can both be answered without reference to the text:

8 The phrase 'tentatively identified' on line 36 is closest in meaning to
a identified after careful study
b identified without certainty (correct answer)
c occasionally identified
d easily identified

9 The word ‘perspective’ on line 46 is closest in meaning to
a sense of values
b point of view (correct answer)
c calculation
d complication

Both terms, “tentatively identified” and “perspective”, belong to a fairly standard academic lexis; the questions require test takers to recognize standard definitions. For a student with a European background, in higher education, or aspiring to higher education, this lexis should not be problematic; if anything, he or she might assume that the obvious answer is wrong (since MCQ options are intended to distract), and be persuaded to choose a different answer, on the assumption that the term might have acquired a different meaning in a specific context.

Q2, which asks students why the writer includes the information that dinosaurs “had flourished for tens of millions of years and then suddenly disappeared”, also appears to be problematic, since at least two of the options (c and d) seem credible.

The listening section offers the tapescripts of two texts, a dialogue (between a male basketball coach and a female member of the basketball team) and a monologue (an extract from a lecture about a novel by Wilkie Collins). The dialogue has lots of discourse markers built in, to highlight the informal tone (*yep*, *wow*, *well*, *oh*, *okay*, *good*), and the five questions are on understanding the main points (Qs 1-2) and recognizing communicative functions (Qs 3-5). These questions do not appear to present any particular difficulty.

The monologue (approximately 900 words) seems more demanding. It focuses on what is generally considered to be the first modern detective novel in English, *The Moonstone*. As such, at least two answers might be accessible to test takers who are familiar with the novel;

Q7 In what way is the Moonstone different from earlier novels featuring a detective?
a in its unusual ending
b in its unique characters
c in its focus on a serious crime
d in its greater length

Q8 According to the professor, what do roses in *The Moonstone* represent?
a A key clue that leads to the solving of the mystery
b A relief and comfort to the detective
c Romance between the main characters
d Brilliant ideas that occur to the detective

A major flaw, which would be easy to put right, is the key to the listening section, which gives answers for questions 15-25; but for each tapescript questions are numbered, respectively, 1-5 and 1-7.

In the speaking part of the sample, the directions remind readers that in the actual test they would be listening as well as speaking, and in some cases, reading, listening and speaking. The first two tasks are responses to invitations to "speak coherently and clearly about familiar topics". The first of these is to remember a pleasant and memorable event at school. This may seem deceptively simple, but for many candidates (and not just those with unpleasant or indeed traumatic memories of school) it could be problematic to remember a single pleasant event. The second task, which invites a comparison between two modes of spending time with friends (at home, or in a café or restaurant) seems more accessible, and better structured.

The remaining tasks integrate reading and listening tasks with a spoken outcome, and are more university-oriented. They get students to speak about social facilities on campus, the psychological notion of 'flow', problems students are facing following a course in calculus, and two definitions of 'tool' which are presented in an extract from a biology lecture. Tapescripts, complete with phonological reductions such as *gotta* in the extract from a lecture, replace the listening files of the actual exam.

The short reading text on flow is potentially problematic: it is presented as an extract from a psychology textbook, and students of psychology may well find it easier than other test takers to respond to the question "Explain *flow* and how the example used by the professor illustrates the concept". In the same way, in the extract from the biology lecture, the two definitions of tool, as used by animals in nature, will probably come more easily to students familiar with the topic.

The last section, writing, includes a reading and listening to writing activity (on the vote counting system in the US) and a free-standing essay on the topic of what makes a good teacher. The two questions are both provided with two sample answers, both of which are at the top end of the 1-5 scale, the first sample scoring four points, and the second five. This is useful feedback. However, there seems to be a large gap in level between the first two sample answers (on the vote counting system), the second one being deficient in organization, and containing a large number of formal errors not usually acceptable in an academic writing context. It begins: "The leture (*sic*) disagreed with the article's opinions" and continues in the same vein. If this deserves a mark of 4/5, one might wonder what a score of 3 looks like.

3.2.3 The Online Sampler

One of the main functions of the on-line sampler is to get would-be candidates to have a feel of the exam, both in using the interface, and in the timings of tasks. Like the PDF material, it does not offer a complete test, but it samples rather differently: there are three reading texts (compared to only one in the PDF), but only one speaking task; the listening tasks mirror the PDF (one 'campus' type dialogue, and an extract from a lecture) and there are two writing tasks, one based on a written and spoken input, the other being the free-standing listening task.

The interface is uncluttered and user-friendly, with a time bar indicating how much time is left to listen, or to answer, but moving from one section to another can be a laborious task, as one always has to move through a pre-programmed sequence; for example, users cannot go directly from the reading to the writing section, but have to click their way through listening and speaking, and they cannot return to a previous section. A drop-down menu could have rectified this and made browsing easier. Some feedback is given; students can use a button to reveal the correct answer to the reading and listening questions, and the speaking task is provided with two sample answers.

The three reading tasks deal with science and technology (windpower and botany) and prehistoric art (cave paintings). The first, on three theories behind the cave paintings in Lascaux, is noticeable because it begins with a non-standard sentence:

> "In South-West France in the 1940's playing children discovered Lascaux grotto"

in which *playing*, to indicate progressive aspect, rather than a compound structure (*playing field*, *playing card*), would normally follow the noun. The structure does not compromise understanding in any way, but it is curious since it suggests that the writer might not be a native speaker; whereas all the certifications described in this chapter are based on native speaker models of the language. We shall return to the theme of non-native input in chapters 6 and 7.

As we saw with the reading text in the PDF sample, here, too, some reading questions can be answered by candidates without actually reading the text. In the second passage, for example, (on wind farms), the first questions appeals to basic understanding of what is meant by 'wind farm':

Q1 Based on the information in paragraph 1, which of the following best explains the term wind farms?[2]

a Farms using windmills to pump water
b Research centers exploring the uses of wind
c Types of power plant common in North Dakota
d Collections of wind turbines producing electric power

Logical inference might help the candidate to choose the correct answer in other cases, again without reading the text, as in the following question on cave paintings:

Q8 According to paragraph 4, why do some scholars believe that the paintings were related to hunting?

a Because some tools used for painting were also used for hunting
b Because cave inhabitants were known to prefer animal food rather than plant food
c Because some of the animals are shown wounded by weapons
d Because many hunters were also typically painters

in which options a, b and d take for granted that a lot was known about the prehistoric cave dwellers, while only c reasons from the evidence which emerges from the paintings themselves.

As with the PDF reading, here, too, there are a number of vocabulary items the meanings of which test takers are expected to infer from context. Some of them, however, are likely to be recognizable to European candidates because they are cognate with words in their own languages (e.g.: *methods*, *emit*, *accompanied*, *massive*) and, if so, not ideal items in a test of reading.

One question (n. 6) related to the third passage (opportunist *vs* competing plants) asks "Which of the sentences below best expresses the essential information in the highlighted sentence in the passage?"; but no sentence appears to be highlighted. As with the incorrect answer key in the PDF, this could easily be corrected.[3]

There are two listening tasks, preceded by rather lengthy instructions, which include a rationale about what the listening section does. The first task involves listening to a conversation between a professor and a student who has missed a class. The dialogue is clear, and the five questions straightforward. At one point the professor offers to lend the student a video tape, asking her if she has a VCR at home – dating the passage,

2 'wind farms' is not highlighted in any way, through italics or quote marks.

3 The sentence is however highlighted in a PDF version of the sample material. URL `http://toefl.uobabylon.edu.iq/papers/ibt_2015_1821899.pdf` (2017-11-01).

probably, to the pre-DVD 1990s, and impacting negatively on the image of the test.

The second listening text is, following the pattern we have already seen, much more academic in style. It is an extract from a lecture, it is long, and describes a quite complex process of crystallization. There are six questions.

The third section has two speaking tasks. The first one asks candidates to talk about on-campus accommodation for students. We are provided with a single, mid-level, sample response. The candidate speaks carefully and slowly, in clear accurate English. He answers the question well, makes virtually no formal errors, produces language such as:

> "In my opinion it would be in the better interest of a first year student to live in a dormitory on campus but I wouldn't make it a requirement but make it a personal choice."

and yet is called 'not fluent' because he speaks slowly.

He seems to have been knocked back only on speed of response. He has (in the opinion of the author) addressed the question appropriately, and thoughtfully, and the judgement we read "He does provide some relevant information but in general the topic is not sufficiently developed to score at the highest level" thus seems unfair. If this is a bad performance, or 'mid-level', it would have been more useful to have a sample of a good performance. It would also been useful to have a numerical score for 'mid-level': does the performance score 3/4 or 2/4?

The problem of the sample responses is compounded with the second speaking activity, in which test takers read an introductory sentence (in this case about taming herd animals), listen to an extract from a lecture on the same topic, and then have to explain how the behaviour of horses and antelopes as herd animals relates to their suitability for domestication. The sample performance is (rightly) flagged as low level, while poor pronunciation and a background hiss make it extremely difficult to follow. But this is of little help to the would-be test taker; far more useful would have been to provide an example (or examples) of a good response, at a mid-to-high level.

The final section contains two writing tasks. The first of these is a carefully structured test of writing from a reading and listening input. Candidates are given three minutes to read a short text (approximately 250 words, on the altruism of meerkats), after which the text is removed and they listen to a lecturer refuting some of the information given in the text. This second part lasts for about two minutes. The writing task is to summarize the lecture, and show how it sheds doubt on the reading passage. However, instead of being given the opportunity to write the text, candidates are then shown three responses, at different levels (high, mid, and low). There is no comment on these sample responses, but there does appear to be a clear gap in level between them, in terms of content, accuracy, and range.

The free-standing task, in contrast, offers test takers the chance to write using the test interface. They have 30 minutes to complete the task (a 300 word essay on whether or not telling the truth should always be the most important consideration in a human relationship). One useful interactive feature is the word count, which charts students' progress as they write. However, there are no sample answers to compare with one's own.

3.2.4 TOEFL Sample Material: Concluding Remarks

Although there is plenty of sample material, in a range of formats, it is badly organized, and it is not possible to do a complete practice test. The website shows signs of age, as does the material (some of which has been left untouched for years), while newer pages and texts have been added. The result is a lack of guidance for potential test-takers coming to the site hoping to have a clear idea of how the test is structured and what they will have to do.

3.3 The IELTS Sample Material

3.3.1 Critique of Sample Material for IELTS: Overview

The sample material is easily accessed through a user-friendly website. A drop-down menu "About the test" on the uncluttered homepage takes students to pages headed "two types of test" (which distinguishes between the academic version of the test and the general training version), "test format", where students can click on one of the four section headings (listening, reading, writing, speaking) to find out more about the test structure, and a third page, "test format in detail" which uses the same interactive section headings, but this time by clicking on them students find out more about the tasks, the question types and the scoring. The fourth page has links to sample material for each section of both versions of the test. There are other pages, too, about scoring, test development, fairness and security (amongst other things) but the first four pages give enough information for potential candidates to have a very good idea of what to expect in the test itself.

3.3.2 The Sample Material in Detail

There are nine sample listening texts, which take the form of mp3 files. This compares with the four texts of the actual exam. Along with the

questions and the answers, tapescripts of the texts are also supplied, but, inexplicably, only for the first seven tasks.

The first seven are short texts, some very short, and are representative of the first three listening tasks, which include two conversations and a short monologue, such as an announcement. Texts eight and nine are much longer (more than five minutes each), as is the final monologue in the actual exam.

Why does IELTS offer so many listening texts? The obvious answer is to offer potential test takers as much practice as possible. But since one of the most publicized features of the test is the range of (native speaker) accents it uses, this could also account for the wide sample. We hear UK, US and Australian accents in the samples, although some of them (such as the male 'American' in sample 4 and the female 'American' in sample 5) appear to be British English speakers attempting to put on American accents. This lack of authenticity is even more obvious in the first sample, in which a British English speaker plays the role of a Kenyan man (who is presumably not a mother tongue speaker of English) wanting to ship goods back to Africa.

The listening samples are also organized so that each one gives practice with different question types, such as matching, sentence completion, multiple choice, and labelling. The seven reading samples are similarly organized. Topics range from dung beetles to rocket science and pollution from cars, the risks of cigarette smoke, and agriculture and the environment. Like the TOEFL sample, much of the material seems dated: the reference to the 1986 Round Table, for example, on multilateral trade agreements, suggests that this is a recent event the results of which have not yet been felt.

There are other ways in which the samples are likely to be less than satisfactory, or even confusing, for would-be test takers. One of the samples (on the dung beetle) is used twice (samples one and seven) to be exploited through different question types. Another sample, on agriculture and the environment (sample six) is recycled in part in the next text (sample seven), which begins: "All these activities may have damaging environmental impacts". It thus begins with a reference to part of text which is not shown, although the introductory rubric reads: "The text preceding this extract explained how subsidies can lead to activities which cause uneconomical an irreversible changes to the environment". A student coming to this sample might wonder if it is standard practice for texts to begin like this one *in medias res*; there is no answer to this question.

A further curiosity is the choice of (very) different fonts for the different reading texts. There seems to be no reason for this, unless it is to indicate that they come from different sources, and thereby hint at 'authenticity'. But they are not facsimiles, and although we read in the introductory material that "texts are taken from books, journals, magazines and newspapers, and have been written for a non-specialist audience", no credits are given to indicate the actual sources – if indeed they have not been specially written for the exam.

The overall effect then, of the sample listening and reading material, is to give candidates an idea of the type of questions they will have to answer, but not the feel of a complete reading paper, which would have only three long texts, but to each of which would be appended two or more question types.

The third part of the test, writing, is adequately covered in the sample. In the actual exam candidates have to write a 150 word report synthesizing or summarizing material from a visual input such as a graph or chart. There are two example questions. This is followed by a 250 word argumentative essay, in which the writer is invited to agree or disagree with an opinion, and provide arguments in support of their choice. Here, too, there are two sample questions. However, (unlike the TOEFL pages) there are no sample answers.

What is most striking about this material is that it is decades out of date. Both of the graphic input questions present data from the 1990s, while the first essay question reads:

> The first car appeared on British roads in 1888. By the year 2000 there may be as many as 29 million vehicles on British roads. Alternative forms of transport should be encouraged and international laws introduced to control car ownership and use. To what extent do you agree or disagree?

To refer to the year 2000 as if it belonged to the distant future, rather than a rapidly receding past age, is likely to lead to a moment of disbelief for the would-be candidate looking at this material. Admittedly it could be justified as providing an example of the kind of writing task a candidate could be faced with. But to use it as an actual test item today would be unthinkable – how would the candidate begin to answer it? – and one can only wonder why a more updated sample has not been provided.

The speaking part of IELTS is a free standing test which can be taken on a different day from the rest of the exam. The sample material covers all three phases, the introduction, the 'long turn', in which the candidate talks about a topic which is provided by the examiner, and the discussion. Separate PDF files are provided with the questions, and tapescripts and mp3 listening files with the partial answers of one candidate. It is not clear why we are not given the candidate's complete performance; after all, the complete speaking test lasts only for around twelve minutes. A further failing is that we are not provided with any indication of how the candidate is rated. Although he speaks clearly, makes few formal errors, and provides thoughtful, intelligent answers to questions, there are long pauses; any would-be candidate listening to this sample would probably want to know if they are penalized for the pauses.

At this point, we can refer to the band descriptors for speaking (public version) which are available on the IELTS website. Fluency, we note, is one of the assessment criteria, but 'content-related hesitation' does not

prevent a high mark (band 8 or 9) from being awarded. Interestingly, this seems to be in contrast with the approach to fluency in the TOEFL listening sample which we referred to earlier in this chapter, and in which performance was penalized because of the slow delivery.

The question is, whether or not the long (eight seconds!) pause in the sample material is 'content related'; the candidate seems to be thinking of something more to say in the 'long turn', (so hesitating as he searches for content), rather than experiencing any particular language problem, until the silence is broken by the examiner with a prompt. There are other long hesitations in the same file, and as such it seems like a strange choice to offer as a sample candidate performance.

3.3.3 IELTS Sample Material: Concluding Remarks

The sample material does not do justice to the overall well-designed website. As with the TOEFL website, we find material that is incomplete, or dated; and for the subjective parts (writing and speaking) there are no examples of candidate performance (writing) or evaluation of performance (speaking). It is difficult to understand why this should be, since these could have been provided fairly easily.

Like TOEFL the IELTS website does not offer a complete version of an exam for practice, but, again like the TOEFL website, it offers a range of practice materials, including complete tests, for sale.

3.4 PTE Academic Sample Material

3.4.1 Critique of Sample Material for PTE Academic: Overview

The PTE is the most recent academic certification on the market (2009). It has an uncluttered home page making it easy to find and access the sample material through the drop-down menu. The "Test taker" menu leads to a "Preparation" page, where one option for "free materials" is given alongside a range of materials which are for sale: "scored practice test", "sample questions", and course books.

The "free materials" link offers four features: two PDF documents "Test tutorial" and "Top Tips", a "Skills video" which is a collection of short YouTube clips showing the range of tasks the candidate has to perform in the test, and an "Offline practice test". This latter brings together sample questions and answers, with comments, for the productive skills, on sample performances of test takers. Since the same sample items (or at least, some of them) turn up in all four blocks of materials, we shall focus

primarily on the Offline practice test, which offers the most materials.

The "Test tutorial" and "Top Tips" documents are similar in form and size, the former being a 38 page document, the latter having 40 pages. The tutorial focuses mostly on test procedure, while the "Top Tips" pages pepper the sample tasks with brief exam-taking strategies such as "Use punctuation to help you decide when to pause when you read" or "Skim the text before the reading begins"; the tips frequently border on the trite, such as "Make good use of the 40 second speaking time". Some of the tips are also to be found in the tutorial document, and some of them are self evident, such as "Use correct punctuation for writing tasks" or "Don't click NEXT before you have completed the task and are ready to move on". Given the overlap of function and content, these two documents could probably have been more usefully combined as a single document giving procedural information and exam strategies.

The sample videos have been uploaded onto YouTube. Rather misleadingly called "Skills" videos, they are brief, approximately half minute clips which familiarize potential test takers with the procedures for the many different tasks. However, the clips fade out after the instructions have been completed, or during sample student responses; they are not intended as practice material.

3.4.2 The Offline Practice Test

The "Offline practice test" is more than a test, since it offers multiple items for some of the shorter task types, such as "repeat sentence", "describe image" and "answer short question". Readers see screenshots of the tasks, and on a later page are given the answers (for objective type questions) or can read or listen to sample candidate answers.

The sample responses are a strong feature of the practice test, since for each task three responses are given, illustrating three key levels B1, B2 and C1 of the CEFR. It is interesting that no reference is made to the Pearson "Global Scale of English" (see chapter 2, 2.3), so we do not know the exact score for each response. However, the Framework levels will probably be far more meaningful for prospective universities, which are likely to discard B1 candidates as below level, to view B2 candidates as potential students, and C1 candidates as fulfilling all language requirements.

The samples are either written, or, for speaking activities, available on audio files. Besides the attribution of a CEFR level, each response is described in four or five lines. This is likely to give useful feedback about the test to would-be candidates, who can identify those features which are clearly below level. One comment, for a B1 level response to a 'describe image' task, can suffice:

> Two basic elements of the graph are described, but the main idea is not discussed. While there are a few phrases spoken at a natural rate, fluency is negatively affected by multiple hesitations and long pauses. Incorrect pronunciation of consonants might require listeners to adjust to the accent of the speaker. There is limited control over simple grammatical and lexical structures. This response lasts for 31 seconds.

This response is below level because the task has not been fulfilled, it is not sufficiently fluent, consonant production is problematic, and there are grammatical and lexical inaccuracies. The reference to "incorrect pronunciation" is interesting, since it suggests that the need for the listener to "adjust to the accent" is a negative feature of the response, whereas listeners always have to adjust to accents, whether native or non-native, and the sample texts which students have to listen to in this offline test contain a range of British, American and Australian accents.

The range of questions on offer is helpful, but also points to problems. For example, one of the speaking activities, "Answer short questions" is more of a listening comprehension and vocabulary check than a speaking task. Candidates listen to definitions and identify the word. This in itself is not necessarily an inappropriate task in an academic test, but it could be relabelled, or re-presented, as a listening task. The problems arise in the variety of the nature of the task. In some sample questions, students have to identify one of three given words, such as:

> Which is the longest – a decade, a millennium, or a century?

making it a three-option multiple choice question, whereas in other questions, candidates are not given the target word, but have to work it out from themselves, as in:

> What key mineral makes seawater different from freshwater?

or

> If a figure is hexagonal, how many sides does it have?

This last question could easily have been turned into "If a figure is hexagonal, does it have five, six or seven sides?" making it into a qualitatively different type of question.

There also seems to be a labelling problem with the "Retell lecture" task. The first of three samples comes across as an extract from a lecture. The pauses, added emphases, and overall intonation patterns all give it an authentic feel. The other two samples, however, are extracts from inter-

views, both involving two speakers, an interviewer, who takes two turns, and an interviewee, making them dialogues rather than monologues, and the title "lecture" a misnomer.

A further problem with items arises in the "Highlight incorrect words" in the listening section, in a sample task which features in the "Tutorial" but is not repeated in the offline test. This concerns possible test bias (which we also noted in a TOEFL reading passage). In this task, test takers have to listen and highlight the words which are different from the text they have in front of them on the screen. In the following introduction, however, the story of Amundsen's quest for the north-west passage may be familiar to many European students, enabling them to identify the incorrect words for the wrong reasons:

> When explorer Roald Amundsen set out to find the Northwest *Pasture*, his official mission was scientific – a search for the magnetic *south* pole. (italics added)

A further limitation of the sample material is to be seen in the quality of some of the written texts. Here are two extracts from the short texts used as examples for the "fill in the blanks" task in the reading section:

> Up until our research the predominate wisdom in the scientific community was that umami was not a separate sense.

> Peering into the future seldom produces a clear picture. But this is not the circumstances with bio-energy.

"Predominate" as an adjective? All learner dictionaries give this word as a verb, and only by going to a big dictionary, such as the Oxford Shorter, do we find the entry "Now rare. [App. a mistaken form for predominant]". "Circumstances" a singular noun with a plural marker? Or should it have been "These are not the circumstances"?

These two errors – or slips – seem all the more problematic in that both "predominate" and "circumstances" are the target words in the activity, which test takers have to select to complete the texts. Here, the danger is that they discard the correct answers for the right reasons: their superior (to the item writer's) knowledge of the language.

3.4.3 PTE Academic Sample Material: Concluding Remarks

Of the three certifications, the PTE website probably offers the most complete information to potential candidates. As we noted, the provision of sample responses, with comments, at three different levels, for all the

productive tasks, is a strong feature. In addition, in keeping with the user-friendly home page, clarity of design is also a feature of the downloadable booklets. These are written in an appropriate style for learners of English, addressing them directly, and highlighting important information in bold.

The sample material is packaged in a zipped file (which may take some time to download). Apart from the problematic items, which we have commented on, one noteworthy feature is the disparity in the number of samples per task type. For most task types there are two examples, but there are six "read aloud" and "describe image" questions, and no fewer than ten examples of "repeat sentence" and "answer short question". Why, one wonders, does Pearson want us to have ten examples of such a straightforward activity as "Please repeat the sentence exactly as you hear it"? The unintended message that may be inferred by potential candidates is that this activity has to be repeated several times in the actual test. Perhaps their interests would be better served if this file were to be labelled "Sample questions and answers" rather than "Practice Test".

4 An Experiment in 'Co-Certification'

Abstract In chapter 4 we report on a 'co-certification', a project developed jointly by the University of Ca' Foscari Venice and one of the best-known examining boards operating in Italy, Trinity College London. The rationale behind the project (which still today appears to be unique in its genre) was to adapt part of an existing international certification, *Integrated Skills in English* (levels B2 and C1) to suit the needs of a local institution: in this case, by introducing writing tasks which were more appropriate for Italian university students than the generic, 'politically correct' tasks for the international market. The project, which continued in its original form from 2004 until 2014, was doubly attractive to candidates, not only for reasons of content validity, but because of its dual function, since it could be used both as an external certification, but also to replace in-house university exams.

4.1 Background: the Growth of Certification in the New Millennium

With hindsight, the turn of the millennium seems to have ushered in a new era of language teaching, learning and assessment in Europe, and sanctioned English as the default foreign language to be taught in schools. The coming of age of a communicative approach to language teaching, the publication of the Common European Framework, and the introduction of foreign languages in primary schools, all played their part in this phenomenon. That the choice of foreign language usually fell on English reflected the fact not only that English had become the world's preferred lingua franca in virtually every domain of human activity, ranging from sport to academia, but also that in Europe it had resoundingly taken over from French as the main working language in the EU.

In Italy, it was also the moment when language certifications began to make their presence felt in schools and universities. Protocols signed by the Italian Ministry of Education (MIUR) and the examining boards made it possible for certifications to be used in the public sector,[1] while projects such as *Progetto Lingue 2000* (for schools) and *Campus One* (for higher education) provided the organisational frameworks. Schools were able to

1 For an updated list of recognized examining boards see http://hubmiur.pubblica.istruzione.it/web/istruzione/dg-personale-scolastico/enti-certificatori-lingue-straniere (2017-11-01).

obtain European funding through the *Programma Operativo Nazionale*[2] to prepare pupils for language certification, and many schools and universities became exam centres for certifications, a role which had previously been the exclusive domain of private language schools. Cambridge ESOL and Trinity College London quickly established themselves as the main providers of certifications to the secondary sector, with Cambridge PET (B1) and FCE (B2) the preferred tests for upper secondary level, and the Trinity GESE suite (a test of spoken English) at lower secondary level.

In the universities the arrival of language certifications can be seen as an offshoot of the 1999 reform[3] which introduced, among other things, the three year first degree, known as the *laurea triennale* or *laurea breve,* the two year second level *laurea specialistica* (later renamed *laurea magistrale*), and the European Credit Transfer System (ECTS). Certifications, especially those linked to the CEFR, came to be used for gate-keeping functions such as providing proof of a minimum entrance level to specific degree courses. This soon settled down, in the second half of the first decade of the new millennium, to a nationwide requirement of B1 to access first level degree courses, and more recently, B2 to access second level degree courses. We shall discuss the reasons behind these choices in chapter 6.

Certification could also be used to substitute existing in-house exams, or part of them. This was an attractive possibility in faculties (such as science and economics) where language courses were compulsory components of degree programmes, but were not seen as core to the curriculum. To replace an English language exam with a certification, for those students who could afford to pay for it – certification could only be used as an option to an existing exam – brought benefits to test takers and test users alike. In one fell swoop, the student passed a university exam and gained an internationally recognized certification, while the university saved on the costs of administering its own exam.

4.2 The Generic Nature of Global Certification

In the language faculties, however, it was a different story. For a degree in modern languages, with its special emphasis on literature, linguistics and translation, the scope for certification was more limited; after all, given that teaching and assessment is at the heart of any university degree course, why should language specialists relinquish the assessment of student progress in their own discipline? At most, certification, with

2 For the rationale behind the PON see http://www.istruzione.it/pon/ilpon.html#sec_pro (2017-02-02).

3 http://www.miur.it/0006Menu_C/0012Docume/0098Normat/2088Regola.htm (2017-10-26).

its newly found Framework-related parameters, could perhaps be used to substitute the generic components of language courses.

This reluctance to engage with certification in language faculties was compounded by the generic nature of certification. TOEFL, IELTS, and PTE, as we have seen, are intended for university access in English speaking countries for students from anywhere in the world, and as such have to take steps to avoid culture-related bias, for example in the choice of texts used. As a result, texts which might have a specifically European cultural context, and which could be of interest to students in European universities, are not used. In contrast, Framework-related certification in Europe, such as the suites developed by Cambridge ESOL and Trinity College London, seemed in their target language and task types to be aimed at younger learners, in the 16 to 18 age group, rather than at university students, reflecting the large numbers of test takers in schools.

Some of the doubts that test users may feel towards certifications in a university context are listed by Balboni (2012a, 115). They range from theoretical (is it really possible to certify a 'level'?) to methodological (certifications encourage teaching to a generic test, instead of testing the unique curriculum of an institution), from ethical (how valid are the validation processes?) to sociological (certifications are big businesses which often operate in near monopolies). But the same author also acknowledges the importance of certification in a "recognition-based society" (Balboni 2012a, 115), and even sees ways of formally harnessing them to university entrance requirements, such as offering discounts on registration fees to applicants who have certification, and who thereby absolve universities from having to make their own (costly) initial assessments (Balboni 2012b, 121).

4.3 Co-Certification Conceived

At the Faculty of Languages at the University of Ca' Foscari Venice the advent of certification was viewed with interest. At the time, Ca' Foscari had the largest modern language faculty in Italy, offering degree courses in 42 languages, with the biggest concentrations of students in oriental languages (especially Chinese and Japanese) and the major European languages.[4] The English language teaching programme had undergone a considerable overhaul, with the introduction of a new, Framework-related, syllabus, integrated across the three years of the new *laurea breve*. In addition, the exam was no longer harnessed to a literature syllabus, which may have fuzzed the language learning objectives in the exam in the old, pre-reform, four year degree course now known as the *vecchio ordinamento*.

4 With the Gelmini law of 2010, university faculties were replaced by departments.

The new syllabus had been partly informed by a student survey of final year *vecchio ordinamento* students,[5] in which a resounding majority of respondents said they believed that a new syllabus should be linked to specified internationally recognized levels. They also believed that speaking and writing should be the primary focus of language teaching, followed by reading, listening, grammar and phonology (in that order). The syllabus thus took up many of the *can do* statements in the Framework, adapted some of them slightly, and articulated them as year by year attainment targets. These targets were set at B2 (for the end of the first year), and presumed that the second year would be a year of consolidation or maintenance at around B2+/C1-, with students required to have reached C1 at the end of the third year, when they graduated. The new syllabus inevitably took a skills-based approach, with writing featuring prominently and speaking targets also listed; astonishingly, speaking had not previously been formally assessed as an independent skill.

In this context of major change, newly developed CEFR-related certification was seen as a potential ally. It took a direct approach to testing skills, and there was a substantial convergence of content with new university programmes. But it was too generic, aimed at younger learners, and not sufficiently academic to be considered equivalent to a university exam. If, however, (went the reasoning in the Faculty of Languages), a local version of an international certification could be created, responding to the needs and profiles of Italian university students, then (judging by the feedback from the student questionnaire) it would have positive washback and it would reinforce the new syllabus as an alternative, but equivalent, means of assessment.

It was with this possible scenario in mind that in 2004 the Dean of the Faculty approached Trinity College London with a proposal to adapt their new *Integrated Skills in English* ISE3 exam, set at level C1 of the CEFR, so that it could be used as an in-house university exam, equivalent to the general language part of the final year exam in English language, while retaining its value as an external certification. The choice fell on the ISE suite for a variety of reasons. In the first place, it was the first exam suite to have been linked from the outset to the newly published CEFR, rather than to have tweaked its existing exams to the level descriptions of the Framework (as was the case with Cambridge ESOL certification). Secondly, Trinity had a performance-based, whole-skills approach to language assessment which seemed to sit well with the focus on productive skills of writing and speaking in the new syllabus. Furthermore, Trinity College already had a strong presence in Italy, especially through its Graded Examinations in Spoken English (GESE) which were popular in schools,

5 Reported in Newbold 2004.

and this could provide a guarantee of local support and assistance. The rationale behind this choice is described in detail in Newbold (2009).

The proposal found an interested interlocutor, and after a series of meetings to discuss financial, theoretical, and especially operational aspects of the project, an agreement was reached to produce a "co-certification", to be developed and administered jointly by Ca' Foscari and by Trinity College. In retrospect this seems to have been quite an unusual arrangement. As far as we are aware, there are no other examples of co-certifications involving a major examining board and a local institution to produce a tailor-made version of a certification for local consumption. But much as the idea might appeal, there are a number of reasons, especially involving the need for a clear definition of roles, why embarking on a co-certification may prove difficult, as we shall see.

In principle, any form of collaboration which brings test developer and test user together is likely to work in the interests of fairness, as Kunnan (2000) points out, since, while the test developer has the duty to produce materials which do not discriminate, the test user has a monitoring function. But whereas the examining boards, as we saw in chapter 3, provide extensive information about tests to test users as well as to test takers and teachers, there seem to be no official channels (such as forums on the boards' websites) for test users to provide test developers with feedback. Test development is more research-led than user-informed.

It thus came as something of a surprise to find a major examining board to be a willing partner in this small-scale project. The first, crucial, hurdle to overcome was the establishment of roles. A three page contract was drawn up, premised on "the common interest of both parties to organize English language exams for students of the University", and asserting that "the organization of such exams is compatible with the institutional aims of both parties". The collaboration which was envisaged involved the pooling of specific resources and competences, and was tersely expressed as follows:

> 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.[6]

This provided the basis for a working partnership in which the University would make suggestions to adapt the international version of the exam,

6 Agreement dated 2004-09-16.

while Trinity College would remain responsible for the entire assessment process. The project would be financed through the fees paid by students (which would not be higher than fees for the international version), and made visible by the logo of both institutions on the certificate awarded to successful candidates.

4.4 A Construct for the Co-Certification

The ISE exam which the University proposed to adopt and to adapt was noteworthy for its portfolio component. In addition to a controlled written exam, and an oral exam, part of the final mark was reserved for a portfolio of three short texts which candidates wrote in their own time, and which they then discussed with the examiner during the oral. This was a new departure for high-stakes certification, since the portfolio texts, by definition, could not be secure in the way that the products of an invigilated exam normally would be. In short, there was nothing to guarantee that the portfolio texts were entirely the candidates' own work. This presumably was why it counted for only twenty per cent of the final mark, while the controlled written exam was more heavily weighted, at thirty per cent.

The interest for portfolios as an alternative form of assessment was undoubtedly linked to the appearance of the European Language Portfolio (ELP) at the same time as the Framework, and which had "a documentation and reporting function, as well as a pedagogic function" (Lenz 2004, 24), designed to promote learner autonomy and self-assessment. This approach to assessment as an ongoing, dynamic process involving the learner had clearly been attractive to test developers at Trinity College and it was also of interest to the University, where it was felt that few students would be tempted to cheat, and where there were checks in place, through feedback forms filled in by teachers who read draft copies of the texts, to ensure where possible that the work was their students' own.

However, there were doubts about the contents of portfolio texts. Although the first two tasks, "correspondence" and "factual writing" both seemed appropriate within the context of the new university syllabus, the third task, "creative writing", was not. Examples of creative writing from previous ISE exams suggested they would appeal to younger learners, and reflect a genre which was a long way from the academic writing which the new syllabus intended to foster. Here, it was felt, there was a need for a university-specific writing task, which we would call 'critical writing'. Recent research, such as Hyland (2002) on rhetorical options open to writers, and Stapleton (2005) on critical writing and interpreting websites as pre-requisites for critical writing, was illuminating, as was Swales' less recent (1990) but well-established excursion into genre studies.

A non-exhaustive list of underlying constructs, or language functions,

was drawn up, which it was felt would contribute to the target language domain of critical writing:

- evaluating
- exemplifying
- contrasting and conceding
- effective organization
- making comparisons
- using persuasion
- using a formal register
- coming to an effective conclusion

This was then articulated in a Framework-like descriptor for critical writing. Starting with the overall description of written production at C1 level in the CEFR:

> Can write clear, well-structured texts of complex subjects, underlying 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.

the *can do* statement for C1 critical writing for the proposed co-certification eventually emerged, after some fine tuning, as:

> 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.

The italicized additions were made to ensure that students from backgrounds other than literature and the humanities would not be excluded; the co-certification was to be open to all students enrolled at the University.

The introduction of the critical writing task was the only change made to the international version of the exam. However, since the controlled written exam replicated some of the writing functions of the portfolio, including the creative writing element, here, too, creative writing was replaced by a critical writing task. Moreover, the fact that the student would speak to the examiner about the portfolio ensured that the 'university element' of the co-certification would be maintained across the three parts of the assessment process. Some of the portfolio questions deliberately had a local European or Italian dimension, which it was felt would make them more accessible, but also more interesting, to students at the University. Here is

a typical list of portfolio critical writing tasks (students choose one only):

1. "Who does the English Language belong to? Its native speakers or anyone who uses it?" Write an essay evaluating both sides of this debate and concluding with your own opinion from your perspective as a learner.
2. Write a critical review of the writer, who in your opinion, best captures the fragmented, globalized dimension of the world in which we live.
3. 'Italian design' is appreciated around the world, but what is it and what characteristics does it display? Write an overview investigating the phenomenon, referring to different examples.
4. Italy has been described as a country where reforms do not happen, although everyone agrees that they should. Write an essay commenting on this paradox. Try to identify the root of the problem and support your views with relevant examples.
5. Too much money is spent by local municipalities on cultural events which are of no great significance or utility to local communities. Discuss this statement indicating how far you consider it to be true, referring to a local context you know well.

The first version of the co-certification was held in 2004, for which there were 38 enrolments, and 33 passes. After this, it was held on a yearly basis, settling down to an average of nearly fifty candidates per year. From 2007 a B2 level of the co-certification was also introduced, with a simplified critical writing construct. The descriptor reads:

> 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 different viewpoints, using appropriate exemplification and showing evidence of effective structuring.

Seventy five students took the first edition of the new B2 level. As with the C1 level, Trinity College provided the university with numerical values for each part of the test, which made it easy to convert results to the Italian system of a mark out of 30, in which 18 is a pass and 30 the highest score obtainable. Although when asked for feedback in a survey (see 4.6 below) most students said that they were interested in the certification for its external value, a large majority of students used the certification, with the converted mark out of thirty, to replace a university exam. For them, the fact that they had also acquired an internationally valid certification was an added bonus; the immediate objective was to pass a university exam. The certificate itself, however, only displayed the grade ("pass", "merit" or

"distinction") which is typical of international certification, as well as the double logo, of the University, and of Trinity College. A separate list of converted scores for internal use was drawn up and published by the University.

4.5 Coping with Crisis and Redefining Roles

A collaborative project such as the co-certification we are describing is based on trust, but also flexibility. The University had relinquished any part in the assessment process, and relied on the examining board not only to provide a fair assessment, but also to report the results rapidly so that they could be used to substitute in-house exams. Trinity College had accepted the idea that the content input for the critical writing questions in the portfolio was the exclusive concern of the University. The initial agreement was that the University would supply ten portfolio tasks, of which Trinity College would choose five, perhaps on the basis of the language that the tasks might be supposed to generate. Trinity also reserved the right to tweak the questions slightly, for example if they did not fit the house style. The controlled written exam, however, was to be produced entirely by Trinity College, which meant that they would provide their own critical writing task, following the style of the portfolio questions, but without such a markedly 'local' character.

This arrangement worked for the first few years, during which time the co-certification settled down as regular yearly event with around fifty candidates for the higher ISE3 (C1) level and slightly smaller numbers for the ISE2 version. In 2009, however, six of the ten suggested questions for ISE2 were rejected by Trinity, together with five of the titles at ISE3, while the wording of three of the remaining questions was called into question. The reasons given were:

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

Although the first point was not controversial, and could be easily rectified by Trinity, the remaining objections were more so. The fact that similar topics (such as university reforms and cinema) recurred in successive years might simply reflect that they were central to the test takers' experience as university students; so long as they focused on different aspects, it was felt, they were not problematic. In contrast, the fact that some titles were considered to be unlikely to elicit language at the required levels was a harder issue to address, and required clarification. Although much has been written about eliciting spoken language through a range of different formats (see May 2010 for a comparison of formats chosen by different

examining boards), there is a dearth of research connecting free-standing essay titles with language elicited.

The most controversial point, however, was the notion that the items writers at the University had chosen topics which were not appropriate for their students. It called into question the initial agreement on which the co-certification was based, namely, that the topics for essay titles should be chosen by the University, and that, therefore, by implication, the University could judge whether or not a topic was 'appropriate'. One question at ISE3 level, for example, required students to reflect on the way in which, in the space of a single generation, Italy had changed "from a country of emigrants to a country of immigrants". Why was this considered to be inappropriate?

All examining boards need to be sensitive to controversial or potentially offensive topics, and as a consequence issue guidelines to their own item writers and examiners. This includes the need to avoid bias in areas such as:

Gender and sexual orientation
Race
Class
Culture
Religion
Nationality

Since they operate in a global market, the net needs to be cast wide; what is not offensive in one culture might be considered to be so in another. For many test developers and item writers, this warning is interpreted as a taboo; better to avoid a text, or an essay title, which explores these areas than to risk offence. The topic of immigration, which might overlap with issues of race, culture, and nationality, had been weeded out as 'inappropriate'.

However, in a local context perspectives change. For the University team, to engage with the topic of immigration into Italy in a critical writing task seemed not only appropriate but also stimulating and linguistically challenging. Indeed, the culture-specific setting of the co-certification meant that there were very few areas from the list which were likely to cause offence *a priori*; offence could lie in a biased or stereotyped approach to the topic, but not as an inherent feature of the topic. The 'added value' of the co-certification, it was felt, lay precisely in the possibility of offering themes which might not be available in the more anodyne international version, and which could be sensitively explored by test developers and test takers alike.

Another problem which arose at about the same time was of a completely different nature. The controlled written question, we said, was provided by Trinity along the lines of the portfolio questions. Unlike the portfolio questions, however, for which candidates had a choice, and which could be written at home, using a variety of resources such as dictionaries and the Internet, the controlled written exam offered no choice and no resources.

Thus, when students found themselves asked, in the controlled exam, to describe and reflect on a painting they knew well, many found themselves in difficulty. A question which could have been used for the portfolio – students would have been able to look at the picture as they wrote – turned out to pose major problems for anyone without photographic recall. The protocol for high stakes test administration (and this included the co-certification, which followed the standard procedure for the international version) forbids feedback from invigilators about the exam questions. Students were thus left to their own devices to cope with the question and somehow produce a coherent answer. The best solution was probably found by those students who mentally conceived an imaginary painting and described it, thereby paradoxically turning the question into a creative writing task, of the kind which the co-certification had been developed to replace.

With the spirit of collaboration strained by incidents such as these, the University sought a crisis meeting with the examining board to clarify and if necessary re-define the roles, if the co-certification was to survive (reported in Newbold 2012b). The importance attached to the meeting by Trinity was indicated by the presence of their CEO, as well as all the team who had been working with the University on the project. This, it was felt at the University, was a sensitive response to local needs in a project which may have offered Trinity College some research insights into testing in an academic context, but certainly no great financial rewards. It ended with a reassertion of the previously agreed roles, and a number of resolutions:

- The University team would be more attentive to house style in item writing;
- Trinity College would arrange an item writing training day with one of their senior item writers;
- The University would double the number of item writers (from two to four) and provide twenty portfolio titles per level each year from which Trinity would make a selection of five;
- Trinity College would show the University their chosen question for the controlled written exam, and the university could change or modify it if necessary.

4.6 Working for Washback

Swain (1985) concludes her well-known list of guiding principles for a good communicative test with the advice “work for washback”. In the first chapter of this volume we identified washback as a ‘local’ manifestation of the wider phenomenon of impact, and suggested that examining boards, by nature of their international role, are primarily interested in impact, and the connection between their certifications and language policies worldwide, and how certifications impact on life beyond the classroom, prompting a number

of commissioned impact studies, such as Wall (2008) on changes in teaching practices across central and eastern Europe in the light of structural changes to TOEFL, and Gribble et al. (2016) on the way in which IELTS interfaces with language skills required in the workplace in Australia.

Washback, in the definition of Alderson and Wall (1993), is confined to "the way that tests are perceived to influence classroom practices, and syllabus and curriculum planning" (117). In Venice, as explained above, the co-certification was introduced at a time of syllabus change in the light of a major reform to the Italian university system, and specifically the degree structure. The co-certification, with its focus on the productive skills, mapped well onto the new Framework-inspired syllabus, so that, although the in-house exam structure was quite different from that of the co-certification, students preparing for the co-certification would be refining those language skills prescribed for the syllabus. Over time, the co-certification, especially because of its 'whole text' approach to writing, began to shape teaching and in-house tests; the most recent (2016) revision of the in-house test of English at B2 level has seen the abandonment of an objective part (sentence correction and multiple choice testing of listening and reading) for a test of listening, reading and writing through paraphrase, summary, and a free-standing critical writing activity.

Although most students majoring in English in the language department (for whom the co-certification was originally devised) continued to do the in-house test, which had no cost for them, around thirty percent chose the comparatively expensive option of the co-certification. From the beginning they were asked why; 21 out of 39 candidates replied that "it is an opportunity to get an internationally recognized certificate in English". Only six said they chose it primarily as an alternative to the in-house exam. In fact, the pass rate (nearly 90% the first year, 81% over the first twelve years) was consistently higher than the in-house exam, which, rather than indicating that the co-certification was easier, may reflect the motivational levels of students. The exam was seen as an investment, and students prepared for it accordingly; whereas the in-house exam, which has no cost, and can be done up to three times during the same academic year, is often taken by students just to 'get an idea' of the level.

In addition, the exam itself was seen as a motivating experience. Ahead of its time, the co-certification introduced short presentations on subjects chosen by candidates as part of the oral exam. For the co-certification sometimes, although not always, candidates would choose topics which were related in some way to their university experience. The co-certification also gave the opportunity to talk about their chosen subject to an unknown but benevolent native speaker. Today, this performance-based approach to learning and assessment has become commonplace in Italian universities, especially in second level courses where class numbers are smaller, and is usually appreciated by students.

4.7 Reform

In 2014 Trinity embarked on a major overhaul of their ISE exams suite. This may have been prompted in part by concerns of reliability and fairness connected to the take-home portfolio (see chapter 1); but the changes also reflected a desire to integrate reading and writing more closely in the controlled written exam, and to separate interactive listening from independent listening, by introducing a free-standing listening task, in the oral exam. Thus, in the written exam, a "long reading" for comprehension was to be followed by four short texts on a related topic, but representing different genres (one of which contained graphic material), the contents of which were to be synthesized in a summary. The final, free-standing writing task was to remain. In the oral exam, listening would continue to be assessed as part of the "spoken interaction" construct, but the exam was to conclude with candidates listening to a recorded passage (such as an extract from a lecture) which the live examiner activated, and then asked the candidate questions about.

These features brought the ISE suite more into line with overtly academic certifications such as TOEFL and IELTS, which include similar reading-to-writing tasks and pre-recorded listenings. The free standing writing task, no longer "creative writing", but renamed "extended writing", covers a range of possible output genres, not just essays, but also reports, reviews, and correspondence. However, all three sample papers posted on the Trinity College website[7] require an argumentative type of essay for the new ISE 3, suggesting that this is the default writing task at this level. Similarly, the free standing or independent listening task, we are told in the specifications for the new exam, will contain content "generally of a discursive nature", such as might be found in "lectures, complex discussions, debates, podcasts, radio programmes and documentaries".[8]

These changes are reflected in the declared objectives of the revised certification, the first of which is to certify that candidates are suitably qualified for "entrance to university where a specified level of English is required for study", followed by "progression to a higher level of English study" and "preparation for further or higher education, where English-medium teaching or Content and Language Integrated Learning (CLIL) methodology may be in use". This latter objective reflects the huge rise in English medium instruction in universities in Europe and elsewhere over the last decade, and we shall return to it in the next chapter.

7 http://www.trinitycollege.com/site/?id=3196 (2017-03-27).

8 Taken from ISE specifications document ("Speaking and Listening"), 47. URL http://www.trinitycollege.com/site/?id=3196 (2017-03-27).

The new Integrated Skills exams, at both B2 and C1 levels, were thus more academic, and more university-oriented, and in this sense closer to the kind of content required for the co-certification. However, if the co-certification was to survive as a local version of the new exam, it would have to adopt the same structure, and a meeting with Trinity management and the Head of Research was held to redefine what this would mean. The main contribution to the co-certification, as we have seen, had been the portfolio questions for "critical writing", and the portfolio had now disappeared.[9] Was there anything left in the new version of the exam which could be usefully changed for a co-certified version? Was there any reason not to accept the new international version as suitable for the needs of the University of Venice? After all, the co-certification was a niche product, costly and demanding to organize. Perhaps it was time to bring the project to an end?

However, ten years after the appearance of the first co-certification, a lot had happened in the way in which English was being used in the universities. The original co-certification was intended to cater for specialists in English; but over the years it had attracted an increasing number of candidates from other departments, such as economics, science, and oriental languages. The needs and profiles of potential candidates had changed, too. The co-certification was being used to access courses, as well as to exit them, and a growing number of courses delivered through the medium of English meant that it was not only students majoring in English who might be interested in focused, high level certification. In the space of a decade, language certification had come of age in the universities, and had permeated across disciplines and courses.

In the end an agreement was reached that the University would take time out, to consider if it would be possible to adapt the new exam in any significant and useful way, or whether there would be any point in slightly tweaking an exam which already had an academic slant. The last exam of the original co-certification was thus administered in 2014, and there was no exam in 2015. A new co-certification was eventually implemented in 2016 and is described in chapter 6. First, however, we shall look closely at the changed circumstances in the use of English in European universities which made an update to the co-certification not only possible, but desirable; and which could, in the long run, have implications for all international English language certification.

9 However, Trinity College proposes a "Portfolio toolkit for teachers", downloadable from the website, as part of a process-oriented approach to preparing students for the written exam.

5 The Spread of English as an Academic Lingua Franca in Europe

Keywords This is a key chapter in the book since it analyses the rapid growth of English as a lingua franca in Europe, and the slow but necessary acknowledgement of its significance by European institutions. After an overview of the kinds of lingua franca interaction which have become an everyday reality in Europe, we focus on universities, and student and teacher mobility, noting that even those students who do not themselves go on mobility may need to interact with international students, or attend lectures given (in English) by non native speakers. In addition, new certification needs have been further driven by the growing phenomenon of English medium instruction, and the recent appearance of first level courses, as well as master's degrees, delivered entirely through the medium of English. These courses are likely to set minimum levels of competence in English for applicants: but which English? We argue for a new rationale for assessing English as a lingua franca to access higher education in Europe.

5.1 What Is English Lingua Franca?

The term "English as a lingua franca" (ELF) began to gain currency at the start of the new millennium, acknowledging the undeniable fact that English had become the preferred language of international communication. Of course, the spread of English around the world was a much earlier phenomenon, inextricably linked to British colonial expansion and American economic clout, and the subsequent adoption of English as a working or official language in institutions such as the United Nations organization and the European Union. But it is probably true that the process received its greatest impetus, in Europe at least, during the 1980s. The end of the "short century", to use the term defined by Hobsbawm (1995), saw the levelling of the Berlin wall separating East and West, the beginning of the age of the Internet, a spurt in the process of globalization, and an intensification of English language teaching, including (in Italy and elsewhere) the introduction of primary foreign language teaching.

Curiously, this period had provided the background for a number of prophetic dystopian visions of the future which focused on changes to the English language. The most well known is probably Orwell's *1984* with its controlled language, called "newspeak", which has been made the official language of the western superpower known as Oceania. Anthony Burgess

also imagined the 1980s[1] as the setting for his 1962 novel *A Clockwork Orange* and its invented, Russian-based, slang called "Nadsat". But the most comprehensive futuristic vision of English as a lingua franca is to be found in H.G. Wells *The Shape of Things to Come* (1933) in which the First Basra Conference, which determines the emergence of the Modern State, also paves the way for Basic English (a simplified version of English proposed by C.K. Ogden in 1925 as a teaching tool) to become the world's lingua franca:

> [English] had many natural advantages over its chief competitors, Spanish, French, Russian, German and Italian. It was simpler, subtler, more flexible and already more widely spoken, but it was certainly the use of Basic English which gave it its final victory over these rivals. (431)

All of these visions provide insights into language use and language change, but none of them capture the defining feature of the lingua franca, which is its variability. Although, as we shall see, it may be possible to identify formal features (phonological, grammatical or lexical) which regularly recur in ELF interaction, and which are not part of any standard description of English, it is the users who shape the content and co-construct the language. Seidlhofer (2011, 7) defines ELF as:

> Any use of English among speakers of different first languages for whom English is the communicative medium of choice, and often the only option.

This is a broad definition, since it can include a native speaker as one of the participants in the interaction. Most researchers, however, focus on what happens in interactions in which both, or all, participants are non-native speakers, since the dynamics change considerably when a native speaker, applying native speaker norms, is involved. In this book, we take the term *ELF* to refer to the use of English by non-native speakers.

What makes ELF any different from EFL, a term which has been in use for decades, and which refers to "English as a foreign language"? For MacKenzie (2015) it is

> an outlook or an *attitude*: while EFL learners make *mistakes* (or errors), ELF users are said to show a lot of *variety*: instead of restricting themselves to the realizations of native English speakers, they exploit unused latent possibilities of English morphology, syntax and phraseology.

1 Reported in *The Guardian*, April 13 2015. URL `https://www.theguardian.com/books/2015/apr/13/100-best-novels-clockwork-orange-anthony-burgess` (2017-10-27).

This insight is illuminating, but it does not reveal the whole picture. The EFL learner and the ELF user may indeed be one and the same person, but the EFL learner can be simply described in terms of language acquired or learnt (for example, through a Framework-related proficiency test), whereas the ELF user needs to be described in terms of a range of pragmatic, multilingual, extra-linguistic, and cross-cultural competences which include accommodation, negotiating strategies, code-switching, and cultural referencing (Archibald, Cogo and Jenkins 2011). If the learner is traditionally seen (e.g. Selinker 1972) on a chronological axis, situated at a specific point somewhere on an impossible journey towards a destination – native speaker competence – which can never be reached, the ELF user is best seen synchronically, moving freely across a plane, reinventing the communicative act in every interaction, with the help of his or her interlocutor.

5.2 Research into ELF

Language use in ELF is meaning-focused. It is less likely to be form-focused than learner English, and less likely to be used as a badge of cultural identity than it might be by a native speaker. It needs to be transparent, not only because of the limited resources (in English) that the speaker might be able to deploy, but also because of the potential limited resources of the interlocutor. Thus economy of language, simplification, and overgeneralization of rules are common. A first wave of ELF research focused on formal features, such as lexicogrammar (Seidlhofer 2001, 2003) and phonology (Jenkins 2000). Seidlhofer (2001, 149) notes commonly recurring non standard features, such as the absence of the third person *s*, wrong prepositions in V + PREP combinations (*spend money to* something), or *who* used as a default relative pronoun. Jenkins proposes a "core phonology" which includes those phonemes and other speech phenomena (such as nuclear stress, but not word stress) which she believes are essential for comprehension, and relegates others (such as interdental fricatives and stress timing) which she thinks are not; an opinion which seems to be supported by the fact that there are native speaker varieties of English which also lack these features.

But ELF is not a single describable variety of English. If it were, it might be taking its place alongside emerging new varieties in a World Englishes paradigm, in what Schneider (2007) refers to as the "endonormative phase" in his dynamic model of postcolonial Englishes, and groping its way towards a set of fixed norms. This of course is not the case – ELF is a variable, not a variety, and its norms, if it has any, are fluid. Thus the focus in ELF research began to move away from language features to user strategies, from product to process. Early work on speech accommodation theory by Giles (1973) and Coupland and Giles (1988) provided the ELF movement with a powerful tool

for describing interaction between speakers with very different language competences. The research effort grew with studies of signalling strategies (Cogo 2010), repetition (Cogo 2009), paraphrasing (Kaur 2009), idiom creation (Pitzl 2009) intonation (Pickering and Litzenberg 2011), using shared linguistic resources such as cognates (Hulmbauer 2011) and other ploys to co-construct meaning in ELF.

Basso (2012), who spent a year in an international humanities faculty in Venice where students came from a range of language backgrounds, and where the official language was English throughout the campus, confirms the use of strategies such as these in the co-construction of meaning, but also indicates how a growing awareness of the role and nature of ELF on the part of students, and their own self-awareness as users of ELF, but also as speakers of other languages, can contribute to successful communication. The role of ELF users as multilinguals offers new research perspectives for ELF, and one which Jenkins (2015) has developed in her attempt to 'reposition' ELF research, widening the context to include multilingual phenomena such as translanguaging. She suggests that ELF research is now entering, or should be entering, a third phase, "ELF 3", after having had "an uneasy sense that ELF research was becoming too self-contained, too repetitive, and was lacking the cutting edge it had previously had" (Jenkins 2015, 62). She also appears to be addressing criticism, such as O'Regan (2014) who sees ELF research as reifying ELF as a stable form (a criticism which has long been levelled at ELF researchers, but which is difficult to sustain), but also, as Jenkins seems to be admitting, in need of continuous and vigorous re-theoriziation.

5.3 The Reality of ELF in Europe

More than a decade of ELF research has gone hand-in-hand with the relentless advance of English as a lingua franca in Europe. The way had been paved, we suggested, in the 1980s. The expansion of the European Union to include former Eastern Bloc countries accelerated the growth of English as the main working language of the Union, in place of French, while the intensification of English language teaching in schools contributed to the steady growth of young Europeans able to communicate efficiently in English. In the 2012 Eurobarometer survey *Europeans and their Languages* it is the youngest age bracket interviewed (15 to 24 year olds) which contains the highest percentage (27%) of respondents who answer "very good" to the question "Is your English very good, good or basic?".[2]

2 *Europeans and their Languages*, 27. Report published by the European Commission (June 2012). URL http://ec.europa.eu/public_opinion/archives/ebs/ebs_386_en.pdf (2017-04-10).

There are economic implications associated with this development, too. In 2013, as the UK was preparing for the Brexit referendum, the German President Joachim Gauck called for the British to stay in Europe – and for English to become the official language of the Union.[3] This was not just an emotional plea, but a reminder of the high costs of translation which the Union has to bear – around €350m per year[4] – for an army of more than two thousand translators shuttling between the 24 national languages all of which have equal status, and are thus entitled to translations of all documents. A single official language would drastically cut costs in one fell swoop. Ironically, with the UK voting to leave the EU in the 2016 referendum, and Ireland having promoted Irish (Gaelic) as the official working language of the Irish republic, there were subsequent calls from members of the European Parliament to have English removed as an official language of the EU, although this possible future scenario was excluded by the European Commission.[5]

The English used in European Institutions has been the object of a number of studies. In her full-length report *Euro English: Assessing Variety Status* Mollin (2006) weighs up the evidence for and against the emergence of "Euro-English" as a separate variety of the language (which, on the basis of ELF research quoted above, would take it beyond the sphere of 'pure' ELF interaction). Using a corpus of spoken and informal written English entirely taken from the proceedings of European institutions, Mollin finds plenty of examples of deviant forms (deviant, that is, from native speaker norms), such as omission or wrong use of articles, ubiquitous tag *isn't it?*, and wrong prepositions, noting, on the way, only a very small number of instances of missing third person marker. Investigating possible emerging new forms and functions, such as the use of *already* as a focus particle due to mother tongue influence, she finds inconclusive evidence for attributing variety status to Euro-English. In contrast, an English native speaker translator at the European Parliament, Jeremy Gardner, lists 89 words which have assumed a different meaning in European use, such as *actor* (= someone who does something) or *control* (= check). In the preface to the 2016 edition of his *Misused English Words and Expressions in EU Publications* Gardner also speculates on the possible influence of Euro-English on Standard English, citing *working group* as probably gaining currency in the UK, to the detriment of *working party*.

3 Reported in *The Guardian*, February 22 2013. URL `http://www.theguardian.com/world/2013/feb/22/german-president-pleads-britain-stay-eu?INTCMP=SRCH` (2017-04-10).

4 According to *The Guardian*, April 24 2013. URL `https://www.theguardian.com/world/2013/apr/24/europa-english-official-language-eu?INTCMP=SRCH` (2017-04-10).

5 Reported in *The Irish Times*, June 28 2016. URL `https://www.irishtimes.com/news/world/europe/european-commission-rejects-claims-english-will-not-be-eu-language-1.2702734` (2017-06-30).

Mostly, however, ELF in Europe is the default contact language used by Europeans when travelling, as tourists, for business, or on mobility programmes in an educational context. It serves a short term need; anyone transferring permanently to another country in Europe would do well to learn the language spoken there. Today, it is the norm for tourists in any country in Europe (and possibly elsewhere) to address a shopkeeper, or just a passerby for directions, in English, without any preamble such as "Do you speak English?"; and they are likely to be understood. Similarly, any self-respecting tourist destination is likely to offer notices, advertisements, signs and warnings, in English as well as the local language. Graffiti in ELF is part of the urban scenery of Europe, too, the language of pop art, protest, and messages of unrequited love, offered to the world in the lingua franca, often with non-native spellings or inflections, which may make them more memorable. "Regina still miss iù" reads a long-surviving scrawl on an overpass at the beginning of the causeway from the Italian mainland to Venice; "We don't going back" proclaimed a banner wielded by economic migrants from Sub-Saharan Africa who set up camp on a beach in the south of France in 2016.

But the use of ELF in Europe runs deeper than superficial contact and slogans; the twin motors behind ELF in Europe are education policies and the Internet. A great deal of ELF communication takes place off the streets, online, in blogs, chatrooms and using social media where the distinction between oral and written codes is blurred, as is frequently the identity, and also the native language, of the interlocutor. This phenomenon, with its implicit user/learner paradigm, is discussed in detail in Vettorel (2014). For Mauranen (2012, 33), it is the Internet which is co-responsible for the "explosive expansion" of English in the mid nineties, and the contamination between native and non-native users is likely to shape the English of the future: if today's users are the first generation, "by the time the third generation learns English, we may expect English already to show clear traces of lingua franca influence".

5.4 ELF in Schools

Implicit in early criticism of the ELF movement was the suspicion that there was a pedagogical agenda dictating the research, and that a road map was being laid out for language planners, syllabus designers, and publishers. After all, since one of the aims of any education system must be for its pupils to be able to communicate successfully with the outside world, setting the world's lingua franca firmly in place as a teaching objective would seem to make sense. But as we have seen, ELF is not a describable variety which can be taught, or learnt, and ELF research would consequently be better seen by language planners as providing insights into the nature of non-na-

tive speaker interaction in English, and informing decisions about language teaching, rather than outlining a syllabus of simplified English. The priority for teachers, and students, should be to become "ELF aware", rather than to "teach" or "learn" ELF (Sifakis 2014, Sifakis and Tsantila forthcoming, Sifakis 2017). This may mean teachers taking a different approach in the classroom than they would when teaching other foreign languages.

The special role of English in educational systems in Europe is evident in the sheer number of pupils learning the language. Many EU countries now have nearly 100% of pupils in the primary sector learning English, and by 2014 there was an average of 94% of *all* secondary school pupils in the EU studying English.[6] The nearest rivals (French, 23%, Spanish and German both with 19%) are clearly in a different league. The significance is, or should be, clear: English is learnt 'for a different reason' from other languages. The European Commission, however, has been slow to recognize this. From its beginnings, the EU has promoted the learning of member state languages, to enable citizens to move, work and study freely in the Union. This policy had settled down, by the turn of the century, to a "mother tongue plus two" mantra; in 2002, the Barcelona meeting of the EU Council recommended that children should start learning two foreign languages from an early age. As recently as 2008 a five page document entitled *Council Resolution on a European Strategy for Multilingualism*[7] sets out the rationale for schools to promote multi-lingualism in schools, to include subject learning in secondary schools through the use of a foreign language, known as CLIL (Content and Language Integrated Learning), but makes no mention of the role of English as a lingua franca.

Behind the EU's policy of fostering multilingualism lay the twin objectives of safeguarding minority languages and promoting multiculturalism. As the Union grew, so too did the challenges posed by these objectives, as the number of languages brought in by new member states also grew. But at the same time, Europe was having to come to grips with an unprecedented and vast influx of asylum seekers and economic migrants. Over the last two decades Italy has found itself in the front line of a wave of immigrants arriving from Asia, the middle East, and Africa, bringing with them a wide range of lingua-cultural systems. At the time of writing this wave shows no sign of abating. It has led to a further dimension for ELF in Europe, as a contact language between Europeans and non-Europeans, the traumatic and unequal encounters of which have been described in detail by Guido (2008, 2012).

6 http://ec.europa.eu/eurostat/statistics-explained/index.php/Foreign_language_learning_statistics (2017-11-02).

7 http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32008G1216(01)&from=EN (2017-11-02).

Over time, the children of non-European immigrants have swelled the ranks of Italian classrooms, reaching the current figure of 10%[8] of all enrolled pupils, with the largest number (around 35% of the total of 815,000 foreign pupils) in the primary sector. In areas of high density immigration, however, the percentage of immigrant pupils in primary classes is much higher, frequently exceeding the 30% limit imposed by the ministry in 2010.[9] This has of course posed challenges to teachers, since children need to be integrated into society, and this can only be done through the acquisition of Italian as a second language. However, many would-be immigrants come from Anglophone areas, such as West Africa, and the Indian subcontinent, where new Englishes have flourished since independence, and for whom English is the language of choice in international interaction. Older children, who have attended school in their country of origin, may well have been used to English as a medium of instruction, or exposed frequently to a variety, or varieties, of 'postcolonial' English.

This means that in many multicultural classrooms English, or rather, a potential context for ELF, is a latent linguistic resource waiting to be put to good use by teachers working towards cross-cultural understanding, a platform for integration for immigrant children, and the opportunity for real foreign language use for Italian children almost all of whom, as we noted above, are learning English. In each case, English has to be reinvented by participants on the basis of the limitations, and potential, of their own, and their interlocutors', competences.

This scenario has been described by Lopriore (2015) reporting a longitudinal study of primary school language learning in Europe.[10] She notes that the phenomenon (of migrant children in primary language classes) "has partly contributed to affect the type of second language learning and acquisition processes young learners undergo since migrant children, when at school, are exposed to and use more than one language to learn and we may hypothesize that elements of ELF begin to emerge" (161). The fact that teachers are themselves non-native users of English, she suggests, adds a further resource to the effort of co-constructing meaning in a polylingual class of young learners.

At secondary school level many young Europeans, as we saw in the Eurobarometer survey quoted above, consider themselves to be competent users of English. In Italy, a level of B1 on the CEFR (the first level of the "independent user" bracket, originally known as "threshold level") is the target set for age 16, at the end of the two year *biennio* cycle at the beginning of the secondary school, and by this age many pupils will be using

8 http://www.istruzione.it/allegati/2016/Rapporto-Miur-Ismu-2014_15.pdf (2017-04-05).

9 Ministry of Education (MIUR) circular no. 2, 2017-01-08.

10 ELLiE, (Early Language Learning in Europe).

the language outside school, on the Internet, on holiday, or on school trips and exchanges. They are exposed to more English, and more non-native English, than any previous generation, and this fact may, paradoxically, be seen as a problem by some teachers rooted in native speaker standards. There is no shortage of research showing that teachers feel the need to base their teaching on a standard model, even if they are well aware of the existence and importance of ELF (Groom 2012, Vettorel 2015, Soruc 2015), but young learners know from experience that native speaker standards (or rather, the standard grammar of EFL textbooks) are not needed for successful international communication.

Potentially, this might seem like a recipe for confrontation in the classroom, but it can also be harnessed by teachers as a source of reflection on language variety, the nature of a lingua franca, and the relationship between accuracy and communicative efficiency. One of the more interesting challenges facing English language teachers in secondary schools is to open their classrooms to the reality of ELF, for example through twinnings or projects shared with international partners, which may involve real time communication through the Internet at school, but which students could continue in their own time and space. Grazzi (2015) describes one such project from the perspective of developing intercultural communicative competence.

5.5 ELF in Higher Education

It is however in higher education, in Italy, as elsewhere in Europe, that the speed of change in the spread of ELF has been most apparent. Unlike the evolution of ELF in schools, it has been a top-down process, willfully imposed by the governing bodies of universities as part of a strategy of internationalization. The impetus for internationalization is largely due to the 1999 Bologna process, which created the premises for the recognition of degrees across Europe through the credit transfer system (ECTS) and the adoption of a two level degree programme, to include a three year first level degree, followed by a one or two year second level degree (which together replaced the old four year *laurea* in Italy). One of the main objectives was to promote student and teacher mobility, but the Bologna process opened up a number of other important possibilities, such as the recognition of diplomas by prospective employers, and a more competitive stance to attract international students, which until the turn of the century had been largely the prerogative of the US and the UK.

In the new 'internationalised' European university English has two main functions: as a medium of instruction (EMI) and as a lingua franca in everyday campus life in international encounters. English is the language of choice as the teaching language (or 'medium of instruction') in most in-

ternational universities, or universities which have pretensions to being 'international', and there is a long tradition of English medium universities in non-native English countries across the world. In Europe, Holland was the trailblazer, and Italy a comparative late starter. But over the last ten years, the number of foreign students enrolling in Italian universities has more than doubled, and although the national percentage (around 4%)[11] is still low compared to Germany or France,[12] it continues to grow rapidly, while some of the more prestigious universities in the north of the country, such as the Polytechnics of Milan and Turin (20%) and the Bocconi University in Milan (15%), have much higher percentages. Many of these international students will have chosen the university because they offer courses in English in their chosen disciplines. There are currently 276 degree courses offered in English in 54 Italian universities,[13] over 90% of which are at master's level.

The international dimension of a university is completed by students, academics, and administrative staff on mobility. Every year around 20,000 European students choose Italy as their destination on the Erasmus programme, and many of them will expect to follow courses in English. They will also expect to communicate with their peers, and their teachers, in English. Thus local students, even those who do not attend EMI courses, or go on mobility themselves, may frequently find themselves having to interact with foreign students, or to listen to a visiting academic, in English.

In 2010 a needs analysis carried out among final year students from all four faculties at the university of Ca' Foscari Venice[14] (Newbold 2012a) found that most students had needed English for research purposes, online or in books and articles; while sizeable minorities had also had to write e-mails, attend seminars or one-off lectures, or interact with foreign students as part of their everyday lives at the university (see tab. 1 below). Nearly a decade later, these figures are likely to be much higher. At the time, they gave support to the decision which had recently been taken by the University to require an entrance level of B1 English for all incoming students, irrespective of the course they were enrolling for; a requirement made by most Italian universities at the same time.

Table 4. Needs analysis of final year undergraduate students at the University of Venice: percentage of students per activity.

11 http://www.rivistauniversitas.it/Articoli.aspx?IDC=2986 (2017-04-25).

12 But most foreign students will be following courses delivered in French.

13 http://www.universitaly.it/index.php/cercacorsi/universita?lingua_corso=en (2017-04-25).

14 Economics, Languages, Humanities, and Science.

What have you needed English for during your time as a student?	
Reading textbooks & articles.	70%
Using internet for research.	53%
Watching film and video.	23%
Participation seminars in English.	21%
Writing emails.	19%
Interaction with foreign students.	18%
Interaction with foreign lecturers.	9%
Writing letters.	2%

More recently, the University has ratcheted up the entrance requirements, so that all incoming students for second level degree courses (*lauree magistrali*) now have to produce evidence of a B2 level in English. At the same time, it has widened the net of courses delivered in English to include first degree courses, with a 3 year BA (*laurea triennale*) in Philosophy, International Studies and Economics being introduced in 2015.

Another indicator of the process of internationalization is to be found in the care taken in the development of English language versions of university websites. All Italian universities with an international vocation offer links to an English version on their home page, the main function of which seems to be the marketing of courses to prospective new students, as well as to respond to a need to provide information about the university to students on mobility. In recent years they have become more sophisticated, to include well-made video testimonials of international students. Jenkins (2014), in her study of *English as a Lingua Franca in International University*, devotes a chapter to university websites, and notes a preference for native speaker norms, as well as, in Europe, remnants of a diffidence and opposition to EMI entrenched in an orthodox European philosophy of multilingualism. She quotes, in particular, the episode of the Polytechnic of Milan which had just announced (in spring 2014) that all of its courses would in future be held in English, causing an uproar among academic staff (many of whom felt they would not be have the competence to deliver their courses in English) and subsequent legal action. This led to a court ruling that the move was unconstitutional, followed by a rectification from the Constitutional Court that universities could exercise autonomy in their choice of language, so long as the national language (Italian) was "not completely sacrificed".[15]

The Milan uprising, however, was atypical, a consequence of the sudden and drastic changes which the University management was attempting to

15 Reported in *La Repubblica*, February 24 2017. URL `http://milano.repubblica.it/cronaca/2017/02/24/news/corsi_in_inglese_al_politecnico_via_libera_della_consulta_purche_non_sacrifichi_totalmente_l_italiano_-159117927/` (2017-04-21).

implement. Three years on, the website shows that the majority of undergraduate courses continue to be delivered in Italian, while most second level courses are in English. At the same time, decisions taken elsewhere in Europe seem to signal that the rise of ELF in academic is unstoppable. In 2013 an EU commission found that knowledge of English was a necessary pre-requisite in European higher education. With this recommendation, the EU was breaking new ground by naming the unnameable, by recognizing that English is a pre-requisite for intra-institutional communication in Europe:

> Higher education institutions should develop and implement holistic internationalisation strategies as an integral part of their overall mission and functions. Increased mobility of students and staff, international dimension of curricula, international experience of faculty, *with a sufficient command of English* and a second foreign language and intercultural competences, transnational delivery of courses and degrees, and international alliances should become indispensable components of higher education in Europe and beyond.[16]

The appeal for plurilingualism is still there, but it is in second place, as "a second foreign language". In the same year, France introduced legislation to make it possible to use a language other than French as the medium of instruction in state universities, opening the floodgates to EMI in a country which has a history of legislating against the use of English in public life.[17]

5.6 ELFA: English as a Lingua Franca in Academic Settings

English as an academic lingua franca (or ELFA, as it has come to be known) functions, as it does in a non-academic context, as a complex second-order language contact between similects (Mauranen 2015, 38). But since it also involves contact between members of a community of practice – albeit a very large one – it seems reasonable to assume that it might exhibit, or develop, traits which are characteristic of that community. This was the research question behind the establishment in 2008 of the 1m word ELFA corpus at the University of Helsinki, which has recently been flanked by the 1.5m word WeELFA corpus of written academic English at the same

16 Recommendation 12 made by the High Level Group on the Modernisation of Higher Education. URL `http://europa.eu/rapid/press-release_IP-13-554_en.htm?locale=en` (2012-04-21) (italics added).

17 The 1994 *loi Toubon* outlawed the use of English in public documents, advertising and university lectures, and introduced quotas for the number of foreign language songs which could be broadcast by public media.

university.[18] The spoken corpus contains both monologic samples (extracts from lectures given by NNS) and dialogic samples, such as seminars and conference discussions, across a range of disciplines. The written corpus includes research blogs, unedited research papers, and examiner reports. It does not include published materials, even if at least 80 per cent of the world's peer-reviewed academic articles are now published in English.[19] The reason for this is that many published articles are edited by mother tongue proof readers. Some publishers make it a contractual requirement of their authors to have their work edited in this way, with the result that many publications are hybrid co-constructions which might display native-like features (such as sentence-level formal accuracy) but also non-native features in their higher level rhetorical organization. This is especially true for publications in the humanities, but is probably less so for scientific publications which adhere to a more rigid framework.

Unsurprisingly, the corpora show features of lexical simplification, lexical creation, redundancy reduction and creation, and the regularization of irregular verbs, all of which are attested in other corpora. At the same time, Mauranen (2015) shows a fairly close match between the most common three word phrases in the ELFA corpus and a corresponding native corpus of academic English (MICASE[20]). Given the comparatively formal nature of academic discourse, a notable feature of the ELFA corpus turns out to be high productivity in morphological manipulation, yielding examples like *intrevent*, *introducted*, *addictation*, *devaluaized.* Uncountable or mass nouns, often conveying abstract notions, which are a stock-in-trade of academic reporting, frequently crop up in plural forms (*advices*, *informations*, *evidences*), often fulfilling a communicative need as they do so – *evidences*, for example, may be used to indicate more than one source of evidence, whereas the monolithic, uncountable form *evidence* cannot do this. A native speaker, constrained by the one-form-only of the mass noun, would need to think hard to convey the idea of "more than one incidence of evidence". The corpus also throws up examples of what can be seen as the quintessential ELFA verb, *to discuss about*. This non-standard form, derived from analogy with *talk/speak about*, has slipped into international conferences everywhere, with Mauranen claiming that it is also attested in native speaker English (Mauranen 2015, 40).

Academic ELF also exists in the grey area of the Internet, sharing some of the features of both spoken and written codes, in e-mails, blogs, calls for papers, and abstracts. Depending on a number of factors, such as

18 For an overview of the project see `http://www.helsinki.fi/elfa` (2017-10-27).

19 `https://www.theatlantic.com/science/archive/2015/08/english-universal-language-science-research/400919/` (2017-04-21).

20 Michigan Corpus of Academic Spoken English.

time constraints, but also, perhaps, a growing sense of awareness of ELF and a shared tolerance level within the community, non-standard forms are rife. Here are just three of the more interesting forms the author has come across in personal email communications over the last two weeks:

ex 1

From a conference abstract:

Weeping, mourning, praying, crying out were parts of a behavioural pattern, a visible, **hearable**, and ritualized performance taking place in private houses, streets and public spaces.

ex 2

From a call for papers for an international conference:

This is a **gentleman reminder** about the call for abstracts for the next [...] conference.

ex 3

Message from the same organizing committee as Ex 2:

We have a serious problem with the abstract submission system. The webmaster **tries** to solve this problem.

Whether one considers these as 'errors' or simply examples of ELF depends of course on one's perspective as a reader. Most intended readers of this kind of email communication are unlikely to spend much time lamenting the lack of formal correctness, since rapidity of communication and communicative effectiveness are the writers' aims. From a teaching/learning perspective, *hearable* should have been corrected to *audible*, *gentleman* to *gentle* and the progressive form *is trying* is required in Example 3. But from an ELF perspective, *hearable* is an easily understood example of lexical creativity, while *gentleman* may come from an analogy with the expression *gentleman's agreement*, rather than the NS collocation *gentle reminder*, giving an extra layer of genteelness to the expression. Only with the choice of aspect for the verb (simple instead of progressive) does there appear to be a net loss rather than gain: the present simple has kicked in as a default all-purpose present tense. However, there is no loss of intelligibility. Moreover, the language context in which all three non standard forms find themselves suggests that the writers are competent users of English.

The dictates of real time communication may lead to unchecked errors

for native speakers, too, such as when the phonological dimension of a word or phrase interferes with the written form, further blurring the distinction between NS and NNS, as in the following examples:

ex 4

Secondly I will describe the Australian English (AusE) from a general **point of you**.

ex 5

We plan to get the work done in September (I need to check if this is possible) with a **few** to an event before [……] leaves.

Curiously, in both examples, it is the same word, *view,* which has failed to materialize. The curiosity is compounded by the fact that the first has been written by a NNS student, the second by a NS member of faculty.

We have already mentioned the attention given to English by European university websites. When we take a closer look at the University of Venice website we find that this extends to punning and other kinds of wordplay, suggesting, perhaps, that the academic community is in control of English as a creative resource for international communication. Before accessing the English language version, on the main Italian pages, we find the titles of lectures and seminars in English and projects and messages to students wilfully code-switching and punning:

ex 6

A production at the university theatre:

Friendly Feuer - una polifonia europea.

ex 7

An introduction to archeology:

Welcome to the Dig! Strategie e nuove professioni per un'archeologia pubblica

ex 8

A presentation of second level degree courses:

Lauree magistrali: **postgraduate opportunities**

ex 9

An invitation to participate in a sponsored run:

Y.our Future Run

These examples, downloaded on the same day,[21] give an idea of the extent to which the Italian user of the website, within a context of Italian, before engaging with ELF and the international community, is exposed to English. Whereas examples 6-8 are best seen as instances of translanguaging – the conscious exploitation of shared language resources (Garcia and Wei 2014) – example 9 is more complex. Presumably the intention is to indicate that 'your' future (referring to potential sponsors) is also 'our' future (referring to the university). It is hard to imagine this kind of word manipulation to be successful in a native speaker context, perhaps because of the phonological clash between 'your' and 'our', but it is increasingly common in an international environment.

The final dimension of academic ELF is of course, published research. The 'finished product' nature of academic publications, and the need for clarity and lack of ambiguity, means that publications continue to be the object of standards-based scrutiny, and that articles in most important international journals display little variation from native speaker texts. Indeed, attention to the structure of academic texts in English has had a huge boost from the work of, among others, Swales (1990), Swales and Feak (2004), and Hyland (2003, 2015), and has led to the development of courses in writing "English for academic purposes", on line and in timetabled classes, in universities everywhere, and a healthy branch in the ELT publishing market. For the moment, then, native speaker English provides a norm, and only a few publications[22] specify that they do *not* require a native speaker editing process.

This component of "academic ELF" – arguably the most important – is the one which, as we have seen, lies furthest from the typical ELF construct of a dynamic and on-going co-construction of meaning, and has attracted least attention from ELF researchers. However, the sheer volume of published research, not to mention the increased pressure on any self-respecting journal to require two peer reviewers of each article, means that a 'native speaker control' or 'near native speaker' control of all material now being published is simply not possible, and it is legitimate to suppose that here, too, over time, native speaker norms may begin to be superseded by the more fluid strategies of transmitting knowledge which are the stock-in-trade of competent second language users.

21 2017-04-25.

22 The *Journal of English as a Lingua Franca* is one of these.

5.7 ELF, EMI and the Role of Certification: Towards a Rationale

In this chapter we have examined the phenomenal rise of English as a lingua franca, and as a medium of instruction, in European and especially Italian universities. We have seen that it has multiple facets which impinge on 'stay at home' students just as they do on students and teachers on mobility. We have also noted that the phenomenon permeates spoken interaction, email communication, and university websites, and may contribute to shape academic publishing in the future.

For the student, therefore, success at university is at least to some measure dependent on a range of strategic competences in English, which will be needed for research, for interacting with other students, for attending lectures given by visiting academics, and even for browsing their own university website. It is hardly surprising, then, that universities, mindful of the disastrous drop-out rates of the past, which in Italy at least were attenuated after the 2000 reform, have sought to offset this danger by setting entrance level requirements, usually B1 (for a first level degree course) and, increasingly, B2, to access second level or PhD courses, since at this level students may well be required to make academic presentations in English.

But what kind of tests are being used to assess these competences, which, besides 'traditional' skills such as reading comprehension include the pragmatic competences needed for oral interaction between NNS as well as a range of digital literacies? Universities usually give incoming students a choice: to provide a recognized certification, such as those we discussed in chapters 2 and 3, or to do an in-house test. The latter tends to be an objective, computer-delivered test which is easy and (comparatively) cheap to administer, although it requires universities to have large numbers of PCs available at the same time for multiple delivery of tests. This kind of test is also reliable, but it is likely to be limited in scope, and may be confined to testing formal accuracy of de-contextualised grammar and vocabulary items. A more complete test will include listening, and a more integrated and appropriate test (perhaps at B2 level) might include the productive skills of speaking and writing – but at a cost, to the university (or to the student).

Most in-house tests are rigorously native-speaker norm based; typically, they require test-takers to choose between 'correct' and 'incorrect' forms, encouraging, or reinforcing, a behaviour which has only limited use in real ELF interaction. There is no guarantee that a high scorer on such a test will be an effective ELF user; vice versa, a low performer may turn out to have good communicative skills when speaking to contemporaries from other countries.

Certification may provide a more streamlined passport to the university, but it has a cost. At present, not more than twenty percent of enrolling

students are likely to have an acceptable certification.[23] In addition, as we saw at the end of chapter 2, in spite of claims made by boards such as IELTS and TOEFL as to the international nature of their certification, this is primarily intended for non-native speakers who intend to study in a native speaker environment, where there will be a premium on formal accuracy, and where native speakers may not be willing, or able, to make too many concessions to foreigners. But in Europe, where English has a different role, where participants find themselves on an equal footing as partners in communication, where English is a vehicle for information exchange rather than a transmitter of culture and cultural values, a qualitatively different approach to assessment seems to be needed. In the final chapter of this book we will attempt to articulate the rationale for a more 'ELF friendly' certification, and look at the form such a certification might take. But first we shall return to the co-certification and its revised (2016) version which, for the first time, we believe, led to the incorporation of an ELF element into a construct for an international exam in English.

23 Based on data from the Centro Linguistico di Ateneo, Ca' Foscari University of Venice.

6 Co-Certification Revisited

Abstract In 2015 Trinity College overhauled its *Integrated Skills in English* suite, to bring it more in line with other academic certification, notably by introducing a reading to writing task based on multiple input texts, different text types, and an independent listening task. This had repercussions on the co-certification (chapter 4); if it were to continue, the University would have to adopt the same structure. The revision was seen as an opportunity to update the co-certification by introducing an "ELF element" – listening to a non-native lecturer – as the independent listening task. In this chapter we report the results after two administrations of the certification, and note that, for most candidates, the "ELF task" seemed realistic and unproblematic.

6.1 Envisaging an ELF Element for the New Co-Certification

One of the new features of the revised Integrated Skills certification was to be a free-standing pre-recorded listening task (chapter 4). In the previous version, listening had been tested only as an interactive component of the oral exam, in conversation, and in collaborative tasks, reflecting the performance-based approach of Trinity College exams. The new format marked a change in direction, aligning the Trinity exam more closely with academic certifications, and their target language domains, by acknowledging the importance of listening to monologue, and related academic skills such as note-taking and summarizing (whether orally or in writing). The rationale for this 'expert listener' construct hypothesized by Trinity drew in part on the work of Field (2012, 2013) into cognitive validity, and it can be seen as complementing the socio-cognitive framework (Khalifa and Weir 2009) which lies behind the new reading to writing part of the certification.

This free standing listening attracted our attention as a part of the new exam which could be easily adapted in an 'ELF-aware' co-certified version, and which could reflect students' needs as ELF users in a European context. As we noted in the previous chapter, a 2010 survey had shown that more than twenty per cent of all students looking back over their experience as full-time students in the period 2007-2010 had been expected to participate in seminars, or to listen to lectures, in English, as part of their course. A decade later this percentage would surely be much higher. But it would also be true that most visiting lecturers giving these seminars or talks, to non-native speakers of English, would themselves be non-native speakers. This fact is not however captured in the specifications of the new international version of the certification, where we read:

Accents
May include varieties that can be processed using southern British and General American as a point of reference[1]

In this, too, the ISE exam follows substantially the same line as IELTS and TOEFL by offering a variety of accents, all of which, however, are native speaker accents. Yet some of these accents are less likely to be encountered on a regular basis by European students than (say) French, or German accents, in a context of English lingua franca.

We thus proposed to adapt the specifications for the listening task at C1 level, while keeping the structure and level of difficulty the same as in the international version. Two of the specifications, *topic* and *accent,* needed to be changed, in the interests of task authenticity, while all the others – *speech rate*, *syntactic complexity*, *processing* and *task outcomes* – could be left intact. Our revised specifications for the free standing listening (co-certification) became:

Topic Information generally of a discursive nature. Could be expositional, summative, or procedural. The context would always be academic, such as an extract from a lecture or a seminar.

Accent Fluent non-native speaker of English.

We also made slight changes to the rest of the exam, (adding "education" and "higher education" to the list of possible topics of conversation in the oral, and continuing to provide the input for the final, free-standing writing task in the reading-to-writing paper).

As far as we were aware, this was the first time that non-native speaker accents were to be exclusively used in a high stakes listening test. It also offered potential research questions, such as:

- Is understanding a non-native speaker more problematic than understanding a native speaker?
- If so, why? If not, why not?

We could imagine that familiarity with a particular accent might make it more accessible to the listener, just as we could imagine that entrenched attitudes towards some accents might make them less accessible. In any case, although we did not expect to get any definitive answers to such questions, we hoped that a judiciously administered post-exam survey could elicit some interesting insights.

1 ISE specifications document, 47. URL `http://www.trinitycollege.com` (2017-01-24).

6.2 Test Development

A series of meetings with the Trinity College research and development team brought the project into clearer focus. Although we would have preferred to use extracts from real lectures, it would have been problematic (and extremely time consuming) to find authentic texts which had the right level of information density for the two-and-a-half minute intensive listening tasks we had in mind. In this respect, the co-certification would be no different from other certifications: we would use specially written texts following guidelines which would be drawn up by the team at Ca' Foscari, and mediated by Trinity College. Texts would be supplied by Ca' Foscari, but recorded in London in the recording studios regularly used by Trinity College, by expert non-native speakers identified by Trinity College or by the recording studios. Here too, we would have preferred to use colleagues from the University, with different mother tongues, whom we knew to be competent users of English; but we accepted that for organisational reasons, and comparability of our version with the international version, the uniform recording conditions offered by the studios were a positive feature.

The agreement, then, was to supply Trinity College with enough texts for two administrations of the certification (2016 and 2017), which would cover the two year renewable contract which had been a feature of the partnership since 2004. Firstly, however, a training session was arranged for the university team of four item writers who would produce the texts, with input from a senior item writer from Trinity. Each member of the team was invited to supply, in advance of the meeting, sample texts at levels B2 and C1. A rationale for writing was drawn up, focusing on how the texts in the co-certified version might differ from the international version, such as in the choice of topics, and how they could be made similar to real extracts from lectures, for example by (limited) use of signposting, redundancy, and hedging, as well as by focusing on a specific mode of delivery, 'procedural', 'expository' or 'summative'.

The meeting produced a consensus of opinion on some points, such as the need to limit the use of non-transparent idiomatic language, and long noun phrases more suited to written tasks, and the possibility that, given the large number of cognates with Italian words in academic texts, the B2 level texts could be more lexically dense than their counterparts in the international version. One useful activity was for each writer to read their own text aloud and note where they stumbled, and why, and to reflect on the nature of hesitations, stumblings, and self-repair in the actual delivery of a lecture. The main discussion focused, perhaps not surprisingly, on bridging the gap between a written text, and the immediacy of live oral performance.

In the end, because of the time involved in preparing and editing the texts, we agreed to limit the ELF input, at least initially, to the higher C1 level certification, and to provide Trinity College with forty texts by the

end of the summer (2015), allowing us time to edit our work, and Trinity time to process and record the texts for the spring 2016 session. Drawing topics from the humanities and the social sciences, we aimed to produce texts which would be accessible to European students, especially Italians, or international students in Europe, especially in Italy. So topics ranged from young peoples' voting habits in Europe, to ancient Greek science, to young writers in Wales (although we avoided texts which focused on traditional aspects of British culture). For each text we provided a sample gist question, which test takers had to answer after a first listening, and then the four or five main points which we expected them to be able to report after a second listening (during which they were allowed to take notes). This sequence, of course, followed the standard procedure for the international version.

Of the forty texts we wrote, ten were rejected by Trinity on the grounds that they were more suited to a B2 level test than C1. This was useful feedback: what these ten texts seemed to have in common was a more conversational style, and more self-reference, than the others, rather than an obviously simpler lexis or structure. Trinity also suggested some style and content changes to the other texts; however, some of the suggestions, especially those concerning content, seemed to be dictated by the 'default' position of the international examining board, and the need to avoid topics which referred (even superficially) to religion or politics. Thus we were invited in one text to change *Christmas* to *Birthday*, in another *church attendance* to the rather meaningless *religious attendance*, and to avoid altogether the topics of migration and the division of Cyprus (branded as "sensitive") which were the subject of two other texts. In actual fact, such topics would be unlikely to cause offence to university students, and indeed, at Ca' Foscari, could be of particular interest to students of International Relations, a heavily subscribed master's level course which regularly provided candidates for the co-certification.

The proposals made by Trinity were reminiscent of the crisis reported in 4.5, and so, as before, we had to remind our partners of the content rationale for the local version, before proceeding to the recording of the thirty mutually agreed texts for the C1 co-certification.

6.3 The Recordings

The recording studios engaged four component non-native speakers of English to read the texts. All of them had been living for some time in the UK, all of them had noticeable non-native accents, but (to the native speaker author of this book) these were in no way difficult to understand. All of them used vowels which approximated to native English vowels, especially in their use of diphthongs, while one of them had acquired a glottal stop reminiscent of Estuary English in words like *about* [ə'baʊʔ] and

but [bʌʔ], and made use of syllabic consonants, e.g. at the end of the word *written*. Nonetheless, they also all retained characteristic features of their mother tongue, such as the velar fricative /x/ (for the Spanish speaker), and nasalization of some vowels (for the French speaker).

The mother tongues were Italian, Spanish, French and Catalan. Ideally, we would have preferred a sample of accents from across Europe, including, for example, a native speaker of German (which has more native speakers than any other European language). Instead, we had only speakers of Romance languages from southern Europe. Furthermore, one of them was Italian: an accent with which, we presumed, most of our test takers would be familiar. However, despite the problem of potential bias (Harding 2012), there was a strong validity argument for including an Italian accent, precisely because this would be the most common non-native accent to which our students would be exposed, in English taught programmes for which most lecturers would be local faculty using English, or in international events held at the University. Two of the readers were men (Italian and Catalan); two were women (Castilian Spanish and French). We prepared a guide for them, which explained the background to the project, and then went on to give instructions about how to read, as follows:

> You have been asked to read the text because you are a competent user of English whose mother tongue is not English. The listening texts which you produce will, we hope, be accessible to students not only because of the content, but also because they are familiar with the accents and speech habits of Europeans using English.
>
> As far as we are aware, this is the first time that a major examining board has used non-native speakers (NNS) in a test of English, and so we are keen to collect as much data as possible about the processes involved in NNS-NNS interaction, especially in the context of a test.
>
> In particular we would like to ask you
>
> a. to read the texts in as natural way as possible, in your 'best' English, without unnaturally exaggerating either your mother tongue accent, or any English accent;
> b. to imagine that you are speaking to an audience of about 100 students, most of whom will be Italian, a few of whom will be from other countries, none of whom will be native speakers of English;
> c. if you wish to make any very small changes to the text (adding words like *so* or *and*) to do so;
> d. if you make any small 'errors' (e.g. of pronunciation or grammar) and self correct, please leave the correction (i.e. don't re-record the text);
> e. if you are aware of any small 'errors' (e.g. of pronunciation or grammar) only at the end of the recording, please leave them (i.e. don't re-record the text).

The instructions were intended to encourage as far as possible a uniform approach to reading, as well as to create the impression of live performance. In actual fact, on listening to the recordings, we found numerous hesitations, self-corrections, and errors, in phonology, word stress, and organization of tone units. This latter was the most common error of all, with all readers making inappropriate pauses, in the middle of nominal groups or between verb and object; an error type which, perhaps more than others, indicated that the speaker was in fact reading (and was probably not very familiar with the text).

Partly because of this incorrect chunking, nuclear stress was sometimes compromised, as in:

ex 1

"One of the group's keys to success" (instead of success).

ex 2

"Some two and a half thousand years ago" (instead of two and a half thousand).

For Jenkins (2000) this is an error of "core phonology" which risks compromising intelligibility. There were also word stress errors (for Jenkins, "non-core", and so potentially unproblematic for the listener), for example in compound nouns, where the stress moved to the second element, as in *love story*, *travel writer.* Most word stress errors occurred with low frequency words (*consequently*, *delicacy*, *infamous*, *refuge*), while others involved selecting the wrong form of words with two pronunciations (*process* and *record*, both nouns, were articulated as if they were verbs).

Phonological errors were infrequent, and included /'kɒmræd/ for *comrade* /hɒl/ for *whole*, and /'ɔːʧɪd/ for *orchard*, and included several mispronunciations of proper nouns, such as the names of places and people, for which some speakers used a default mother tongue pronunciation (for *Pythagoras*, *France*, *Vatican*).

There were also noticeable errors in the interface between phonology and morphology, such as the omission – or addition – of plural "s", as in:

ex 3 "is interesting to university student" (instead of "students")

ex 4 "banks and local governments" (instead of "government")

ex 5 "they are out of sights and also out of mind" (instead of "sight")

The words appear in their correct form in the text being read, of course. Example three we might suppose to be phonologically induced, in which the reader reduces the final consonant cluster in “students”; in examples four and five, however, the additional “s” may have been induced by knowledge of grammar (selecting “government” and “sight” as count nouns, the former perhaps prompted by the plural marker in “banks”, the first part of the co-ordinated noun phrase).

There are a lots of hesitations and false starts, such as:

ex 6 “history of art and ah, ah, architecture”.

ex 7 “despite the presen, despite the presence”.

which occasionally lead to apologies:

ex 8 “the future of art, sorry, the future of art restoration”.

There are also misreadings with self corrections:

ex 9 “and the attempt to evangelize ends here.... ends there”.

ex 10 “which is now being a reality, which is now becoming a reality”.

ex 11 “the most controversial area is what to, is to what extent...”.

Some of these slips have the feel of performance errors which might be made by any speaker (whether native or non-native) in a lecture. But perhaps the most interesting errors were those grammar errors which passed unnoticed by the readers, as if they had subconsciously adjusted the text to fit an internalized grammar, and which are hardly noticeable even to the most attentive listener reading simultaneously from the script:

ex 12 "*on* the front line” (instead of “*in* the front line”).

ex 13 “it *is largely consisted* of” (instead of “it largely *consists* of”).

ex 14 “back *in 1940’s*’” (instead of “in *the* 1940s”).

ex 15 “working in *the* job for which they are qualified” (instead of “*a* job”).

In the end, we felt we had a corpus of texts which, although featuring numerous hesitations, slips, and stress errors – not one of the thirty texts was completely free of these – they would nonetheless be accessible to our students, and in some cases, the performance errors would be familiar to

students from their own experience of listening to non-native lecturers.

We were also interested in feedback from the readers themselves, especially their own estimates of how 'authentic' the texts felt, and how they rated their readings of them. Each reader completed a feedback form (appendix 1), in which two stated that they had some experience in lecturing in English themselves. The feedback revealed considerable disagreement in their opinions. Two (including one of the former lecturers) felt that the texts seemed to be "authentic"; two felt that they were not. Three found them difficult to read, because of time constrictions and/or the lengthy sentences; two said they were aware that they had made errors "typical of non-native speakers", which they identified as vowels, the failure to articulate the inter-dental fricative, and intonation. In fact, none of the speakers seemed (to the author of this book) to have problems with the inter-dental phonemes (which for Jenkins 2000 are "non core"). Two believed that the texts would have been easier to understand if read by native speakers; two did not. However, when asked if they thought that non-native speakers would understand them as easily as native speakers would, three were in agreement.

There was only one question which produced a unanimous response. All four readers answered "yes" to the question "Do you think your reading of the texts sounded natural?" The word "natural" had been offered in the questionnaire with no explanation, but clearly was understood to mean something different from "like a native speaker". Clearly, too, the four readers were unanimous in their confidence that there can be a 'naturalness' to lingua franca communication, which transcends the 'naturalness' of native speakerism, and which is 'naturally' fluid and variable, making conscious or non-conscious use of nonstandard features, which do not necessarily compromise intelligibility but may actually promote it. We shall return to this idea when we consider the feedback from the test takers in the following section, and their comparison of native and non-native speaker intelligibility.

6.4 Test Administration and Test Taker Feedback

The data which we present in this section comes from the first two administrations of the new co-certification (ISE 3) in the spring of 2016 and 2017. The exam comes in two parts, "reading and writing", and "speaking and listening". The reading and writing part is allocated a fixed date, concurrently with the international version, with which it shares most of the exam material. The date of the speaking and listening part is chosen by the test centre (i.e., in this case, the University), usually a month or so after the written part.

In 2016 there were 29 candidates for the co-certification at ISE 3 (C1) level, a lower number than usual, perhaps because it was the first admin-

istration after a gap year; in 2017 the number grew to forty, closer to the average number of candidates for the first decade of the project, from 2005 to 2014. We thus have data for 69 candidates over a two year period.

In the new certification, the two parts (reading and writing, and speaking and listening) are certified separately, making it possible for candidates to fail one part of the exam, but to receive a certificate for the other part. A candidate passing both parts will thus receive two partial certificates, and an overarching certificate for the four skills when both parts of the exam are passed. In all, 64 candidates passed the speaking and listening part; 57 passed the reading and writing. Of the five who failed the speaking and listening, only two went below the minimum score for the free-standing listening, which therefore appears to have been the easiest section of the whole exam. This is confirmed by the number of candidates (fourteen) earning a distinction for listening (compared to nine distinctions for speaking).

Why should this be so? Firstly, we need to clarify that this was the shortest part of the test, carrying the least weight. The tasks (identifying the topic, and then, after the second listening, listing the main points) were probably more straightforward than the interactive speaking and listening tasks in the same exam, in which the candidate had to assume a persona in response to a cue from the live examiner, and take the initiative, by making suggestions, giving advice, and generally being imaginative. This sort of 'empathetic' listening is quite different from the focus on content required in the independent listening task. Trying to understand the content of lectures (in English or not) is part of the day-to-day reality of being a university student; engaging with strangers in role plays is not.

The feedback from students shed further light on the results. All 69 students completed a short, one page form with eight questions (appendix 2), all of which concerned the independent listening task. Sixty three students said that they had not found the content difficult, while sixty eight thought that the speaker spoke clearly. This almost unanimous response was in spite of the numerous errors, hesitations, and false starts which we noted above. It would seem, then, that performance imperfections do not necessarily impede communication in lingua franca, if the content is accessible. Clarity was presumably enhanced by an appropriate speed of delivery: sixty one students believed the speaker spoke "at about the right speed".

There was more variation of responses when it came to making judgements about the speakers' accents. Eleven students thought the accent had interfered with their understanding; fifty-eight did not. Of the eleven supposed comprehension problems, five were caused by the native speaker of French, and four by the Catalan; the Italian and Spanish speakers, in contrast, each caused problems in only one case. Given the marked accents of all four speakers, these results seem to bear out findings that communication can be successful in the face of noticeable or strong accents (Levis 2005, Derwing and Munro 2015).

The case of the French speaker merits a reflection. In spite of the fact that there were fewer performance lapses in her recordings, her accent was the most problematic, which could have been due to the nasal vowels which we have already mentioned. One student commented that the accent was "too thick", another that it was "very strong" while a third reflected:

> I couldn't stay focused on what the speaker was saying because I was being distracted by the accent.

This is an interesting comment, because it suggests that it was not so much the intelligibility of the phonology as the listener's own attitude, or low tolerance level to a marked accent, which errected a barrier to understanding. Nonetheless, the authors of these comments both passed this part of the test.

Forty three students said they were familiar with the accent in the recording they listened to, and a similar number (44), unsurprisingly, recorded that they did not think the speaker sounded like a native speaker of English. Perhaps the most interesting feedback of all came in the answers to question 7, which compared the accent of the non-native speaker in the recording with the accent of the live native speaker examining conducting the exam with the student. The question read

> In comparison with the accent of the examiner the speaker of the recorded listening text was
>
> EASIER / MORE DIFFICULT / NEITHER EASIER NOR MORE DIFFICULT
>
> to understand.

A large majority (78%) thought that the recorded text was neither easier nor more difficult (55%) or even easier (19%) to understand than the native speaker who was with them in the examination room. Only 18 students (26%) found the non native speaker more difficult to understand than the native speaker. Given that the native speaker examiner (British, male), spoke clearly and used an accent unmarked by regional inflections, this is perhaps surprising. After all, the examiner had ways of making himself understood – such as repetition and the use of non-verbal language – which the recorded voice did not have. Again, students' comments are illuminating; predictably, those who found the recording more difficult referred to an "unfamiliar accent", or a "foreign accent", but also to the fact that "we couldn't see the gestures and expressions". For those who found the recordings easier, reasons given included being "accustomed to recordings, not used to talking with native speakers", "I'm more familiar with stranger (*sic*) accents", and the self reflective: "I think it's psychological: if I know someone is a non-native speaker I feel closer to him".

The majority found no difference in difficulty between native and non-native speaker; some students felt the need to explain why:

> "I have many foreign friends around Europe, so I'm used to different accents."

> "I listen to both native and non-native speakers regularly."

This is a timely reminder, not only of the increased mobility of university students across Europe, but also of the fluid nature of ELF communication, which, in its widest sense includes interaction with native as well as non-native speakers, and one of the defining characteristics of which is the ability to cope with variability.

Only a few students added any additional comments on the listening task (question 8), mostly to comment on the accessibility of the accent, or to approve of the perceived rationale behind the test:

> "I consider the British accent more difficult to understand but I can imagine that the aim of this task is not to make the exam more difficult but to test our understanding of the foreign accent."

> "I found the speaker's accent really understandable. His hesitations did not influence the clarity of the speech."

> "The non-native speaker's level of English was good enough to be understood easily. As most of English speakers nowadays aren't natives I think it's a good test."

However, at least one student questioned the validity of using non-native speakers:

> "I believe it's nice to hear a non-native speaker speaking, but probably not for an English exam."

6.5 Test Results: Unproblematic and Uncontroversial?

The test results (first reported in Newbold 2017b) suggest that for most students the listening part was unproblematic, and even those students who flagged up difficulties related to the accents for the most part demonstrated sufficient understanding of the texts to pass the exam. The potential issue of fairness which Harding (2012) raises – namely that a candidate might be at unfair advantage if he or she shares the same first language as the speaker, does not seem to arise. For Harding, reporting research carried

out in Australia, and which made use of Chinese, Japanese, and Australian English accents, the evidence of unfair L1 advantage is not conclusive. But, he suggests, the problem is avoided if the accent is written into the test construct, or diluted, if the listening test uses a range of accents.

In the co-certification, the Italian accent was part of the construct of the "fluent non-native speaker"; in a meeting held for candidates before the exam, in which the structure of the new exam was explained, students were told that they would hear a European accent. Most, but not all, students recognized the Italian accent when they heard it, just as most correctly identified the French accent,[2] although one student wrote "I think the speaker was – or pretended to be – a Spanish woman, so since Spanish is quite similar to Italian, and we have a similar accent, it was really easy to understand her".

Compared with the results for the generic, international version of the independent listening task, the co-certification results are particularly interesting: the pass rate of 97% is matched by 84% for test takers of comparable age (i.e., university students) in the rest of Italy and 72% for candidates worldwide. This comparison, obviously, should be treated with caution, given the small number of candidates for the co-certification.

The new co-certification is not a 'test of ELF', nor was it meant to be, but it is certainly an 'ELF-aware' test in its attention to local needs for the writing part, and, especially, the non-native speaker recordings in the listening. But it is also a small scale project, relying on limited resources, and with an uncertain future; the need to pre-test all items, to align them with the test production procedure for the main international suite, is problematic for a certification which has a small catchment area. But whatever the future of the co-certification, this second, latest version has shown that a language test which looks beyond native speaker models is not only feasible and valid, it can also be uncontroversial for the test taker and the recognizing institution, and potentially generate good washback for the development of future teaching programmes.

More problematic is the development of a 'full-blown' test of ELF – if indeed, such a thing is possible or even desirable. The receptive skills are one thing, the productive skills quite another. If intelligibility, rather than nearness to a native speaker model, is to become the yardstick by which success is measured, then new modes of measuring will be required to assess speaking, and possibly writing.

In the next and final chapter, we shall look at possible future directions for language assessment in general, and high-stakes certification in particular, in the light of the growing need to assess competence in using English as a lingua franca.

2 This emerged in informal feedback after the exam; students were not required to guess the accent when completing the feedback form.

7 The Shape of Certification to Come

Abstract This final chapter offers a reflection on possible future directions for English language certification. The major problem to solve (or to attempt to solve) seems to be not so much *what* to assess (fifteen years of ELF research have offered lots of insights into this) as *how* to do it. After discussing a series of problems related to rater rubrics, and the notion of error, we consider the format that future ELF aware certification may take, concurring with Harding and MacNamara that an add-on ELF component currently seems the most practical way of incorporating an ELF element into mainstream certification. We conclude that the development of ELF certification is likely to be slow and painstaking, it may combine local and global elements, but in the long term it is inevitable, since the demand for valid and reliable certification of competences in the use of the world's lingua franca is destined to grow, perhaps for many more years.

7.1 The Need for New Approaches

In the new world order envisaged in *The Shape of things to come* Wells does not make it clear whether citizens need to certify their level of Basic English to access the jobs market (which seems to be controlled by their "educational guardians"), or any other position in society which will require them to use the lingua franca. Perhaps it is no longer necessary to do so; Basic English appears to be an easily acquired lingua franca in the new "body of mankind" which has become "one single organism" (444). In Wells's brave new world of well-behaved citizens the acquisition of the world language has in fact proved most difficult for the native speakers of English, who require special training "to restrict themselves to the forms and words" needed for successful lingua franca communication. This is an interesting reflection on the role of the native speaker in lingua franca interaction, and it raises questions not only about *what* ELF 'certification', if it is ever to exist, should attempt to certify, but *who* should be taking the test. The native speaker vs non-native speaker is just one of a number of dichotomies that the test developer, or examining board, will need to address in the preparation of any test of ELF.

After a lifetime in language testing, in a "State of the Art" interview, James Dean Brown (Salmani Nodoushan 2015, 139) argues that there are (at least) fourteen approaches to testing English language proficiency ("whatever that may be"), six of which are "top down" and eight of which are "bottom up". Only one of these (the first top down approach) is rooted in a native speaker

model approach. The others (top down) he labels as "truth-in-advertising", "multiple world Englishes", "English as a lingua franca", "global standard English", and "functional approaches", while the bottom up approaches include "the effective communicator", "scope of proficiency", "scale of range", "intelligibility", "resourcefulness", "symbolic competence", "intercultural communication skills", and "performative ability".

The list is useful, not because it is exhaustive, but because it is long. Apart from the "top down" approaches which might loosely correspond to the agendas of language planners and curriculum designers, the list of "bottom up" approaches suggests a wide range of user-focused competences, most of which could be of interest to an ELF test designer. For example, "effective communicator" suggests developing tasks which have a measurable outcome in terms of successful communication, "intelligibility" suggests a focus on perception rather than (native speaker like) production, and "resourcefulness" could include a raft of strategies (such as paraphrasing, self-repair, and requests for clarification) which have been described in the ELF literature and which tend to facilitate successful outcomes in ELF interaction.

In an early (2006) publication Elder and Davies outlined a number of tasks which might feature in a test of ELF, such as avoiding native speaker-centric lexis, listening to non-native speakers, and participating in a role play with a speaker from a different lingua-cultural background. The first of these seems conceptually problematic, since it requires raters to look for an absence of something, and evaluate it positively, while they are asked to overlook non-standard features which do appear but which do not impair communication. The second (listening to non-native speakers) was the focus of the project described in chapter 6; the third, task-based interaction, was already being used in the form of paired assessment, such as the speaking tasks in the Cambridge exams. In any case, the authors themselves conclude their proposal by warning "against moving too quickly to assess ELF before it has been properly described".

A more structured approach has been put forward and experimented by Harding (2015), who took an information gap activity, carried out by two participants from different lingua-cultural backgrounds, one of whom was the "information provider", the other the "information receiver". Ten raters were invited to observe ELF features relating to accommodation, negotiation of meaning, and discourse maintenance. Although they agreed broadly on which of the two participants performed better, Harding concludes that it was not clear how the holistic rating scale they were using was actually being interpreted.

This kind of information gap task has been familiar since the communicative language testing revolution announced by Morrow (1979), and it brings with it a series of rater-related problems which, as Harding acknowledges, will need to be addressed if such a thing as an ELF test is to be developed. If the main focus of ELF assessment is to be spoken interaction – the co-

construction of meaning between two or more participants who have different mother tongues – then the major challenge for examining boards will be to develop reliable rating scales to evaluate this interaction. Of course, every area of language activity can be undertaken in an ELF context, and an ELF assessment could thus be extended to include listening, reading, writing, and spoken (monologic) production, all of which could be relevant to an assessment for academic purposes. But it is not the *what* to test which is the primary problem for the ELF-aware test developer; this should be directly linked to the target language use domain envisaged, which (for English in academic contexts) emerges clearly in needs analyses such as the one we described in chapter 5. Rather, the problem is *how* to assess the one-off, unique, never-to-be-repeated performance moment of any ELF interaction through an assessment tool (such as a holistic grid) which is nonetheless fixed, stable, and (ideally) potentially reliable.

7.2 Re-Thinking Rating

Paran and Sercu (2010) analyse four aspects of language education which they consider to be "untestable", yet worthy of testing: literature and literary competence, learner autonomy, CLIL and inter-cultural competence. To these could be added ELF, but there is a difference. Paran and Sercu (2010) take a process, learning-based approach to strands which have come to occupy important positions in school curricula, and for which evidence of acquisition and/or progress would be useful. ELF, as ELF researchers are at pains to point out, is use of English beyond a learning context (see chapter 5). The strategies that ELF users bring to bear in interaction may of course be fostered in language classrooms, by 'ELF aware' teachers, but they may also develop in users independently of any formal learning process. Indeed the familiar (and perhaps cosy) environment of the classroom is at odds with the unpredictable nature of ELF interaction, and any test of ELF interaction would need to guarantee a degree of unpredictability in the task it attempts to assess. This is just one aspect of the "rating problem", and it concerns the identity of the participants, as well as the nature of the task. We turn now to consider briefly some of the areas in which an examining board engaging with ELF interaction would need to rethink existing communicative tests.

7.2.1 The Identity of Participants

By definition, participants in any ELF interaction do not share the same native language. If the paired assessment model is to be used, this is likely to cause logistic problems for examining boards, especially if a traditional

format is used, i.e. with the test takers physically in the same room together; one of the participants would have to be brought in from a different lingua-cultural community to the local one. In relatively stable monolingual communities which are still the norm in Europe, this would be difficult.

Of course, existing communicative-type paired assessments could similarly be criticized when they somewhat unnaturally invite candidates who have the same mother tongue to converse in English, which may cause unexpected comprehension problems for the native-speaker examiner.[1] But interactive tasks in a traditional communicative test are primarily designed to elicit appropriate language, and not to sample a range of pragmatic strategies which enable ELF communication to take place.

Alternative formats could include setting up a video interaction using the Internet – but this would require negotiating criteria for matching test takers, and bring into play a number of variables related to the use of technology – or to revert to one-to-one interaction, between examiner (or facilitator) and candidate, in which the examiner is herself part of the meaning-construction process. This takes us to the next aspect of the problem, the need for empathetic raters.

7.2.2 The Empathy of Raters

More than a decade ago, House (2003, 573) suggested that

> the yardstick for measuring ELF speakers' performance should [...] be an 'expert in ELF use', a stable multilingual speaker under comparable socio-cultural and historical conditions of use, and with comparable goals for interaction.

The monoglot native speaker, it is implied, would be at a disadvantage for assessing ELF interaction. This would probably be a consensus view for most ELF researchers and 'ELF-aware' teachers today, although Canagarajah (2007, 927) points out that there is "nothing stable about the multilingual speaker". This is not to assert that a trained native speaker rater would be unable to make judgements about the effectiveness of strategies used by test takers, but by referring to "comparable goals for interaction" House seems to be alluding to the collaborative nature of meaning making;

1 The author was once told the following anecdote by an examiner who had attempted to make a paired assessment in Naples. The two candidates chatted away comfortably in 'Neapolitan' English, fluently, respecting time limits and turn taking, clearly understanding each other, and thereby achieving a degree of communicative success, but the examiner understood little or nothing of what was being said, and consequently found it difficult to rate the candidates' performance.

whoever is doing the rating also needs to be part of this process, whether she is interacting directly (in an interview) with the test taker, or simply observing performance. In short, raters need to be empathetic participants and/or listeners. This is at odds, of course, with a traditional view of an examiner as detached, impartial, and objective.

Other queries also arise about the identity of raters. What if they were to share the same mother tongue as one of the test takers? Would that compromise fairness and impinge on test validity? Examining boards would need to draw up a recruitment and training policy for raters, define the competences required, develop scoring rubrics, and implement a validation process to ensure a degree of inter rater-reliability. The starting point could be the trialing of a holistic grid, such as the one suggested by Harding (2015).

7.2.3 The Need for Evidence

Harding tentatively suggests a check list of strategies for a holistic rubric organized under the principle competence areas of "accommodation", "negotiation", and "maintaining smooth interaction". The first of these includes making oneself intelligible and adjusting to the interlocutor's speech or style. "Negotiation" lists four well documented ELF strategies, clarification, self-repair, repetition, and paraphrasing; the final area of discourse management includes turn-taking and politeness.

This is a good start, but other strategies could be added. In many ELF interactions, progress is anything but smooth; communicative success, if it is achieved, is achieved against the odds (Newbold 2015a, 214), and it may involve such diverse ploys as explicit or implicit requests for help, the use of body language, or specific references to shared cultural resources. In short, there can be a messiness to the negotiation of meaning which should not be mistaken for lack of competence(s), but an attempt to harness all possible resources.

On the other hand, interaction may indeed be 'smooth'; so seamless, in fact, that there is nothing to observe in the way of self repair, repetition, paraphrasing, or any other criterion which may be taken from a taxonomy of pragmatic strategies for ELF communication. What happens when test takers converse with no apparent need to resort to accommodation or repair strategies? How would communicative success be measured in these cases, with little or no evidence of ELF strategies being employed? This is an eventuality which test developers would need to anticipate. In Harding's information gap activity, participants were presumably chosen because of their very different lingua-cultural backgrounds – one a native speaker of Thai, the other of Spanish. The lingua-cultural gap may close when both or all participants come from the same geopolitical area, such as the European

Union which has been the main focus of this book, and this may make communication easier. In an international test of ELF which included spoken interaction, how would an examining board match test takers?

This begs another question about traditional levels of language competence. Like the lingua-cultural gap, a mismatch of levels of fluency (however we might define this) is likely to cause more strain for participants, and as a result more opportunity for resorting to ELF strategies for both participants; could 'mismatch' be a criterion for pairing test takers? In a test of ELF (if it is ever to exist) should test takers be required to supply information about their presumed level on a well-known scale (such as the Common European Framework) when they enrol for the exam?

7.2.4 The Problem of Levels

In a criterion referenced, task-based, communicative test success is ideally measured in terms of outcomes. To take a simple example from real life: if an information receiver R is able to get to the railway station on the basis of directions provided by information provider P, then the interaction can be considered as having a successful outcome. From this perspective, a 'purely' communicative test can have only two possible outcomes: success or failure. Indicating a degree of success – or even more grotesquely, a degree of failure – would be difficult and irrelevant.

In a hypothetical rating rubric for ELF interaction, even if we are to focus on evidence of ELF strategies which facilitate a successful outcome, rather than the outcome itself, there will be a problem of identifying levels. Luoma (2004, 80) suggests that the norm (to guarantee a degree of inter-rater reliability and therefore consistent results) is from four to six levels of performance, but she is referring to both holistic and analytic grids in traditional tests based on a standard model of the language. When it comes to the 'untestable' areas of language ability, the would-be ELF tester might find Sercu's (2010, 29) discussion of three possible levels for measuring intercultural competence (basic, intermediate and full) relevant, although not transferable in any acritical way, to the ELF context.

In short, the problems of rating ELF interaction seem insurmountable. Wherever we focus our attention on rubrics or on levels of performance, on raters or on the test takers themselves, we find questions but no obvious answers. However, so far we have been considering a hypothetical stand-alone test of ELF; a test which *only* measures a yet to be defined ELF construct. The prospective changes when we think in terms of ELF assessment as an add-on element to a more traditional (Framework related) test. This is the conclusion reached by Harding and McNamara (2017):

It seems more likely that ELF is at least in the short term not going to

replace more static proficiency constructs, but rather would function as an add-on in contexts of language assessment where ELF competences are expected to come into play (which may be all situations).

We shall return to this idea of the 'add on' in the section on test formats below. First, however, we need to consider another rater-related dichotomy, about which examining boards attempting to assess ELF would need to issue guidelines, and which, for many teachers preparing students for tests would be crucial: the notion of 'error' in international communication.

7.3 Rethinking Errors

The notion of error in language teaching and testing is traditionally, and often unquestioningly, equated to a deviance from native-speaker norms. References to native speakers may be built into rating scales, and there are numerous references in the CEFR to native speakers. Notoriously, concepts such as not "unintentionally amusing or irritating" or "keeping up with native speakers"[2] are built into the scales for spoken interaction, suggesting that native speaker-like proficiency, and indeed, native speaker-like behaviour, should be the wider target language domain as a testing objective. However, it should be remembered that the CEFR was developed not with a single language (English) in mind, but as a functional description which could be used for all European languages, and it was never intended to describe levels of competence for a lingua franca.

The case of English is doubly exceptional: not only because of its use as a lingua franca, but also because of the emerging paradigm of world Englishes, which embraces variability in all aspects of language use (phonology, syntax, lexis, discourse management, etc.). English does not have one 'standard' version, but many native and second language speaker norms, and a growing awareness of this variability, and the choices to be made about which English to teach – and consequently test – have become a major subject for discussion in training courses and publications for the ELT (English Language Teaching) profession. (Newbold 2017a).

For would-be language certifiers, one possible approach to error would be to discard *any* deviation in production from *any* native speaker norm, at least if these deviations were not considered to undermine comprehension; but this would be problematic, not only because of the subjective judgements involved (on the part of the rater, who may not always be sure

2 CEFR Descriptors for Level B2 include:
conversation: Can sustain relationships with native speakers without unintentionally amusing or irritating them.
informal discussion: Can keep up with an animated discussion between native speakers.

that understanding has taken place), but also because an initial lack of comprehension (and awareness of such) will often be the trigger for those ELF repair strategies which raters would be looking for, and which are a necessary part of the co-construction of meaning. In any case, it would be advisable for examining boards to establish a public policy on errors, including a definition of error, and the part played by errors (if any) in the assessment process. We shall briefly consider how these might vary from one aspect of language use to another; these considerations could be addressed in a policy document on errors which could be incorporated into test specifications.

7.3.1 Phonology

Outlining a 'lingua franca' approach to testing pronunciation, Sewell (2017, 238), writing from Hong Kong, suggests that the challenge "lies in navigating the local/global polarity". This observation seems particularly pertinent in the light of the research by Basso (chapter 5) who found that, for the majority of European students in an international campus in Venice, the most difficult accents to understand were North American (i.e., native speakers of English) and South East Asian (speakers whose mother tongues were Chinese and Japanese). Although this research did not have pronunciation as its main focus, we might speculate that the comprehension problems are linked to two concepts which Sewell refers to: "intelligibility" and "functional load". The first of these is taken to mean the quantity of understandable speech; the second, the extent to which specific phonemes are used contrastively (an indication of which can be given by the number of minimal pairs a phoneme contrast is required to keep apart). In the case of the north American speech, unfamiliarity with accent, coupled with speed of delivery and lexical load, could have made understanding difficult, whereas in the case of the Japanese and Chinese speakers problems of perception may have been more exquisitely phonological.

Another interesting factor Sewell refers to (243) is the possibility that "written language and worldwide literacy operate as centripetal forces on pronunciation". This also seems relevant in the context of English as an academic language. The notable mismatch between spelling and pronunciation, as well as the rhythms of stress timed language, which are features of native speaker English, are often eroded in lingua franca interaction. Stress timing is not part of Jenkins' "core phonology", and it is easy to see why: careful syllable-timed speech makes perception less, not more, difficult, and it may be adopted as an accommodation strategy.

7.3.2 Syntax and Morphology

Language testers, and especially examining boards delivering high stakes tests, have come to be seen as guardians of standards, and it would perhaps not be unfair to assert that this role has been promoted by a testing culture which has developed around the notion of errors, and especially grammar errors. Generations of test takers have been tricked into selecting erroneous forms in an array of objective test types, from multiple choice to cloze, from true/false to sentence rewriting. One reason is that such tests (or parts of tests) are easy to create and easier to score. But they belong to the written domain. Unsolicited grammar errors in spoken production, and in spoken interaction, may be captured in analytic scoring grids, but (as we saw in chapter 2) grammatical accuracy is likely to be seen as just one of several assessment criteria, and probably not the most important one.

Grammar errors do not usually compromise intelligibility, but they are harshly viewed by the academic community. In the 2006 study by Mollin, in a survey of 435 European academics, 95% responded that omission of the third person "s" (in the example sentence: "Do you know where she live?") was unacceptable, making it the most despised error of all. Yet, taking the long term view, it is arguably a fossil structure, the only morphological inflection left of a once highly inflected verb system, and doomed to disappear.

In a lingua franca context the focus changes again, since grammar may be manipulated to enhance meaning. A sentence such as

ex 1

I will go to Rome if you will come with me.

mirrors structures in many other languages, while emphasis may be achieved by left dislocation – Mauranen (2010) provides a number of examples of this from the ELFA corpus – such as

ex 2

This problem, I'll come back to it in a minute.

Reduplication, which has only a limited use in standard English, but is a feature of some world varieties, as well as other languages (including Italian), might also be used for emphasis, instead of an intensifier:

ex 3

It's a small small problem.

These are just a few instances of deviation from a standard which could be used to inform a new approach to errors in a test of ELF.

7.3.3 Lexis

We referred to lexical creativity in chapter 5 as a major focus of ELF research and an effective strategy for creating meaning. In fact, much lexical creativity takes place at the interface of grammar and lexis, through the manipulation of morphology. Like many ELF strategies, it cuts across the divide between native and non-native speakers. For example, the word *involvable* was recently used by an Italian post doc student in conversation with the author, to refer to a motivating classroom activity, in which everyone could take part:

ex 4

It's a very involvable activity.

A Google search[3] asks the information seeker if they didn't in fact mean *insolvable*, and when the offer is turned down, returns a count of just 2,230 hits for *involvable*, some of which are clearly in a non-English context. The meaning that was inferred was both "motivating" and "not difficult to participate in"; which was confirmed by the person who had coined it. It seemed to the author (and still seems) not so much an error as an economic and elegant term for a useful concept.

This kind of creativity shows considerable language awareness. It demonstrates knowledge of lexis (*involve*) and knowledge of word formation processes (affixation). There is nothing in the example to indicate that it is a non-standard form used by a non-native speaker, rather than a term invented by a native speaker to plug a gap. In a context of ELF assessment, it would be an observable strategy promoting communication.

More problematic, from an assessment point of view, is to sanction lexical choices which seem to hinder communication, as in the case of "unilateral idiomaticity" (Seidlhofer 2011, 134). Communication breaks down when a word or words (whether used idiomatically or not) are not familiar to the interlocutor, but it is at this moment that ELF accommodation strategies can kick in, and the channel of communication be re-opened. An ELF assessment grid, rather than simply noting errors and breakdowns, should be observing if and how these are transformed into opportunities for co-operative meaning making.

3 Search made on 2017-07-03.

7.4 Rethinking Test Formats

So far in this chapter we have been discussing spoken interaction, which lies at the heart of ELF usage, and which probably poses most challenges for any hypothetical "test of ELF". But a test of ELF, or a more realistic ELF aware test, may include other skills, and may embrace many formats.

To start with, ELF may be manifested in different ways, and allow for more or less variability, and consequently require a more or less rigid test format. Basic English, with which we began this chapter, is an example of a controlled natural language (Kuhn 2014), with a prescribed word list and specific rules for meaning creation (through the combination of words in the list). Similarly, there are areas of professional use of English today in international contexts, such as so-called "Seaspeak", for maritime communication, and "Airspeak", for air traffic controllers and pilots, the main aim of which is the avoidance of ambiguity. In these contexts language needs to be carefully regulated and assessed, not least because human lives daily depend on the successful communication in English between non-native, and also native, speakers.

This is not the kind of lingua franca use we have in mind for certifying competences in academic English. Rather, beyond the challenge posed by the assessment of spoken interaction, future certifications may not look very different from existing certifications described in some detail in this book, and extend, as they always have done, to reading, writing, and spoken (monologic) production. They could, however, be made ELF aware in the choice of texts for reading and listening components, and in their assessments of written and spoken production.

Reading components, for example, could include texts by non-native writers. These could be literary, academic, formal or informal, depending on the underlying construct for reading skills; they could be carefully sourced or specially written, published or unpublished, from a "world English" variety, or from the "expanding circle", to use Kachru's well known (1985) model. In a one topic, multi-text approach which has been adopted in the new international version of the Trinity College *Integrated Skills in English* suite (chapter 6), one text could be by a non-native writer. Similarly, tests of listening could incorporate non-native voices, such as the extract from a lecture in the updated co-certification (chapter 6), but also genuine short ELF interactions which might be relevant to the overall aims of the test. These latter might not be very different from the "extracts from life on campus" which are a feature of the TOEFL test, with the difference that both participants would be non-native users of English.

In the productive skills the problem of native speaker norm returns, and with it, the problem of rating. In writing, especially formal writing of an academic nature, it is harder to justify deviations from native speaker norms. But the advent of computerized testing of writing may alleviate

these, since test takers could switch on spell and grammar checks to reduce low-level formal inaccuracies. After all, in most non-testing contexts of writing, writers would normally be able to make use of tools (such as dictionaries, grammars, and style guides) to help them; it is thus what they can do *with* such tools, rather than without them, which should be of greater interest for assessment purposes and provide most information for the test user. This would allow an empathetic ELF user/rater to shift her focus to higher level aspects of discourse management, such as structural cohesion and clarity of argument; an objective which might also be within the range of some future (non-native) machine marking system.

The assessment of spoken production seems to us to be particularly important in the context of ELF. More than ever, English, or rather, ELF, plays a role in the professional lives of non-native speakers, and universities can provide a training ground for future professionals who may have to give reports in meetings or address audiences, by offering opportunities to hone their presentation skills. With the reform of the university system in Italy, and the introduction of the *laurea magistrale*, student presentations have become a staple feature of many courses, and may be used as part of a continuous assessment process. Certification provides an excellent opportunity for an ELF-type presentation, of a topic chosen by the candidate, and addressed to a putative non-native speaker audience. The skills which might feature on a check list for raters could include, to give just a few examples, voice control (speed, volume, use of pauses), repair strategies, and discourse management features such as signposting. Newbold (2015a, 219) suggests that these could be usefully assigned to a higher order of categorization for rating purposes: *control* (of voice, lexis, etc), *range* (of repair strategies, etc.) and *alignment* (or ways in which the speaker connects to the audience).

Of course, all of these skills would be part of the stock-in-trade of a competent native speaker, but none of them belong exclusively to the native speaker domain; they cut across the language divide, and there would thus be no point, indeed, no meaning, in including "native speaker like" behaviour on the assessment check list. Rather, a hypothetical future test, or certification, of "speaking to an international audience" could be aimed at both native and non-native speakers; and the native speakers (as Wells predicted) might find it more difficult to score highly on such a test than their battle-hardened, ELF-using, non-native counterparts.

Such a test could be a free standing 'certification' of spoken production in its own right, of interest to prospective employers in an international jobs market. We have already referred to a test of spoken interaction as a possible 'add-on' component to an otherwise conventional certification. At this point it seems that a modular approach to certifying ELF competences, whichever skill(s) we are interested in, is likely to be the most practical, for at least three reasons. Firstly, it would keep ELF and

non ELF approaches to rating separate, allowing a generic component to be linked to a framework such as the CEFR. Secondly, it recognizes that some skills might relate to a 'specialized' ELF construct (academic writing, interacting with patients in a healthcare context, etc) and could also be offered as 'add-ons' or stand alone tests. Thirdly, a modular approach would allow local versions of a test, on the global/local interface, so that, for example, a European test of ELF for academic purposes might include both local and global elements; it might offer 'local' contents but look for global ELF strategies in the test taker.

Versioning certifications obviously has a cost for examining boards, but allows them to reach more candidates. This approach has long been adopted by IELTS (chapter 2), which offers an "Academic" and a "General Training" version of the exam, in which the listening and speaking parts are the same for all test takers, while reading and writing are different. A recent switch to a more modular approach has been made by Trinity College in the ISE exam, (chapter 6), so that the reading and writing exam, which is done on a different day from the listening and speaking, is now certified separately, making it possible for a test taker to have a certificate for just one part of the exam, and consequently, in the case of failure of one part of the exam, to re-sit only that part, with a subsequent reduction in the fee.

The greater flexibility offered by a modular approach would also allow test users to make informed choices about which elements would supply the information they were interested in, by adapting those modules most relevant to a local context, and in this way, mirroring the fluid nature of ELF itself. The modular approach, one could maintain, is more 'ELF aware' than a 'one size fits all' certification. The test format, of course, is not the test construct, but it could grow naturally out of it.

7.5 Conclusion: Evolution, not Revolution

In this book we have tried to show that, although the certification industry has grown enormously over the last two decades, it has still to address the underlying cause of that expansion: the unprecedented growth of a genuinely global lingua franca, and the need for reliable independent measurements of what ELF users can do with it. We have noted the aspirations and also the shortcomings of existing tests, and we have presented a small local project of an 'ELF aware' certification, only to return, in this chapter, to the fundamental problem of rating, to which we have offered tentative approaches but no real solutions.

If the primary focus of the book had been "assessing ELF" we might have managed to write most of it without referring to any certifications. We would have discussed a range of more alternative approaches, such as

peer assessment, self assessment through reflective feedback, or continuous observation-based assessment such as the project described in Tsagari and Kouvdou (forthcoming). Assessment which involves reflection on the part of all participants is likely to be richly formative as well as informative, and to fulfil an essential role in any ELF-aware language programme.

But certifications are here to stay, and they are important. They have a function in today's globally mobile society because they provide independent assessments which (as we have seen) prioritize fairness, reliability and security. They drive a large sector of the English language teaching and publications market, as well as providing a high stakes gate-keeping function for immigration services, potential employers, and higher education. In short, they have a controlling function which is more apparent than ever before (and which has more than a faint analogy with the controlled global society portrayed in *The Shape of Things to Come*).

This is why examining boards need to reconsider constructs, to invest in ELF assessment research, to be able to stay in touch with emerging new language needs. In the long term, to do so would make commercial sense, and assert an ethical role which not-for-profit organisations typically subscribe to. If they do not, then other locally-based organizations may emerge to do so. Indeed, a strong case could be made for locally developed tests which combine specific professional, vocational or academic content with specific international settings, such as a university access test for European University students.[4]

So far, the major examining boards seem to have shown little interest in engaging with the phenomenon of ELF, beyond the co-certification project described in these pages, although it is to be presumed that they are aware of the issues involved. Whatever the future of English language certification, formal ELF assessment is likely to come about slowly, piecemeal, perhaps through more small-scale projects, and assisted by developments in technology. When communicative language teaching was being theorized, in the late nineteen seventies, Keith Morrow (1979, 156) concluded his seminal article "Communicative Language Testing: Revolution or Evolution?" by speculating that "there is some blood to be spilt yet". Four decades later, there is not yet much evidence of blood having been spilt in the testing profession (as far as the author is aware), but rather an ongoing consensus which has evolved out of different assessment traditions and, more recently, the CEFR. The time is now ripe to move a bit further along the communicative route. As if to underline the urgency, an email alert has just arrived on the author's desktop which reads "Once You Go Global,

4 For the form such a test might take, see Newbold 2015b.

There Is No Coming Back".[5] To engage with ELF in a language certification also means going global, and to follow English along its evolutionary path as a hybrid, many-faceted tool of communication, and from which there is indeed no going back.

5 On closer inspection, it turns out to be an invitation to a webinar organized by IATEFL, the International Association of Teachers of English as Foreign Language.

Appendix 1
Feedback sheet for readers

Please write brief answers or circle the appropriate responses.

1 What is your mother tongue? .....

2 Have you lectured or given a lesson in English before? YES / NO

3 Did the texts seem to be 'authentic' (i.e. similar to a real university lecture)? YES / NO

4 If not, can you briefly say why not?

.....

5 Did you find them difficult to read? YES / NO

6 If yes, can you briefly say why?

.....

7 Do you think your reading of the texts sounded (reasonably) natural? YES / NO

8 If not, can you briefly say why not?

.....

9 Are you aware of having made any 'errors' typical of non-native speakers? YES / NO

10 If so, which?

.....

11 Do you think non-native speakers would find it easier to understand these texts if they were read by a native speaker? YES / NO

12 Do you think non-native speakers will find your readings as easy to understand as a native speaker would? YES / NO

13 Do you think you would have used simpler language if you had given the lecture? YES / NO

Thank you for providing this feedback!

Appendix 2
Post-exam feedback sheet for test-takers

Please answer these questions about the recorded listening task. This won't take long: Circle the answers which seem most true for you or write short answers where appropriate.

1 Did you find the content of the listening text difficult? YES / NO
If so, can you say why?

.....

2 Do you think the speaker spoke clearly? YES / NO

3 I think the speaker spoke TOO QUICKLY / TOO SLOWLY / AT THE RIGHT SPEED?

4 Did the speaker's accent interfere with your understanding? YES / NO

5 Did the speaker sound like a native speaker of English? YES / NO / DON'T KNOW

6 Are you familiar with the speaker's accent? YES / NO

7 In comparison with the accent of the examiner the speaker of the recorded listening was
EASIER / MORE DIFFICULT / NEITHER EASIER NOR MORE DIFFICULT to understand.
Can you say why?

.....

.....

8 If you have any other comment about this listening task, please write here:

.....

.....

Please return this form to (.....)

Thank you! **Your feedback will help us to develop the co-certification**

References

Alderson, C.; Wall, D. (1993). "Does Washback Exist?". *Applied Linguistics* 14, 115-129.
Alderson, J.C. (2007). "The CEFR and the Need for More Research". *The Modern Language Journal*, 91(4), 659-3.
Alderson, J.C. (2009). "Test Review: Test of English as a Foreign Language: Internet-based Test (TOEFL iBT)". *Language Testing*, 26(4), 621-31.
Alderson, J.C.; Clapham, C.; Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: CUP
Archibald, A.; Cogo, A.; Jenkins, J. (2011). *Latest Trends in Elf Research*. Newcastle: Cambridge Scholars.
Austin, J.L. (1975). *How to Do Things with Words*. Oxford: Oxford University Press.
Bachman, L.F.; Palmer, A. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
Bachman, L.F.; Palmer, A. (2010). *Language Assessment in Practice*. Oxford: Oxford University Press.
Balboni, P.E. (2012a). "'Sapere una lingua': dall'idea intuitiva al significato scientifico" [online]. Balboni, P.E.; Deloiso, M. (eds.), *La formazione linguistica nell'università*. Venezia: Edizioni Ca' Foscari, 75-116. DOI 10.14277/978-88-97735-13-7.
Balboni, P.E. (2012b). "Strategie operative per la formazione linguistica dello studente universitario". Balboni, P.E.; Deloiso, M. (eds.), *La formazione linguistica nell'università*. Venezia: Edizioni Ca' Foscari, 117-130.
Basso, N. (2012). "Dealing with the Unexpected: the Use of ELF in an International Academic Context". Ludbrook, G.; Newbold, D. (eds.), *English Lingua Franca: Contexts, Strategies and International Relations.* Venice: Cafoscarina, 21-40.
Bridgeman, B.; Powers, D.; Stone, E.; Mollaun, P.; (2012). "TOEFL iBT Speaking Test Scores as Indicators of Oral Communicative Language Proficiency". *Language Testing*, 29(1), 91-108.
Brooks, L. (2009). "Interacting in Pairs in a Test of Oral Proficiency: Co-constructing a Better Performance". *Language Testing*, 26(3), 341-66.
Brown, A. (2003). "Interviewer Variation and the Co-construction of Speaking Proficiency". *Language Testing*, 20(1), 1-25.
Burgess, A. (1962). *A Clockwork Orange*. London: Heinemann.

Canagarajah, S. (2007). “Lingua Franca English, Multi-lingual Communities and Language Acquisition.” *The Modern Language Journal,* 91, 923-9.

Canale, M.; Swain, M. (1980). "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing." *Applied Linguistics,* 1.1, 1-47.

Carroll, J.B. (1983). “Psychometric theory and language testing”. Oller, J. (ed.), *Issues in language testing research*. New York: Newbury House, 80-107.

Chapelle, C.A.; Douglas, J. (2006). *Assessing Language Through Computer Technology*. Cambridge: Cambridge University Press.

Chapelle, C.A.; Enright, M.K.; Jamieson, J.M. (eds.) (2008). *Building a Validity Argument for the Test of English as a Foreign Language*. New York: Routledge.

Cogo, A. (2009). “Accommodating Difference in ELF Conversations: A Study of Pragmatic Strategies”. Mauranen, A.; Ranta, E. (eds.), *English as a Lingua Franca: Studies and Findings*. Newcastle: Cambridge Scholars, 254-73.

Cogo, A. (2010). “Strategic Use and Perceptions of English as a Lingua Franca”. *Poznan Studies in Contemporary Linguistics*, 46(3), 295-312.

Coleman, J.A. (2006). “English-medium Teaching in European Higher Education”. *Language Teaching*, 39(1), 1-14.

Corbett, A. (2005). *Universities and the Europe of Knowledge: Ideas, Institutions and Policy Entrepreneurship in European Community Higher Education Policy, 1955-2005*. Basingstoke: Palgrave Macmillan.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning Teaching, Assessment*. Cambridge: Cambridge University Press.

Council of Europe (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment* [online]. URL `https://rm.coe.int/1680667a2d`.

Coupland, N.; Giles, H. (1988). “Introduction: the Communicative Contexts of Accommodation”. *Language & Communication,* 8(3-4), 175-82.

Cronbach, L.J.; Meehl, P.E. (1955). “Construct Validity in Psychological Tests.” *Psychological Bulletin,* 52.4., 281.

D’Este, C. (2012). “New Views of Validity in Language Testing” [online]. *Educazione Linguistica.Language Education*, 1, 61-76. DOI `10.14277/2280-6792/5p.`

Derwing, T.M.; Munro, M.J. (2015). *Pronunciation Fundamentals: Evidence-Based Perspectives for L2 Teaching and Research*. Amsterdam: John Benjamins.

Elder, C.; Davies, A. (2006). “Assessing English as a Lingua Franca”. *Annual Review of Applied Linguistics*, 26, 282-304.

Field, J. (2012). “The Cognitive Validity of the Lecture Based Question in the IELTS Listening Paper”. Taylor, L. (ed.), *IELTS Collected Papers 2:*

Research in Listening and Reading Assessment. Cambridge: Cambridge University Press, 391-453.
Field, J. (2013). "Cognitive Validity". Geranpayeh A.; Taylor, L. (eds.), *Examining Listening: Research and Practice in Assessing Second Language Listening*. Cambridge: Cambridge University Press.
Figueras, N. (2012). "The Impact of the CEFR". *ELT journal*, 66(4), 477-85.
Figueras, N.; Noijons J. (2009). *Linking to the CEFR Levels: Research Perspectives*. Arnheim: Cito.
Figueras, N.; North, B.; Takala, S.; Verhelst, N.; Van Avermaet, P. (2005). "Relating Examinations to the Common European Framework: a Manual". *Language Testing*, 22.3, 261-79.
Fulcher, G. (2000). "The 'communicative' Legacy in Language Testing". *System,* 28(4), 483-97.
Fulcher, G. (2003). "Testing Second Language Speaking". London: Routledge.
Fulcher, G. (2015). "Re-examining Language Testing". London: Routledge.
Fulcher, G.; Reiter, R.M. (2003). "Task Difficulty in Speaking Tests". *Language Testing*, 20(3), 321-44.
Garcia, O.; Wei, L. (2014). *Translanguaging: Language, Bilingualism and Education*. London: Palgrave Macmillan.
Gardner, J. (2016). *Misused English Words and Expressions in EU Publications, European Court of Auditors* [online]. URL `http://euenglish.webs.com/` (2017-04-20).
Giles, H. (1973). "Accent Mobility: a Model and Some Data." *Anthropological Linguistics,* 15, 87-105.
Grazzi, E. (2015). "ELF and the Development of ICC". Vettorel, P. (ed.), *New Frontiers in Teaching and Learning English*. Newcastle: Cambridge Scholars, 179-204.
Green, A. (2014). *Exploring Language Testing and Assessment*. London: Routledge.
Gribble, C.; Blackmore, J.; Morrissey, A-M.; Capic, T. (2016). "Investigating the Use of IELTS in Determining Employment, Migration and Professional Registration Outcomes in Healthcare and Early Childcare Education in Australia". *IELTS Research Reports Series*, 4, 1-58.
Groom, C. (2012). "Non-native Attitudes Towards Teaching English as a Lingua Franca in Europe". *English Today*, 109, 50-7.
Guido, M. (2008). *English as a Lingua Franca in Cross-cultural Immigration Domains*. Bern: Peter Lang.
Guido, M. (2012). "ELF Authentication and Accommodation Strategies in Cross Cultural Immigration Domains". *Journal of English as a Lingua Franca*, 1/2, 219-140.
Harding, L. (2012). "Accent, Listening Assessment and the Potential for a Shared-l1 Advantage: A DIF Perspective". *Language Testing*, 29(2), 163-80.

Harding, L. (2015). "Adaptability and ELF Communication: the Next Steps for Communicative Language Testing?". Mader, J.; Urkun, Z. (eds.), *Language testings: Current trends and future needs*. Canterbury: IATEFL.
Harding, L.; McNamara, T. (2017). "Language Assessment: The Challenge of ELF". Jenkins, J.; Dewey, M.; Baker, W. (eds.), *Routledge Handbook of English as a Lingua Franca*. London: Routledge.
Hobsbawm, E. (1995). *Age of Extremes: the Short Twentieth Century, 1914-1191*. London: Michael Joseph.
House, J. (2003). "English as a Lingua Franca: a Threat to Multilingualism?". *Journal of Sociolinguistics,* 7(4), 556-78.
Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
Hulmbauer, C. (2011). "Old Friends? Cognates in Communication". Archibald, A.; Cogo, A.; Jenkins, J. (ed.). *Latest Trends in ELF Research*. Newcastle: Cambridge Scholars, 139-62.
Hyland, K. (2002). "Options of Identity in Academic Writing". *ELT Journal,* 56(4), 351-8.
Hyland, K. (2003). *Second Language Writing*. New York: Cambridge University Press.
Hyland, K. (2015). *Academic Publishing: Issues and Challenges in the Construction of Knowledge*. Oxford: OUP.
Jenkins, J. (2000). *The phonology of English as an International Language*. Oxford: Oxford University Press.
Jenkins, J. (2014). *English as a Lingua Franca in the International University*. London: Routledge.
Jenkins, J. (2015). "Repositioning English and Mulitlingualism in English as a Lingua Franca". *Englishes in Practice*, 2(3), 49-85.
Jenkins, J.; Leung, C. (2013). *English as a lingua franca* [online]. John Wiley & Sons, Inc. DOI: 10.1002/9781118411360.wbcla047.
Kachru, B. (1985). "Standards, Codification, and Sociolinguistic Realism: the English Language in the Outer Circle." Quirk, R.; Widdowson, H. (eds.), *English in the World: Teaching and Learning the language and the literature*. Cambridge: Cambridge University Press.
Kachru, Y. (1997). "Culture and Argumentative Writing in World Englishes". Smith, L.E.; Forman, M.L. (eds.), *World Englishes 2000*. Honolulu: University of Hawaii Press, 48-67.
Kaur, J. (2009). "Pre-empting Problems of Understanding in English as a Lingua Franca". Mauranen, A.; Ranta, E. (eds.), *English as a lingua franca: Studies and findings*. Newcastle: Cambridge Scholars, 107-23.
Khalifa H.; Weir, C. (2009). "Examining Reading: Research and Practice in Assessing Second Language Learning". *Studies in Language Testing*, 29. Cambridge: Cambridge University Press.
Khalifa, H.; ffrench, A. (2008). "Aligning Cambridge ESOL Examinations to the CEFR: Issues and Practice" [online]. Paper delivered to the 34th

Annual Conference, International Association for Educational Assessment. URL http://www.iaea.info/documents/paper_2b711ec02.pdf (2017-06-12).
Khalifa, H.; Vidakovic, I. (eds.) (2014). "Research Notes 58". *Studies in Language Testing*. Cambridge: Cambridge University Press.
Kuhn, T. (2014). "A Survey and Classification of Controlled Natural Languages". *Computational Linguistics*, 40(1), 121-70.
Kunnan, A. (ed.) (2000). "Fairness and Validity in Language Assessment. Selected Papers from the 19th Language Testing Colloquium, Orlando, Florida". *Studies in Language Testing*, 9. Cambridge: Cambridge University Press.
Lenz, P. (2004). "The European Language Portfolio". Morrow, K. (ed.), *Insights from the Common European Framework*. Oxford: Oxford University Press, 22-31.
Levis, J.M. (2005). "Changing Contexts and Shifting Paradigms in Pronunciation Teaching". *TESOL Quarterly,* 39, 369-77.
Lopriore, L. (2015). "Young Learners in ELF Classrooms". Vettorel, P. (ed.), *New Frontiers in Teaching and Learning English*. Newcastle: Cambridge Scholars, 159-77.
Luoma, S. (2004). *Assessing Speaking*. Cambridge: Cambridge University Press.
MacKenzie, I. (2015). "Will English as a Lingua Franca Impact on Native English?". *Studies in Variation, Contacts and Change in English*, 16.
Mauranen, A. (2010). "Features of English as an Academic Lingua Franca". *Helsinki English Studies*, 6, 6-28.
Mauranen, A. (2012). *Exploring ELF: Academic English Shaped by Nonnative Speakers*. Cambridge: CUP.
Mauranen, A. (2015). "What is going on in Academic ELF?". Vettorel, P. (ed.), *New Frontiers in Teaching and Learning English*. Newcastle: Cambridge Scholars, 31-52.
May, L.A. (2010). "Developing Speaking Assessment Tasks to Reflect the 'social Turn' in Language Testing". *University of Sydney Papers in TESOL*, 5, 1-30.
McNamara, T.; Roever, C. (2006). *Language Testing: The Social Dimension*. Oxford: Blackwell.
Merrifield, G; GBM & Associates (2016). "An Impact Study into the Use of IELTS By Professional Associations in the United Kingdom, Australia, Canada and New Zealand, 2014-15" [online]. *IELTS Research Report,* 7, 1-35. URL https://www.ielts.org/teaching-and-research/research-reports/ielts_online_rr_2016-7 (2017-04-20).
Messick, S. (1975). "The Standard Problem: Meaning and Values in Measurement and Evaluation". *American psychologist*, 30(10), 955.
Messick, S. (1989). "Validity". Linn, R.L. (ed.), *Educational Measurement*. New York: Macmillan, 13-103.

Messick, S. (1994). "The Interplay of Evidence and Consequences in the Validation of Performance Assessments". *Educational Researcher,* 23(2), 13-23.
Messick, S. (1996). "Validity and Washback in Language Testing". *Language Testing,* 13(3), 241-56.
Messick, S. (1998). *Consequences of Test Interpretation and Use: the Fusion of Validity and Values in Psychological Assessment*. Princeton, New Jersey: Educational Testing Services.
Mollin, S. (2006). *Euro-English: Assessing Variety Status*. Tubingen: Guntter Narr Verlag.
Morrow, K. (1979). "Communicative Language Testing: Revolution or Evolution?". Brumfit, C.J.; Johnson, K. (eds.), *The Communicative Approach to Language Teachin*g. Oxford: Oxford University Press, 143-57.
Newbold, D. (2004). "Which English for a Modern Languages Faculty?" *Didattica delle lingue straniere: testing e multimedialità*. Venice: Cafoscarina, 89-105.
Newbold, D. (2009). "Co-certification: a New Direction for External Assessment?". *ELT Journal*, 63(1), 51-9.
Newbold, D. (2012a) "The Role of English Lingua Franca in a University Entrance Test". Ludbrook G.; Newbold D. (eds.), *English Lingua Franca: Contexts, Strategies and International Relations*. Venezia: Cafoscarina, 103-0.
Newbold, D. (2012b). "Local Institution, Global Examination: Working Together for a 'Co-Certification'". Tsagari, D.; Csépes, I. (eds.), *Collaboration in Language Testing and Assessment*. Frankfurt: Peter Lang.
Newbold, D. (2015a). "Assessing ELF in European Universities". In Vettorel, P. (ed.), *New Frontiers in Teaching and Learning English*. Newcastle: Cambridge Scholars, 205-6.
Newbold, D. (2015b). "Engaging with ELF In an Entrance Test for European University Students". In Bayyurt, Y.; Akcan, S. (eds.), *Current Perspectives on Pedagogy for English as a Lingua Franca*. Berlin, Walter de Gruyter and Company, 205-2.
Newbold, D. (2017a). "Towards a (painful) Paradigm Shift? Language Teachers and the Notion of Error" [online]. Num. spec., *Altre Modernità*, 118-132. DOI 10.13130/2035-7680/8306.
Newbold, D. (2017b). "Co-certification: a Close Encounter with ELF For an International Examining Board". *Journal of English as a Lingua Franca*, 6(2), 367-88.
Norton, J. (2005). "The Paired Format in the Cambridge Speaking Tests". *ELT journal*, 59(4), 287-97.
O'Regan, J.P. (2014). "English as a Lingua Franca: an Immanent Critique". *Applied Linguistics*, 35(5), 533-2.
O'Sullivan, B. (2002). "Learner Acquaintanceship and Oral Proficiency Test Pair-task Performance". *Language Testing*, 19(3), 277-95.
Orwell, G. (1949). *Nineteen eighty-four*. London: Secker & Warburg.

Papageorgiou, S. (2007). “Relating the Trinity College London GESE and ISE Exams to the Common European Framework of Reference: Piloting of the Council of Europe Draft Manual”. Unpublished final project report. London: Trinity College.

Paran, A.; Sercu, L. (eds.) (2010). *Testing the Untestable in Language Education*. Clevedon: Multilingual Matters.

Pickering, L.; Litzenberg, J. (2011). “Interaction as a Pragmatic Resource in ELF Interaction”. Archibald, A.; Cogo, A.; Jenkins, J. (eds.), *Latest trends in ELF research*. Newcastle: Cambridge Scholars.

Pitzl, M-L. (2005). “Non-understanding in English as a Lingua Franca: Examples from a Business Context. Idiom and Metaphor in ELF”. In: Maauranen A.; Ranta E. (eds.), *English as a Lingua Franca: Studies and Findings*. Newcastle: Cambridge Scholars, 298-322.

Pitzl, M-L. (2009). “‘We Should not Wake Up Any Dogs.’ Idiom and Metaphor in ELF”. Maauranen A.; Ranta E. (eds.), *English as a Lingua Franca: Studies and Findings*. Newcastle: Cambridge Scholars, 298-322.

Reinalda, B.; Kulesza-Mietkowski, E. (2005). *The Bologna Process: Harmonizing Europe's Higher Education*. Opladen: Barbara Budrich.

Richards J.C.; Rodgers, T.H. (1986). *Approaches and Methods in Language Teaching*. Cambridge: Cambridge University Press.

Roever, C. (2001). “Web-based Language Testing”. *Language Learning & Technology,* 5(2), 84-94.

Salmani Nodoushan; M.A. (2015). “Language Testing: the State of the Art. An Online Interview with James Dean Brown”. *International Journal of Language Studies,* 9(4), 133-43.

Schneider, E. (2007). *Postcolonial English*. Cambridge: Cambridge University Press.

Seidlhofer, B. (2001). “Closing a Conceptual Gap: the Case for a Description of English as a Lingua Franca”. *International Journal of Applied Lingustics*, 11(2), 133-58.

Seidlhofer, B. (2003). “A Concept of International English and Related Issues: from ‘real English’ to ‘realistic English’?” [online]. Council of Europe. URL `http://www.coe.int/t/dg4/linguistic/source/seidlhoferen.pdf` (2017-11-02).

Seidlhofer, B. (2011). *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.

Selinker, L. (1972). “Interlanguage”. *International Review of Applied Linguistics in Language Teaching*, 10(1-4), 209-32.

Sercu, L. (2010). “Assessing Intercultural Competence: More Questions Than Answers”. Paran, A.; Sercu, L. (eds.), *Testing the untestable in language education*. Clevedon: Multilingual Matters, 17-34.

Sewell, A. (2017). “Pronunciation Assessment in Asia’s World City: Implications of a Lingua Franca Approach in Hong Kong”. Isaacs, T.; Trofi-

movich, P. (eds.), *Second Language Pronunciation Assessment*. Bristol: Multilingual Matters, 237-55.
Sifakis, N.C. (2017). "ELF Awareness in English Language Teaching". *Applied Linguistics*, amx034. DOI 10.1093/applin/amx034 (2017-11-03).
Sifakis, N.C. (2014). "ELF Awareness as an Opportunity for Change: a Transformative Perspective for ESOL Teacher Education". *Journal of English as a Lingua Franca*, 3(2), 317-15.
Sifakis, N.C.; Tsantila, N. (eds.) (forthcoming). *ELF for EFL Contexts*. Clevedon: Multilingual Matters.
Soruc A., (2015). "Non-Native Teachers' Attitudes Towards English as a Lingua Franca". *Hacettepe University Journal of Education,* 30(1), 239-51.
Spolsky, B. (1976). "Language Testing: Art of Science?" Paper read at the 4th International Congress of Applied Linguistics, Stuttgart, Germany.
Spolsky, B. (1985). "The Limits of Authenticity in Language Testing". *Language Testing*, 2(1), 31-40.
Stapleton, P. (2005). "Evaluating Web-Sources: Internet Literacy and L2 Academic Writing". *ELT journal*, 59(2), 135-43.
Stoicheva, M., Hughes, G.; Speitz, H. (2009). *The European Language Portfolio: an Impact Study = Report of the 8th International Seminar on the European Language Portfolio* (Graz, 29 September - 1 October) [online]. Graz: Council of Europe. URL https://rm.coe.int/16804595a9 (2017-06-12).
Swain, M. (1985). "Large Scale Communicative Language Testing. A Case Study". Lee, Y.; Fok, A.; Lord, R.; Low, G. (eds.), *New Directions in Language Testing*. Oxford: Pergamon, 35-46.
Swales, J. (1990). *English in Academic and Research Settings*. Cambridge: Cambridge University Press.
Swales, J.; Feak, C.B. (2004). *Academic Writing for Graduate Students*. 2nd ed. Ann Arbor: University of Michigan Press.
Taylor, L. (2004). "IELTS: Some Frequently Asked Questions" [online]. *Research Notes,* 18, November, 14-17. URL http://www.cambridgeenglish.org/images/23135-research-notes-18.pdf (2017-10-23).
Taylor. L. (2005). "Washback and Impact: the View from Cambridge ESOL" [online]. *Research Notes,* 20, May, 2-3. URL http://www.cambridgeenglish.org/images/23138-research-notes-20.pdf (2017-10-23).
Tsagari, D.; Kouvdou, A. (forthcoming). "Towards an ELF-Aware Alternative Assessment Paradigm in EFL Contexts". Sifakis, N.C.; Tsantila, N. (eds.), *ELF for EFL Contexts*. Clevedon: Multilingual Matters.
Uysal, H.H. (2010). "A Critical Review of the IELTS Writing Construct". *ELT Journal,* 64(3), 314-20.
Vettorel, P. (2014). *English as a Lingua Franca in Wider Networking*. Berlin: De Gruyter.
Vettorel, P. (2015). "Primary School Teachers' Perceptions: Englishes, ELF and Classroom Practices – Between 'correctness' and 'communicative

Effectiveness'". Vettorel, P, (ed.), *New Frontiers in Teaching and Learning English*. Newcastle: Cambridge Scholars, 129-55.
Wall, D.; Horák, T. (2008). "The Impact of Changes in the TOEFL Examination on Teaching and Learning in Central and Eastern Europe: Phase 2, Coping with Change". *ETS Research Report Series*, 2(105).
Weir, C.J., Vidaković, I.; Galaczi, E.D. (2013). *Measured Constructs: a History of Cambridge English Examinations, 1913-2012*. Cambridge: Cambridge University Press.
Wells, H.G. (1933). *The Shape of Things to Come*. London: Hutchinson.
Weltens, B. (1989). *The Attrition of French as a Foreign Language*. Dordrecht Providence: Foris Publications.
Weltens, B.; Cohen, A.D. (1989). "Language Attrition Research. An Introduction". *Studies in Second Language Acquisition,* 11(2), 127-33.
Widdowson, H.G. (1978). *Teaching Language as Communication*. Oxford: Oxford University Press.
Xi Xiaoming (2010). "How Do We Go About Investigating Test Fairness?". *Language Testing,* 27(2), 147-70.

What are the challenges posed to English language examining boards by the phenomenal growth of English as a lingua franca?
This volume takes a critical look at existing international English language certification, which assesses test takers on the basis of the proximity of their performance to native speaker models. It describes a pilot project to develop an 'ELF aware' certification for higher education, and concludes that it may be necessary to introduce new assessment criteria to reflect the ability of users of English to communicate successfully in an international environment.

Università
Ca'Foscari
Venezia