Language Teaching and Testing
Selected Works of Renowned Applied Linguists

世界知名语言学家论丛（第一辑）

Series Editor: Rod Ellis

J. Charles Alderson

语言测试

A LIFETIME OF LANGUAGE TESTING

Language Teaching and Testing
Selected Works of Renowned Applied Linguists

J. Charles Alderson

语言测试

A LIFETIME OF LANGUAGE TESTING

世界知名语言学家论丛
第一辑

Editorial Advisors: 杨惠中　庄智象

Series Editor: Rod Ellis

Contributors to the Series:

Charles Alderson

Gaby Kasper

Paul Nation

Peter Skehan

Rod Ellis

Zoltan Dornyei

出　版　说　明

　　“世界知名语言学家论丛（第一辑）”由上海外语教育出版社约请国际知名学者、英语教育专家 Rod Ellis 教授担任主编。丛书作者均为国际应用语言学领域耳熟能详的权威专家。丛书中的每一本聚焦应用语言学领域的一个特定主题，收录一位在该研究领域最有建树和影响力的语言学家一生中最重要的经典文章，如：Rod Ellis：《语法学习与教学》；Paul Nation：《词汇学习与教学》；Charles Alderson：《语言测试》，等等。书中的每篇文章经由精心挑选，既有对某一领域理论主题的深入阐述，又探讨了对第二语言教学和测试颇具意义的话题；除了作者一生的代表性作品外，还有不少新作，体现了作者的思索过程和研究轨迹，也展示了应用语言学领域发展历程中理论和研究逐步完善的一个个精彩镜头。

　　相信本套丛书的出版定能为国内应用语言学研究提供一个新的平台，带来新的启示，进一步推动我国语言学研究的发展。

Preface

This book is a collection of articles for the series *Language Teaching and Testing* — *Selected Works of Renowned Applied Linguists*. This series collects articles written by a number of leading applied linguists. Each collection focuses on a specific area of research in applied linguistics — for example, on grammar learning and teaching, vocabulary language learning and teaching, language testing (this book), and task-based language teaching. The aim of each book is to bring together older and more recent articles to show the development of the author's work over his/her lifetime. The articles are selected to address both theoretical issues relevant to a particular area of enquiry and also to discuss issues of significance to the teaching or testing of a second language (L2). As a whole, the series provides a survey of applied linguistics as this relates to language pedagogy and testing.

Each book begins with an autobiographical introduction by the author in which he/she locates the issues that have been important in his/her lifetime's work and how this work has evolved over time. The introduction also provides an outline of the author's professional career. The rest of the book consists of chapters based on articles published over the author's lifespan.

Each book, then, will contain articles that cover the author's career (over thirty years in some cases). Not surprisingly there are likely to be shifts (and possibly contradictions) in the author's positioning on the issues addressed, reflecting the changes in theory and research focus that have occurred in the specific area of enquiry over a period of time. Thus, the articles will not necessarily reflect a consistent theoretical perspective. There is merit in this. Readers will be able to see how theory and research have

developed. In other words, each book provides a snapshot of the kinds of developments that have occurred in the applied linguistic field under consideration.

This book brings together a collection of Charles Alderson's articles on language testing. Alderson has been a leading figure in the field of language testing for more than thirty years and in 2008 received a Lifetime Achievement Award from the International Language Testing Association. He has sole-authored or edited 18 books on different aspects of language testing, including the testing of foreign language reading ability, English for specific purposes testing, the washback effect of testing, testing metalinguistic knowledge, computer-based testing, and diagnostic testing. In addition, he has published over 90 articles in leading journals of language testing and applied linguistics. As well as an abiding interest in the principles and politics of language testing, he has been closely involved in the practical business of designing English language tests — for example, version D of the English Proficiency Test Battery (the 'Davies Test'), the International Language Testing System (IELTS) and, more recently, DIALANG (an online diagnostic test) and ELPAC — a test of English for air traffic controllers. This collection is a representative sample of Alderson's work, starting from his early work on the cloze procedure and ending with his current work on computer-based diagnostic testing. It will prove stimulating reading not just for language testers but for anyone interested in the nature of second language proficiency.

Rod Ellis
Auckland
September 2010.

Contents

Section Eleven   PRINCIPLES

Section Twelve   POLITICS

Section Thirteen   REFLECTIONS

Section One
INTRODUCTION

Looking Back on a Lifetime in Language Testing

EXPERIENCES

Like most educated people, my life has been dominated by tests and testing, yet I became a language tester more by chance than by design. One of my earliest memories goes back to the 1950s in England, when I was taking a practice test for the 11-plus examination. In those days, pupils nearing the end of primary school took an intelligence test which determined what sort of secondary school they would go to in the coming academic year. If they passed the test with high marks, they went to a grammar school. Somewhat lower marks allowed them to go to a technical high school and the secondary modern school was the destination for those who 'failed' the test. I vividly remember one particular item on the practice test. The task was to fill the gap in the following sequence of words:

> tale take __ like live

My first thought was that the missing word must be *lake*. However, this was a word that my grandfather, who spoke with a Lancashire accent in a Northern English dialect, used when he was talking about playing. *Lake* meant *play*. But I could not believe that the testers would be testing

the language that my grandfather used, and so I chose another word instead, and got the item wrong. I had not thought of the Standard English meaning of the word *lake*, and so I was penalised for knowing how my grandfather spoke.

Nevertheless, I eventually passed the 11-plus test and went to the local grammar school in my home town, Burnley, in north-east Lancashire. In my second year, I was publicly ridiculed by my art teacher for speaking in the local accent, where *look*, *cook* and *book* rhymed with the vowel of *hoot*. That humiliation had a lasting inhibition on my ability to speak unself-consciously, and I have ever since been hypersensitive to my accent and especially aware of how (not) to pronounce certain vowels and diphthongs. One teacher's prescriptive judgements had a lifetime's effect.

In Chapter 10 of this volume, I recount how, in the same school, my history teacher obliged me and my fellow pupils to dictate to each other his written notes on topics that would likely appear on the Ordinary-Level history examination, which notes we then had to learn by heart before the exam. I never understood what good he thought learning those notes by rote would do us, but, as I look back, I realise that I had experienced an early example of (negative) test washback.

Nor did I understand why my German teacher cheated in my Advanced-Level German examination by passing me a note with a word in German which he thought (wrongly, as it happened) that I did not know. Why did he cheat? Was it because of the power of the test? Was he aware that this was unethical? Do tests make teachers do things that they would not normally do? This was not only a puzzle about possible test impact but also an early introduction to ethics and the power of tests, which have, of course, been ongoing concerns for the language testing profession.

Having passed my Advanced-Level examinations and been awarded a scholarship in French and German at Oxford University (where the College Principal's main concern in my interview was not my academic

ability, but what sport I played and whether I might be able to represent my college in the annual championships) , I took a temporary job teaching in a primary school before going to Oxford. Despite having no teacher-training certificate, I was required to write termly reports on the progress of the children in my class — yet more ' experience ' of assessment, although at no point had I been taught how to write a formative assessment, what to say, how to observe and record student behaviour and achievements.

The experience of being an undergraduate at Oxford confirmed my impression of examinations as being arbitrary and highly subjective, in many ways mere rituals, a way of impressing on one the need to take study seriously, even if the contents of the examinations, termly tests and tutorial interviews were largely whimsical and unrelated to anything one might need to know in the real world. Testing was a game with its own absurd rules and procedures.

This impression was contrasted with, and yet confirmed by, the management recruitment procedures I experienced in 1967 in my final year at Oxford. Looking for a job in the real world beyond Oxford that I wished to join, I was subjected to two very different experiences. First, the ordeal of a three-day assessment process for Coates, Paton and Baldwin, textile manufacturers, where I was given (yet another) IQ test, a personality test, an interview with an occupational psychologist, and a group task with other applicants in which I was observed by the same psychologist. I also had to give a ten-minute lecture on any subject I wished, which had to be prepared overnight and then delivered in front of the other candidates, who were keen to show their sharpness and who subjected me to an intense grilling. I was not offered a job.

Contrast that with the interview I had that same year with Reckitt and Colman, food manufacturers: a 30-minute chat and then conversations over lunch. I was asked questions about why I preferred cycling as a sport to rugby, whether I was a team player rather than an individualist. Finally I was asked what I thought about the whole process

and I said that it was amateurish in contrast with the first experience. I passed the 'test' and became product manager for Colman's mustard.

All in all, then, my experience of (language) testing was that it was rather unreal, unregulated, non-standardised, unreflective, yet unquestioned, unchallenged, just 'done'. It was unprofessionalised; indeed, there appeared to be no such profession.

A DEVELOPING INTEREST

Nevertheless, after two years working at Reckitt and Colman in management and marketing, I came to realise how much I missed studying and dealing with language as a topic, and so I decided to go back to the world of education, where, like so many others, I had been successful, especially in tests and examinations. Facing a choice of teaching German in the UK or English in Germany, I decided my English was better than my German and so I went to Germany to teach English, without any training whatsoever, at Düsseldorf University.

Düsseldorf was a new university, with an English department that was one year old. There was no curriculum for the English language courses, and I had no experience teaching English as a Foreign Language, and indeed no qualifications at all in teaching. Yet the first task I was given, to my surprise, was to develop a test to put the new students into language classes that did not yet exist. The only experience I had of language tests was translation tasks at Oxford University and the Ordinary and Advanced-Level examinations I had taken years before in German, French and Spanish.

Fortunately, the colleague with whom I was to share an office, Gerold Deffner, who quickly became my very good friend, was an experienced teacher of English, and together we put together my first English language test. The test was administered, the students were put into four different groups, and my career as a language teacher was launched.

Soon after that, I came across Robert Lado's 1961 book *Language*

Testing. I read about the principles he was expounding of a contrastive approach to test content design, and I even learned that what we had created was called a placement test. I learned a little about item analysis, and was eventually able to analyse the results of my test, at first calculating means and standard deviations and then item facilities and discriminations. It was a sobering, but at the same time, an exciting experience, and my interest in a more objective approach to creating tests that had demonstrable quality was awakened.

Soon it became clear to me that I needed to study more about language teaching, not only the practice, the methodology and the textbooks, but also the theoretical principles behind the practices advocated by colleagues and what little I had read. Somewhat to my surprise, I was accepted to study for an MLitt in Applied Linguistics at Edinburgh University (without having to take any sort of test) and in October 1969, I travelled from Germany to Scotland, to be told that I had been accepted onto a research degree (of which I was completely unaware), and that, before I could begin my thesis on 'Psycholinguistic-sociolinguistic approaches to language teaching', which I had naïvely put forward as my major interest, I would have to take a set of courses for the Diploma in Applied Linguistics. Not only was this not a setback, it provided me with the opportunity to explore a range of topics through the courses offered, one of which was research design and another was language testing. The teachers were Alan Davies, Ruth Clark, Elizabeth Ingram and Tony Howatt.

I could not have had a better introduction to language testing with a better set of teachers, and by the end of the nine-month Diploma (which, fortunately, I passed), I realised the folly and pretentiousness of my former research topic. I decided to switch to a topic within language testing, since I had been challenged by the theory of test construction, excited by the introduction to rigorous statistical analysis of test data, enthused by the growing perception that such analyses could help one understand what one had created and what its flaws might be, and