Tests, Testers and Testing: The British Council and language testing 1941-2013

Professor Cyril J. Weir AcSS
Historical Literacy

Nescire autem quid antequam natus sis acciderit, id est semper esse puerum.

Not to know what happened before one was born is always to be a child.

—Marcus Tullius Cicero  *De Oratore* (XXXIV)
There is nothing new under the sun.
Is there a thing of which it is said, 'See, this is new'?
It has already been in the ages before us.

[Ben Sages, Chapter 1, verses 9-10]
Testing oral proficiency

Spolsky writes: ...pride of place for a direct measure of oral language proficiency is usually granted to the oral interview created by the Foreign Service Institute (FSI) of the US State Department Developed originally between 1952-56...

Assessment Scales

Glenn Fulcher describes how the first attempt to assess second language speaking can be found in the work of the Rev. George Fisher who became Headmaster of the Greenwich Royal Hospital School in 1834.

“In order to improve and record academic achievement, he instituted a “Scale Book”, which recorded performance on a scale of 1 to 5 with quarter intervals. A scale was created for French as a second language, with typical speaking prompts to which boys would be expected to respond at each level. The Scale Book has not survived.”
A Standardised Scale

Thorndike developed a standardized scale for the measurement of quality in the handwriting of children and also one for the handwriting of women in 1908 (Thorndike 1911, 1912). Instead of estimating a scale based simply on connoisseurship as was often the case in the United Kingdom, Thorndike took a large sample of student handwritten scripts and used 200 teachers to rank these scripts in order.

From the data he created a scale upon which he placed each script. He then provided a set of exemplar scripts at various levels to operationalise a scale from an absolute zero base with scale points defined, and their distances defined (1912: 295-299). Teachers were asked to compare their student’s scripts with those samples on the scale and identify the closest match to give the level.
Robeson, F. E. (1913) in *A Progressive Course of Precis Writing* quotes from the criteria for marking the precis test set by the Oxford and Cambridge Schools Examination Board:

...The object of the précis is that anyone who had not time to read the original correspondence might, by reading the précis, be put in possession of all the leading features of what passed. The merits of such a précis are (a) to include all that is important in the correspondence, (b) to present this in a consecutive and readable shape, expressed as distinctly as possible, and as briefly as is compatible with distinctness.
Assessment Literacy

Familiarity with how a construct was measured in the past provides us with a valuable perspective when developing 'new' tests or critiquing existing ones.
First Lesson

There is nothing new under the sun, but there are lots of old things we don't know. **Ambrose Bierce**, *The Devil's Dictionary*

Examination boards/test developers need to preserve/write their own histories and archive important documents for posterity…
British Council and testing in C20th

Stage 1  (1936 – 1959) **Traditional Approaches to Language Testing**

1941  The UCLES-British Council Joint Committee for overseeing Cambridge English exams

1945  The Diploma of English Studies (DES)

1954  *Knowledge of English Form* for screening applicants for UK Universities by BC abroad

1958  *OSI 210 Assessment of Competence in English Form*

Stage 2  (1960 onwards) **Psychometric-structuralist Approaches to Language Testing**

1963-5  *English Proficiency Test Battery (EPTB)*, (Davies PhD University of Birmingham 1965)

Stage 3  (1971 onwards) **Communicative Approaches to Language Testing**

1980  *ELTS Test* (Carroll 1978, Munby 1978)

1989  *IELTS Test* (Alderson, Clapham and Hargreaves)

2012  *APTIS* (O’Sullivan)
Stage 1 (1941 – 1959) Traditional Approaches to Language Testing: The Garden of Eden

1934 British Committee for Relations with Other Countries established (1936 becomes the British Council)

1941 The UCLES-British Council Joint Committee for overseeing Cambridge English exams

1945 The Diploma of English Studies (DES)

1954 Knowledge of English Form for screening applicants for UK Universities by BC abroad

1958 OSI 210 Assessment of Competence in English Form
1941-1993 The UCLES-British Council Joint Committee

The 1941 agreement between the British Council and UCLES makes it clear how close the working relationship was to be between the two organizations in running the Cambridge English examinations until their formal parting of ways in 1993.

From 1941 the British Council would be a close partner in the development of Cambridge English examinations for over fifty years.

It provided advice to Cambridge based on its growing experience and expertise in English Language Teaching, as well as helping to deliver the examination around the world.
A New D Level

- As well as offering advice on their existing examinations (LCE and CPE), the British Council took an active role in encouraging Cambridge to accept a new examination, the Diploma in English (at a level above CPE), into the Cambridge suite of English language examinations for overseas students. The Diploma of English Studies (DES) first appeared in December 1945.
- The origins for this exam can be found in the demand for an advanced, specialist examination in English Literature and background studies at post-Proficiency level in the British Institute in Cairo, Egypt.
- The Council was keen on such an exam ‘to be regarded abroad as a valuable qualification for foreign teachers of English’
The Birth of EAP Testing

In 1954 an attempt was made to regularise BC screening procedures for students coming to study in Britain by the introduction of a **Knowledge of English** 4 point scale for use in determining proficiency in each of the 4 skills. This “test instrument” was used to measure adequacy of English of overseas students competing for official awards. The form was issued to British Council officers to make subjective assessments in country.
Knowledge of English Form 1954

(2) Speaking of English

(a) Speaks fluently and accurately and is easily intelligible

(b) Speaks intelligibly, but is not fluent or altogether accurate

(c) Speaks haltingly, and is often at a loss for words and phrases

(d) Does not speak English
1958: OSI 210 Assessment of Competence in English Form

- Addition of a sheet of instructions about procedures and suggestions to the examiner about materials + instruction about level of proficiency necessary for eligibility to study in Britain

- 1961/2 the rating form was administered to over 2000 applicants but considerable dissatisfaction with it from BC officers

- 1961 enquiry into reliability of this test (see Perren 1963) because found to be unsatisfactory for receiving institutions, particularly in the case of students only staying for one year where there was very limited time for improvement for those with language deficit. A clearer picture of their abilities was required
Stage 2 (1960-)

Psychometric-structuralist Approaches to Language Testing: The Vale of Tears

1960 Robert Lado's visit to the UK  

1963-5 English Proficiency Test Battery (EPTB), (Alan Davies’ PhD University of Birmingham)
1965 English Proficiency Test Battery: The Davies Test

First administered one year after TOEFL, the EPTB was a traditional set of standardised tests in a multiple choice format, focusing on the receptive skills of reading and listening together with a knowledge of grammar and pronunciation.

**Reading component** of EPTB included subtests of grammatical structure, reading speed and comprehension (both science and non-science texts).

**Listening component** comprised subtests of phonemic discrimination (in isolation and in context), intonation and stress (in conversation), and comprehension (again, both science and non-science texts).
EPTB

Although the EPTB developers readily acknowledged the importance of writing and speaking skills, the practical problems of testing these skills (e.g. the requirement for skilled examiners), combined with the British Council’s need for a test which could be taken in a short period of time, meant that tests of speaking and writing could not at the time be included in the EPTB battery
Construct Validity: the sine qua non of language testing

Alan Davies (1984:68) wrote in the first issue of the journal Language Testing:

“…in the end no empirical study can improve a test’s validity... What is most important is the preliminary thinking and the preliminary analysis as to the nature of the language learning we aim to capture.”

Davies (1977b, p.63) had argued earlier:

“… it is, after all, the theory on which all else rests; it is from there that the construct is set up and it is on the construct that validity, of the content and predictive kinds, is based"
Stage 3 (1970’s-) Communicative Approaches to Language Testing: The Promised Land

1989 IELTS Test (Alderson, Clapham & Hargreaves)
2009 International Language Assessment (ILA) (O’Sullivan and Weir)
2012 APTIS (O’Sullivan)
Language as communication

- The 1970s saw further movement from teaching language as a system to teaching it as a means of communication as is detailed in Widdowson’s 1978 *Teaching Language as Communication.*

- The emphasis was no longer on linguistic accuracy, but on the ability to function effectively through language in particular contexts of situation.
ELTS

Davies describes EPTB as being “swept away” by the communicative revolution. He sees ELTS as taking up a strong communicative approach in its use of “real life examples of specific language uses.”

In the mid-1970s a project was established to develop a replacement for the EPTB which would address some of the problems the current test was facing and which could also take account of the significant changes that took place in the 1960s and 1970s in approaches to language learning and teaching.

The new communicative competence paradigm brought with it a much greater emphasis on the use of language skills in context. There was a shift towards a much more contextualised, language-in-use oriented approach to testing comprehension of spoken and written language.
The design of the new E.L.T.S. battery was said to reflect more closely a change in emphasis in language teaching from an atomistic approach to a broader sociolinguistic one. It also took account of developments in English for Specific Purposes (E.S.P.) where the concern was no longer with catering, in a single test, for the needs of all users regardless of the purposes for which the language is required.
The **ELTS** test offered a choice of six modules covering five broad areas of study of UK tertiary education, plus one non-specific area. The six modules were:

- Life Sciences
- Social Studies
- Physical Sciences
- Technology
- Medicine
- General Academic
Coverage

- G1 General Reading
- G2 General Listening
- M1 Study Skills
- M2 Writing
- M3 Individual Interview
1989 IELTS

In 1981 the British Council and UCLES EFL (now known as Cambridge English Language Assessment) commissioned Edinburgh University to conduct a validation study. Following this report (1986), the ELTS Revision Project, under the academic direction of Professor Charles Alderson of Lancaster University, was set up to oversee the design and construction of the revised test.
IELTS 1989

• From 1989 IELTS candidates took two non-specialised modules, Listening and Speaking, and two specialised modules, Reading and Writing. The non-specialised modules tested general English while the specialised modules were intended to test skills in particular areas suited to a candidate's chosen course of study. Specialised reading and writing modules (incorporating a direct link between the reading and writing activities) were available in three discipline fields which linked together related fields that had previously been separate modules in the ELTS battery, as shown below:

• Module A – Physical Science and Technology
• Module B – Life and Medical Sciences
• Module C – Business Studies and Social Sciences
1995 Revisions

• The field-specific Reading and Writing Modules A, B and C were replaced with ONE Academic Reading Module and ONE Academic Writing Module. Details of the research behind this change to the test design can be found in Clapham (1996) who concluded that the different subject modules did not appear justified in terms of accessibility to specialists.

• In addition, the thematic link between the reading and writing activities was also removed to avoid confusing the assessment of reading ability with that of writing ability.

• Measures were introduced to gather data on test performance and candidate background so that issues of fairness relating to test use and users could be more effectively monitored.
The SiLT Construct Project (2003-2013): the post communicative era. Beyond the Promised Land
The SiLT Construct Volumes: the post communicative era

These volumes established socio cognitive frameworks for evaluating the various components of construct validity in each of the broad skills areas of reading, writing, listening and speaking and then applied these to Cambridge ESOL examinations at 5 different proficiency levels. The focus was no longer just on language e.g. communicative functions, but also on the cognitive processes required for operating at different levels of proficiency.
Construct Validity

CONTEXT VALIDITY

COGNITIVE VALIDITY

TEST TASK

PERFORMANCE

SCORING VALIDITY
Cognitive Validity

- Do the cognitive processes required to complete test tasks sufficiently resemble the cognitive processes a candidate would normally employ in non-test conditions, i.e. are they **construct relevant** (Messick 1989)?

- Are the range of processes elicited by test items sufficiently comprehensive to be considered **construct representative** of real-world behaviour i.e., not just a small subset of those, which might then give rise to fears about construct under-representation?
Context Validity

• Cognitive processing in a writing test never occurs in a vacuum but is activated in response to the contextual parameters set out in the wording of the writing task.

• Context validity for a writing task addresses the particular performance conditions, the setting under which the task is to be performed (e.g. purpose of the task, time available, length required, specified addressee, known marking criteria as well as the linguistic demands inherent in the successful performance of the task)
Context Validity

• Are the characteristics of the test task an adequate and comprehensive representation of those that would be normally encountered in **the real life context**?
Scoring Validity

• How appropriate and comprehensive are the criteria employed in evaluating test output?

• How far can we depend on the scores which result from applying these criteria to test output?
British Council 2009 ILA

The International Language Assessment (ILA) is a placement test which helps the British Council to place new students at the appropriate level for their course of study in its learning centres around the world test.

Since the test was expected to distinguish learners at 7 levels (from CEFR levels A0 to C2) in a limited time, the plan called for an innovative design. It includes test papers for grammar (based on an empirically derived model of progression), vocabulary (using the British National Corpus as a base), and reading (employing a new model of reading progression based on Khalifa and Weir 2009).
Aptis is an innovative English assessment tool based on the socio cognitive approach. It is designed to help organisations and institutions identify standards of English and select the staff or students with the right skills.

Aptis aims to be responsive and can be customised for different domains such as business, travel and tourism, or education. Clients choose when, how and in what way to test, to ensure they get the information they need.

Core concepts include: Construct validity, test localisation, impact by design, economic value.
People and ideas have an important influence on English language testing in the UK
cyril.weir@beds.ac.uk

CRELLA: www.beds.ac.uk/crella