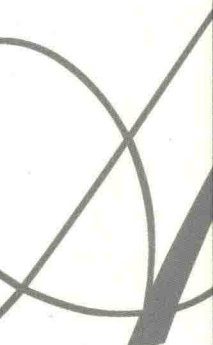


剑桥语言测试研究丛书



**An Empirical
Investigation of the
Componentiality
of L2 Reading in
English for Academic
Purposes**

**学术英语阅读能力
构成成分实证研究**

Cyril J. Weir, Yang Huizhong, Jin Yan 编

剑桥语言测试研究丛书

**An Empirical
Investigation of the
Componentiality
of L2 Reading in
English for Academic
Purposes**

**学术英语阅读能力
构成成分实证研究**

Cyril J. Weir, Yang Huizhong, Jin Yan 编

图书在版编目 (CIP) 数据

学术英语阅读能力构成成分实证研究 / (英) 西里尔·威尔, 杨惠中, 金艳编.

—上海: 上海外语教育出版社, 2018

(剑桥语言测试研究丛书)

ISBN 978-7-5446-5316-9

I. ①学… II. ①西… ②杨… ③金… III. ①英语—阅读教学—教学研究 IV. ①H319.4

中国版本图书馆CIP数据核字 (2018) 第082313号

This is a reprint edition of the following title published by Cambridge University Press:

An Empirical Investigation of the Componentiality of L2 Reading in English for Academic Purposes (ISBN: 9780521653817)

© Cambridge University Press 2000

This reprint edition for the People's Republic of China (excluding Hong Kong, Macau and Taiwan) is published by arrangement with the Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

© Cambridge University Press and Shanghai Foreign Language Education Press 2018

This reprint edition is authorized for sale in the People's Republic of China (excluding Hong Kong, Macau and Taiwan) only. Unauthorised export of this reprint edition is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of Cambridge University Press and Shanghai Foreign Language Education Press.

本书版权由剑桥大学出版社和上海外语教育出版社有限公司共同所有。本书任何部分之文字及图片, 未经出版者书面许可, 不得用任何方式抄袭、节录或翻印。

此版本仅限中华人民共和国境内销售, 不包括香港、澳门特别行政区及中国台湾。不得出口。

图字: 09-2017-927号

出版发行: 上海外语教育出版社

(上海外国语大学内) 邮编: 200083

电话: 021-65425300 (总机)

电子邮箱: bookinfo@sflap.com.cn

网址: <http://www.sflap.com>

责任编辑: 蔡一鸣

印刷: 上海书刊印刷有限公司

开本: 635×965 1/16 印张 19.5 字数 409千字

版次: 2018年12月第1版 2018年12月第1次印刷

印数: 2 100册

书号: ISBN 978-7-5446-5316-9 / G

定价: 60.00元

本版图书如有印装质量问题, 可向本社调换

质量服务热线: 4008-213-263 电子邮箱: editorial@sflap.com

出版说明

语言测试作为语言教学的重要组成部分，一直是社会关注的焦点。近年来，我国学术界对语言测试研究的关注也在逐渐升温。国际上，学者们对语言测试的研究已经达到较高的水平，无论是从理论框架的形成到复杂数学模型的建构及其应用方面都取得了实质性的进展，而且在对有国际影响力的外语考试（如雅思、托福等）的信度、效度的研究方面也取得了不俗的成果，出版了不少优秀的著作。

英国剑桥大学出版社出版的“*Studies in Language Testing*”是其中较有影响力的系列之一。丛书由剑桥语言测评中心策划推出，从测试领域的各个研究视角入手，展示了该领域研究的设计、数据收集和数据分析等环节的基础理论与实践操作，反映了测试领域的最新研究动态。

外教社从中精选9种，引进出版。这些图书中，有对语言测试听说读写等方面的效度和质性研究，也有对测试反拨效应及其对教学影响的研究，还有对雅思考试和欧洲语言共同参考框架的分析思考。希望丛书的出版可以为我国语言测试研究领域的发展和 innovation 提供具有借鉴意义的启示，打开更广阔的研究思路。

*An empirical investigation of
the componentiality of L2
reading in English for
academic purposes*

Cyril J. Weir, Yang Huizhong, Jin Yan

Series Editor's note

This volume represents an interesting and important study in the assessment of reading for academic purposes. The approach adopted is very methodical and follows a clear development and validation pattern thus acting as a valuable case study for anyone developing language tests.

The project was based in China and financially supported both by the British Department for International Development (DFID) and the Chinese National College English Testing Committee. DFID has supported numerous English Language Teaching and Testing projects throughout the world and was particularly active in China during the 1980s and 1990s where it also helped in the development of the College English Test (CET) now widely used at the tertiary level in China. A forthcoming volume in this series will focus on the validation of the CET.

English Language Testing in China is characterised by projects like this one which demonstrate a genuine interest in developing effective and validated measures of English Language ability. Alongside CET the Public English Testing System (PETS) has been developed recently. This project, funded by the DFID and the State Education Commission (SEC), was carried out by the National Educational Examinations Authority (NEEA) and the EFL Division of the University of Cambridge Local Examination Syndicate (UCLES). Taking place between 1997 and 2000 the PETS project developed a five-level system of English language tests that aimed to rationalize a much larger number of tests developed and implemented in a more haphazard manner over many years. The project was driven by the stated aim of the Chinese government to raise the standards of English throughout China. Testing is seen as an important tool in achieving this aim and for the first time the assessment of writing, speaking and listening are integral parts of a testing system in China from the outset. Direct criterion-referenced assessment poses substantial logistic difficulties in China with its enormous population and it is to the credit of the authorities that they were prepared to undertake such an initiative.

Michael Milanovic
Cambridge
May 2000

Preface

This study reports on an empirical investigation of the componentiality of L2 Reading in English for academic purposes. The focus on careful reading in the theoretical literature has meant that we have somewhat ignored expeditious reading behaviours such as skimming, search reading and scanning in the teaching and testing of reading. We have theories of careful reading but very little on how readers process texts quickly and selectively, i.e., expeditiously, to extract important information in line with intended purpose(s).

Furthermore, because of a focus on the local level, e.g. word recognition or syntactic parsing, only limited attention has been paid to careful reading at the global level, i.e. comprehension of the main ideas in a text or of the discourse topic; the macropropositional as against the micropropositional level of text. In addition to careful reading at the local level, we felt it was important to explore further four kinds of reading: Search reading, Skimming, Scanning, Careful Reading (at the global level).

The development of the Advanced English Reading Test (AERT) for university students in the People's Republic of China under the auspices of the National College English Test Committee was the vehicle for investigating the nature of and the relationships between these different types of reading at the macro level. The project was developed at CALS, University of Reading with colleagues from Shanghai Jiao Tong University in China. A number of CALS staff worked on the project with the Chinese members of the development group and in particular Rita Green made a valuable contribution to the project activities reported below.

This volume reports on the methodological procedures that led to the development of this important test and discusses the results of the empirical investigations that were carried out to establish its validity both *a priori* and *a posteriori*. As such it offers a blueprint for those wishing to research in the area as well as generating data on these different reading styles of interest to both testers and teachers of reading in an additional language as well as researchers.

Acknowledgements

Many people were involved in the Advanced English Reading Test project (AERT) that was carried out by College English Test personnel in the People's Republic of China and academic staff in the Testing and Evaluation Unit (TEU) at CALS University of Reading between 1995–1998. Particular thanks are due to Luo Peng of Shanghai Jiao Tong University, PRC and to Rita Green, Hanan Khalifa, Amos Paran and Eddie Williams of CALS, UK for their numerous contributions to this project. Luo Peng helped greatly in generating the data that Chapter 4 is based on and in the proof reading.

We would like to acknowledge the financial support given by the Department for International Development (DFID) of the British Government and also by the National College English Testing Committee, PRC. Without their support, this project would not have been possible. We are also grateful to DFID for their permission to publish this material. Lastly, we must thank Barbara Wickham of the British Council who helped make this project the success it was.

Contents

Series Editor's note	IV
Preface	V
Acknowledgements	VI
Chapter One	
Introduction	1
Chapter Two	
Establishing the parameters of EAP reading	14
Chapter Three	
Synthesising the <i>a priori</i> validation data	51
Chapter Four	
Operationalising the specification: test development	55
Chapter Five	
The <i>a posteriori</i> validation of the prototype AERT version 1: quantitative studies	73
Chapter Six	
The <i>a posteriori</i> validation of the prototype AERT version 1: qualitative studies	93
Chapter Seven	
Conclusions and recommendations	118
References	122
Appendices	128
Subject Index	298

1

Introduction

Background: the Chinese operational context

Reading in English at undergraduate level in China

Before 1985, the required speed of reading in English for Chinese university students was 17 words per minute. A survey by the Ministry of Higher Education (later the State Education Commission) showed that only one third of university graduates acquired this 'reading ability'.

In 1985, the National College English Teaching Syllabus (NCETS) was introduced by the State Education Commission. In this syllabus, the English course, which is compulsory for all university students across the country, is divided into six bands. All the students must meet the requirements of Band 4, which include a reading speed of 50 wpm for careful reading and 80 wpm for quick reading with 70% comprehension. The requirements of Band 6 are aimed at students who have successfully completed Band 4 study, the target reading speed being 70 wpm for careful reading and 120 wpm for quick reading with 70% comprehension. Band 4 and 6 together constitute the basic grounding stage of the College English Course. This course focuses primarily on the development of students' linguistic competence with only limited attention being paid to the development of language skills and strategies.

Despite this focus on linguistic competence, the publication of this syllabus has had a positive impact on English language learning at tertiary level in China. This can be seen in the data available on the College English Test (CET) based on the NCET syllabus which was inaugurated in 1987.

A recent 4 year research study has shown the CET to be a valid and reliable measure of general linguistic competence (Yang and Weir forthcoming). It has had a powerful backwash effect on the numbers learning English at the foundation stage in the university system with the candidature growing to over two million by 1997. Since the inception of the test ten years ago there has been an improvement in the language competence of university students as attested by institutional and national performance levels.

However, an important stated aim of English language teaching at the tertiary level in China is to improve access to scientific and technical literature

through reading in English. Unfortunately linguistic competence is not the same as performance ability in language skills and strategies in reading. If providing students with the latter is the ultimate goal of English language teaching as specified in NCETS, then further steps need to be taken.

It was evident that many students and also their university authorities saw the foundation stage of English study as the end of English language learning and teaching. Furthermore, the end-users of the CET tended to misinterpret the value of CET and often expected too much from the certificate holders in terms of performance skills and strategies. Clearly the system was not providing university graduates with the requisite skills and abilities to access foreign academic and technical literature through the medium of English.

Rationale for the reading test project

It should be noted that to help achieve this criterial goal for English teaching in Chinese universities an EAP reading course is already stipulated in the syllabus for the undergraduates having finished foundation stage study after year 2. The problem is that the course is not given due attention and is neglected in many universities and institutions by university authorities, teachers and students.

The lack of an adequate and appropriate assessment tool is seen as a major reason for this neglect. A widespread tradition in China, as everywhere else, is 'what is tested is taught' and consequently 'what is not tested is not taught'. In this situation, an Advanced English Reading Test (AERT) could be the necessary catalyst for encouraging the achievement of the aim of the final stage of the college English course in China.

This project, therefore, set out to help achieve that end by developing an Advanced English Reading Test (AERT). The main benefits from this test would be:

- the availability of an appropriate tool for universities and teachers to monitor and evaluate students' performance in EAP reading
- the exertion of a much-needed positive backwash effect on English teaching in years 3 and 4 of college education in China in the sense of actually encouraging the teaching of reading skills and strategies where none may take place at present.

Monitoring the impact of the reading project

Currently demographic statistics on the College English Test (CET – 4 and 6) are collected as a matter of routine and there is no reason why the same could not be done for the AERT. Scores by centre, region, institution, gender, etc., are made available annually and can be compared to previous national and institutional averages.

Rigorous attempts are made to statistically equate the various forms of the CET test through IRT and an anchor test so comparisons can be made between

cohorts from year to year. In this way it is possible to monitor whether there can be said to be improvement at a national or institutional level over a period of time. A similar system could be set in place for the AERT.

Empirical data can also be collected from institutions (staff and students) and end-users by self report (questionnaire survey and interview) in order to triangulate the descriptive statistics emanating from test administrations.

Such data will enable the authorities to monitor the impact of the AERT on the gain in reading ability in years 3 and 4 of Chinese universities. The most effective design would collect data relating to the state the students are at when they begin year 3 and where they get to by the end of year 4. Implementation data which established the amount and nature of mediated instruction in reading that went on within institutions might enable some useful pedagogical lessons to be drawn.

Failing such comprehensive evaluation, a comparison from year to year of exit behaviour would provide useful though less comprehensive data. At the very least it might show whether the test was having an impact in terms of gains in population reading scores across the various types of reading being tested over a period of time. The value added for the Chinese economy might then be investigated. This would involve end users calculating what progress in terms of performances at different levels on the test would contribute financially to their organisations. The enhanced value of improvement in the different types of reading measured by the test might then be estimated.

Background: developing tests to measure the construct of reading in English for academic purposes

The specification and operationalisation of the construct

In the past ESL examination specifications were either absent or extremely limited. Typically one met a spuriously circular argument relating examinations to the textbooks used to prepare the students for the examinations. The textbooks were viewed as a benchmark for establishing what levels such as intermediate and proficient meant. These textbook writers would conversely refer to the examinations as their point of reference for both the content and level for their coursebooks.

In recent years a number of the major examination boards have taken a more principled and systematic approach to the development of tests. The University of Cambridge Local Examinations Syndicate (UCLES) is a good example of this (see handbooks produced for each examination by UCLES and *User's Guide* prepared for Council of Europe). They have attempted to provide clear specifications for each of their major examinations and establish systematic development procedures to faithfully implement these.

It is clear that a reading test/examination is only as good as the texts and tasks that are used to operationalise the construct that it is intended to measure. Inadequacies or limitations in the texts and tasks employed will constrain the value of any comprehension test. Given the huge potential test population and national importance of the Advanced English Reading Test (AERT) in China it was imperative that we develop maximally valid operationalisations of what we believed to be the important elements of the construct of EAP reading in the form of texts and associated tasks. In order to develop a construct-valid reading test we had to develop effective, efficient and replicable *a priori* and *a posteriori* procedures for test development. In this book we lay out a comprehensive set of procedures for the development of an advanced reading test in English as a Second Language.

Our initial studies indicated that reading research in this century has been constrained by a narrow view of reading (in the main focusing on careful reading at the local level) and by serious limitations in the tests used as research instruments (see Urquhart and Weir 1998). It was therefore imperative for us to develop a comprehensive specification with a sound theoretical and empirical base and implement this as faithfully as possible in the AERT. As well as helping us to develop a valid and reliable operational test, such research might also provide data which cast some light on the componentiality of the reading construct. Only through the development of a valid and reliable set of tests could we hope to resolve the issue of whether reading is a unitary activity or whether it is made up of separable components, for example: expeditious types of reading as in search reading, skimming, scanning for specifics, and careful reading at the global and local levels. Such a test might shed light on the relative contribution of the posited skills and strategies to the overall picture of a student's reading ability. It could also tell us about the relationship between the test components and inform us of the relative weaknesses and strengths of our students. Whether for formative or summative purposes such diagnostic evaluation might impact on whole educational systems as well as individual classrooms.

The data from the test development procedures described in this book are all grist to the construct validity mill. They can all shed light on what it is we are measuring and how well we are doing this. The more of these we can embrace in our research investigations into EAP reading the more valid and reliable the resulting data on academic reading ability.

Urquhart and Weir (1998: Chapter 5.3) outline a principled set of procedures for investigating the componentiality of reading (see Figure 1.1 below). The *a priori* and *a posteriori* procedures outlined there were based on the development work behind the AERT in China and are discussed in detail in this volume. Such a methodology we feel is generic and should for the most part apply to all reading situations.

Figure 1.1

A methodology for investigating the EAP reading construct

A priori validation

Stage 1: Specification of the construct

The reading strategies and skills and the conditions under which these activities are performed might be established through:

- target situation analysis of target population's reading activities;
- theoretical literature review of reading processes;
- research literature review concerning the componentiality of reading;
- document analysis: EAP reading course-books/EAP reading tests.

Stage 2: Development of pilot tests to operationalise EAP reading specification

Systematic textmapping of appropriate texts:

- to establish the consensus information recoverable according to type of reading employed i.e. careful versus expeditious;

Produce pilot version of test(s):

- decide on most appropriate format in relation to operations;
- allow for attrition in texts and items by trialling extra;
- ensure intelligibility of rubrics;
- empirically establish timing;
- consider order of questions/process dimension;
- check layout;
- trial on small samples. Produce first draft of mark scheme;
- moderate tasks and mark scheme in committee.

A posteriori validation

Stage 3: Analysis of data on the test

Trial on reasonable sample

Item analysis

Establish item:

- facility values;
- discrimination indices;

Estimates of reliability

- marker reliability.

Estimates of internal validity

- internal consistency;
- correlations;
- principal component analyses;
- level of subtests: Means, t tests and cross-tabulation.

Estimates of external validity

Establish what items are testing through:

- qualitative expert judgement of items;
- qualitative introspection/retrospection by test takers;
- feedback from test takers (Interview /questionnaire).

Revise

- administrator's instructions;
- items;
- timing;
- rubrics;
- mark schemes;
- re-trial any new items.

[Source: *Reading in a Second Language*, Urquhart A. H. and Weir C. J. (1998). Longman]

Each of these stages is discussed in detail in the following chapters in this book as we describe the steps we took to try to achieve construct validity in our measuring instruments in terms of operationalising the skills/strategies underlying the various types of reading of which we wished to establish the empirical existence.

In the next chapter, we will examine in close detail the *a priori* procedures which led to the specification of AERT. These constituted the first stage in the development of a systematic approach to developing our test of English for Academic Purposes reading ability. On the basis of the specification we then developed AERT using a principled set of procedures described in Chapter 4.

A brief outline of these *a priori* procedures employed to define the construct of reading both theoretically and operationally is provided in the next section.

A priori validation

Establishing the parameters of the reading construct

To establish a specification of operations and performance conditions to be tested we pursued a number of avenues in the development of the Advanced English Reading Test (AERT) for undergraduates in China. We carried out:

- a review of research and theories of reading;
- analysis of EAP reading tasks in Course-books;
- analysis of EAP reading tasks in Public Tests;
- needs analysis.

The investigation of the theoretical construct of EAP reading and the development of a test for this required systematic research. The research began with a review of the literature on existing theories of the reading process. The study of various models of reading shed light on the construct of reading from various aspects though it was noticeable that these theories are mostly premised on only one of the identified types of reading, that is, careful reading. Nevertheless such research drew our attention to the importance of reader-driven processing at the text level and the importance of goal setting, as well as more traditional text-driven processing at the word/sentence level.

We also examined the empirical (largely test-based) research literature concerning the divisibility of the reading construct into components and the salient performance conditions under which these types of reading are performed. This product focused empirical research relating to the componentiality of reading ability points to at least a bidivisible view of reading with vocabulary loading on a separate factor in addition to general reading comprehension in nearly all cases. The data suggest that a partially divisible view of reading is preferable to a unitary view.

The empirical study also involved a survey of Chinese undergraduates' EAP reading needs as viewed by subject teachers of advanced reading in English in China.

For a view of what and how EAP reading is currently taught and tested, a survey was made of all the major EAP reading course books and EAP tests available in terms of operations (skills/strategies) and performance conditions (length of texts, reading speed, etc.).

The findings from these various strands of enquiry resulted in specifications of the types of reading to be included in the test as well as the conditions under which these reading activities should be performed. The types of EAP reading our research indicated we should include are listed below:

- careful reading for global comprehension of main ideas;
- careful local reading for understanding at the word level;
- expeditious reading (reading quickly, selectively and effectively) for global comprehension of main ideas;
- expeditious reading at the local level (scanning for specific details).

In addition our research suggested the following performance conditions had an important effect on performance on these tasks and must be carefully considered by the test developer:

- length of text must be appropriate for intended type of reading and time allocated be consonant with this;
- time allowed to complete to be empirically determined for each reading type tested;
- strict enforcement of such time controls at the passage/reading type level;
- nature of text must be accessible across three broad discipline areas;
- rhetorical organisation of texts appropriate for reading type;
- overtness of text organisation (markers of importance, textual signposting) for expeditious tasks;
- nature of vocabulary: appropriate degree of specialisation;
- topic familiarity: low to medium familiarity.

The test designed in line with the specified guidelines is expected to maximally operationalise the construct in question. Once appropriate operations and conditions are established these have to be implemented in a test.

Textmapping of main ideas in chosen texts

Having established the skills and strategies we wanted to test located appropriate texts which lent themselves to the testing of these and which met the specified performance conditions, we moved on to constructing the test items.

The first step was to process the texts to establish the macro-propositions which might be extracted in line with the specified type of reading to be performed on that text. This utilisation-focused approach seeks to establish the main ideas of a passage through expeditious or careful 'textmapping' procedures.

1 Introduction

Urquhart and Weir (1998) describe how in each textmapping an attempt should be made to replicate a single type of reading on a single text, e.g. *reading a text slowly and carefully to establish the main ideas*. The product of the particular reading of a text can be compiled in the form of a spidergram or as a linear summary. This is first done individually and then the extent of consensus with colleagues who have followed the same procedure is established. The objective of the procedure is to examine whether what we have decided is important, in line with the specified type of reading activity, and matches what colleagues consider important.

Urquhart and Weir (1998) see this as a crucial first step in trying to ensure the validity of our tests. The answers to the questions developed equate with the important information in the text that could be extracted by the particular type of reading being assessed. An ability to answer the items should indicate that the candidate has understood the passage in terms of successful performance of the specified operation(s).

Procedures such as textmapping should enable us to determine in a principled fashion the content we might wish students to recover from a text according to the type of reading employed. However, the format, which acts as the vehicle for testing reading activities may constrain the operations and conditions we attempt to include. So as well as carefully specifying the latter we need to consider carefully the method we are going to use so as to minimise the influence of method on measurement of the trait. The cardinal rule remains however: we must first decide what types of reading we want to test and develop systematic procedures for deciding the micropropositions and/or macropropositions that we would expect candidates to extract from texts in performing these types of reading. Only then can we consider test formats and decide which will most faithfully mirror the procedures and allow the appropriate propositions to be extracted.

The mapping procedure will provide the content to be extracted for each of the types of reading in the test. It will also show whether each passage is suitable for its intended reading purpose. Where it is possible to produce a consensus textmap, this then needs to be converted into appropriate test items in the format selected. Where consensus is not achieved or the textmapping produces too few items these texts must be rejected!

A posteriori validation

Once the first version of the test was ready *a posteriori* validation procedures were applied at the trialling stage to determine statistically whether the test was working in the way it was intended to and to closely examine the construct underlying the test. The *a posteriori* statistical procedures employed in the two trials of the AERT in China are discussed in detail in Chapter 5 below. They indicate that the components in the test do seem to be measuring differing parts of the reading construct and that students do perform differentially on different types of reading.