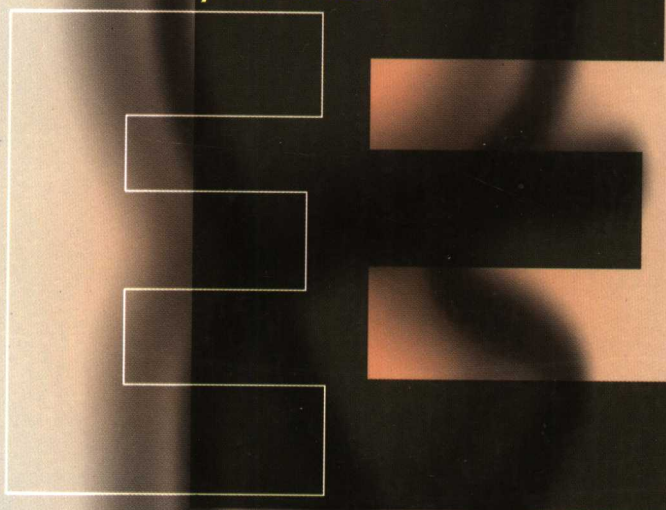


CET-6和TEM-4的对比研究

周越美 著

Exploration into Comparability of
Two National EFL Tests
(CET-6 and TEM-4) in China



W
外教社

上海外语教育出版社

H310.42

97

CET-6和TEM-4的对比研究

Exploration into Comparability of

Two National EFL Tests

(CET-6 and TEM-4) in China

周越美 著

4



外教社

上海外语教育出版社

图书在版编目(CIP)数据

CET-6和TEM-4的对比研究 / 周越美著.

—上海: 上海外语教育出版社, 2004

ISBN 7-81095-528-4

I. C… II. 周… III. 英语—水平考试—对比研究
—中国 IV. H310.42

中国版本图书馆CIP数据核字(2004)第000396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@slep.com.cn

网 址: <http://www.slep.com.cn> <http://www.slep.com>

责任编辑: 江雷红

印 刷: 上海江杨印刷厂

经 销: 新华书店上海发行所

开 本: 850×1168 1/32 印张 8.625 字数 231 千字

版 次: 2004年12月第1版 2004年12月第1次印刷

印 数: 2 100 册

书 号: ISBN 7-81095-528-4 / H · 194

定 价: 15.00 元

本版图书如有印装质量问题, 可向本社调换

ACKNOWLEDGEMENTS

A book of this kind on exploratory quest owes a great deal to many individuals: to those whose work is explicitly presented and discussed, to those who grant the birth of its printed version, a revision of my doctoral dissertation of the exploration, and to those whose encouragement, help and support have extended well beyond what can be expected from the likes of me, an explorer and researcher.

In this last connection I am very much grateful to Professor Zou Shen, supervisor of my Ph. D. dissertation. Her enlightening and detailed suggestions and comments have drawn my attention to a number of deficiencies and have made many things clearer for me. Without her heuristic urge, for instance, the preliminary version of the research could not have been submitted and even accepted for network presentation by the committee of 24th International Language Testing Research Colloquium in 2002.

I would also like to thank the following professors: Zhang Yanli, Mei Deming, He Zhaoxiong and Li Ji-an, professors from Shanghai International Studies University (SISU); Zhou Guoqiang and Yang Huizhong, professors from Jiaotong University; Wang Xinqiu, Tang Zhenbang, Hu Ruizhang, professors from Shanghai Normal University; Tan min, Wang Lili, Wang Xiaoqun, Feng Runmin and Tan Jing, professors and my colleagues from Shanghai University of Finance and Economics (SHUFE). All of them have offered me invaluable advice and help, in one way or another, in the process of the study.

My sincere thanks should also be extended to many of my friends and schoolmates in SISU who never hesitated to offer me help, academic or technical, for the study. They include Zhang Yadong, Xi Zhong-en, Liu Baoquan, Liu Qin, Dang Zhengsheng, Hai Fang and others.

I am even much indebted to several hundred students and their teachers in SISU and SHUFE. They endured the trouble of providing data on the sample tests.

And finally, I wish to thank my family who showed me affectionate patience in the long-time endurance of my cries, anxieties, lost weekends and holidays, and endless hours before the computer.

ABBREVIATIONS

CET	College English Test
CET-4	CET Band Four
CET-6	CET Band Six
CETC	College English Test Committee
CLA	communicative language ability
COM	comparability-oriented model
CPE	Certificate of Proficiency English
EFL	English as a foreign language
ETS	Education Testing Service
FCE	First Certificate of English
MET	Matriculation English Test
SHUFE	Shanghai University of Finance and Economics
SISU	Shanghai International Studies University
SNU	Shanghai Normal University
TEM	Test for English Majors
TEM-4	TEM Band Four
TEM-8	TEM Band Eight
TMF	test method facets
TOEFL	Test of English as a Foreign Language

CONTENTS

Preface	1
Chapter 1 Theoretical and Empirical Survey	4
1.1 Research methodology	5
1.1.1 The paradigm dialog	6
1.1.2 Overview of paradigms or methodological traditions	11
1.1.3 Guidelines for quantitative and qualitative studies	13
1.2 Test assessment	15
1.2.1 Conceptualization of communicative language ability (CLA)	16
1.2.2 Test method facets (TMF)	35
1.3 Previous comparability research	44
1.3.1 Type one — three cases	46
1.3.2 Type two — other cases	51
1.4 A tentative model for comparability research	52
1.4.1 Procedure-observation of the former three cases	53
1.4.2 Procedure-observation of the other cases	55
1.4.3 A new model — COM	58

Chapter 2	<i>Preliminaries</i>	64
2.1	Rationale and objectives of the study	66
2.2	Study design	67
2.3	Instruments	69
2.3.1	CET-6	69
2.3.2	TEM-4	71
2.3.3	Questionnaire	74
2.3.4	Content rating	75
2.4	Procedures of administration, scoring, and data organization	84
2.4.1	Site selection	85
2.4.2	Site coordinators	86
2.4.3	Test administration	87
2.4.4	Scoring	88
2.4.5	Data collection	89
Chapter 3	<i>Sample Subjects</i>	92
3.1	Group A — typical TEM-4 takers	94
3.2	Group B — typical CET-6 takers	95
3.3	Other characteristics of the sample subjects	96
Chapter 4	<i>Descriptions and Analyses of the Test Scores</i>	100
4.1	Test performance	101

4.1.1	Overview of scores	101
4.1.2	Mean comparisons	104
4.1.3	Other comparisons	117
4.2	Reliability considerations	119
4.2.1	Reliability estimates	120
4.2.2	Aspects that affect reliability estimates	124
4.2.3	Some other systematic measurement facets	129
4.3	Comparability of abilities measured	133
4.3.1	Across-test factor structure	137
4.3.2	Within-test factor structure	145
4.3.3	Within-component factor structure	149
4.3.4	Discussion	156
4.4	Interchangeability of CET and TEM test scores	157
4.4.1	Quartiles/percentiles	159
4.4.2	Linear equating analysis	162
4.4.3	Discussion	170
Chapter 5 Test Content Analyses		177
5.1	Characteristics of the test content	178
5.1.1	Constitution of content input	179
5.1.2	Quantity of content	181
5.1.3	Nature of content	193
5.2	Comparisons of content ratings between CET and TEM	208
5.2.1	Listening parts	217

5.2.2	Reading parts	218
5.2.3	Vocabulary and structure/grammar parts	220
5.2.4	Composition parts	221
5.2.5	Short-answer questions and cloze parts	222
5.3	Discussion	224
5.3.1	Significant distinctions	224
5.3.2	Paired rating comparisons	227
Chapter 6	<i>Conclusions, Reflections and Implications</i>	232
6.1	Conclusions and findings	232
6.1.1	Review of the COM model	233
6.1.2	Findings from the analyses	234
6.1.3	Significance of the study	236
6.2	Recommendations and implications	238
6.3	Reflections on further research	239
Appendix 1	Tests and administration times	241
Appendix 2	CET sample used in the study	243
Appendix 3	TEM sample used in the study	247
Appendix 4	Information questionnaire	254
Appendix 5	Rating table	256
Appendix 6	Factor structures	259
Appendix 7	Overview of sentence types of the two tests ...	261
References	262

PREFACE

For more than a dozen of years the College English Test (CET) and the Test for English Majors (TEM) have been in operation nation-wide, and their scores have been increasingly used as measures of proficiency in English as a foreign language (EFL) throughout China. The number of test candidates has increased every year; about 4,500,000 took the CET and about 100,000 the TEM in 2002. The large numbers of test takers along with the extensive use of test scores stimulate a tendency in China that a great number of younger individuals' careers or education decisions are considerably affected. These decisions include award of diplomas before graduation, applying for admission to an educational program, or seeking employment as well as advancement in a career.

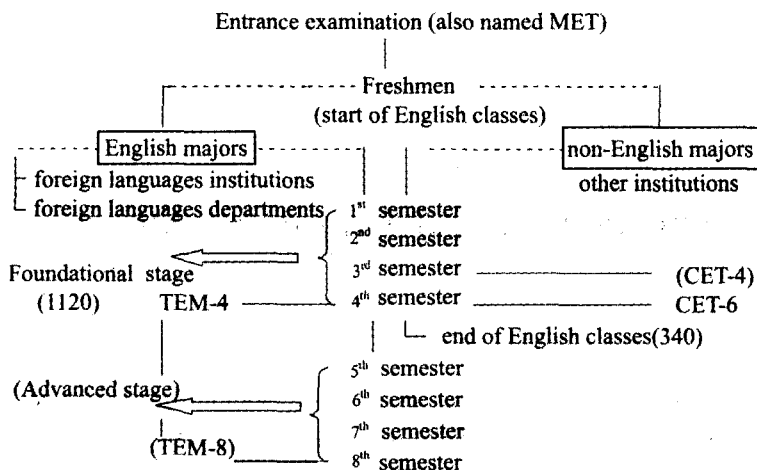
What is the target of the tests? The National College English Tests-Band Six (CET-6), one in the CET test battery, was originally designed for college students after they completed the sixth and highest level of English study for non-majors (CET designer group, 2000:1). Some institutions have made the CET-6 certificate a prerequisite for graduation (Yang, 2002:4). The test has been sponsored by the Higher Education Department of the PRC Ministry of Education (Yang, 2002:1) since its inception in 1987. Twice a year the test is administered by the National College English Testing Committee of China (CETC), a testing service centre now based in Shanghai Jiaotong University.

The Test for English Majors-Band 4 (TEM-4), one in the TEM test battery, is designed for students majoring in English language and literature; it is given near the end of the first two years' foundation stage of a four-year degree programme. The test is organized every May by another testing centre located in Shanghai International Studies University, also under the auspices of the Ministry of Education.

As each test is claimed not to be solely confined to certain textbooks, they are in essence proficiency tests (Yang, 1998: 10; Zou, 1998-b:2), though devised in accordance with the requirements of the respective national EFL teaching syllabuses (Yang, 2002:1; Zou, 1998:1). The only distinction between the two syllabuses seems to lie in the fact that the English majors have much more class instruction (about 1120 class hours) than the non-English majors (about 340 or even less). The fourth semester is the normal time for the majority of both types of students to take their respective test. CET-6 test takers need to have a pre-intermediate form of the test, that is CET-4; TEM-4 students can take an advanced test (TEM-8) two years later. Figure 1 illustrates this.

The entrance examination, known also as MET (Matriculation English Test), refers to the yearly national test of English, among the other summative tests of Chinese language, chemistry or mathematics for high school graduates, planning to enter colleges and universities. Only those who achieve acceptable scores in these examinations are admitted to major in English or in other humanities and science disciplines. Four semesters later, they are expected to take the TEM-4 or CET-6 tests.

Figure 1 Relationship between school years and the two EFL tests



In recent years, however, the CET-6 has been taken also by **English majors**, while the TEM-4 also attracts **non-English majors** in some parts of China. Thus the distinction between the two tests has become blurred, and they even seem to be practically regarded as something equivalent or at least interchangeable. Until now, no official or research explanation for the exchange has been presented. It was this phenomenon that aroused the author's curiosity to explore the relationship between the two tests. This interest has now resulted in a research undertaking of the present study.

CHAPTER 1

Theoretical and Empirical Survey

It is well known that large-scale standardized language tests, such as CET and TEM, are nowadays often backed up by theoretical conceptions, which resulting from research work are generally based on “abstract beliefs of what language is, what language proficiency consists of, what language learning involves and what language users do with language” (Alderson, Clapham & Wall, 1995:16). Each test is actually the operationalisation of such beliefs or theories, whether explicit via test specification or relying on intuition (Alderson, Clapham & Wall, 1995: 19). Therefore, any discussion about language tests can hardly be significant enough without theoretical support.

As a theoretical as well as empirical background, therefore, three aspects will be surveyed in the chapter: evolution of research methodology, conceptions and their evolution of test assessment, and review of comparability research. All the survey will finally bring about a research model for comparability study, recommended by the author, which is to lend itself to the process of the discussion.

In other words, embarking on a unified project of quest for test comparability like this will inevitably include a variety of empirical methodologies, ranging from the more commonly-used quantitative exploration about test scores to the traditional as well as modern qualitative analysis of test content and test-taking or test-administrating process. To back up all these, some historical evidence or important traditions and guidelines are to be reviewed in passing.

Apart from methodology issues that are instrumental to the overall study, assessment theories concerning language competence and test methods will then be traced, as they are particularly relevant to the topic of discussion.

Other than the theoretical side, discussions about cases of comparability study will follow and their characteristics described for background information.

Based on all these theories and considerations, an operational model is finally suggested, which functions not only as a summary of the literature review but also an operational framework to guide the subsequent discussion.

1.1

Research methodology

As has been claimed, both qualitative and quantitative methods will be applied in this comparability study about CET-6 and TEM-4. These two types of research traditions, once called research paradigms, are equally important for the present purpose. As Davis once pointed out, "the history of scientific inquiry should convince us that, in either the physical or social

sciences, research paradigms do not actually compete in scientific discourse" (quoted from Lakatos, 1978); rather, different paradigms can serve for different purposes (Davis, 1995). The combined utilization of the two approaches for the study is thus believed hopefully to generate results of optimal objectivity as well as plain credibility.

This belief of the author has actually resulted from an extensive search for information about the evolution of methodological theories. In social science in general, a broad range of research perspectives has evolved over nearly a century of scholarly interest in applied linguistic (including testing and SLA) issues. The various approaches employed by applied linguistics for research are the consequence of the particular philosophical and theoretical considerations that they have consciously or unconsciously adopted. In this way, a number of parallel research movements have developed and some of them have tended to remain separate (Davis, 1995 : 427). There have been exceptions, though. A good case in point is the so-called dialog between supporters of quantitative research methodology and qualitative research methodology dozens of years ago.

1. 1.1 *The paradigm dialog*

During the 60s and 80s of the last century, there were two events that exerted a far-reaching influence upon educational research (Liu, 1989). One was the birth of the computer, and the other was a major debate in education and psychology between researchers for positivistic, quantitative methodology

and naturalistic (later relabeled as constructivist), qualitative methodology (Lynch, 1996:13). The debate was just a typical example of research movements, first separate and then related to some extent. This was because both sides finally realized that “at the core of this debate is a discussion of the epistemological basis for research methodology” (Lynch, 1996:13) in general.

Before and during the debate, there was a strong tendency to favour a quantitative and experimental approach to conducting research. Henning noticed in 1986 that many articles exemplified quantitative research, and he called this “a positive development — a kind of coming of age of a discipline” (1986:704). The quantitative research began with the ontological (as opposed to epistemological) assertion that the reality is objective and facts can and must be separated from values (Lynch, 1996:15). With such a philosophical stand, researchers attempted to gain objective data by controlling human and other extraneous variables and thus gained what they considered to be reliable, hard data and replicable findings with statistical analyses. Their findings were usually generalized beyond the sample subjects to those throughout the population from which the sample was drawn (Davis, 1995:428). These characteristics thus comprised the positivistic paradigm, stemming, as it did, from logical positivism.

As an alternative approach to this inquiry, the qualitative approaches, or the naturalistic paradigm, have been seen to challenge the traditional authority of positivistic research. Stemming from phenomenology and the interpretive approach to social inquiry that developed in the late nineteenth century, this naturalistic inquiry was shaped by the belief that reality is not objective, or that there can be no meaningful separation of