

Cambridge University Press

978-0-521-54250-0 - Assessing Academic English: Testing English proficiency

1950-89 - the IELTS solution

Alan Davies

Excerpt

[More information](#)

1 The 1950s and 1960s: the English Proficiency Test Battery

Introduction

In this volume I discuss attempts in the UK since about 1950 to represent proficiency in academic English by means of language test instruments. By proficiency in academic English I mean the ability to operate successfully in the English used in the academic domain. Such uses of English vary along several dimensions: the receptive–productive, the spoken–written, the general–specific, to name the most obvious. How far these variations in the use of academic English impact on proficiency is a matter of debate, as we shall see. Representing that proficiency in a language test requires a decision on the language content of the test, its language sample. As we explain in this volume, there are differing views on how to make that decision. In academic language proficiency testing, where the domain consists of large areas of the language, it is just not possible to test everything. And so the test constructor must sample the domain and face up to the question of how to make rational choices.

Alongside the increase in the numbers of second language English speakers over the years requiring to be tested for their academic proficiency, there are two related stories to tell. One is the debate about content and method of testing. The other concerns the growing attention to means of test administration, delivery and analysis. The first story reflects very closely the changing views about the nature of language and of language learning over this period: the emphasis has shifted to and fro between language form and language use. Thus, in the period before large-scale tests began, the focus (in the so-called traditional stage) was on language use (translation, essays, literature, summarising); then, in the so-called structural period, attention shifted to language form (grammar, phonemic discrimination, vocabulary). This was followed by the various communicative approaches, moving from the extreme of specific language domain use to the present compromise which still privileges language use but as a general approach to academic study. At all stages, some kind of balance has been maintained, such that in the traditional period there was some attention to, for example, vocabulary; in

Cambridge University Press

978-0-521-54250-0 - Assessing Academic English: Testing English proficiency

1950-89 - the IELTS solution

Alan Davies

Excerpt

[More information](#)

1 The 1950s and 1960s: the English Proficiency Test Battery

the structural period, to reading comprehension, and in the period of communicative tests some attention was usually paid to grammar. The second story reminds us that large-scale language testing is both a practical and an applied enterprise in which delivery and administration are as important as any more theoretical concerns.

This volume argues that language testing has matured over the period under discussion, moving from a dependence on fashions dictated by linguistics, applied linguistics and language teaching studies to an independent confidence in itself, as we see to some extent in tests such as the current International English Language Testing System (IELTS). The argument is therefore as much one about the development of a discipline as of the history of a particular test. What this means, in our view, is that when, as must happen, changes are made to IELTS, these will come about more as a result of development within language testing, less as a direct reflection of external influences.

The need for proficiency tests

Between the early 1950s and the early 1960s the number of overseas ('international') students from non-English-speaking countries in UK higher education institutions rose fivefold (from 12,500 to 64,000). In 2003/4 there were some 300,000 overseas students in UK higher education, over 95% of them from non-English-speaking countries.

Higher education in all English-speaking countries is currently experiencing a sharp increase in international (or foreign or overseas) students; for example, in 2001, in New Zealand, the University of Waikato had 1,000 students from the People's Republic of China (PRC) at the undergraduate and postgraduate level. Such students present special linguistic and cultural challenges to their receiving institutions.

The problems facing the institutions which have admitted these students include unprepared admissions officers, a shortage of interpreting staff in the international offices, the failure of institutions to provide adequate English language support even for those students who have tested out at the appropriate admissions level on tests such as the IELTS test or the USA Test of English as a Foreign Language (TOEFL), a lack of trained teaching staff to mount in-session English language courses, and a shortage of teaching space in which to conduct these courses. The influx in New Zealand is such that new students may wait months before they can be given the proficiency test (usually IELTS) they need for admission. In addition institutions lack understanding of the very different cultural expectations these students may bring with regard to what independent study means. The institutions may also lack awareness of these students' very real problems, problems of isolation, of culture clash, of inadequate language proficiency and consequently, very often, of unsatisfactory academic progress.

Cambridge University Press

978-0-521-54250-0 - Assessing Academic English: Testing English proficiency

1950-89 - the IELTS solution

Alan Davies

Excerpt

[More information](#)*Introduction*

These are not new problems: they may loom large in New Zealand in recent years because the country has (over) admitted international students without adequate planning. And because New Zealand is a small country, an influx of this order is very salient. But the problems, not always perhaps as dramatic, have been experienced in English-speaking countries over the last 50 years, indeed since the end of the Second World War. Future historians are likely to chart the rise and rise of English worldwide from the mid-1940s, when the USA became the major English-speaking super power. English then started to become the language that all educated people needed to possess, and the main vehicle for serious academic study in many disciplines, above all in the sciences (Crystal 1997). Before long, in both the UK and the USA (and later in Canada, Australia and New Zealand) it came to be felt that for the sake both of the students themselves and of the receiving institutions, some form of English language admissions test was necessary (Davies 1965). Otherwise, because of inadequate proficiency in English, the institutions and the students would waste time and effort. Looking back, it is interesting to wonder why institutions took such intervention as the provision of an English language test for granted. After all it would have been possible, though it might not have been very humane, to admit students with or without English language proficiency at the required level. The students would then be personally responsible for their own progress in English. The fact that institutions have not done so probably indicates that they do care about the welfare of their students – and/or that they are unwilling to accept large numbers of failures.

By the mid-1960s both the UK and the USA had English language proficiency tests in place. But that simultaneous test development was about the only similarity: in terms of test content and test implementation the practices were very different. In this volume we chart the history of the UK experience; while we refer in passing to the USA history, that is not our concern and in any case is described elsewhere (see for example Spolsky 1995). Our purpose is to describe how academic English language proficiency testing in the UK moved from the British Council subjective measure (Perren 1963b), through the English Proficiency Test Battery (EPTB) (Davies 1965) to the English Language Testing Service (ELTS) (Criper and Davies 1988) and then to the International English Language Testing System in 1989, which has itself evolved since the early 1990s (UCLES 2005). We refer in passing to two other British based English proficiency tests, the English Language Battery (ELBA) and the Test in English for Educational Purposes (TEEP). We shall also try to explain these changes from an applied linguistics perspective, in other words placing them in the context of educational, linguistic (including sociolinguistic and psycholinguistic), psychometric and methodological influences.

English language tests (or examinations) for second/foreign language learners were available in the UK from the early part of the 20th century. The

Cambridge University Press

978-0-521-54250-0 - Assessing Academic English: Testing English proficiency

1950-89 - the IELTS solution

Alan Davies

Excerpt

[More information](#)

1 The 1950s and 1960s: the English Proficiency Test Battery

University of Cambridge Local Examinations Syndicate (UCLES) made a start with its Proficiency examination in 1913 (see Weir and Milanovic 2003), followed by other bodies such as the Royal Society of Arts, Trinity College London, London Chamber of Commerce, and later the University of London, the University of Exeter, the Joint Matriculation Board (JMB) of the northern universities and the Associated Examinations Board (AEB). UCLES itself (now known as Cambridge Assessment) has increased its provision over the years until today. But with the exception of the JMB and the AEB, the purpose of these tests/exams has been to provide a (foreign) language certification in the English language itself rather than as a means of studying another subject through the English language medium.

The USA

In the USA the University of Michigan was a pioneer. Between 1946 and 1958 Robert Lado and his colleagues in the English Language Institute there (where Charles Fries was Director) produced some of the best known early tests in the field: the test of Aural Comprehension, the English Language Test for Foreign Students (still being revised for use today), and the Michigan Test of English Language Proficiency. These Michigan tests (and the TOEFL test which was later modelled on them) exemplified Lado's theory of language testing (Lado 1961). They were objective paper and pencil tests and targeted a student's problems. Lado maintained that a learner has language learning problems. The task of the tester is to find those problems and test them. Lado was more precise: a learner's problems, in his view, were essentially those of the major points of contrast between a learner's first language and his target language. (This was the 'contrastive language hypothesis' on which the technique of error analysis was based.) In addition to appearing to be linguistically based, Lado's theory demanded validation by an external criterion; and the procedures employed in validation were sophisticated. And in spite of his attachment to the contrastive language hypothesis, Lado would appear to have been a universalist, committed to the view that language learning is not basically different country to country, context to context.

Contrastive analysis was, it eventually became clear, an inadequate basis on which to determine learners' problems once the test was made available to students from different language backgrounds, and in any case only partially adequate as a complete explanation of second language learning. The theory of contrastive analysis was, as it were, overtaken by the impracticality of providing for all possible cross-language problems. Lado himself recognised the dilemma, conceding as early as 1950 that the rigours of contrastive analysis were, in the proficiency field, an ideal, unrealisable except for a few major languages:

Cambridge University Press

978-0-521-54250-0 - Assessing Academic English: Testing English proficiency

1950-89 - the IELTS solution

Alan Davies

Excerpt

[More information](#)

The USA

the task of preparing separate tests for all language backgrounds is so enormous that we may never hope to have such tests except for a limited few languages. A practical solution to this problem may be that of keeping separate norms for the various national groups of students that take the tests (Lado 1950:66).

This ‘practical solution’, which incidentally has been consistently practised by Educational Testing Service (ETS) for the Test of English as a Foreign Language (TOEFL), is more psychometric than linguistic. Certainly it has nothing to do with contrastive analysis since the test on which the separate norms are to be provided must already have been constructed on a non-contrastive or unilingual basis, that is on an assessment of some of the major problems generally involved in learning English.

Other US based tests in use at this time were the English Examination for Foreign Students (EEFS 1947, 1951, 1956) – a very long test, 5 hours in total testing time – and the Diagnostic Test for Students of English as a Second Language (Davis 1953), lasting 1 hour and targeting written English structures. It appears that this short test was meant to replace the under-used (and over-long) EEFS. Several other tests were developed at the American University Language Center (AULC) by Davis and colleagues (Croft, Freeman, Harris and Jones): the English Usage Test for Non-Native Speakers of English, the Rating Language Proficiency in Speaking and Understanding English (the Aural/Oral Rating Sheet) (Harris 1959) and A Vocabulary and Reading Test for students of English as a Second Language (Harris 1960). The English Usage Test and the Aural/Oral Rating Sheet are together commonly known as the AULC test (Harris 1961) and were mandatory for many CIA participants and other State Department grantees. In 1961 the AULC made their Listening Test available and then with TOEFL (Harris 1964), a true battery of tests appeared. The AULC tests were in many ways precursors of TOEFL; Harris was a very important contributor to this development and indeed it was at the American University Language Center that TOEFL was housed in the early days.

The first TOEFL was made up of five parts: Listening Comprehension, English Structure, Vocabulary, Reading Comprehension and Writing Ability (the link with earlier AULC tests is very obvious). What was noteworthy about TOEFL from the start was the nationwide co-operation both in preparing and administering the test: from an administrative point of view, TOEFL had the mark of being a great breakthrough in language proficiency testing. But this positive judgement did not hold for the linguistic analysis. For example, it was unclear on what basis the vocabulary items in Section 3 (Vocabulary) were selected, the method of questioning in Section 4 was traditional in the extreme (and the passages semi-literary), and Section 5 was more a test of style than of writing.

Cambridge University Press

978-0-521-54250-0 - Assessing Academic English: Testing English proficiency

1950-89 - the IELTS solution

Alan Davies

Excerpt

[More information](#)

1 The 1950s and 1960s: the English Proficiency Test Battery

An important (and early) contribution to proficiency test development in the USA was the Investigation of the Teaching of a Second Language. This study was praiseworthy for its attempt to ask basic questions right at the start (and formulate these as hypotheses): thus just ‘what does . . . reading knowledge mean when applied to the actual process of reading?’ (Agard and Dunkel 1948:26). Agard and Dunkel were concerned to evaluate the new experimental programmes which put a high premium on oral–aural proficiency. It should be remembered that the focus of their attention was on the acquisition of a foreign language in institutions which offered foreign language teaching: the tests of interest to them, therefore, were achievement rather than proficiency. They were interested in tests which would rank the native speaker and the non-native speaker on the same scale, a vain ambition they found.

By the early 1960s it had become clear that there was a widespread need for English language proficiency tests and that collaboration would make sense. In May 1961 (CAL 1961) an international conference on testing the English proficiency of foreign students, sponsored by the Center for Applied Linguistics of the Modern Language Association of America, the Institute of International Education and the National Association of Foreign Student Advisers, was held in Washington. The major recommendations of this conference (and of a second held in January 1962) led in the USA to the setting up of a national Council on the Testing of English as a Foreign Language: this in turn led to the initiation of the TOEFL programme under David Harris in 1963. One of the more illuminating contributions to the conference was made by the psychologist, J B Carroll, who, focusing on the need for external validity and for moving beyond contrastive analysis, stressed the importance of both a discrete and an integrative approach in the specification of language proficiency.

The British Commonwealth

At that Washington meeting, Norman Mackenzie reported on a number of English language tests in use in the British Commonwealth, mentioning relevant work in New Zealand, Canada, India and Central Africa and, most importantly, Australia. The Commonwealth Office of Education (COE) in Canberra had designed a test to assess the English proficiency of students attending Australian universities under the Colombo Plan. The rationale behind the COE test was explained by Coppock (1961): first, there was the native speaker standard (‘students should be able to comprehend readily the speech of native speakers of English and their speech should be readily comprehended by native speakers of English’); and, second, teachers’ experience of learners’ common errors in English (‘an analysis of experience in intensive English tutorials provided for some of these sponsored Asian students had

Cambridge University Press

978-0-521-54250-0 - Assessing Academic English: Testing English proficiency

1950-89 - the IELTS solution

Alan Davies

Excerpt

[More information](#)

The British Commonwealth

already revealed certain aspects of English pronunciation and sentence structure as being of particular difficulty for many of them’).

The intention was to tap four kinds of ability:

1. The ability to hear accurately and to produce accurately and distinctly the significant sounds of English, particularly those which are used in contrast, to signal difference in meaning.
2. The ability to understand English when spoken at normal speed with correct word stress, sentence rhythm and intonation in the stream of speech.
3. The ability to recognise, comprehend and use a reasonable range of English structures and vocabulary.
4. The ability to recognise, understand and use a reasonable range of common structures including idiomatic expressions in written English.

Daphne Keats, working in the University of Queensland, had found that in terms of predictive validity, reading comprehension in English was the most important skill for the university success of Asian students (Keats 1962). This was an important finding and has influenced later practice: since then all English proficiency test batteries have of necessity included a reading comprehension component. Keats also found that there was no linear relationship between increase in proficiency and time spent in-country: she reports that longer attendance at the University of Queensland did not raise test scores.

Also in 1961 a Conference on the Teaching of English as a Second Language was held at Makerere University College, Uganda. English language testing was one of the topics discussed. A number of papers were given on the topic including those by Coppock (1961) and Rackham (1961). The Stevens contribution was particularly illuminating. He argued that a language test could be devised according to a set of language test principles soundly evolved from a theory of language: he took the view that the most important principle in language testing was the foundation of a test on linguistic categories (Stevens 1961:11).

Stevens explained that tests of spoken English had been experimented with in West Africa, including components of reading, comprehension and conversation. Their significance (McCallien 1958, Stevens 1960) lay in their attempt to base their theoretical standard or norm not on Received Pronunciation (RP) but on Educated West African Pronunciation of English. These tests had been designed on the basis of a bilingual comparison between this model of West African pronunciation and various well-known local West African accents – taking us back to the contrastive analysis construct while offering a compromise between the extreme L1–L2 comparison and the common error approach. What the tests did was to focus on common problems for all West African second language learners of English, by

Cambridge University Press

978-0-521-54250-0 - Assessing Academic English: Testing English proficiency

1950-89 - the IELTS solution

Alan Davies

Excerpt

[More information](#)

1 The 1950s and 1960s: the English Proficiency Test Battery

hypothesising that speakers of various West African languages had problems in common when learning English (Davies 1965).

The UK

In the UK, the best known proficiency tests were the University of London Certificate of Proficiency in English for Foreign Students and the two Cambridge examinations, the Certificate of Proficiency in English (CPE) and the Lower Certificate in English (see Weir and Milanovic 2003 for more details of the history of CPE). Although all three were labelled ‘proficiency’, they are more properly regarded as ‘achievement’ (or ‘attainment’) since they were intended to be taken at the end of a rigorous course in English: the London examination mainly for foreign teachers of English, the Cambridge examinations for those who came to the UK primarily to study the English language. There is thus an important distinction between these examinations and those which are more usually described as Proficiency Tests in English as a Second/Foreign Language, since none of these was intended (even though they might have been so used) for students wishing to use their English in order to study some other subject. Neither London nor Cambridge published norms nor had they undertaken any validity studies. Both were somewhat traditional, giving a central place to written composition, insisting (in London) on a knowledge of phonetics and (in Cambridge) on a translation, though notice had been given that this was to become optional.

Teachers who had worked in ESL/EFL settings outside the UK found that on their return to live in the UK they were able to make use of that particular language teaching experience to work with the growing numbers of international students in the UK itself. One of these was George Perren who had carried out research on issues to do with English proficiency testing as a Simon Fellow at the University of Manchester. Perren had previously worked for a number of years in teacher training in East Africa and later became the first Director of the British Council’s English Teaching Information Centre (ETIC) in London. He wrote:

It was desired to construct and administer tests of English ability to West and East African students in Britain in order to discover:

1. to what extent their work in technical or academic courses in Britain was significantly handicapped by weaknesses in English;
2. in which aspects of English such weaknesses are most prevalent;
3. how weaknesses might best be overcome either by different teaching overseas, by preliminary courses in Britain, or by remedial courses which could be taken concurrently with other studies (Perren 1963a:2).

To this end he set up a battery of six tests which he applied both to foreign students and to native English-speaking students. The tests were:

1. A test of articulation (ability to produce the primary phonemes of English).
2. A test of phonemic discrimination (ability to distinguish between English phonemes).
3. A test of auditory comprehension of prosodic features (stress and pitch in a recorded dialogue).
4. A test of reading comprehension (single sentences).
5. A test of hearing (combining the features isolated in Tests 1–3).
6. A test of reading comprehension (two texts).

This experimental work laid the necessary foundations for subsequent academic English proficiency testing in the UK. It was based on sound theory, set out the features in isolation to be tested and constantly looked towards a satisfactory method of validating control over these features. Perren reported very high validity correlations for combined test scores with teachers' estimates used as the criterion. He concluded that his Tests 3 and 5 looked worthy of development, that a test of reading speed might well be added to future batteries, and, above all, that it was still not clear what constituted a truly valid assessment of an overseas student's English, what criterion, in fact, should be used to establish the validity of a proficiency test. This comment of Perren's has echoed down the years. Language testing remains caught between a rock and a hard place, the rock of achieving a valid and reliable test which meets the practical constraints of usability, and the hard place of specifying exactly what it is the test is meant to predict, the criterion. Like gamblers, proficiency testers predict: unlike gamblers they have no certain outcomes.

The British Council

The British Council, agency of the British Government charged with cultural and educational exchanges, had by 1954 developed a test instrument to measure the adequacy in English of the growing numbers of overseas candidates for official scholarships and fellowships. This test instrument was in fact a rating form, the Knowledge of English Form; it was issued to guide Council officers in making their subjective assessments in-country. The rating form consisted of a 4-point scale for each of the four skills, speaking, listening, reading and writing. By 1958 there was concern about the reliability of this method of assessment leading to the institution of an amended version which became known by its reference number, OSI 210, and the title changed to 'Assessment of Competence in English Form'. Among other

Cambridge University Press

978-0-521-54250-0 - Assessing Academic English: Testing English proficiency

1950-89 - the IELTS solution

Alan Davies

Excerpt

[More information](#)

1 The 1950s and 1960s: the English Proficiency Test Battery

changes were the addition of a sheet of instructions about procedures and suggestions to the examiner about materials, the replacement of the 4-point scale by a 10-point one and an instruction on the rating form about the level of proficiency to be attained in order to qualify for study in Britain. In 1961/2 the rating form was administered to 2,000 overseas applicants. But there was still considerable dissatisfaction (among Council officers and others) with this method of assessment (Perren 1963b).

One of the reasons for the concern in the British Council about the inadequacy of its English testing procedures was, as we have seen, the growing number of overseas students entering British universities. Equally important was their length of stay: many were short-term (for example three months), for most students rarely more than 12 months. In the circumstances, there was usually too little time for these students and visitors with inadequate English to improve sufficiently after arrival in the UK. It was different for PhD students, with 3–5 years to look forward to. Something extra was needed to help the short-term students and their receiving institutions. It was essential that they should have adequate English on arrival. Hence the need for a valid English language proficiency test.

So great was the dissatisfaction, both among Council officers overseas and receiving institutions in the UK with the existing British Council procedure for assessing English language proficiency, that in 1961 it was decided to set up an enquiry into the reliability of OSI 210, first by retesting in the UK on the same form and then by applying some of the existing American and Australian test materials to a sample of overseas students in the UK with a view to checking their applicability to British requirements.

‘The whole investigation would, it was hoped, draw attention to existing tests, clarify their shortcomings, and provide basic material which would help in the construction of new tests.’ The report on the investigation concluded that the problem remains of ‘producing a true sample of linguistic skill [. . .] which can be reliably scored’ (Perren 1963b:28). What should be aimed at was ‘functional load’ in communication, not contrastive analysis and not frequency. To that end, in 1962, a project funded by the British Council was set up in the University of Birmingham.

Relevant development work into academic language proficiency testing was already under way at the University of Edinburgh’s School of Applied Linguistics by Elizabeth Ingram. There the English Language Battery (ELBA) grew out of the common need to assess foreign students’ English, particularly at the university level (see Chapter 3 and Appendix 8.1). The American tests had been tried but found to be too American in content and too easy at the desired level. The Ingram battery consisted of nine tests:

1. Phoneme recognition.
2. Sentence stress.