

意愿与现实： 中国高等院校统一招生 英语考试的反拨作用研究

The Intended Washback Effect of the National
Matriculation English Test in China:
Intentions and Reality

● 亓鲁霞 著

外语教学与研究出版社
FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

意愿与现实： 中国高等院校统一招生 英语考试的反拨作用研究

The Intended Washback Effect of the National
Matriculation English Test in China:
Intentions and Reality

◎ 亓鲁霞 著

外语教学与研究出版社

FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

北京 BEIJING

744003

(京)新登字 155 号

图书在版编目(CIP)数据

意愿与现实:中国高等院校统一招生英语考试的反拨作用研究/亓鲁霞著.
—北京:外语教学与研究出版社, 2004.3
ISBN 7-5600-4050-0

I. 意… II. 亓… III. 英语—高等学校—入学考试—研究—中国—英文
IV. G634.413

中国版本图书馆 CIP 数据核字(2004)第 016394 号

意愿与现实:

中国高等院校统一招生英语考试的反拨作用研究

亓鲁霞 著

* * *

责任编辑:申 葳

出版发行:外语教学与研究出版社

社 址:北京市西三环北路 19 号 (100089)

网 址: <http://www.fltrp.com>

印 刷:北京市鑫霸印务有限公司

开 本:880×1230 1/32

印 张:13.375

版 次:2004 年 5 月第 1 版 2004 年 5 月第 1 次印刷

书 号:ISBN 7-5600-4050-0/H·1994

定 价:18.90 元

* * *

如有印刷、装订质量问题出版社负责调换

制售盗版必究 举报查实奖励

版权保护办公室举报电话:(010)88817519

ABSTRACT

The present study, which adopted the approach of use-oriented language testing proposed by Shohamy (2001), investigated the intended washback effect of the National Matriculation English Test (NMET) in China. Specifically, it aims to find out whether the test has fulfilled its mission to induce an intended effect, which is to make English language teaching in secondary schools shift from its previous focus on language form to an emphasis on language use. In the study, factors that contribute to the test's success or failure in this mission were scrutinized. A multi-method design was contrived, using interview, observation, and questionnaire to collect data from eight NMET constructors, six English inspectors, 388 secondary school teachers, and 986 students.

The results show that the NMET has largely failed to achieve the intended washback effect. The failure is attributable mainly to an inherent conflict between the two functions of the test. One function is to select students for higher education. The other function is to bring about changes in teaching and learning, which is the intended washback of the test per se. While injecting high-stakes into the NMET and empowering it to effect educational reform, the selection function also imposes constraints on the test design and production, and thereby hinders realization of the intended washback effect. In addition, the pressure arising from the testing process encourages teaching and learning mainly to achieve higher scores rather than to develop the desired ability to use language effectively. These findings demonstrate that a high-stakes test is a powerful encouragement to "teaching to the test", but an ineffective means to promote learning

or development of the educationally desired knowledge and ability.

Other factors that have been found to interact with the NMET to shape teaching and learning include teachers' own language proficiency and learning experience, and misuses of test results to evaluate schools and teachers.

On the basis of the findings of the present study as well as other washback studies, it is argued that the potentiality of a test to achieve strong intended washback effects is determined, to a large degree, by the stakes attached to the test. Neither low-stakes tests nor extremely high-stakes tests produce high intended washback effects. The tests that are likely to succeed in bringing about intended effects on teaching and learning are those whose stakes are at the right level. The stakes are high enough to make users pay attention to the message encoded in the tests but not so high as to trigger intense test preparations at the expense of teaching and learning to the objectives laid down by the curriculum.

The study also suggests ways to modify the NMET with a view to facilitating intended washback effects and minimizing unwanted side effects on teaching and learning.

ACKNOWLEDGEMENTS

The study reported in this book is my Ph. D. research carried out at the City University of Hong Kong. Many people have contributed in various ways to the completion of the research and the book. Sincerely, I would like to express my heartfelt thanks to the following people.

First of all, I would like to thank my two supervisors, Dr. Alastair Allan and Dr. Jonathan Zhu for their advice and support, both in the early and later stages of my research. Thanks are also due to the other two members of my supervision panel, Dr. Mathew Peacock and Dr. Ken Rose, both of whom made valuable comments on my ongoing work. I am also grateful to Professor Charles Alderson, Professor Liz Hamp-Lyons, and Dr. Angel Lin, the examiners of my Ph. D. orals, for their critical and constructive comments on an earlier draft of the present book.

My thanks also goes to Professor Li Xiaojun, one of the designers of the National Matriculation English Test, who generously shared information with me, and also Dr. He Zhou for his time and helpful discussions. Dr. Carol MacLennan also provided valuable comments and constructive feedback. Other colleagues and friends who have helped in one way or another include Dr. Ouyang, Dr. Liu Jianda, Ms Ou Aihua, Dr. Rodney Jones, Dr. Betty Li, Dr. Charles Man, Dr. Jin Jianbin, and Dr. Wu Xudong.

The participants of the study—the test constructors, the English inspectors, teachers, and students—were kind enough to set aside much of their precious time to talk to me in the interviews and complete the questionnaires. This was much appreciated. Some

ACKNOWLEDGEMENTS

teachers also let me sit in on their classes to carry out observations. Their generous help and cooperation have contributed significantly to the completion of this study.

I am grateful to the Drs. Richard Charles Lee and Esther Yew-pick Lee Charitable Foundation which awarded me the scholarship that made my Ph. D. studies possible. City University of Hong Kong has not only offered me a place to do the studies but also provided excellent service in every way, the library service in particular.

My family members have been extremely considerate and supportive. My husband frequently discussed my research with me, spot-checked my transcripts and coding, gave me useful feedback, and encouraged me to persevere during the course of the research. He also shouldered all family responsibilities during my absence from home. My 99-year-old grandmother, my elderly mother, and teenage daughter accepted my absence and neglect of their interests without complaint during the years of my studies.

Without the guidance, support, and help from all the people above, it would not have been possible for me to complete the research reported here. I am deeply indebted to them all.

TABLE OF CONTENTS

| | |
|-----------------------|----|
| List of Tables | ix |
| List of Figures | xi |

| | |
|--|----------|
| Chapter One Introduction | 1 |
| 1.1 Research Problem | 2 |
| 1.2 The Setting of the Study | 4 |
| 1.2.1 The Educational System in China | 4 |
| 1.2.2 Key Schools | 5 |
| 1.2.3 Curriculum and Teaching Materials | 10 |
| 1.2.4 The Role of Examinations | 11 |
| 1.2.5 The National Matriculation English Test as a Selection Tool | 12 |
| 1.2.6 The National Matriculation English Test as a Change Agent | 14 |
| 1.3 Significance of the Study | 17 |
| 1.4 Definition of Terms Referring to Test Effect | 18 |
| 1.5 Organization of the Thesis | 19 |

| | |
|---|-----------|
| Chapter Two Review of the Literature | 21 |
| 2.1 Empowerment of Tests; Origin of Washback | 21 |
| 2.2 Role of High-Stakes Tests in Educational Reform | 24 |
| 2.2.1. Test as a Spur for Action | 24 |
| 2.2.2. Test as a Model for Instruction | 26 |
| 2.2.2.1 Finely tuned models | 26 |
| 2.2.2.2 Roughly tuned models | 28 |
| 2.2.3 Criticisms of High-stakes Tests | 29 |
| 2.3 Conceptualization of Washback/Impact | 30 |

| | |
|---|----|
| 2.3.1 The Debate over Washback/Impact as an Aspect of Validity | 30 |
| 2.3.2 Use-oriented Testing and Washback | 33 |
| 2.3.3 The Scope, Pattern and Mechanisms of Washback | 36 |
| 2.4 Empirical Research on Test Impact or Washback | 41 |
| 2.4.1 General Education | 41 |
| 2.4.2 Language Education | 48 |
| 2.5 A Critique of Approaches and Methods in Washback/Impact Studies | 59 |
| 2.5.1 Research Approaches | 59 |
| 2.5.2 Research Methods | 64 |
| 2.6 Summary | 66 |

Chapter Three Method 71

| | |
|--|----|
| 3.1 Conceptual Background | 71 |
| 3.2 Research Design | 73 |
| 3.3 Participants | 76 |
| 3.3.1 Test Constructors | 76 |
| 3.3.2 English Inspectors | 77 |
| 3.3.3 Teachers | 78 |
| 3.3.3.1 Sampling teachers for qualitative data collection | 78 |
| 3.3.3.2 Sampling teachers for quantitative data collection | 81 |
| 3.3.4 Students | 82 |
| 3.3.5 The Researcher | 83 |
| 3.4 Instruments | 85 |
| 3.4.1 Interview | 85 |
| 3.4.1.1 Data collection procedures | 85 |
| 3.4.1.2 Data analysis | 88 |

1

TABLE OF CONTENTS

| | |
|--|-----|
| 3. 4. 3. 7. 3 Measurement of frequency in which different content areas were practiced in class | 121 |
| 3. 4. 3. 7. 4 Measurement of frequency in which different content areas were practiced after class | 124 |
| 3. 4. 3. 7. 5 Measurement of beliefs on different aspects of writing | 126 |
| 3. 4. 3. 7. 6 Comparison of teachers' and students' responses to the same items | 127 |
| 3. 5 Checklist | 129 |
| 3. 6 Summary | 129 |

Chapter Four The General Intended Washback and the Actual School Practice—Analysis and Findings I

| | |
|--|-----|
| | 131 |
| 4. 1 Test Constructors' General Intentions and Major Measures Taken to Fulfill Them | 132 |
| 4. 1. 1 General Intentions | 132 |
| 4. 1. 2 Measures Taken to Realize Intentions | 140 |
| 4. 2 The Focal Points of the Senior III English Course | 146 |
| 4. 2. 1 The Chronological Structure of the Senior III English Course | 147 |
| 4. 2. 2 The Conceptual Structure of the Senior III English Course; Focus on Linguistic Knowledge and the NMET | 150 |
| 4. 2. 3 A Profile of the Senior III English Course | 155 |
| 4. 3 The Senior III English Course in Miniature—Classroom Observation Results | 156 |
| 4. 3. 1 Purposes of Classroom Observation and Lessons | |

TABLE OF CONTENTS

| | |
|--|-----|
| Observed | 156 |
| 4. 3. 2 Linguistic Knowledge and NMET Skills Drilling | |
| Reflected in Teaching Content, Method, and Materials | 159 |
| 4. 3. 2. 1 Review lessons | 164 |
| 4. 3. 2. 2 NMET section training and mock tests | 168 |
| 4. 3. 2. 3 New lessons | 171 |
| 4. 3. 3 Overt Attention on Meaning, Form, and Test-taking | |
| Strategies in the Classes Observed | 173 |
| 4. 4 Extending the Generalizability of the Findings about the | |
| Senior III English Courses—Questionnaire Results | 177 |
| 4. 4. 1 Teaching Materials | 177 |
| 4. 4. 2 A Comparison of Teachers' and Students' Responses to Items about Activities and Exercises in and after Class | 181 |
| 4. 4. 3 Focus of Content Areas of Teaching | 184 |
| 4. 5 A Comparison between the Test Constructors' General | |
| Intention and School Practice | 190 |
| 4. 6 Summary | 193 |

Chapter Five The Intended Washback Effect of the NMET Writing Task—Analysis and Findings II

| | |
|---|-----|
| | 195 |
| 5. 1 Characteristics of the NMET Writing Task | |
| 5. 1. 1 Communicative Features of the Writing Task | 195 |
| 5. 1. 2 Encouraging Process Writing | 197 |
| 5. 2 Writing Behavior and Beliefs—Interview Results | |
| 5. 2. 1 Writing Practice | 198 |
| 5. 2. 2 Views on Writing | 199 |
| 5. 3 Context of Writing Neglected by Teachers and Students— | |

TABLE OF CONTENTS

| | |
|--|-----|
| Questionnaire Results | 203 |
| 5. 4 Where Does Attention Go in a Writing Class? —Classroom | |
| Observation Results | 211 |
| 5. 4. 1 Writing Activities in Class | 212 |
| 5. 4. 2 Explicit Focus of Attention on Different Aspects of | |
| Writing | 214 |
| 5. 4. 2. 1 An example of unsuccessful teaching of | |
| writing appropriately in relation to the | |
| rhetorical context | 215 |
| 5. 4. 2. 2 An example of stressing accuracy in writing | 219 |
| 5. 5 Summary | 223 |

Chapter Six Factors that Hinder the Intended Washback—

Analysis and Findings III

| | |
|---|-----|
| 6. 1 The Selecting Function Hinders the Directing Function | 227 |
| 6. 1. 1 Restrictions on Test Format | 228 |
| 6. 1. 2 Necessity of Norm-referencing | 231 |
| 6. 1. 3 Restrictions on Vocabulary | 233 |
| 6. 2 The Selecting Function Induces the Evaluating Function | 234 |
| 6. 3 Relationship between the NMET's Three Uses and Senior | |
| III English Teaching and Learning | 235 |
| 6. 3. 1 The Selecting Function of the NMET Reinforces | |
| the Goal to Raise Scores | 237 |
| 6. 3. 2 The Evaluating Function of the NMET Reinforces | |
| the Goal to Raise Scores | 240 |
| 6. 3. 3 The Directing Function of the NMET Is Enfeebled | 243 |
| 6. 4 Teaching to the Test and the Intended Washback | 245 |
| 6. 4. 1 Operationalizing Test Construct | 245 |
| 6. 4. 2 Teaching to the Test Content | 251 |
| 6. 4. 3 Teaching to the Test Format | 253 |

TABLE OF CONTENTS

| | |
|--|------------|
| 6. 4. 4 Automatizing Test Performance | 254 |
| 6. 5 Summary | 256 |
| Chapter Seven Discussion and Conclusions | 259 |
| 7. 1 Limitations of the Study | 259 |
| 7. 2 Discussion | 262 |
| 7. 2. 1 Intended Effect of the NMET | 262 |
| 7. 2. 2 Undesirable Prolonged Test Preparations | 265 |
| 7. 2. 3 Finely Tuned Model Tests and Teaching | 267 |
| 7. 2. 4 High Stakes Hindering Intended Washback Effects | 269 |
| 7. 2. 5 Redefining the Stakes of Tests | 271 |
| 7. 2. 6 Assumption of a Linear Relationship between the Stakes of a Test and the Strength of Washback | 277 |
| 7. 2. 7 Three Conditions for Testing Intended Washback Effects | 278 |
| 7. 2. 8 Inverted U-curve for a Relationship between Test Stakes and Intended Washback | 283 |
| 7. 2. 9 An Analysis of IELTS' and TOEFL's Washback in Light of the Three Conditions | 287 |
| 7. 2. 9. 1 The intentions of IELTS and TOEFL designers | 288 |
| 7. 2. 9. 2 Why was one IELTS preparation course more communicative than the other? | 289 |
| 7. 2. 9. 3 Why were TOEFL classes less communicative than non-TOEFL classes? | 294 |
| 7. 2. 9. 4 Summary | 296 |
| 7. 3 Conclusion | 296 |
| 7. 4 Implications | 297 |
| 7. 4. 1 Theoretical Implications | 297 |
| 7. 4. 2 Practical Implications | 301 |
| 7. 5 Contributions of the Present Study | 305 |

TABLE OF CONTENTS

| | |
|--|-----|
| 7. 6 Further Studies | 305 |
| References | 309 |
| Appendix A Interview Guide | 323 |
| Appendix B Observation Scheme for Senior III English Lessons | 327 |
| Appendix C Observation Scheme for Writing Lessons | 331 |
| Appendix D English Version of Questionnaire for Teachers | 335 |
| Appendix E English Version of Questionnaire for Students | 343 |
| Appendix F Chinese Version of Questionnaire for Teachers | 349 |
| Appendix G Chinese Version of Questionnaire for Students | 357 |
| Appendix H Checklist for Application of General NMET Item Setting Principles | 363 |
| Appendix I NMET (1995) | 383 |
| Appendix J Words for Memorization | 401 |
| Appendix K Sample MC Items Used in Class to Complement Reading Exercises in the Senior III English Textbooks | 405 |
| Appendix L A Sample Page of NMET Item Writers' Guidelines (1998) | 409 |

LIST OF TABLES

| | | |
|----------|---|-----|
| Table 1 | Promotion rate of school graduates by level | 5 |
| Table 2 | Studies on impact in general education | 42 |
| Table 3 | Studies on washback in language education | 50 |
| Table 4 | A multi-method research design for the washback project | 74 |
| Table 5 | Participants of the Study | 76 |
| Table 6 | Factor analysis results of 14 items on activities and exercises in class | 122 |
| Table 7 | Reliability of items on different content areas of teaching in class | 123 |
| Table 8 | Factor analysis of 8 items on activities and exercises after class | 125 |
| Table 9 | Reliability of items on different content areas practiced after class | 125 |
| Table 10 | Factor analysis of items on views on different aspects of writing | 126 |
| Table 11 | Mean ratings by teachers and students on the items on writing | 128 |
| Table 12 | The National Matriculation English Test | 143 |
| Table 13 | The chronological structure of the Senior III English course | 148 |
| Table 14 | Lessons observed | 158 |
| Table 15 | Materials used in the observed classes | 160 |
| Table 16 | Teacher and student behavior in class | 161 |
| Table 17 | Percentage of class time on different content areas | 162 |
| Table 18 | Attention to form, meaning, and test-taking strategies | 175 |

LIST OF TABLES

| | | |
|----------|--|-----|
| Table 19 | Materials used in Senior III English courses | 178 |
| Table 20 | How Senior III English Textbooks were used | 179 |
| Table 21 | Teachers' and students' mean ratings on activities and exercises in and after class | 182 |
| Table 22 | Frequency of different class activities | 185 |
| Table 23 | Mean of content areas of teaching | 186 |
| Table 24 | Comparisons of different content areas of teaching | 186 |
| Table 25 | Frequency of off-class activities and exercises | 187 |
| Table 26 | The amount of NMET skills drilling in a school year and how it was done | 189 |
| Table 27 | Views on writing of the test constructors, teachers, and students | 200 |
| Table 28 | Items on beliefs about writing | 204 |
| Table 29 | Nine teachers' writing views (counts) | 205 |
| Table 30 | Teachers' and students' writing views | 207 |
| Table 31 | Frequencies of item non-response | 209 |
| Table 32 | Class time spent on different writing activities | 212 |
| Table 33 | Explicit attention to different aspects of writing | 214 |
| Table 34 | The two goals of ELT at secondary schools | 236 |
| Table 35 | Students' reasons for having a tertiary education ... | 239 |
| Table 36 | Effect of students' NMET scores on teachers' professional life | 243 |
| Table 37 | Teaching to the test for higher scores | 244 |
| Table 38 | The stakes of different tests | 273 |
| Table 39 | Capacity of tests with differential levels of stakes to meet the conditions for intended washback | 282 |