

Rethinking English Language Certification

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

David Newbold

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

<p>TOEFL</p> <p>Part 1: Reading (60-80 minutes) The Reading section includes 3 or 4 reading passages. There are 12 to 14 questions per passage.</p> <p>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.</p> <p>Part 3: Speaking (20 minutes) 6 tasks. 2 independent speaking tasks. 4 integrated speaking tasks which develop ideas from reading and/or listening inputs.</p> <p>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).</p>

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 Pearson⁶ 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”?) 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