牛津应用语言学丛书

Measured Words

客观语言测试

Bernard Spolsky

上海外语教育出版社

牛津应用语言学丛书

客观语言测试
Measured Words

Bernard Spolsky 著

上海外语教育出版社

上海市版权局
著作权合同登记章
图字:09－1999－025号

牛津应用语言学丛书

Measured Words

客观语言测试

Bernard Spolsky 著

上海外语教育出版社出版发行
（上海外国语大学内）
深圳中华商务联合印刷有限公司印刷
新华书店上海发行所经销

开本 880×1187 1/32 13印张 567千字
1999年4月第1版 1999年12月第3次印刷
印数:1500册

ISBN 7－81046－571－6
H·582 定价:26.00元

Oxford University Press
Walton Street, Oxford OX2 6DP

Oxford New York Athens Auckland
Bangkok Bombay Calcutta Cape Town
Dar es Salaam Delhi Florence Hong Kong
Istanbul Karachi Kuala Lumpur Madras
Madrid Melbourne Mexico City Nairobi
Paris Singapore Taipei Tokyo Toronto

and associated companies in
Berlin Ibadan

Oxford and *Oxford English* are trade marks of
Oxford University Press

ISBN 0 19 437201 4

First published 1995
Second impression 1996

Set by Wyvern Typesetting

This edition of *Measured Words*,
originally published in 1995, is published
by arrangement with Oxford University Press.

本书由牛津大学出版社授权上海外语教育出版社出版。

出 版 前 言

这是一部讨论语言测试问题的学术专著，作者 B·斯伯尔斯基是以色列巴依兰大学的英语教授，从事语言测试研究多年，在这方面有丰富的经验和丰硕的成果。本书汇集了翔实的史料，介绍了客观语言测试在20世纪的发展，为研究人员了解英语测试的历史以及展望未来可能的发展前景提供了可贵的背景资料。

全书分为两大部分。第一部分题为“客观语言测试的历史”，论述了测试的驱动力，回顾了1913－1935年间出现的新型语言测试，心理测量学的传播及主导地位的取得，30年代语言测试技术的新发展，二战期间服务于国家利益的语言测试之诞生，语言测试的专业化和1954－1965年间语言测试日趋成熟的过程。作者通过回顾语言测试发展的历史，揭示了测试走向客观的必然趋势。第二部分题为“TOEFL考试以及美国英语测试的兴起”。这部分中作者回顾了TOEFL考试的酝酿、诞生及发展，论述了公共测试和学院测试的兴起，并分析了机构、政治和经济等因素对测试理论的影响。

在书中，作者指出自第一次世界大战以来，心理测试原理及实践就逐渐风靡世界客观语言测试领域并被广泛接受。然而所谓的客观测试，包括其最成功的典范——TOEFL考试，真是纯粹的客观测试吗？作者认为几乎现代所有重大的测试项目都表明答案是否定的；尤其是TOEFL考试更清楚地表明，在影响语言测试发展的方方面面中，往往是经济和政治因素，而非测试理论和技术，扮演了更为重要的角色。作者写作此书的目的就是要扩展语言测试研究者的视野，希望他们能够将语言测试的发展置于特定的历史、社会和政治背景中加以分析。语言测试具有社会性，研究者决不能脱离其存在的环境孤立地进行研究。

此外，作者还认为人们对测试发展的认识中还普遍存在着误区，例如：认为测试理论和方法的进步主宰着测试实践的发展。然而情

况并非如此。如果这样,我们就无法解释为什么测试学上理论的突破往往无法被立即接受、被完全付诸实践。实际上,往往是对外部的、非理论的、机构性和社会性的因素进行的深入分析,方能更有力地解释实际语言教学和测试的发展。

因此,作者的结论是:在语言测试教学领域中,学者们不能只依靠专业理论,而是还要考虑到包括经济学、政治学和社会学在内的各种因素,才能更有效地把握测试的背景。

本书论述的内容并非语言测试理论和实践的全史,而是高度机构化、产业化的测试的局部历史。这是一部人类如何发展那些着重可靠性、高效性和商业可行性的测试的历史。

本书是一本颇有分量、富有启发性的语言测试专著,适用于攻读应用语言学和外语教学理论的硕士生和博士生,可作为他们的教材或必读参考书;对于从事语言测试发展研究和应用的学者,也有相当高的参考价值。

本社编辑部

Acknowledgements

The opportunity to work on this book came with a sabbatical leave from Bar-Ilan University. Once again, I am deeply grateful to the institution for its wise and generous sabbatical policy, and to my departmental colleagues and students who graciously and efficiently dealt with my absence from teaching. In particular, I acknowledge the extra burden placed on doctoral students who accepted electronic rather than personal guidance.

Atmosphere and facilities to work were provided by a Mellon Fellowship at the Institute of Advanced Studies of the National Foreign Language Center at the Johns Hopkins University, Washington DC, just a block from the building where TOEFL started out. I want to express my deepest thanks to the director of the Center, Professor Richard Lambert, and his colleagues for their hospitality and intellectual stimulation, and to the staff members of the Center for their constant and willing help and support.

The three libraries where I did the bulk of the research were the Library of Congress, Georgetown University Library, and the George Washington University Library. For access to Georgetown University Library, with its vital collection transferred from the Center for Applied Linguistics, I am grateful to Dean James Alatis, with whom I am also happy to acknowledge more than twenty years of friendship. I especially thank him for handing over to me, from his personal papers, a number of key documents from the early years of TOEFL which he rescued.

During the course of this study, I have been permitted to read and copy archival material at the Educational Testing Service, the Ford Foundation, the University of Cambridge Local Examinations Syndicate, and the College Board. Without these institutions' willingness to open up these records, much of this book would have remained speculation. While I may have bared some of the skeletons in their cupboards, I hope my admiration for their devotion and service to the field of language testing continues to shine through.

A number of other scholars have also searched their own archives or memories. I thank in particular: John B. Carroll, Leslie Palmer, David Harris, Sydney Sako, Robert Lado, and John Roach, all of whom deservedly have starring roles in parts of the story that follows. I am grateful to William G. Shephard for valuable leads to written and unwritten archives at the University of Cambridge Local Examinations Syndicate.

In my struggles to fit all the fascinating or important data I found into a book, I was especially grateful to Henry Widdowson and a number of anonymous readers who challenged or encouraged me to overcome my reluctance to sculpt and pare a large amount of data into a readable form and manageable size. I also acknowledge, with deep gratitude, John Carroll's reading of an early draft and his effort to help me correct some of the errors and biases in it, and Alastair Pollitt's perceptive commentary on a near-final draft that has enabled me to grasp more clearly what I am trying to say.

While I may have for a little been tempted to look elsewhere, I appreciate the continued association with Oxford University Press and the encouragement, co-operation, and efficiency it has guaranteed from Cristina Whitecross and her colleagues. I wish in particular to thank Antoinette Meehan for painstaking editing of a complicated manuscript.

I have been particularly fortunate that my work on this book has been paralleled by my wife's own studies of the relevance of scepticism to understanding the literature and painting of the Renaissance, and our conversations and shared readings have played a major role in my own clarification of the issues I have been working on. Although we have not yet sat down to the joint authorship that more than thirty years of sharing family and careers might have been expected to induce, in writing this book, in particular, I have sensed the intellectual overlap that matches other common values and pursuits.

The author and publisher are grateful to the College Board, the Educational Testing Service, and the Ford Foundation for permission to reproduce material held in their archives.

Preface

When, somewhat early in my career in language testing, I was first honoured with an invitation to be a plenary speaker at an International Congress of Applied Linguistics, the dignity of the event so affected me that I allowed myself to make *ex cathedra* pronouncements on the history of the field. Although my speculations then seem to have been quite well received and are widely cited without complaint, I have from time to time returned with some anxiety to the paper that I presented in Stuttgart and wondered if the notions in it would stand up to more careful scrutiny, or if the data would demolish the theory. Thus, this book is both an exploration and an expiation.

It is also a small contribution to the professionalization of a field in which it has been a pleasure to work. Without knowing our past, we have all enjoyed regular discoveries of round objects or other similar novelties. I hope I will not be felt to be spoiling the fun of colleagues who have provided such a sympathetic fellowship. I acknowledge a debt, as any historian of language testing must, to my predecessors in the field and to the countless people who suffered or gloried in the tests that they gave or inspired. This book is dedicated to the students who, over my years of teaching, have provided me with opportunities to try out my burgeoning ideas, and justified the paid employment that has allowed me to continue my research.

Bernard Spolsky
Jerusalem 1994

Acronymns

ASTP	Army Specialized Training Program
CEEB	College Entrance Examination Board
CITO	(Centraal) Instituut voor Toestonwikkelling (National Institute for Educational Measurement)
CPE	Certificate of Proficiency in English
ETS	Educational Testing Service
EUROCERT	English Proficiency Certification Program
FCE	First Certificate in English
FSI	Foreign Service Institute
(I)ELTS	(International) English Language Testing Service
IIE	Institute of International Education
MLA	Modern Language Association of America
TOEFL	Test of English as a Foreign Language
TSE	Test of Spoken English
TWE	Test of Written English
UCLES	University of Cambridge Local Examinations Syndicate

Contents

1 Prolegomena

Read this carefully before starting the test

'There is no blessing to be found', the Babylonian Talmud remarks (in Treatise *Ta'anit*, 8B), 'in something that has been weighed, or in something that has been measured, or in something that has been counted'. None the less, the last century has seen a determined effort to weigh, gauge, and count not just obvious and visible physical objects but also unseen forces and conjectured abstract concepts. The flowering of modern scientific language testing has been one facet of the attempt to measure an aspect of human ability, and a further application of the rationalistic Cartesian search for certainty to an area perhaps better left for a healthy humanistic scepticism.

Since the days of World War I, psychometric principles and practices have come to dominate the testing of foreign language proficiency, and a movement that initially blossomed in the United States has spread throughout the world. As long as testing was confined to helping students learn or to determining the qualifications of individuals seeking employment, there was a strong ethical case to be made for it, as the ends justified the means. But, from its beginnings, testing has been exploited also as a method of control and power—as a way to select, to motivate, to punish. The so-called objective test, by virtue of its claim of scientific backing for its impartiality, and especially when it operates under academic aegis and with the efficiency of big business, is even more brutally effective in exercising this authority. Clothed in the respectability of psychometric objectivity, and with powerful institutional support, the Test of English as a Foreign Language (TOEFL) was able to capture the market and become industrialized. It is only by taking full account of the institutional or political context that one can appreciate how the psychometric controversies have distracted attention from more serious social (or anti-social) motivations and impact.

This point can be illustrated by any of a number of modern language testing programmes. One might choose the pioneering work of Henmon and his associates in the late 1920s, the development of the Foreign Service Institute Oral Interview in the 1950s, the Modern Language Association Cooperative and Proficiency tests created in the 1960s, the British work resulting in the International English Language Testing Service test battery in the last few years. Interesting as all these are, I found that it is the early history of TOEFL that best demonstrates the tendency for economic and commercial and political ends to play such crucial roles that the assertion

of authority and power becomes ultimately more important than issues of testing theory or technology.

My main intention in this study, then, has been to widen the current perspective by looking at one aspect of the field of language testing in its historical, sociological, and political context. Most recent books on language testing have been written ahistorically, to put it politely, as if the field rose Venus-like out of the waves of applied linguistics sometime after 1960.[1] While this book was in manuscript form, I gave a copy to a colleague who was just preparing to teach a course he had labelled 'A history of language testing'. His first surprised comment was that he had planned to start with 1961, a year I reach half-way through this book. Clearly, one writer's history is another's pre-history.

Not only are our horizons restricted, but there has been another limitation in our understanding. Most historical references read as though advances in methodology and theory had been the driving force behind the development of language teaching and language testing. We regularly talk and write (I know because I have done it) in terms of progress and periods. We see the Audio-Lingual Method as the result of the application of structural linguistics and Skinnerian learning theory. We interpret the cognitive approaches as products of the theoretical revolutions of transformational generative grammar. We regard the notional-functional syllabus as related to theories of pragmatics and communicative competence. We lament the failure of British applied linguists to agree a model in place of Munby. We propose three periods of language testing, one traditional, a second modern or psychometric-structuralist, a third as post-modern or psycholinguistic-sociolinguistic (Spolsky 1977, 1981a).

With such a restricted outlook, almost new historical in its egocentricity, we have difficulty in recognizing why a theoretical breakthrough (especially if it is one that we have just proposed) does not immediately win absolute acceptance and total implementation. Only recently has a handful of scholars—Richards (1984), Pennycook (1989, 1990), and Phillipson (1992)—forcibly diverted attention to 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.

Without this reminder, we too easily forget, for instance, the enormous power of institutional inertia: it is much easier to think up reasons against a change than to provide arguments in its support. There are, of course, important and valuable reasons why institutions resist change. They are explainable as much through their history as through the logic of their present operation. They function because of their constancy, their imperviousness to other than minor changes. It is much easier to come up with a new theory than to find a way of fitting its implementation into an existing establishment. The decisions more often represent political compromise

than theoretical principle. In debate, statements of principle serve as rhetorical devices or rallying cries rather than as the basis for empirical proof or logical argument.

A clearer account of a field depends on willingness to look carefully not just at the history of the ideas that underlie it, but also at the institutional, social, and economic situation in which they were and are actualized. In the field of language pedagogy, for instance, the development of language laboratories was a commercial offshoot of innovative technology rather than an answer to theoretical needs. The enormous growth of the demand for English language teaching throughout the world is explained, at least in part, by the hugely profitable language teaching industry and related publishing (and testing) businesses, where new theories hold interest as sales pitches. The move towards a European economic community clarified, as no theoretical approach would have done, the requirement to define language teaching goals as precisely as did the notional-functional syllabus.

In the study of fields like language testing and teaching, scholars need to be ready to draw not just on the obvious theoretical disciplines that underpin applied linguistics, such as the various language sciences and education, but also on fields like economics, political science, and sociology that furnish methods of investigating the context in which language and education exist.

I should stress that this book is not a general or complete history of language testing theory and practice but a history of some highly institutionalized and industrialized tests and test batteries. Because it focuses on the *objective* language test in some institutional-industrial contexts, it is essentially a history of the attempts to develop tests that place their highest value on technical reliability, efficiency, and commercial viability. It thus does not attempt to chart the evolution of the kind of post-modern testing that many testers (among whom I number myself) have come to favour as an alternative to the model described here. It is because TOEFL marks the acme of this development in language testing, as well as revealing the forces that led to it, that the second part of the book is so narrowly focused. Because of this focus, events since the institutionalization and industrialization of TOEFL are only sketched. The main emphasis is on testing in America, where industrialization and objectivity have been most developed, but parallel British progress is also described.

Language testing is of particular interest because of the various competing factions that contribute to it. One of the reasons for my continuing fascination has been the way that it constantly forces practical and theoretical issues into fruitful tension. The needs of the tester regularly challenge the theorist, just as the findings of the theorist repeatedly tempt the tester. While it is fairly easy to come up with new assessment procedures, it remains difficult to explain exactly what is being measured,[2] a situation that guarantees a continuing productive stress. As if this first cause of strain