

Language Testing

<http://ltj.sagepub.com/>

Book reviews : Spolsky, B. 1995: Measured words: the development of objective language testing. Oxford: Oxford University Press. xiv + 402 pp

Antony John Kunnan

Language Testing 1996 13: 116

DOI: 10.1177/026553229601300108

The online version of this article can be found at:

<http://ltj.sagepub.com/content/13/1/116>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Language Testing* can be found at:

Email Alerts: <http://ltj.sagepub.com/cgi/alerts>

Subscriptions: <http://ltj.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://ltj.sagepub.com/content/13/1/116.refs.html>

>> [Version of Record](#) - Mar 1, 1996

[What is This?](#)

Spolsky, B. 1995: *Measured words: the development of objective language testing*. Oxford: Oxford University Press. xiv + 402 pp.

This pioneering study, with particular emphasis on American and British institutional-industrial contexts, was written, in Spolsky's words, 'to widen the current perspective by looking at one aspect of the field of language testing in its historical, sociological, and political career' (p. 2). The author argues for a book of this kind because he states that there is both a time restriction and a limitation in understanding in recent language testing books. The time restriction, Spolsky states, is the question of how far back the history of the field should go because recent books have been written ahistorically, 'as if the field rose Venus-like out of the waves of applied linguistics sometime after 1960' (p. 2). The limitation in understanding persists because, Spolsky claims, advances in methodology and theory have been seen as the driving force behind the development of language teaching and testing and not 'some of the external, non-theoretical, institutional, social forces that, on deeper analysis, often turn out to be much more powerful explanations of actual language teaching practice' (and testing practice) (p. 2).

This 20-chapter book is a historical narration structured into two parts: Part 1 covers the history of the objective language test and Part 2 the TOEFL and the rise of the transatlantic English testing industry (presented in roughly chronological order). Part 1 begins with brief notes on several historical language tests: 'the first political modern language test', the Shibboleth test (as recorded in *The Book of Judges* in *The Bible*), which was 'a single-item, objective, oral, phonological test, individually administered' (p. 15) and the 42,000 who failed it were slaughtered on the spot; the Chinese Han dynasty (201 BCE to 8 CE) written examinations on classical Confucian doctrine and their anonymous marking; the 18th century French examinations for the Parisian *lycees*, the *baccalaureat*, and the *grandes ecoles*; the 19th century Cambridge public oral disputations called the *tripos*; the first British Indian Civil Service examinations for service in India; the Oxford and Cambridge local examinations for secondary school students; and Harvard's annual oral examinations in Latin, Greek, Hebrew, 'Rhetoricke, Logike and Physicke'. Thus, Spolsky's early sweep of examinations shows that from early times to more recent times examinations provide, in Foucault's words, 'a normalizing gaze, a surveillance that makes it possible to qualify, to classify and to punish' (cited by Spolsky, p. 15).

Spolsky records that by the end of the 19th century, written examinations in a traditional form had become stable and entrenched in Europe and in the US 'as essential instruments in the control of education and in the certification of qualification for employment and further education' (p. 22). Key professionals during this period Spolsky identifies are British statistician Edgeworth, who published papers on the effect of error and chance in written examinations; American scholar Cattell, whose interest

in experimental psychology was questioned by Galton; and British statistician Spearman, whose field of 'correlational psychology' was fast developing. Spolsky also notes Thorndike, whose attempts in the 1910s were the earliest to write objective and standardized tests. These attempts led to a published call for 'objective psychological testing' from a group of American teachers of modern languages. An example of this 'objective' testing was the Army Alpha tests used just after America entered World War I. Spolsky argues that these founding thinkers and their concerns made it possible for the modern language test to grow in an atmosphere of psychology and psychometrics.

The earliest test intended for ESL in the US was the English Competence Examination prepared by the College Entrance Examination Board in 1930. This test, Spolsky points out, was indirectly the result of racial and immigration concerns among certain Americans. Among them was psychometrist Brigham, who gave evidence in Congress that persons of non-Nordic descent would contaminate the American gene pool. As a result, Spolsky notes, two commissions were set up to study the feasibility and type of examination. In 1929, the detailed outline of the test was formed: four one-paragraph passages for reading followed by true-false statements, a longer debatable passage from Mill or Lowell for reading, a direct dictation, an oral test, and a composition. The test was discontinued in 1935 when the funding from the Carnegie Endowment for International Peace ran out.

In Britain, testing the English of foreigners started earlier than in America, perhaps as early as 1863. According to Spolsky, the formal entry of the University of Cambridge into testing was not until 1913 when it began to administer the CPE examination. The papers for the examination included phonetics, literature, translation (into French, German, Italian and Spanish) and economic and cultural knowledge. In 1937 the test received a boost when it was recognized as the equivalent of the standard of English required of all aspiring students to Cambridge University, British or foreign. The Lower Certificate examination (now the First Certificate in English) was started in 1939 and consisted of a dictation, a two-hour composition and language paper, a literature paper and an oral examination.

Spolsky documents the rising academic status of language testing, marked by research studies conducted at various American universities in the 1930s. Topics examined included teachers' grades and scores on vocabulary and reading, grading of reading passages, the 'citation exercise' in reading, short-answer and essay-type questions, contextualization of vocabulary, testing techniques, relationships between vocabulary and reading, effect of translation tests, scale development, and divergence of requirements for prospective language teachers. These research studies look astonishingly similar to some of the current work conducted at universities worldwide. At about this time, a consumer protection oriented movement was founded by Oscar Buros to evaluate modern standardized tests. In 1938 his first cooperative review volume (called *The 1938 or First Mental Measurements Yearbook*) was published by Rutgers University and

included reviews of ten language tests.

World War II saw language testing in America face a new challenge in the form of oral language testing. Spolsky discusses the Army Specialized Training Program, created in December 1942, which began officially in April 1943 with the objective of developing speaking proficiency in a target foreign language. At about this time, Walter Kaulfers proposed the first comprehensive plan for oral language testing: a uniform, standardized, test with 'tone-control talking machines' (p. 104). Spolsky reports on Kaulfers' measurement scale that went from 0 to 20, with descriptors for five ratings that look as user friendly as some of today's scales. Though his proposals were novel, they were not used until the Foreign Service Institute of the US State Department developed the first direct measure of oral language proficiency sometime between 1952 and 1956.

After the war, a second attempt at developing a test of English for non-native speakers began. This time, Spolsky points out, the initiative came directly from the US State Department. Responding to this interest, the College Board's committee proposed and developed the *English Examination for Foreign Students: Including a Test of Non-Verbal Reasoning*. It was ready for use in 1947 and had seven separate scores from the following sections: reading comprehension, pronunciation, general and scientific vocabulary, auditory comprehension, composition and non-verbal reasoning. This test, too, was abandoned in 1951 after funding for the project ran out.

As a sign of professional development, Spolsky reports that this period saw the first two Ph.D. dissertations on language testing: Villareal's *Test of Aural Comprehension* in 1947 and Lado's *Measurement in English as a Foreign Language* in 1949. Lado also completed three forms of the *Test of Aural Comprehension in English as a Foreign Language*, which was published by the University of Michigan where he was working under the direction of Fries. This was followed by the *English Language Test for Foreign Students* (also known as the Lado Test). Lado's language test construction philosophy, Spolsky summarizes, was an attempt to combine a theoretically-based structural linguistics and a psychometrically-based testing method.

In Part 2 of the book, Spolsky chronicles the idea of the TOEFL, its birth, and the English language testing industry centered around Educational Testing Service (ETS). Spolsky first briefly discusses pre-TOEFL tests: the *Michigan Test of English Language Proficiency* (developed at the University of Michigan by Upshur, Palmer and Harris); the *Diagnostic Test for Students of English as a Second Language* (developed at the American University, Washington, DC); the *Fundamental and Technical English Proficiency Tests* (developed at the Lackland Air Force Base, Texas); and the University of London's *Certificate of Proficiency in English for Foreign Students*. However, Spolsky only briefly hints at the English language testing efforts in Commonwealth countries like India (especially the involvement of 'an American expert' at the Central Institute of English and Foreign Languages, Hyderabad) and West Africa (the innovations suggested by Grieve).

The founding of ETS in 1948, Spolsky argues, 'constituted a decisive stage in the march towards industrialized testing' (p. 155). Established as a non-profit, non-stock corporation, ETS operated as the agent of client boards such as the College Entrance Examination Board, the Graduate Record Examination Board and the TOEFL Policy Council. The emergence of TOEFL in the new industrial context of ETS is the focus of several chapters in Part 2 of the book. Spolsky begins with the May 1961 Conference sponsored by the Center for Applied Linguistics which 'adopted the goal of establishing an omnibus battery testing a wide range of English proficiency that could meet the needs of all US colleges and universities who were considering the admission of foreign students' (p. 217). Carroll's keynote speech at this conference (many professionals mark this paper as the beginning of modern history of language testing; obviously Spolsky sees the history differently) called for integrative and communicative testing as opposed to the discrete-point tests that Lado and others had emphasized. Spolsky notes that Carroll went on to propose that the test should have four subtests: control of English structure, auditory comprehension, vocabulary and reading comprehension, and writing ability. More than four decades later, this structure is astonishingly close to that of today's TOEFL.

In May 1962, the National Council on the Testing English as a Foreign Language invited ETS to develop this test. From then until July 1963, Spolsky notes, the preparations and intense negotiations between the several agencies who were interested in influencing the testing program (namely, the College Entrance Examination Board, the Carnegie Foundation, the Center for Applied Linguistics, the National Association of Foreign Student Advisors, and the Institute of International Education) were political, financial and even personal. In three chapters of reports on meetings and institutional politics, Spolsky shows how institutional, political, commercial and personal factors dominated scientific and theoretical considerations in language testing. (This is no different from, for example, science politics; see Diesing, 1991).

In the final chapters, Spolsky documents the growth of the TOEFL industry in Princeton, the 1975 agreement between the College Board and ETS by which ETS became the 'sole executive agency' for TOEFL, recent research and development activities at TOEFL, the Test of Spoken English, the Test of Written English and the TOEFL 2000 project. Left to the end is a brief summary of recent developments in the Cambridge examinations (the FCE and the CPE), the British Council English Proficiency Test Battery, and the International English Language Testing Service test.

Summarizing the book and its goal, Spolsky states: 'If students of language testing were required to read Edgeworth, Brigham, Henmon, Kaulfers, Roach, Lado, Carroll, and Grieve, as students of physics are taught the historical development of their field . . . I am certain that we would find much more modesty, and many more useful advances' (p. 352-3). His final points are more radical and appear quite suddenly: 'the fundamental flaw of objective modern language testing has been to

presuppose that a language proficiency is measurable and unidimensional ... Language proficiency is more like pain, the external assessment of which has many analogous properties: it varies from person to person, from context to context, and can only be inferred from self-report and observation of impaired performance' (p. 358).

This book is a major effort at documenting and reconstructing the development of language testing through official documents and memos, tests and manuals, conference and meeting reports, and even personal letters. It ought to be read as an historical narration and evaluated as such. Primarily, it succeeds in following the idea that history is seen as a comprehension of plots, and theories are in fact plot summaries (Veyne, 1984). Specifically, he delineates one clear plot summary through this political history: The contests of power among the main players (for example, during the founding of the TOEFL) set aside the idea that the main players always had the noblest intentions in mind. Clearly, this is the most meritorious aspect of the book.

The book is also a traditional narrative in the sense it articulates the traditions and practices of the field so that language testers can find their way around the recent past and to connect current work to that done earlier. By the same token, it is not a critical narrative in the sense that it is an anti-story, a story that problematizes events or alters given ideas of continuity (Rusen, 1987) by, for example, opening our minds to an alternative understanding of events. It is also not a genetical narrative, in the sense that it gives new direction to the language testing story, for example, a subaltern view of testing from the item writers', test takers', or a colonized person's point of view or a deconstruction of historiography (Bhabha, 1983, Spivak, 1988).

One problematic aspect of the book is related to whether a fair overall representation has been given to the central events and players in this narrative. Spolsky's non-uniform degree of generality and degree of simplicity presents a problem. For example, while the founding of the TOEFL program is accorded great detail and complexity, other test programs, such as the Cambridge CPE, are accorded a level of generality and simplicity that makes the comparative reading of power contests and non-professional interests problematic. There are two additional points of lesser importance that warrant attention. First, the subtitle has the word 'objective' in it, though Spolsky discusses both objective and subjective tests (in the usual simplistic manner language testing books use the two terms). Should the word then not have been in the title, or does the word represent a different viewpoint? Second, the title ought to have indicated that this study is restricted to developments in the US and Britain. The book does not include developments in British Commonwealth countries, Canada, Australia, New Zealand, or other regions where language testing, perhaps with different melodies, might have flourished.

To conclude, this is an excellent first history of the development of language testing, a history that needs to be read and discussed so that we can evaluate whether the study accurately maps the real narrative and narrative structure of the lived past.

References

- Bhabha, H.** 1983: Difference, discrimination, and the discourse of colonialism. In Barker, F. *et al.* *The Politics of theory*. University of Essex, Colchester.
- Diesing, P.** 1991: *How does social science work? Reflections on practice*. Pittsburgh, PA: University of Pittsburgh Press.
- Rusen, J.** 1987: Historical narration: foundation, types, reason. *History and Theory*, 26, 4, 87–97.
- Spivak, G.** 1988: Subaltern studies: Deconstructing historiography. In Guha, R. and Spivak, G., editors *Selected subaltern studies*. New York: Oxford University Press.
- Veyne, P.** 1984: *Writing history: Essay on epistemology*. Translated by M. Moore-Rinvolucri. Middleton, CT.

Antony John Kunnan
California State University, Los Angeles

Hammerly, H.M. 1994: *French immersion test (FIT)*. Lexcel Enterprises, P.O. Box 64522, Como Lake Postal Outlet, Coquitlam B.C., V3J 7V7, Canada.

H.M. Hammerly (Simon Fraser University) has distributed a test intended for Anglophone students who have completed at least five years of French immersion programmes in Canada. His intention is to use the test to validate his claims that immersion education has failed to produce adequate language proficiency (see Hammerly, 1991). Since the potential impact of his findings could be great, it seems imperative to give a close look to the test on which claims will be based.

I have reviewed Hammerly's *French immersion test*, item by item, and offer this appraisal of it. The test has two parts to it. The first is called 'Noun gender', and consists of a list of 20 French words with their English equivalent in parenthesis. The respondent is to supply the appropriate indefinite article. Each item is worth one point. The second part of the test is called 'Grammar and vocabulary', and consists of 50 English sentences that the respondent is to translate into French. Each sentence is worth two points, and one point is deducted for each error in grammar or vocabulary. Half a point is deducted for each error in spelling or punctuation that does not affect grammar or vocabulary.

In the first section, there are words for which the gender is to be determined (words such as *mot*, *avion*, *famille*, *rue*, *chanson*, *science* and *fois*). Gender is a notorious problem for all learners of French, not just those in immersion programmes. Hammerly would be well advised to assess the control of gender among learners of French in all types of