Also in this series:

- Verbal Protocol Analysis in Language Testing Research: A handbook
  Alison Green

- Multilingual Glossary of Language Testing Terms
  Prepared by ALTE members

- Dictionary of Language Testing
  Alan Davies, Annie Brown, Cathie Elder, Kathryn Hill, Tom Lumley, Tim McNamara

- Learner Strategy Use and Performance on Language Tests: A structural equation modeling approach
  James E Purpura

- Fairness and Validation in Language Assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida
  Antony John Kunnan

- Issues in Computer-adaptive Testing of Reading Proficiency
  Micheline Chalhoub-Deville

- Experimenting with Uncertainty: Essays in honour of Alan Davies
  Edited by A Brown, C Elder, N Iwashita, E Grove, K Hill, T Lumley, K O'Loughlin, T McNamara

- An Empirical Investigation of the Componentiality of L2 Reading in English for Academic Purposes
  Edited by Cyril J Weir, Yang Huizhong, Jin Yan

- The Equivalence of Direct and Semi-direct Speaking Tests
  Kieran O'Loughlin

- A Qualitative Approach to the Validation of Oral Language Tests
  Anne Lazaraton

- Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002
  Edited by Cyril Weir and Michael Milanovic

  Edited by Cyril Weir and Michael Milanovic

- A Modular Approach to Testing English Language Skills: The development of the Certificates in English Language Skills (CELS) examinations
  Roger Hawkey

- Testing the Spoken English of Young Norwegians: A study of testing validity and the role of 'smallwords' in contributing to pupils' fluency
  Angela Hasselgreen

- Changing Language Teaching through Language Testing: A washback study
  Liying Cheng

- The Impact of High-stakes Examinations on Classroom Teaching: A case study using insights from testing and innovation theory
  Dianne Wall

  Barry O'Sullivan

- Impact Theory and Practice: Studies of the IELTS test and Progetto Lingue 2000
  Roger Hawkey

- IELTS Collected Papers: Research in speaking and writing assessment
  Edited by Lynda Taylor and Peter Falvey
IELTS Washback in Context

Preparation for academic writing in higher education

Anthony Green
For Sachiyo
# Contents

<table>
<thead>
<tr>
<th>Acknowledgements</th>
<th>viii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Series Editors’ note</td>
<td>ix</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>xiii</td>
</tr>
<tr>
<td>1 Wagging the dog? Towards a model of washback</td>
<td>1</td>
</tr>
<tr>
<td>2 Academic writing</td>
<td>32</td>
</tr>
<tr>
<td>3 Pilot studies</td>
<td>68</td>
</tr>
<tr>
<td>4 Main study methods and instruments</td>
<td>93</td>
</tr>
<tr>
<td>5 Main study results I</td>
<td>163</td>
</tr>
<tr>
<td>6 Main study results II</td>
<td>247</td>
</tr>
<tr>
<td>7 Conclusions and implications</td>
<td>296</td>
</tr>
</tbody>
</table>

**Appendices**

- Appendix 1: Research questions and methods summary 316
- Appendix 2: IELTS Awareness Form A 319
- Appendix 3: IELTS Awareness Form B 321
- Appendix 4: Teacher Questionnaire 323
- Appendix 5: Student Questionnaire A 332
- Appendix 6: Student Questionnaire B 342
- Appendix 7: Observation schedule 350

**References** 356

**Index** 380
I would like to express my thanks to the many individuals who have been involved with this volume as research participants, reviewers, colleagues and editors and who have supported and encouraged me in bringing it to print. Too many people have contributed for me to list them all by name, but I would particularly like to acknowledge the help of the following:

My PhD supervisors Cyril Weir and Roger Hawkey for their assistance and insightful criticism throughout the project and Arthur Hughes for getting me started and for encouragement during the early stages.

The British Council for funding elements of the study through the Joint Funded Research Programme and University of Cambridge ESOL Examinations for providing score data and forms of the IELTS Academic Writing test. In particular, I would like to thank Lynda Taylor, Rowena Akinyemi, Nick Saville and Mike Milanovic at Cambridge ESOL for their help with the manuscript and Chris Banks and Nick Charge for help with obtaining data and IELTS test forms.

The Department of Applied Statistics at the University of Reading for assistance with the choice of statistical analyses and to the many members of staff and fellow students at Roehampton and Reading who provided insights and support and who provided invaluable help with collecting the data.

Above all I am most grateful to the many students and teachers who participated in the study and without whose co-operation the research could not have taken place. I hope that this work will be considered worthy of their involvement.

The publishers are grateful to the copyright holders for permission to use the copyright material reproduced in this book. Cambridge University Press for Tables 2.1 and 7.1 from Assessing Writing by Sara Cushing Weigle, 2002. IELTS Australia for Table 2.2 from IELTS Research Reports (Volume 1), 1998.
The International English Language Testing System (IELTS) is a test of English for academic and vocational purposes managed jointly by three partners: University of Cambridge ESOL Examinations (a division of Cambridge Assessment), British Council and IDP:IELTS Australia. The test measures ‘the language ability of candidates who need to study or work where English is the language of communication’ (IELTS Handbook 2006: 4).

IELTS consists of four modules: Listening, Reading, Writing and Speaking. All candidates take the same Listening and Speaking modules, but there is a choice of Academic or General Training Reading and Writing Modules. If a candidate intends to enter undergraduate or postgraduate courses, they are advised to take the Academic Modules. If a candidate intends to continue their secondary education in English, to undertake work experience or training, or to emigrate, they are normally advised to take the General Training Modules. Test scores are reported, both for overall performance and on each of the four modules, in the form of ‘bands’ at nine defined levels from Non User to Expert User.

In recent years the IELTS candidature has grown exponentially, reflecting the increasing recognition of the test by higher education institutions and professional registration bodies around the world and by immigration authorities in Canada, Australasia and the UK. In 2006, over 700,000 candidates took the test and 70% of these entered for the Academic Modules. Most Academic Module candidates reported that they were taking the IELTS test for the purpose of entering higher education.

It is now commonplace to regard validity as a unitary concept with theory-based, context and criterion-related validation processes all having a part to play in contributing evidence to support the interpretation of test scores. High-stakes test providers such as Cambridge ESOL are also concerned with the ethical dimension of testing in terms of the effects and impact of a test on individuals and society and place as much emphasis on social values and social consequences as on traditional validity concerns in any consideration of the validity of test scores.

It is increasingly recognised that examination boards have a major impact on educational processes and on society in general because their examinations have widespread recognition and cash-in value. Such impact is often perceived as occurring at the ‘macro’ or social and institutional level. Effects may also occur at the level of the individual and such intended positive or
unintended negative effects are normally referred to by the term ‘washback’
(or backwash). These effects are normally considered in relation to teaching
practice or learning outcomes.

The study of IELTS impact, a major long-term programme of research
initiated in 1995 by Cambridge ESOL, was among the earliest investigations
into consequential validity. The project addresses a number of the issues at
the macro level and is described in detail in Hawkey (2006). Green’s volume
focuses on micro issues of washback and as such can be seen as a valuable
complement.

Despite widespread lip service to the mantra of washback in the inter-
national testing community, until recently only a limited number of research
studies have been undertaken to investigate the effects of high-stakes lan-
guage tests on teaching and learning. Even fewer studies have followed
Messick’s call and empirically grounded the impact of such effects on learn-
ers’ resultant test performances.

The research reported by Green in this volume relates to work he con-
ducted on IELTS for his PhD dissertation between 2001–2004. Green was
looking at the influence of the IELTS Academic Writing Module on prepara-
tion for academic study and the equivalence between IELTS test preparation
and other forms of English for Academic Purposes directed at university
study. He investigated the general research question: Is the washback model
supported in relation to the role of the IELTS test in the context of preparation
for academic study in the UK? This general research question in turn implies a
number of related questions.

1. Given the commonalities and discrepancies between IELTS and the
EAP writing construct revealed in the literature review, do students and
teachers regard themselves as engaging in IELTS test preparation rather
than university preparation and do such beliefs give rise to practices, in
relation to IELTS, which fail to address the EAP writing construct?

2. Do practices on courses which are not driven by IELTS better reflect
this construct?

3. What are the characteristics of learners on different courses and how do
these relate to the characteristics of the IELTS test-taking population?

4. Do instructional alternatives at points on a continuum from IELTS-
driven to IELTS-unrelated EAP courses result in differential outcomes
in terms of:
   • gains in scores on the IELTS Academic Module?
   • linguistic (lexico-grammatical) proficiency gains?
   • academic awareness and study skills gains?

5. Do facets of learners’ individual differences interact with instructional
differences in predicting outcomes?
Green’s study informed the subsequent development of IELTS, contributing to the revision of the Writing module in 2005, in particular the rewording of task prompts. It confirmed the need for greater accessibility to information on the meaning of band scores; this has been provided in the form of increased information on score processing, reporting and interpretation being made on the IELTS website in recent years as well as the release in 2006 of the ‘IELTS Scores Explained’ DVD package. In addition, Green’s work contributes to a better understanding of the relationship between test scores, periods of study and language gain.

His study has serious implications for all end users of IELTS test results and indeed the results of other high-stakes tests. It raises serious doubts over the conventional wisdom concerning the amount of language support required by EAP learners at different proficiency levels (as measured by the test) in order for them to meet minimally acceptable standards required for English medium tertiary study.

In the broader context, it adds significantly to our knowledge of the complexity of the mechanisms through which washback occurs and provides the field with a model that embraces this. As such it provides a valuable framework for carrying out further research in this oft-neglected area.

Recent validation projects in Sri Lanka (Wall 2005), Australia (Burrows 1998) and Hong Kong (Cheng 2005) have also addressed the question of washback in a variety of settings. Their work has established that washback is not one-dimensional or easily manipulated and that innovations in testing will not inevitably lead to changes in the classroom. It is clear that the interpretation and uses made of assessment procedures are not solely determined by testers, but depend on interactions involving the test and participants (such as test takers, teachers, administrators, materials developers and policy makers) with implications for teaching and educational policy. Green’s work further grounds this critical relationship between the test, the test taker and test use.

Early research into washback from language tests (Hughes 1988, Khaniya 1990) was criticised for a lack of empirical data (Alderson and Wall 1993), relying instead on insights from interested participants. More recent research has therefore triangulated quantitative data with qualitative descriptions of educational practices derived through interviews, questionnaires and observations (Burrows 1998, Cheng 2005, Watanabe 1996). This has allowed the development of theoretical models of washback that recognise a wide variety of moderating variables interacting with test influence (Burrows 1998, Hughes 1993). At the same time, recent research has given insufficient attention to test design and learning outcomes and is therefore unable to relate the influence of tests on learning processes to test score gains. Green’s work helps fill this gap.

A further important contribution of Green’s work is the use he made of sophisticated quantitative and qualitative methods to investigate these
effects. His main study is unique in the comprehensive range of data collection methods (classroom observation, individual and focus group interviews, staff and student questionnaires, document analysis and a range of test instruments) which provide multiple channels for the different voices to be heard. A combination of quantitative and qualitative methods of enquiry provides new insights into how washback to learners may be mediated by washback to the teacher.

Traditional correlation-based techniques such as multiple regression are limited as tools for investigating the wide range of variables implicated in washback, while experimental methods, in attempting to isolate the influence of test method, would distort the complex social reality of the setting (Bailey 1996, Larsen-Freeman 1997). An alternative which allows for the simultaneous investigation of large numbers of variables in interaction in natural settings is provided by a neural network approach (Garson 1998) which allows for non-linearity and is more flexible in handling missing data. This is one of the few studies to employ neural network methods in tandem with traditional linear regression analysis to investigate the multifaceted relationships between presage, process and product variables.

Together with the three earlier volumes in this series on washback and impact – Volume 21 by Liying Cheng on the Hong Kong exam reforms, Volume 22 by Dianne Wall on the O level English examination in Sri Lanka, and Volume 24 by Roger Hawkey on the IELTS impact studies – this volume enriches our understanding of an under-researched area of validity and helps further ground the methodologies for investigating it. Volume 25 also complements two other IELTS-focused volumes in this series: one is a collection of research papers on IELTS Speaking and Writing (Volume 19, edited by Lynda Taylor and Peter Falvey, 2007), and the other is a historical overview of the development of ELTS and IELTS (Volume 23, by Alan Davies, forthcoming).

Cyril Weir
Mike Milanovic
January 2007
Abbreviations

ANOVA Analysis of Variance
ARELS Association of Recognised English Language Schools
ASI Approaches to Studying Inventory
AWL Academic Word List
AWM (IELTS) Academic Writing Module
BALEAP British Association of Lecturers in English for Academic Purposes
BASELT British Association of State English Language Teaching
BICS Basic Interpersonal Communicative Skills
CAE Certificate in Advanced English
CALP Cognitive Academic Language Proficiency
COLT Communicative Orientation to Language Teaching
CPE Certificate of Proficiency in English
CSWE Certificates in Spoken and Written English
EAC English through Academic Contexts
EAP English for Academic Purposes
EGAP English for General Academic Purposes
ELTS English Language Testing Service
ESAP English for Specific Academic Purposes
ESL English as a Second Language
GSL General Service Word List
IELTS International English Language Testing System
IIS IELTS Impact Study
IRT Item Response Theory
LSP Language for Specific Purposes
MANOVA Multivariate Analysis of Variance
MCQ Multiple Choice Question
MDI Measurement Driven Instruction
MGLH Multivariate General Linear Hypothesis
MLP Multi-layer Perceptrons
MSE Mean Squared Error
MTELPMichigan Test of English Language Proficiency
NNS Non Native Speaker
NS Native Speaker
PEPS Productivity Environmental Preference Survey
PEs Processing Elements
Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLSPQ</td>
<td>Perceptual Learning Style Preference Questionnaire</td>
</tr>
<tr>
<td>SILL</td>
<td>Strategy Inventory for Language Learning</td>
</tr>
<tr>
<td>SLA</td>
<td>Second Language Acquisition</td>
</tr>
<tr>
<td>SLC</td>
<td>Nepalese School Leaving Certificate</td>
</tr>
<tr>
<td>SOLO</td>
<td>Structure of Observed Learning Outcomes</td>
</tr>
<tr>
<td>SPQ</td>
<td>Study Processes Questionnaire</td>
</tr>
<tr>
<td>TEEP</td>
<td>Test of English for Educational Purposes</td>
</tr>
<tr>
<td>TL</td>
<td>Target Language</td>
</tr>
<tr>
<td>TLU</td>
<td>Target Language Use</td>
</tr>
<tr>
<td>TOEFL</td>
<td>Test of English as a Foreign Language</td>
</tr>
<tr>
<td>TWE</td>
<td>Test of Written English</td>
</tr>
<tr>
<td>UWL</td>
<td>University Word List</td>
</tr>
<tr>
<td>VLT</td>
<td>Vocabulary Levels Test</td>
</tr>
</tbody>
</table>