Studies in Language Testing 24

Impact Theory and Practice

Roger Hawkey

Series Editors
Michael Milanovic
and Cyril J Weir

UNIVERSITY *of* CAMBRIDGE
ESOL Examinations

CAMBRIDGE
UNIVERSITY PRESS

Impact Theory and Practice

Studies of the IELTS test
and *Progetto Lingue 2000*

Also in this series:

Performance testing, cognition and assessment: Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem
Edited by Michael Milanovic and Nick Saville

The development of IELTS: A study of the effect of background knowledge on reading comprehension
Caroline Clapham

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

European Language Testing in a Global Context: Proceedings of the ALTE Barcelona Conference July 2001
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

Issues in testing business English: The revision of the Cambridge Business English Certificates
Barry O'Sullivan

Impact Theory and Practice

Studies of the IELTS test and *Progetto Lingue 2000*

Roger Hawkey
ESOL Consultant

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

www.cambridge.org
Information on this title: www.cambridge.org/9780521680974

First published 2006

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN -13 978-0-521-68097-4
ISBN -10 0-521-68097-2

Contents

Acknowledgements

I am sincerely grateful to the many people who helped me, in various ways, with the impact studies described here, and with the writing and editing of the book.

I was fortunate enough to be able to meet, at Lancaster, members of the Lancaster University team who played such a key part in the early phases of the study of IELTS impact. From Charles Alderson, Professor of Linguistics and English Language Education in the Department of Linguistics and English Language, Dr Jayanti Banerjee, researcher in language testing and assessment and English for Academic Purposes (EAP), and Tania Horak, Research Associate currently working on the New TOEFL (Test of English as Foreign Language), I received valuable insights and advice on both historical and current aspects of the study of impact. My thanks, too, to Dr Dianne Wall, also at my Lancaster meetings, for sharing her considerable and continuing experience in the study of test impact.

The second group of individuals whose help and support I should like to acknowledge played key roles during the implementation stages of the studies. Dr Raffaele Sanzo, Co-ordinator of the *PL2000* in the Ministry of Education, Italy, not only helped to facilitate the Cambridge ESOL Impact Study, but offered encouragement and wise insights throughout its implementation. Richard Turner, of Cambridge ESOL Research and Validation, who held an expert camcorder during the video-ing of IELTS impact study preparation classes, also played a major role in the data management and analysis for both IELTS and *PL2000* studies. Liam Vint, Cambridge ESOL Country Manager for Italy, turned *PL2000* Impact Study camera person, too, in addition to providing vital contacts and background knowledge throughout planning and implementation stages. Lee Knapp, Cambridge Development Manager, UK, accompanied me on some of my IELTS impact study visits to UK centres and provided me with information and views from his many other IELTS-related contacts. My thanks also to Tony Green, Validation Officer at Cambridge ESOL, who shared data and validating operations with me for the IELTS study. I should like in addition, to express my gratitude to Susan Chapman, of Cambridge ESOL Research and Validation, who gave invaluable administrative support to both studies, and to Rowena Akinyemi and Sally Downes, who co-ordinated the final editing and production of the book.

I am grateful to those commissioned to read and make suggestions on the manuscript, namely Alan Davies, Emeritus Professor, Department of Theoretical and Applied Linguistics, University of Edinburgh and Peter Falvey, Consultant to Cambridge ESOL. To Nick Charge, Subject Manager for IELTS and Liam Vint, my thanks for reading and advising on parts of the book of particular relevance to their work. May I acknowledge, too, Dr Mike Milanovic, CEO of Cambridge ESOL, and Cyril Weir, Professor of English Language Acquisition at the University of Luton, for guidance, beyond their duties as SiLT series editors, on content and emphasis.

Finally I should like to acknowledge the constant interest in the project taken by Nick Saville and Lynda Taylor, Director and Assistant Director respectively, of the Cambridge ESOL Research and Validation Group. Their wise and helpful advice was based both on their interest in and readings of this book.

Last, and in some ways most, I should like to recognise all the many participants in the studies described in this book. To the students and their parents, the test takers, the teachers, the school and centre heads, the parents, the administrators and managers, including Cambridge ESOL Local Secretaries, involved in IELTS and the *Progetto Lingue 2000*, thank you for your co-operation and helpfulness, for your contribution to the good cause of impact study.

Series Editors' note

Cambridge has always taken a great interest in the educational and social impact of its tests and assessments. For many years while testing experts around the world were preoccupied with the quantitative aspects of assessment worrying about impact was not considered to be particularly relevant. The Cambridge attention to this area was probably even thought of as slightly quirky and old fashioned. However, in recent years, the concept of identifying and measuring how tests impact on the environment in which they operate has been recognised as a very relevant concern. Indeed, three further volumes in this series by Liying Cheng, who looks at washback in Hong Kong, Diane Wall, who documents an impact study in Sri Lanka, and Tony Green, who focuses on IELTS, demonstrate the growing importance of impact research as an aspect of test validation. Languages in general and English in particular, are of ever growing importance, not only for economic reasons, but also for social and political ones. Stakeholders in the language assessment process increasingly require evidence on the interactions between examinations, the stakeholders involved and the outcomes expected. The effective conceptualisation of the dimensions of test impact and its systematic study – within the context of test validation research – is one of the ways that will help us to address this requirement better.

This volume is written from the perspective of an international language testing agency although the issues discussed are of relevance in national and local assessment situations. Roger Hawkey, who has now conducted extensive work in the area of test impact, considers its dimensions and why understanding test impact is important. After some discussion of the concepts of impact and washback and how they fit into a broader educational, research and social context, he looks at the role of impact studies in the Cambridge ESOL test development, validation and revision systems, with particular reference to the *Progetto Lingue 2000* in Italy and the study of IELTS impact.

In the fields of language teaching and testing, the concepts of washback and impact, as Hawkey explores in some depth, are a matter of both theoretical and practical differentiation and concern. Through the 1980s and into the early 1990s attention focused on the concept of test washback and as such took a relatively narrow view, focusing largely on the teaching–learning relationship with some attention paid to the role of publishers and course materials. But beyond the learners and teachers affected by the washback of a language test

are a range of other stakeholders on whom an examination has impact even though they do not take the test or teach it. These stakeholders, for example, parents, employers, university admissions officers and others, form what we might refer to as the language testing constituency. Cambridge ESOL has defined this constituency particularly in relation to candidates taking its own examinations but the definition applies in other contexts too. Different tests will have different constituencies and an examination board like Cambridge ESOL will be dealing with numerous and varied constituencies, quite possibly for the same test and at the same time. The stakeholders interact with the test construct, format, conditions and assessment criteria in various ways.

Cambridge ESOL routinely conducts impact studies as part of the test validation process on an ongoing basis. It is our view that an examination board must be prepared to review and revise what it does in the light of how its stakeholders use and feel about its examinations. As educational processes and social needs change it is vital that examinations adapt to meet the requirements of these changes and the study of test impact facilitates this process even if the interrelationships involved are complex and highly context-dependent. Hawkey rightly points out that impact research is an exemplification of the growing importance of evidence-based approaches to education and assessment. Evidence-based education requires policy and practice capable of being justified in terms of sound evidence about their likely effects. Given that education, or indeed assessment, is not an exact science, it is too important to allow it to be determined by unfounded opinion, whether of politicians, teachers, researchers or anyone else. Validation research, including research into test impact, aims to seek out the evidence necessary to develop, redevelop or indeed operate a testing system in an appropriate and ethical manner.

The role of ethics in language testing has risen to the fore much more significantly in the last decade or so. The intention of those concerned with ethical language testing is to implement codes of professionally and socially responsible practice. These codes should provide tighter yet feasible guarantees of test development rigour and probity, with properly-defined targets, appropriate and reliable evaluation criteria, comprehensive, transparent and fair test interpretation and reporting systems, continuous validation processes, and a keener regard for the rights of candidates and other stakeholders (for example, see the ALTE Code of Practice, the ALTE quality assurance work, and the IELTS Handbook and Annual Review).

An ethical approach to language testing is a must in the modern world and test impact studies play an important role in demonstrating that language tests are used ethically. However, impact studies can also help address some of the concerns raised by the critical language testing lobby. The critical language

testing movement characterises tests as, intentionally or not, biased, undemocratic, and unfair means of selecting or policy-changing. It is argued that the main actual impact of language tests is the imposition of constraints, the restriction of curricula, and the possible encouragement of boring, mechanical teaching approaches.

Whether this is the case or not needs to be a matter of research rather than opinion and in such a general context a focus on test impact is an important area of study. It is driven by considerations in the field of language testing of wanting to do the job right and providing the appropriate evidence to back any claims. It is also driven by a broader social, political, educational and even cognitive impetus, and we see again the growing movement in education to develop the notion of basing what we do on sound evidence. Indeed, the evidence-based education manifesto argues that we need a culture in which evidence is valued over opinion, and where appropriate action (or inaction) is valued over action for the sake of being seen to do something. This applies just as much to what critical language testers have to say as it does to the claims of examination boards, education departments, schools and so on.

This volume is intended to provide the reader with an approach to the study of test impact which allows evidence to be gathered and displayed. It documents in some detail aspects of two impact studies that have been conducted in the Cambridge context and as such, we believe it makes a unique and much needed addition to the field. Its focus on the use of international assessments in state systems in the *Progetto Lingue 2000* is relevant as English becomes a core subject in many countries around the world, and it is vital that there is a good understanding of what impact international assessment may have. The focus on IELTS is no less significant as international mobility continues to increase. The extensive IELTS research takes us beyond a narrow focus on the test itself to the broader impact that it has and demonstrates very clearly that IELTS impacts positively on language learning and teaching in addition to its well known measurement attributes.

Two further volumes on IELTS will be published soon after this volume. The first, entitled *IELTS Collected Papers: Research in speaking and writing assessment* and edited by Lynda Taylor and Peter Falvey, documents a range of research studies with a particular focus on speaking and writing. The second, written by Alan Davies and entitled *Assessing Academic English: Testing English proficiency 1950–2005 – the IELTS solution*, documents the development of the testing of academic English from the 1950s to the present day.

Michael Milanovic
Cyril Weir
2005

1 Impact, washback, evaluation and related concepts: definitions and examples

Impact is the main topic of this book, so this opening chapter will attempt to clarify the concept and its implications. It will also consider related terms, for example *evaluation, monitoring* and *washback.*

Context for the discussion throughout will be via reference to the role of impact studies in the test development and validation systems of the University of Cambridge English for Speakers of Other Languages (ESOL) Examinations.

This book will refer to two Cambridge ESOL impact study projects in particular.

One is the study of the impact of the International English Language Testing System (IELTS), an examination 'designed to assess the English language ability of people whose first language is not English and who need to study, work or live where English is used as the language of communication' (www.ielts.org). The second is a study of the impact of the *Progetto Lingue 2000* (Year 2000 Languages Project), a Ministry of Education, Italy, state school foreign language education improvement programme.

This should set the scene for Chapter 2, which considers different approaches to the collection and analysis of impact data, and Chapter 3, on the definition of research objectives and questions. Chapter 4 then traces the development of impact study instrumentation, and Chapter 5 the collection, management and analysis of data. In Chapters 6 and 7, some of the main findings of the studies into IELTS and the *Progetto Lingue 2000* impacts are presented, in their own right and as examples of the outcomes that may be expected from research into the foreign language learning and testing aspects of educational impact. Chapter 8 traces research and other developments related to the two studies, considers lessons to be learned, and suggests approaches for the continuing study of educational impact.

But first some key *terms* need to be defined.

Impact in educational research

Impact of process and product

Taking an *educational evaluation* viewpoint, Weiss defines *impact* as 'the net effects of a programme (i.e. the gain in outcomes for program participants minus the gain for an equivalent group of non-participants)'(1998: 331). She then broadens this somewhat narrow definition by adding that 'impact may also refer to program effects for the larger community', and admitting that 'more generally it is a synonym for *outcome'* [all italics mine]. This wider view of the impact construct is reflected in a definition from *developmental education* studies:

> Impacts (also referred to as effects) may be planned or unplanned; positive or negative; achieved immediately or only after some time; and sustainable or unsustainable … Impacts may be observable/measurable during implementation, at project completion, or only some time after the project has ended. Different impacts may be experienced by different stakeholders (Department for International Development (DFID) Glossary of terms 1998).

Note that this definition of impact appears to include a focus on *processes* as well as *outcomes* or product, a distinction often at issue in impact and evaluation studies. Roy defines process and product studies as follows:

> A study of the product is expected to indicate the pay-off value while a study of the process is expected to indicate the intrinsic values of the programme. Both are needed, however, to find the worth of the programme (1998:71).

Weiss defines a process focus more straightforwardly as the study of 'what goes on while a program is in progress', whereas outcome studies measure and describe the 'end results of the program' (1998:334–335).

In the field of education, *impact studies* most commonly focus on the effects of interventions, including both teaching programmes and tests, on the people participating in them in various ways. Given the formative nature of education and learning, such studies seek to measure and analyse both outcomes, for example test results or subsequent performance on the criteria the test is measuring, and processes, for example the learning and teaching approaches and activities of programmes preparing candidates for a test.

The study of the impacts of the IELTS test is, by definition, a form of summative evaluation, concerned with outcomes such as candidate test performances. But a test such as IELTS, used as an English language qualification for academic studies in English-speaking countries and for immigration, training and employment purposes, also has significant potential impact on processes such as the ways candidates learn and prepare for the test

itself and for their English language activities beyond it (there is further discussion on this below). There are thus formative aspects (intended to provide information to improve programmes or tests) as well as summative aspects to impact studies. As for the *Progetto Lingue 2000* (*PL2000*) Impact Study, of a Ministry of Education project for the improvement of language learning in the state sector, there is a focus on developments in areas where the Project is intended to have influence. These include, of course, teaching/learning processes and foreign language performance outcomes. Impact studies of tests, like impact studies of learning programmes, are likely to be process- as well as product- or outcome-oriented.

Impact studies and evaluation studies

Varghese contrasts *impact* studies with *evaluation* studies, the latter, he feels, tending to focus more closely on the immediate objectives of projects rather than their longer-term development.

> … An evaluation of adult literacy programmes may indicate the total number of persons made literate by the programme. An impact study of the programme will focus on the social implications of the outcomes … It will also ask, for example, whether the reading habits of the community improved (1998:49).

Varghese (49–50) reminds us that impacts are changes (or effects) rather than the achievement of project targets, which are often seen as the focus of evaluation studies.

Weiss defines evaluation as the 'systematic assessment of the operation and/or outcomes of a program or policy, compared to explicit or implicit standards, in order to contribute to the improvement of the program or policy' (1998:330). The term 'evaluation' refers to an overall process; an evaluation study is, after all, an exercise to appraise (that is, measure the value of) an educational programme. Impact may well be *one of the areas* of the programme covered by an evaluation. So, evaluation and impact are linked, with evaluation in some cases tending to include impact, in the sense of programme effects which evaluators want to find out about as part of their evaluation. Why? In order to make proposals to adjust these effects to 'contribute to the improvement of the program or policy' (see Weiss above).

But the evaluation study literature (e.g. Agar 1986, Connell et al 1995, Cronbach 1982, MacDonald 1974, Parlett and Hamilton 1972, Weiss 1998) warns us regularly that the nature of evaluation should not be over-simplified as it was following Scriven's 1967 contrast between summative and formative evaluation. 'Evaluations that focus on outcomes' says Cronbach, 'can and should be used formatively' (1982). Parlett and Hamilton go further with their concept of 'illuminative evaluation' (1972). This has a primary concern 'with

description and interpretation rather than measurement and prediction', with how innovation operates, 'how it is influenced by the various school situations in which it is applied; what those directly concerned regard as its advantages and disadvantages; and how students' intellectual tasks and academic experiences are most affected'. Illuminative evaluation 'aims to discover and document what it is like to be participating in the scheme' (see Murphy and Torrance (eds.) 1987:60–61). Levine's concept of evolving curriculum is similarly dynamic, 'where evaluation is an inherent aspect of the curriculum planning process *(evaluation in planning)'* with 'the evaluation process itself a perpetual and self-developmental inquiry process *(evaluation as planning)*. The curriculum evaluation process that emerges is flexible, yet methodical, open, yet directive, *and* respectful of the diverse, complex curricular visions, needs and constraints encountered in schools and classrooms' (2002:26).

There would seem to be much to learn from these definitions. The impact studies discussed in this book attempt to combine 'description and interpretation' with 'measurement and prediction'. They seek to investigate the influences of 'the various school situations' in which the IELTS test is prepared for and the principles of the *PL2000* are put into practice. They certainly seek to discover 'what those directly concerned' regard as the 'advantages and disadvantages' of the test and the curriculum reform project, and also how students' 'intellectual tasks and academic experiences are most affected'. We shall also see, throughout this book, that the study of the impact of language tests or programmes, like the evaluation process, tends to be 'perpetual and self-developmental' rather than single and monolithic.

Monitoring

Then there is the term *monitoring,* clearly related to both impact and evaluation, and actually suggested by Italian colleagues participating in the study of the impact of the *PL2000*, as a synonym for impact study. Weiss defines monitoring as '[a]n ongoing assessment of program operations conducted during implementation, usually by sponsors or managers, to assess whether activities are being delivered as planned, are reaching the target populations, and are using resources appropriately' (1998:333). Judging by this definition, there is considerable overlap between monitoring and evaluation, but the fact that monitoring takes place only *during* the implementation of a programme may distinguish it.

A further distinction, suggested by Lynda Taylor (2005, personal communication) sees monitoring as primarily descriptive in function, followed by evaluation, which is, naturally, mainly evaluative in function. As will emerge from this chapter (see Figure 1.5 on page 20) the Cambridge

ESOL model of test development includes, following the establishment of the need for a new or revised test, the stages of:

- design and initial specification
- development through trialling, analysis, evaluation and review
- *monitoring*, mainly through routine descriptive data for analysis, until a decision is made, based on particular monitoring information
- reviewing and *evaluating* the test for possible further revision.

Weiss links monitoring with 'process evaluation', and adds a participant dimension:

> … process evaluation is not very different from what is often called monitoring. One key difference is that monitoring is done primarily on behalf of the funders and other high-level officials to hold the program to account (1998:181).

One teacher/administrator participant in the study of IELTS impact seemed to sense that impact studies may be less top-down and judgemental when she described them as more 'user-friendly' than evaluations or monitoring.

In the case of the *PL2000* Impact Study there was, of course, no question of the impact study 'funders', Cambridge ESOL, holding policy-makers, designers, managers or officials of the *PL2000* to account. Rather, the examinations board was concerned with the two-way impacts (see below) of Cambridge exams on participants in the *Progetto,* and of the *Progetto* on these exams. The *PL2000* Impact Study was carried out by Cambridge ESOL as an interested party, selected alongside other international test providers (see below), to provide external certification for students who had been participating in foreign language courses under the *PL2000*. Cambridge ESOL was not the initiator or leader of the foreign language reform project itself; its role, through the *PL2000* Impact Study which it ran with Ministry approval, was to describe impacts rather than to evaluate the Project.

Insider and outsider roles

The question of the 'evaluator-user and insider-outsider interface' is often at issue in the evaluation literature. O'Dwyer summarises as follows:

> Evaluators may remain distant and report findings in their own way with the expectation that these may be used to improve a program; or, may be actively involved in the program, working hand-in-hand with those in a program, or stakeholders to the program, with a view to specifying the evaluation focus according to the needs of the users. The profile which an external evaluator may adopt, therefore, could be of a complete outsider to a program, or, towards the other end of the spectrum, of a close 'insider' in relationship to the clients (2005).

The differences in evaluator roles described here would appear to apply to impact studies as well as evaluations. With the design, trialling and implementational phases of the study of IELTS impact, which will be described in detail in Chapters 3–6 below, the outsider-insider roles included, at various stages, both external consultant teams and individuals commissioned by Cambridge ESOL, and validation and implementation expertise from within the organisation. Cambridge ESOL is, of course, one of the three partners in the IELTS test, along with the British Council and IDP Education Australia : IELTS Australia. Both these latter partners are fully informed of the impact studies and themselves contribute to research in support of IELTS through the IELTS funded-research programme (see this chapter and Chapters 4, 6 and 8). In the *PL2000* Impact Study, it will be seen, relationships between the impact study team and participants such as the case study school teachers and heads were close, though not quite 'insider'.

Impact and washback in foreign language teaching and testing

In this section of the chapter, the concepts of evaluation, monitoring and impact are investigated within the fields of language teaching and testing, where similarities with and distinctions from the general educational literature will be discovered.

In the language teaching and testing literature, the concept of impact as effects or changes still stands but the term co-occurs frequently with the term 'washback' (or 'backwash') and it is the distinction between the two that is often an issue of debate. In the context of studies of the effects of language programmes or tests on those involved, the concepts of impact and washback/backwash are often considered in terms of their:

- logical location
- definition and scope
- positive and negative implications
- intentionality
- complexity
- direction
- intensity, emphasis
- stakes and stakeholders
- relationships with validity and validation
- relationships with the Critical Language Testing view
- role in impact/washback models.

This chapter attempts below to cover all these aspects of impact and washback.

Washback and impact

'Washback and the impact of tests more generally has become a major area of study within educational research' Alderson (2004a:ix) and as the washback and impact net widens, so does the need for agreed labels for the kinds of study we carry out to investigate the effects of tests or programmes *in and beyond the classroom context.* Hamp-Lyons summarises the situation and the terminology well. She finds that Alderson and Wall's 'limitation of the term 'washback' to influences on teaching, teachers, and learning (including curriculum and materials) seems now to be generally accepted, and the discussion of wider influences of tests is codified under the term 'impact' (Wall 1997), which is the term used in the wider educational measurement literature' (2000:586). In similar vein, Bachman and Palmer 1996 refer to issues of test use and social impact as 'macro' issues of impact, while washback takes place at the 'micro' level of participants, mainly learners and teachers.

So the term 'impact' now appears to be used to describe studies which investigate the influences of language programmes and/or tests on stakeholders *beyond* the immediate learning programme context. An impact study might, for example, investigate the effects of a programme or test on school heads, parents, receiving institution administrators, high-stakes test providers (all these stakeholders included in the two impact studies described in Chapters 3–8 below).

Given that the term 'impact' is a word in everyday use in its meaning of 'influence or effect' (e.g. *Oxford School Dictionary,* 1994), it is unsurprising to find the term also apparently used non-technically. When Alderson (2004a: ix), for example, writes: 'We now know, for instance, that tests will have more impact on the content of teaching and the materials that are used than they will on the teacher's methodology', is he using the term in its lay sense, since technically the content of teaching and the teacher's methodology are washback rather than impact matters? Or is he acknowledging that, for some, impact, the broader construct, *includes* washback? Green notes that although 'the terms have been used to refer to the same concept, backwash is distinguished from test impact by Bachman and Palmer (1996:30) who, with McNamara (1996, 2000), Hamp-Lyons (1998) and Shohamy (2001) place washback within the scope of impact' (2003:6). This would presumably mean that one could use the term 'impact' for all cases of influence from a language test or language programme, whether on teaching and learning or on, say, a university's admissions policy.

Andrews, writing on washback and curriculum innovation, appears to acknowledge the fragility of the washback : impact distinction:

> The term washback is interpreted broadly … the present chapter uses *washback* to refer to the effects of tests on teaching and learning, the educational system, and the various stakeholders in the education process. Where the word 'impact' occurs in this chapter, it is used in a non-technical sense, as a synonym for 'effect' (2004:37).

In this book we shall try to be consistent in the use of terms:

- to use 'washback' to cover influences of language tests or programmes on language learners and teachers, language learning and teaching processes (including materials) and outcomes
- to use 'impact' to cover influences of language tests or programmes on stakeholders beyond language learners, teachers, except when it is the influences of a test or programme on learners and teachers *outside* their learning or teaching roles, for example on their attitudes to matters beyond language learning; in this case the book will tend to refer to impact e.g. Research Question 4: What is the impact of IELTS on the participants who have taken the test?

In terms of these definitions, the two studies which are the focus of this book cover both washback and impact. They are called 'impact studies' because of this breadth.

Washback/backwash

Hamp-Lyons notes that washback 'is one of a set of terms that have been used in general education, language education and language testing to refer to a set of beliefs about the relationship between testing and teaching and learning' (1997:295). Another of the 'set of terms' is 'backwash', but it would appear that the terms 'washback' and 'backwash' are used interchangeably in the field. '… to clarify the distinction between the terms backwash and washback', Alderson says (2004a:xi), 'there is none'. Hughes admits that there is interchangeable use of the two terms in his work but adds, (2003:57) 'Where "washback" came from I do not know. What I do know is that I can find "backwash" in dictionaries, but not "washback"'. Cheng and Curtis choose to use the term 'washback' 'as it is the most commonly used in the field of applied linguistics' (2004:5). This book will follow suit, preferring the term 'washback' as it does now appear to be in more common use in the field.

Impact and validity

Saville and Hawkey cite 'the implementation of new national curricula with national achievement tests' (2004:75) in the UK and New Zealand as examples of a growing tendency for tests to be used to provide evidence of and targets for change, thus having more significant influences on the lives of individuals and groups.' Language tests such as IELTS are more and more frequently used in a 'gate-keeping' role in decisions of crucial importance to candidates such as the admission or otherwise to particular programmes, professions or places. They thus earn the label of 'high-stakes' tests. The social consequences of test use are a growing concern.

Messick insists on the inclusion of the outside influences or 'consequential validity' of a test in its validation, 'the function or outcome of the testing, being either interpretation or use' (1989:20). In an interesting personal communication to Alderson, however, Messick warns against too glib a view of the relationship between test washback or impact and test validation.

> Washback is a consequence of testing that bears on validity only if it can be evidentially shown to be an effect of the test and not of other forces operative on the educational scene … Washback is not simply good or bad teaching or learning practice that might occur with or without the test, but rather good or bad practice that is evidentially linked to the introduction of the use of the test (Alderson 1995:3).

Alderson (1995:4) himself takes 'an agnostic position' on the relationship between test impact and test validity. He agrees that 'test consequences are important and may relate to validity issues (bias being perhaps the most obvious case)' but has 'difficulty seeing the washback and impact as central to construct validity' because of the 'myriad factors' impacting on a test: teacher's linguistic ability, training, motivation, course hours, class size, extra lessons and so on. 'This is not, of course, to deny,' Alderson notes in his paper written for Phase One of the study of IELTS impact (see also Chapters 2 and 4), 'the value of studying test impact and washback in its own right, but it underscores the need to gather evidence for the relationship between a test and its impact on the one hand, and of the futility, given current understanding and data, of making direct and simplistic links between washback and validity'Alderson (1995:3).

Green agrees that backwash 'is not generally considered to be a standard for judging the validity of a test', because 'backwash can only be related to a test indirectly, as effects are realised through the interactions between, *inter alia,* the test, teachers and learners' (2003a). Green cites Mehrens (1998) on 'the lack of agreed standards for evaluating backwash' and the fact that 'different stakeholders may regard the same effects differently'. There is interesting food for thought in Messick's 1996 advice, also cited by Green: 'rather than seeking backwash as a sign of test validity, seek validity by design as a likely basis for backwash' (1996:252).

Hamp-Lyons links the increasing importance attached to tests to the washback/impact relationship, claiming that the 'shift from washback to impact suggests a growing awareness by language testers that the societies in which and for which we work are, whether we approve or not, using tests as their levers for social and educational reform' (2000:586). Actually, as Alan Davies points out (personal communication), this is by no means a new phenomenon, being a feature, for example, of the Civil Service examinations in India in the 19th century.

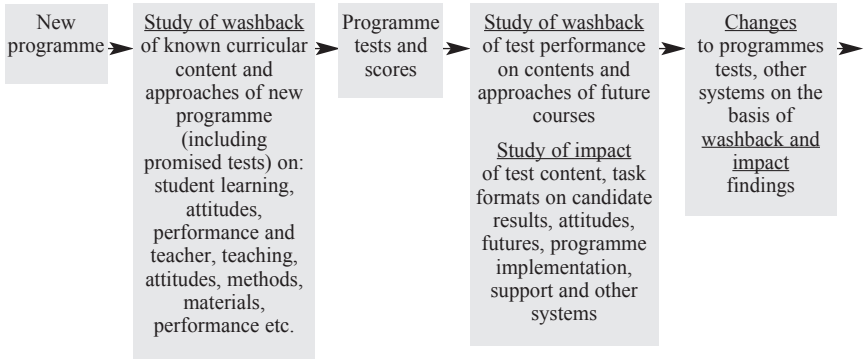
But Bachman (2004) still feels that validity and test use are not necessarily accepted as related in language assessment, despite Messick (1989) and Bachman's own earlier view (1990). In the Bachman and Palmer 1996 definition of test 'usefulness', entailing six qualities: reliability, construct validity, authenticity, interactiveness, *impact* and practicality, 'both the construct validity of our score-based inferences and the impact, or consequences, of test use need to be considered from the very beginning of test design, with the test developer and test users working together to prioritise the relative importance of these qualities' (2004:5). Bachman considers that considerations of validity and impact are thus subsumed 'under a unitary concept of test usefulness'. In Chapter 8, Bachman's case for the articulation of assessment use arguments, in terms of claims, warrants, backing and rebuttals is discussed. These could well feature in a model of test impact study.

Whether impact is intended or unintended, it would seem to be a legitimate and crucial focus of research, both micro and macro, to 'review and change' tests and programmes in the light of findings on, among other aspects of programmes or tests, 'how the stakeholders use the exams and what they think about them' (Saville 2003:60). This is a justification, of course, for studies of the effects of exams as part of the test *validation process*, that is 'the process of investigating the quality of test-based inferences, often in order to improve this basis and hence the quality of the test' (McNamara 2000:138).

The location of impact studies in programme and test development

Figure 1.1 suggests a sequence of washback and impact events in the context of a new educational programme.

Figure 1.1 Sequence of washback and impact occurrences and their study in relation to a new educational programme

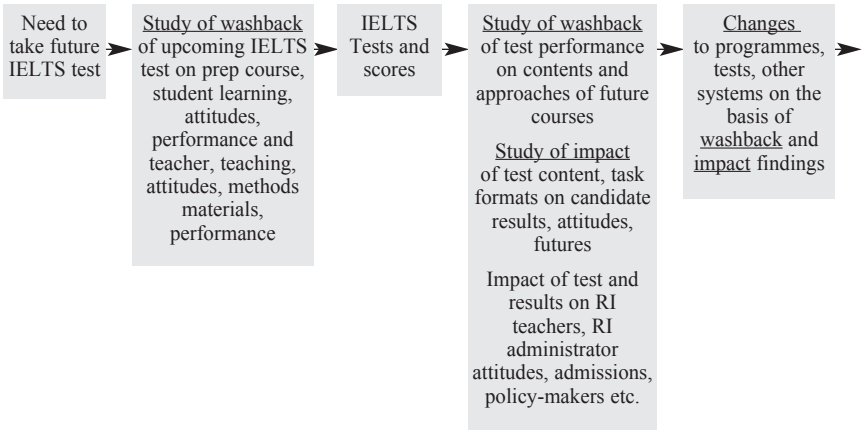


As the chart indicates, washback (that is the effects of an educational intervention, for example a new foreign language programme such as the *PL2000* in Italy, on language teaching, learning, performance) occurs throughout a first implementation of a new language course because the curriculum based on the principles of the Project is known before the delivery of the courses. Washback also takes place in subsequent courses in the new programme because of the experience of the processes of the first implementation and because of the test performances and scores of learners on the tests at the end. The first occurrence of washback in Figure 1.1 is a 'bow-wave' from the programme curriculum and the prospects of the test rather than backwash from a previous programme. The second occurrence is indeed wash*back* from the course and test(s) that have already taken place.

The examples of 'impact' in Figure 1.1 indicate that the term is used to include the effects of a programme or test *beyond* washback, to include stakeholders, domains and systems *in addition to* language learners and teachers and their immediate context. The main point of Figure 1.1. is to see the washback and impact concepts in action and to suggest *where* they occur in the sequence of impact study events. Notice here, as in Figure 1.2, the final arrow indicating continuing washback and impact and the iterative nature of its study.

Figure 1.2 similarly indicates the potential washback (here the effects of an international gate-keeping language test such as IELTS on test preparation programmes) and impact (effects of the test on candidates' futures at receiving institutions (RI) and the admissions policies of those institutions). Again, where washback and impact occur, they may be studied. The data collected on test washback and impact may inform changes designed to improve the test and related systems.

Figure 1.2 Sequence of washback and impact occurrences and their study in relation to a high-stakes test



Washback: negative or positive – intended or unintended?

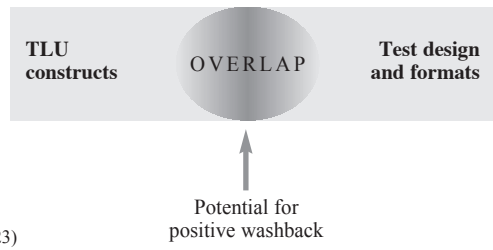
Cheng and Curtis claim that it was Alderson, in 1986, who, in a context of language testing innovation, 'identified washback as a distinct – and at that time emerging area within language testing, to which we needed to turn our attention' (2004:5). Alderson refers to 'potentially powerful influence offsets' (1986:104) from testing to curricular change. Hamp-Lyons (2000) is also informative on both the implications, the scope and the status of the term 'washback' which, she claims, need not imply 'curricular alignment', that is a narrowing of curricula to fit closely and exclusively with the demands of a test. In the Alderson and Hamp-Lyons interpretations (as in, for example, Bachman and Palmer 1996, Davies et al 1999) we see test washback as leading to both positive and negative effects. If this is the case, washback itself should, logically, be regarded as *neutral*. Alderson and Wall (1993:121) suggest as such, with washback a 'metaphor which is useful in that it encourages us to explore the role of tests in learning and the relationship between teaching and testing'. Alderson notes, however, that: 'Although the possibility of positive washback has also often been mooted, there are, interestingly, few examples of this having been demonstrated by careful research' (2004a:x). Perhaps the impact studies reported in this book will demonstrate further such examples.

Green and Hawkey illustrate a historical expectation of *negative* washback from modern language tests, 'the kind of harmful influence that poorly designed tests are said to exert on the classroom' (2004:66) and, it is implied, into the target language use domain (TLU, defined by Bachman and Palmer

1996:44–45 as "a set of specific language use tasks that the test taker is likely to encounter outside the test itself"). Andrews (2004:39) traces the opposite trend, towards positive washback, most clearly, if controversially, exemplified in measurement directed instruction (MDI) 'when a high-stakes test of educational achievement influences the instructional program that prepares students for a test' (Popham 1993, cited in Chapman and Snyder 2000:460). MDI is a clear example of *intentional* washback, with the effect of the test planned in order to influence learning and teaching towards positive outcomes (and, perhaps, processes).

In foreign language teaching and learning terms, in particular in terms of tests (such as IELTS) claiming a communicative language construct, a view of language as purposeful, dynamic, interactive, involving users' grammatical, sociolinguistic and discourse competences, washback might be *intended.* Such tests might be attempting to encourage language learning and teaching where the communicative activities of the target language domain, and language teaching, learning and use on the course are similar or overlap significantly. Figure 1.3 adapted from Green indicates a test washback/construct *overlap* model.

Figure 1.3 Green's model of test washback: target language construct overlap



Source: Green (2003a:23)

An example of the issues raised by this model would, for instance, be a test whose construct and approach led to integrated rather than discrete item test formats, the first closer to the target language use constructs, the latter, perhaps, more reliably scored.

But, of course, the intention of construct/format overlap and an intended positive washback on learning and teaching activity towards target language performance which meets the communicative requirements of the test and of the learner's needs, may be too good to be true. Washback is complex, with a great many independent, intervening and dependent variables. Alderson and Wall (1993) famously state fifteen washback hypotheses, suggesting a test's potential influence on: the teacher, the learner, what and how teachers teach, how learners learn, the rate and sequence of learning, attitudes to teaching and

learning methods. In this interpretation, washback is complex, broad and multi-faceted indeed. Milanovic and Saville also emphasise the broad scope of and complicated relationships within the washback concept, including as it does the 'complex interactions between the factors which make up the teaching/learning context (including the individual learner, the teacher, the classroom environment, the choice and use of materials etc)…' (1996:2).

Stakeholders

For many involved with studies into the effects of language tests or language educational programmes, research questions are raised which concern a whole range of people, usually called 'stakeholders' 'because they have a direct or indirect interest (stake) in a program or its evaluation' (Weiss, 1998:337). Weiss goes on to specify stakeholders as 'people who conduct, participate in, fund, or manage a program', then widens the net still further by adding 'or who may otherwise affect or be affected by decisions about the program or evaluation'.

Rea-Dickins attempts a comprehensive list of potential stakeholders in language testing, including:

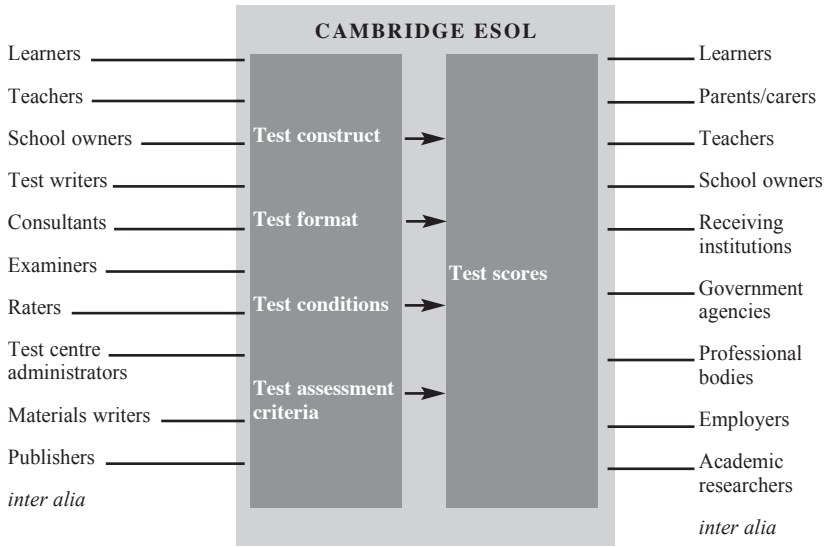
> … language testers, teachers, parents, administrators, teacher educators, sponsors and funding bodies, government bodies, the public, various national and international examination authorities, members of working parties and curriculum committees, test takers (and the larger group of learners of whom they form part) … and to this list we should add test administrators as well as test users, for example university admission officers who need to interpret scores on national and international proficiency tests … (1997:305).

Of course, it is unlikely that *all* these stakeholders would be included in any particular study, and there is always a danger of casting the evaluation or impact study net too widely. Note, however, that the Rea-Dickins stakeholder list already stretches beyond the immediate language learning, teaching, testing context. This is because, if studies of the effects of programmes and tests are limited in their scope to the albeit complex and wide-ranging effects on teaching and learning, they would not cover programme and test effects on the full range of stakeholders whose reactions and attitudes may be relevant to the programme or test's validity and usefulness. As the stakeholder coverage widens, a study may be moving from washback to impact territory.

It will be seen that the studies of IELTS and *PL2000* impact seek information from more than learners, test candidates and teachers, yet by no means from all the stakeholders on Rea-Dickins's list.

The stakeholder concept is certainly an important one in the washback and impact study literature. Figure 1.4 below shows the wide range of stakeholders involved in various ways in the development, construction and administration of high-stakes tests.

Figure 1.4 Stakeholders in the testing community



Source: Taylor (1999)

As Saville indicates, such a 'taxonomy of stakeholders' demands of an exam provider such as Cambridge ESOL, systems that ensure that it can 'review and change what it does in the light of findings on how the stakeholders use the exams and what they think about them' (2003:60). This is a strong justification, of course, for studies of the effects of exams as part of the test validation process, the views of the stakeholders concerned being taken into account along with other impact data, in decisions on test revision or renewal.

Critical language testing and the study of impact

As recently as the early 1990s, Alderson and Buck were lamenting that 'there has not (yet) been a call for public accountability on the part of the examining boards' (1993:21). The 1992 discussion paper by Milanovic and Saville, *Principles of Good Practice for UCLES Examinations*, could, however, be seen as just such a pioneering step:

> It is recognised that UCLES has a responsibility to be held accountable for all matters related to use of its examinations. This involves providing a high quality service to the users of UCLES examinations which meets an agreed code of practice … (1992:5).

Saville and Hawkey note that '[i]n tune with increasing individual and societal expectations of good value and accountability, testers are expected to adhere to codes of professionally and socially responsible practice' (2004:75). This tends to increase the concern of high-stakes exam providers with the *ethics* of language testing. Such codes (for example, of the International Language Testing Association (ILTA), 1997, or the Association of Language Testers in Europe (ALTE), 1994), were intended to improve test development rigour and probity, through measures such as the following:

- properly-defined targets
- appropriate and reliable evaluation criteria
- comprehensive, transparent and fair test interpretation and reporting systems
- continuous validation processes
- a keener regard for the rights of candidates and other stakeholders

(For more information see the ALTE Code of Practice 1994, and the IELTS Handbook 1997/8.)

As on so many key language testing matters, Spolsky himself is instructive on the question of testing ethics, taking, as Hamp-Lyons points out, the stance of a 'humanistic sceptic' (2000:587). Spolsky talks of 'competing sets of forces' namely the *institutional context* of a test, the *need for general acceptance*, and the '*unavoidable uncertainty'* or *'probable error'*. He continues:

> What I have labeled post-modern may be seen as an approach that accepts the equal approach of all three factors … Post-modern testing adds a sincere, ethically driven consideration of the potentially deleterious effects of testing on the test taker, on the instructional process, and on other facets of the social context in which we test (1995:354–357).

'Sincere, ethically driven consideration' of the effects (not only 'potentially deleterious' but also positive) of testing 'on the test taker, the instructional process, and on other facets of the social context in which we test' might be considered an admirable impetus for an impact study. We may recall, however, that the term impact is in itself neutral (see above). The common conviction (on which both the studies described in this book will provide data) is that high-stakes tests always do seem to have washback.

As implied by the overlap model shown in Figure 1.3, there is a *demand* for positive test washback and impact. A test may have optimal validity as a measure of the target ability and thus satisfy the demands of usefulness,

although *not* representing entirely actual performance of that ability. The demand is for test task formats as close as possible to the communicative construct. A further reality, of course, as Green points out, is that 'learners, teachers, administrators and other participants may have competing goals. Thus effects regarded as positive by one constituency within an educational system may be seen as negative by another' (2003:11). Green refers to the disagreements over measurement-driven instruction in the USA (Airasian 1988, Bracey 1987, Popham 1987, Ramirez 1999) to exemplify this.

The apparent non-neutrality of language tests is a concern for the proponents of the *critical language testing* view. Shohamy considers that tests are 'powerful because they lead to momentous decisions affecting individuals and programs … . They are conducted by authoritative and unquestioning judges or are backed by the language of science and numbers' (1999:711). 'Critical testing', Shohamy adds, 'refers to the need to examine the use of tests within a broader social and political critique of aspects of testing as a social and institutional practice' (1999:714). Given the definitions of impact study emerging above, it should be possible to use such studies in the interests of solving some of the problems identified by the critical language testers. Shohamy herself suggests this:

> Studies of the use of tests as part of test validation on an ongoing basis are essential for the [testing] profession. Language tests fall at the crossroads of many conflicts and should therefore be studied, protected and guarded as part of the process of preserving and perpetuating democratic cultures, values and ethics (2001:390).

Shohamy's two themes here, of the study of 'the use of tests' (which she also calls 'research on impact of language tests') as an element in continuous test validation systems and as part of the ethical validation of tests, recur in this book, both in the discussion of impact study policy in general, and in the principles and practices of the studies of IELTS and *PL2000* impact. Later in this chapter such studies are contextualised within the continuous validation systems of Cambridge ESOL.

Alan Davies sounds a note of caution, however, on the ethical and critical language testing case. Davies agrees that the language testing profession should attempt to ensure the validity of its activities, and that it 'may therefore be sensible to equate ethics with validity', but there are limits:

> An ethical perspective for a language tester is, I have argued, necessary. But I have also urged the need in all professional statements of morality for a limit on what is achievable or even perhaps desirable. In my view, therefore, the apparent open-ended offer of consequential validity goes too far. I maintain that it is not possible for a tester as a member of a profession to take account of all possible social consequences (1997:335).

A balance is clearly needed in impact research between two extremes. One would tend to neglect the social consequences or *consequential validity* of a high-stakes test; the other would tend to enquire into its effects on too many stakeholders, or pursue too many intervening variables for clear washback or impact connections to be made.

Good practice and test validation

So, with international examinations of ever-higher stakes (the candidature for IELTS has grown from 50,000 to 500,000 in ten years), and the growing importance of codes of practice, examination boards need to accept responsibility for the impacts of their tests on stakeholders. They do this through systems that ensure good practice in test development and management. For Cambridge ESOL, one way of enhancing these systems was participation, with leading international partners, in the research and negotiation leading to the Code of Practice of the Association of Language Testers in Europe (ALTE). This code of practice (see previous) made strong statements of provider responsibility for exam development, scoring, result interpretation, fairness and transparency.

Saville (2003:65–78) summarises the implications of such good practice as the need to pursue test *validation,* namely to make every systematic effort to ensure that a test or exam achieves:

- appropriacy to the purposes for which it is used
- the ability 'to produce very similar results in repeated uses' (Jones 2001)
- positive influence 'on general educational processes and on the individuals who are affected by the test results' (Saville 2003:73), and
- practicability in terms of development, production and administration.

In the theory and practice of Cambridge ESOL test research and development, these four exam targets are labelled *validity, reliability, impact* and *practicality* (VRIP for short). The overlap with Bachman and Palmer's 1996 six test *usefulness* qualities, *reliability, construct validity, authenticity, interactiveness, impact* and *practicality (*see previous*)* is neither insignificant nor coincidental given the close relationship of Bachman with UCLES. On the 2005 Cambridge ESOL website, reference is still made to activities planned as a follow up to the work of Lyle Bachman and colleagues, on what was known as the Cambridge-TOEFL Comparability Study, carried out between 1987–9.

So impact, the main focus of this book, encompassing (in our interpretation above) and closely associated with washback, is firmly set in the VRIP context. According to Saville, impact studies cover three major groups of stakeholders: the examination developer, the examination taker, and the examination user, that is someone 'who requires the examination for some decision-making or other purpose' (2003:60).

Although there are four components to VRIP, it is clear that they are by no means independent. Test validity, for example, in the unitary construct proposed by Messick (1989:16), *subsumes* reliability, impact and practicality. 'Individual examination qualities', Saville agrees, 'cannot be evaluated independently. Rather the relative importance of the qualities must be determined in order to maximise the overall usefulness of the examination' (2003:61). Weir (2004), in his socio-cognitive framework for validating tests, has a similar perspective. For him, test *validity* is the superordinate category to *theory-based validity* (covering internal language ability processes), *context validity* (the appropriateness of the communicative and administrative context in which the test takers are called upon to perform) and *scoring validity* (the dependability of test results, subsuming the conventional category of reliability). Test validity also subsumes the two post-test validities, *concurrent* and *consequential*, the latter, of course, including the study of the impacts of the test on stakeholders.

But Weir reminds us that test 'validity is perhaps better defined as the extent to which a test can be shown to produce data, i.e. test scores, which are an accurate representation of a candidate's true level of language knowledge or skills'. Validity 'resides in test scores (not tests)' (2004:12). Despite the common shorthand reference to valid tests, it is, Weir notes, 'inaccurate to talk of a test such as TOEFL or IELTS as being valid or not. It is the scores produced by a particular administration of a test on a particular sample of candidates that we are concerned with'. 'Obviously,' Weir adds, 'over time if various versions of a test or administrations of the same test provide similar results then synthetically a case may be made for X or Y test being valid over time and across versions and population samples' (2004:11). McNamara suggests that validity is about 'the relationship between evidence from test performance and the inferences about candidates' capacity to perform in the *criterion* that are drawn from the evidence' (2000:138).

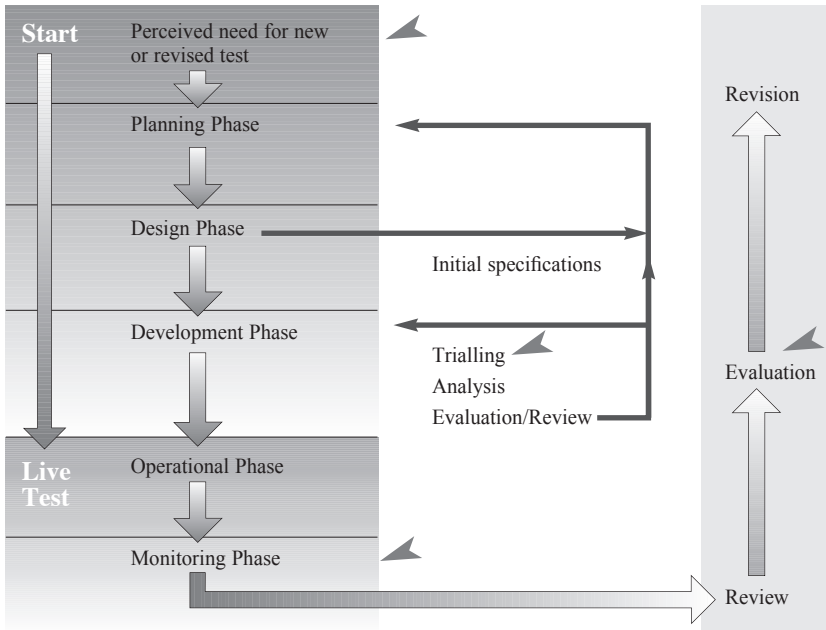
Impact studies and validation

Despite the complexities of washback and impact, and despite the multi-faceted nature of even a unified model of language testing validity, impact studies remain part of the test validation process. The tests need to take account of Messick's consequential validity. As he says, 'the social values and the social consequences [of a test] cannot be ignored in considerations of validity' (1989:20). And, as we have seen above, the study of test impact is but one element in a *continuous and iterative* test validation process or, in Shohamy's words, 'part of test validation on an ongoing basis' (2001:390).

The cycles of exam development, validation and revision are the major theme of three recent volumes in the *Studies in Language Testing (SILT)*

series. Volume 15 traces the continuity, innovation and revisions of the Cambridge Certificate of Proficiency in English (CPE) examination over 90 years, and Volume 16 describes the development of a modular exam, the Certificates in English Language Skills (CELS), developed from several precursor exams. Volume 17 explains Cambridge ESOL's position on language testing for specific purposes and describes the rationale for changes implemented in the revision of the Business English Certificates (BEC). In all three cases, the continuing, iterative and increasingly transparent nature of test development, validation and revision is clear. So, increasingly, is the inter-related nature of validity, reliability, impact and practicality (VRIP), and the superordinacy of validity. As data relevant to the four inter-related validity elements emerge with every administration of an international exam, so the validation of the exam continues through further iterations. Impact data, whether from routine feedback systems, for example Cambridge ESOL's Candidate Information Sheets (CIS), or through test washback and impact investigations such as the continuing IELTS study, the first three phases of which are described in this book, feed into the validation process and contribute, along with other validity, reliability and practicality data, to the exam evolution and revision process.

Figure 1.5 Cambridge ESOL model of the test development process



The standard current Cambridge ESOL model of the test development or revision process is illustrated in Figure 1.5, with indicators added (◄) of the use of washback and impact study data in the process.

The indicators of test impact study data collection and use added to Figure 1.5 suggest the following about their role in the Cambridge ESOL test development process:

- that impact data may inform decisions to revise a test
- that data from impact studies will be part of the test trialling process
- that such data represent part of the test monitoring process, alongside, for example, routine candidate profiling and test performance analyses across test versions and tasks
- that when a test is evaluated for possible revision, data from impact study, based on questions on the evaluation research agenda, will be included in the decision-making.

The aim of a test development project is, after all, to produce a test which will have positive impact, or at least will have minimal negative impact. The validation process, of which impact study is a key element, is to collect adequate evidence to support the test developer's claims about the test's suitability for its intended purpose.

When Saville further defines Cambridge ESOL procedures for the achievement of good practice, he re-emphasises the important role of impact studies:

> The approach now being formally adopted recognizes the importance of validation through the collection of data and the role of research and development in examination processes. In this respect, the principles … will continue to evolve over time as research and development programmes expand (2003:57).

We can envisage impact data of various kinds being taken into account at different stages of the test development cycle, as indicated in Figure 1.5. Washback and impact data would be considered alongside the descriptive statistics, classical and Rasch-based analyses carried out on Cambridge ESOL exams at trial and operational phases (see, for example, Saville: 2003:90–96). It is significant, however, that Figure 1.5 does not yet contain the terms washback or impact, which, as Saville appears to imply, is the most recent element in the VRIP combination.

It would seem that the *principles* of impact study use in Cambridge ESOL test development, validation and revision appeared before *explicit reference* to impact. Clapham describes how, in 1986, the Revision Committee for the English Language Testing Service (ELTS 1980–1989: see Alderson and Clapham 1993, Criper and Davies 1988), the predecessor of IELTS, 'sent out

questionnaires to ELTS users such as test administrators, EAP teachers and testers and receiving institutions such as universities and colleges asking for comments on ELTS and for advice about the content and format of the future test battery' (1996:4). Clapham adds that interviews and discussions with British Council personnel and English for Academic Purposes (EAP) teachers and testers were held, and 1,000 past candidates' test report forms studied. These procedures, one is tempted to say, amount to the study of ELTS *impact*, by any other name. Cambridge ESOL was certainly involved in the study of impact significantly before the term itself was routinely used.

Given the broad scope proposed in this chapter for test washback and impact data and the iterative applications of such data in the test development process outlined in Figure 1.5, a further question arises. Where, if anywhere, do impact data end and the routine post-test monitoring and analysis of test scores and performance data begin? Are test impact and washback data *additional to* routine validation data, and data requested by the test validators because of particular emerging validation, or test revision-related queries? Or could the routine test development and validation systems be regarded as 'intrinsic' impact study, the non-routine additional study data as 'extrinsic'? At present, the 'intrinsic' and 'extrinsic' research do share some areas of data collection, for example, test candidate profiling. The Cambridge ESOL test validation process certainly benefits from both the routine, mainly post-test validation investigations and analyses, and from the more 'extrinsic' but nevertheless planned and systematised study of test impact.

It seems likely, too, that different potentially useful impact data suit different kinds of data collection contexts. Candidate test *performance* data, for example, to be correlated with candidate *baseline data*, with inferences to be drawn, for example, on test task validity, might seem more appropriately handled through routine post-test validation systems. Candidate *attitudes* to the test and to their preparation for it would, on the other hand, suit multi-faceted investigation through the questionnaires, interviews and classroom observations of an impact study.

Saville suggests that, from a validation perspective, Cambridge ESOL must 'be able to monitor and investigate the educational impact that examinations have within their contexts of use'. To this end, procedures are needed 'to collect information that allows impact to be examined'. These procedures should include 'collecting data on: candidate profiles, exam result users and purposes; test preparation courses; public and participant perceptions of the exam'. Saville refers to a 'systematic approach to investigating impact' as 'developed since the mid-1990s' which includes 'procedures to monitor the impact of IELTS as part of the next revision cycle', and representing an example of 'the continuous, formative, test consultation and validation programme pursued by UCLES and which has also been a significant part of the FCE and CPE Revision Projects' (2003:73–76).

The two impact investigations used as case studies in this book illustrate this means of contributing to the development and validation of high-stakes tests in different ways. In the case of the study of IELTS impact, the data are being collected with both a pre- and a post-test validation intent. Along with research data from other sources, the results of our enquiries into candidate profiles, their attitudes, test perceptions, preparation and performances; the experiences and views of their teachers, the nature of test-specific preparation materials and the impact of the test on receiving institutions, are all being used in the monitoring and review of the existing IELTS test. The data are thus *post*-test but the impetus for the study of IELTS impact was, from the beginning (see Saville 2003), intended for use in decisions on potential IELTS *revisions*. In that sense, the study also has a *pre*-test perspective.

The study of the impacts of the *PL2000* also has a test validation aspect. One of the research questions for the study (see Chapter 3) was: 'What impacts is the *PL2000* having on language evaluation and assessment?' Data in response to this question (see Chapter 7) were informative not only on which of the Cambridge ESOL Main Suite exams was suitable at which student English language performance level, but also led to changes in the format and use of certain exams (see also Chapter 7). Note again here both a post-test and test revision element in the study.

This book will also discuss *limitations* on the scope of impact studies, some of them practical, some more a matter of principle. Limitations on both our example studies are described in Chapters 3–7, as is the constant concern with impact study ethics.

This chapter has so far explored the concept of impact, including washback, with reference to related terms in the field of education, including evaluation, monitoring and validation. The means and implications of language programme and test washback and impact have been discussed in terms of their logical research location, definition and scope, positive and negative aspects, intentionality, complexity, stakeholders, and the relationships of washback and impact studies with test standards, ethics and validation. The *intensity* of washback and impact, sometimes also included in lists of their characteristics (e.g. Cheng 1997, Watanabe 2000), would seem more logically left until our actual washback and impact *data* are analysed.

The focus next will be on the role of the studies of IELTS and *PL2000* impact in relationship with other Cambridge ESOL test research and development.

IELTS impact research in context

The impression should be emerging that some impact studies are not to be thought of as monolithic, for example as '*The* IELTS Impact Study', as if there

could be once-and-for-all enquiry which brings closure to the quest for answers to key impact questions on a high-stakes test that is still in operation. As will be seen when the history of IELTS impact study is recounted in Chapter 4, the impact study phases we shall be referring to in this book, and describing in some detail in Chapters 3–6, did indeed have their own planning, trialling, data collection, analysis, and reporting phases. But, as suggested in the previous section, these were only *some* of a range of continuing initiatives designed to assist in the process of validating the IELTS test. The book will therefore tend to refer to 'the study of IELTS impact', or 'IELTS impact study' (with no 'The' and no capitalisation).

The test has been the subject of a number of studies to check its predictive validity, usefulness (see Bachman and Palmer 1996) and positive impacts, and has undergone several revisions since its debut in 1989 (IELTS itself being, of course, a development of the English Language Testing System or ELTS). The most recent modifications to the IELTS were in a major revision in 1995, which implemented three areas of significant change suggested by routine research:

- The three field-specific Reading and Writing modules were replaced with one academic Reading module and one academic Writing module (see Charge and Taylor 1997).
- General training Reading and Writing modules were brought into line with the academic Reading and Writing modules in terms of time allocation, length of written responses and the reporting of scores.

Since 1995 two revisions to the productive modules, Writing and Speaking, have been introduced. The new Speaking test was launched in July 2001, involving a complete revision of both the format and assessment of the test. A change to the assessment of the Writing test was introduced in January 2005. Articles relating to both revisions can be found in Cambridge ESOL *Research Notes* (www.CambridgeESOL.org/rs notes), for example, Shaw August 2002, May 2004, Taylor February, July and November 2001.

So the routine monitoring and evaluation of the IELTS continues, and the validation systems are themselves monitored and modified, including the measures introduced to gather data on test performance and candidate background so that issues of fairness relating to test use and users may be more effectively monitored.

All IELTS *research* activities are co-ordinated within Cambridge ESOL's framework for research and validation, which is designed to cover three broad areas of activity:

- routine operational analyses for exam production, conduct, marking/grading and post-exam evaluation
- instrumental research involving small-scale projects to inform the operational activities

- research projects involving longer term assessment objectives relevant to broader objectives and future developments.

The Research and Validation Group of Cambridge ESOL (established in 1989) is responsible for the provision of quality assurance services for Cambridge ESOL exams. This is achieved through work by specialised staff on test validation matters involving, for example, statistical analyses of candidate performance, test task and item validation, Writing and Speaking test corpus analysis. The group also advises on research design issues, and presents and/or publishes information based on the work carried out. The research and validation programme is designed to ensure that all ESOL assessment products meet acceptable standards in relation to the four essential VRIP qualities. The standard procedures established and implemented by the Group are, it is claimed, themselves evaluated and refined in the light of theoretical, technological and business developments. External stakeholders are kept informed on issues of the quality and fairness of the examinations, and of the relevant research and validation work carried out by Cambridge ESOL.

Routine and non-routine longer term research on IELTS is supplemented by the two other partners in the IELTS test: IDP Education Australia:IELTS Australia, and the British Council, who call for proposals and designate funds for suitable research projects on IELTS. Of the 44 titles on the 1995–2001 IELTS funded research project list (see *Research Notes* 8, May 2002), six contained the word 'impact', five the word 'effect', two 'monitoring' and one 'evaluation'. The labels are interesting given our discussions and definitions, but more important is the further clear evidence that impact-type research is continuous and focused rather than monolithic and one-off. The connection between studies conducted under the IELTS funded-research programme and washback or impact studies is clear – 15 are mainly concerned with the IELTS skill modules (Reading, Listening, Writing, Speaking), 12 with IELTS stakeholders (including candidates, examiners, receiving institutions), and 11 with IELTS preparation courses and candidates' future language-related needs. As the IELTS website (www.ielts.org) rightly claims, 'such research makes an important contribution to the monitoring and test development process for IELTS; it also helps IELTS stakeholders (e.g. English language professionals and teachers) to develop a greater understanding of the test'. Further research under rounds 8–10 of the programme is discussed in Chapter 8, where we look at more recent IELTS impact research-related developments.

As we shall see in Chapter 6, responses to the research questions asked in the study of IELTS impact provide data from learners, teachers and other stakeholders on a wide range of impact and test validation areas. These include:

- candidate profiles, strategies and test attitudes
- perceptions of test motivation and stress

- candidate and teacher perceptions of test relevance, difficulty and reliability
- candidate and teacher views on test preparation course content, approaches, materials and quality.

It is to be expected that such washback and impact data will be given appropriate weight alongside the other continuously emerging information processed through the research and validation systems already established for the IELTS.

Progetto Lingue 2000 impact research in context

The *PL2000* was a radical language education reform project of the Ministry of Education, then *Ministero della Pubblica Istruzione* (MPI, Ministry of Public Education) now *Ministero dell'Istruzione, dell'Università e della Ricerca* (MIUR, Ministry of Instruction, Universities and Research) in Italy. The *PL2000* aimed to provide foreign language education in state schools to meet the communication and certification needs of students as defined by the Common European Framework of Reference for Languages (2001). Courses were to be delivered to homogeneous groups of learners through precisely specified learning modules, using the most technologically informed taught and self-access means. One of the Project's key policies was to encourage the *external* certification of learners' language proficiency through the examinations of providers such as Cambridge ESOL. The *PL2000* is described in further detail in subsequent chapters.

During the 2001–2002 school year, with the encouragement of the Education Ministry in Italy, Cambridge ESOL carried out a study of the impact of the *PL2000*, collecting data from stakeholders including students, teachers, parents, education managers and language testers. Given the fairly broad range of stakeholders covered, the longitudinal, process-oriented nature of the study, and its constructively critical ethos, the study was called an *impact study* rather than a washback study or an evaluation. The study is given the capitalised format, 'The *PL2000* Impact Study' as this was its title from early conceptualisation to completion and final report, and because it was more of a single, self-contained study than Phases 1, 2, 3 (and more) of the study of IELTS impact.

The *PL2000* Impact Study, like the IELTS study, was, however, one element in an evaluation and validation system. It was a study carried out by an external certifying agency, Cambridge ESOL, with the encouragement and approval of the *PL2000*'s originator, the Italian Ministry of Education. The willingness of the Ministry to permit a study by an external examination provider, an officially appointed one (see below), was evidence of an interest in obtaining feedback on the Project from a variety of sources, including, of

course, its own monitoring and research systems. These, according to the Ministry's *PL2000* objectives and procedures statement included:

> … systematic monitoring of *PL2000* at provincial, regional and national levels to support the project and its evaluation, using quantitative and qualitative criteria such as: number of modules and languages taught; number of students involved; time schedules (time organisation of the modules); certifed qualifications and qualification mode (internal/ external to the school); content and organisation of modules; number of collaborative activities with foreign cultural agencies; learners' success (*PL2000* Impact Study Report 2003:11).

In addition, the *PL2000* included a 'national information module for provincial *referenti* to permit regular comparison and verification of *PL2000* trends within their districts and to encourage the sharing of organisational and management strategies' (*PL2000* Impact Study Report 2003:10).

It was clear that the *PL2000* Impact Study, relatively small-scale as it was (see Chapters 3 to 5 and 7 below), produced data and findings of use in areas specified for the Ministry's 'systematic monitoring of *PL2000*'. These, comparing the contents of the *PL2000* Impact Study Report itself (2003:11) with the Ministry's specified target monitoring areas included:

- small-sample quantitative and qualitative impact information on the numbers of students involved in the case study schools
- analyses of teaching/learning module organisation, content, methodologies, media and learners' success.

These data could be added to other Ministry information on crucial focus areas for the evaluation of the *PL2000* and for the development of future initiatives for the reform of school foreign language education in Italy. Chapter 7 will describe, for example, revealing impact analyses of the learning/teaching events in classrooms where courses under the *PL2000* were taking place, and summarise aspects of learner success as seen by the learners themselves, their teachers, their school heads and their parents.

In addition to the information it provided for the Italian Ministry of Education on the impacts of the *PL2000,* the study contributed data to Cambridge ESOL test validation systems. One of the impact study's stated aims was 'to analyse the development of examinations for the certification of language levels as appropriate to the needs of the growing range of language learners in Italy' (*PL2000* Impact Study Report, 2003:13). As Chapter 7 will show, the study provided data on test-validation matters such as:

- student, teacher, school head and parent attitudes to Cambridge ESOL exams including the Preliminary English Test (PET), the First Certificate in English (FCE) and Certificate in Advanced English (CAE)

- teaching/learning : test relationships, including student and teacher perceptions and classroom observation data
- comparisons of *PL2000* and non-*PL2000* exam performances.

An interesting extra insight from the study concerned the effect of the *PL2000* on the exam providers and their exams, an example of *two-way* washback, of a programme on exams as well as of the exams on a programme.

Studies of IELTS and *PL2000* impact

Although the two studies described in this book differ significantly in scale, status, focus and context, both are labelled *impact* studies. Both seek information on the effects respectively of an education reform programme and a high-stakes test on stakeholders beyond, though including, the teaching/learning territory of washback. The *PL2000* study hears the views and attitudes of school heads, project officials and learner parents; the IELTS study contacts receiving institution administrators for their reactions to the test, as well as the candidate/student and teacher stakeholder constituencies. Both studies examine outcomes as well as processes, and are seen as but one element in a continuous and iterative test validation process.

As we look into both studies in subsequent chapters, the labels and significance of constructs such as *impact*, *washback*, *evaluation* and related terms will be revisited. This chapter has attempted to define main constructs and terms and to put the study of impact into perspective and context. Neither the study of IELTS nor of *PL2000* impact is seen as time-bound or self-sufficient. In fact (see also Chapter 8) both studies are already leading to further, related washback and impact research.

2 Impact study approaches

Impact, washback and related concepts are defined and exemplified, and the Cambridge ESOL context for impact studies as part of the test validation process described in Chapter 1. Chapter 2 will survey possible methodologies for the collection and analysis of impact data. It will cover research approaches, sampling, the pre- and post-validation of instrumentation, and data strategies. Chapters 3, 4, and 5 will then examine, respectively: impact study objectives and research questions; instrumentation; and the management and analysis of impact data.

Qualitative and quantitative research approaches

Choices of research approach

The choice of type of research methodology normally involves a consideration of the advantages and feasibility of *quantitative* and *qualitative* research methods. Weiss usefully defines the two general approaches to research, as follows:

> Quantitative Research: *research that examines phenomena that can be expressed numerically and analysed statistically*
>
> Qualitative Research: *research that examines phenomena primarily through words, and tends to focus on dynamics, meaning and context. Qualitative research usually uses observation, interviewing and document reviews to collect data* (1998:335).

Some of the discussion in Chapter 1 of types of *evaluation* reflects the distinctions Weiss makes.

Lazaraton cites in Figure 2.1 (page 30) the generalised research type *continua* of Larsen-Freeman and Long (1991) to summarise aspects of quantitative and qualitative research methodologies, including data collection means, type and focus of data collected and findings derived.

Weiss distinguishes a quantitative and a qualitative research focus thus:

> Quantitative evaluators tend to focus on whether and to what extent change in x causes change in y. Qualitative evaluators tend to be concerned with the process that connects x and y (1998:284).

Figure 2.1: Characteristics of quantitative and qualitative research

Quantitative Research		Qualitative Research
controlled	.....	naturalistic
experimental	.....	observational
objective	.....	subjective
inferential	.....	descriptive
outcome-oriented	.....	process-oriented
reliable	.....	valid
particularistic	.....	holistic
'hard', 'replicable' data	.....	'real', 'rich', 'deep' data

Source: adapted from Charles S Reichardt and Thomas D Cook (eds): *Qualitative and Quantitative Methods in Evaluation Research,* page 10, copyright 1979 by Sage Publications, Inc (in co-operation with the Evaluation Research Society). Reprinted by permission of Sage Publications, Inc.

The characteristics in Figure 2.1 are, of course, by no means mutually exclusive. They may be combined within the same study; they may be present in different strengths along their continua. Cronbach (1982:18–19) talks of a 'reconciliation' between what he terms 'scientistic' and 'humanistic' research and welcomes 'multiple-method approaches' (Saxe and Fine 1979:64) combining qualitative and quantitative elements. In fact, it is nowadays quite rare to encounter evaluation or impact studies that do *not,* to a greater or lesser extent, claim to *combine* qualitative and quantitative research approaches.

We have noted in Chapter 1 the complex nature of washback, and the non-monolithic and iterative nature of impact studies. With such inherent complexity and continuity, it is understandable that the investigation of washback and impact will require a range of approaches. In a recent summary of the methodologies employed in washback research (including the study of the impact of the IELTS test which is described in this book) Watanabe suggests that since it is 'conceptualised on several dimensions', 'the methodology that attempts to disentangle the complexity has inevitably to be multifarious' (2004:20).

The characteristics of impact research thus seem likely to show influences from both ends of Lazaraton's continua, but perhaps with an inclination towards the qualitative end. Watanabe notes that 'qualitative or ethnographic research has been increasingly widely used among researchers in the field of language teaching and learning', such research, he considers, being characterised by strategies 'which are relevant to the research into washback' (2004:20). Watanabe's apparent equation of qualitative and ethnographic research might be disputed; classroom observation and case studies, for example, though regularly used in applied linguistics studies, do not usually constitute fully fledged ethnographic research (Heath 1982:36). 'Ethnography' confirms Lazaraton 'requires a deeper and broader philosophical

and methodological commitment than does simple participant observation' (1995:460).

Duff (2002:1–10), also denies a polarity between quantitative and qualitative research. She notes important 'research practices that can enhance qualitative applied linguistics research' (2002:4), then suggests that such research practices include:

- conducting longitudinal research, when possible
- eliciting participants' perspectives on their own behaviour
- using participants who know each other and have some familiarity with the researcher
- looking for recurrent patterns across data sets
- providing whenever possible methodological, analytical and perspective/epistemological triangulation ('using several methods to reveal multiple aspects of a single empirical reality' Denzin, 1978).

The research practices exemplified here by Duff and worth pursuing in the next few sections with relation to our two studies, clearly derive from the *nature* of impact research rather than from an automatic methodological choice. It is because of the complexity of the inter-relationships between variables such as tests and test-related teaching or lesson plans and lesson implementation (both key areas of attention in the studies of IELTS and *PL2000* washback) that complex and often qualitative research practices are entailed. It is also clear that Duff's research practices such as 'eliciting participants' perspectives on their own behaviour' or 'using participants who know each other', while likely to be enriched and enhanced by the collection of open-ended, qualitative conversational data, will also benefit from more *quantitative* analyses, for Duff's purposes of 'triangulation wherever possible'.

IELTS and *PL2000* Impact Study approaches

It is appropriate here, using insights from both IELTS and *PL2000* impact studies, to examine further the match between impact research objectives, methods and contexts.

The IELTS research (see Chapter 3 for a full specification of its aims, objectives and research questions) seeks answers to questions on the washback and impact of the test on:

- the IELTS test taking population
- classroom activity in IELTS-related classes
- IELTS teaching materials, including textbooks
- other users of the test.

The *PL2000* Impact Study, the aims, objectives and research questions of which are also defined in detail in Chapter 3, sought to:

- ascertain the effects of the *PL2000* foreign language reform measures on English language learner performance in state schools in Italy
- explore the significance of *PL2000* for some main stakeholders
- analyse the effects of the use of external examinations for the certification of *PL2000* learner language levels
- identify areas of potential support for teachers teaching classes under the *PL2000*.

The pursuit of answers in such question areas clearly required 'phenomenological data that represent the world view of the participants being investigated' (Watanabe, ibid.:22), and 'eliciting participants' perspectives on their own behaviour' (Duff above).' It is also commonly noted (e.g. Hawkey forthcoming, Hopkins 1985) that participant perceptions may not always reflect what is actually happening. Thus, the studies of IELTS and of *PL2000* impact elicit perspectives on participants' own and others' behaviour both through more quantitative data, for example, from the administration of the objective-item *Language Learning Questionnaire* (Purpura 1999), and more qualitatively through interviews, focus group discussions and open-ended questionnaire items. The two studies also attempt to *frame* their smaller scale data with broad quantitative comparative data. IELTS band scores among the impact study candidate population, for example, are related to global IELTS candidate results (see Chapter 6). *PL2000* course student results on the same exams are compared with those for candidates from the rest of Italy and for non-Italy candidates (see Chapter 7).

The phases of IELTS impact study described in this book, seeking fairly broad worldwide coverage, were conducted mainly at a distance, though with visits to some participating centres for the collection of face-to-face data. The distance data collection was intended to minimise disruption at the target institutions, given the priorities and pressures affecting those involved with an imminent or recent high-stakes English language assessment such as the IELTS. The *PL2000* Impact Study was undertaken with the permission and encouragement of the Ministry of Education in Italy. It was implemented at a selection of typical schools, again with the promise of minimum disruption to daily educational and administrative life. The institutional conditions, as well as the research focus for both studies, tended to favour the collection of data, in the terms of Figure 2.1, through more 'naturalistic', 'observational', 'descriptive' rather than controlled experimental means.

The *PL2000* Impact Study was longitudinal in the sense that the main data were collected from the selected schools on two visits, in October 2001 and April 2002, that is at the beginning and towards the end of the school year. Differences were then analysed using quantitative data from matched student

and teacher background and attitude questionnaires, and English language proficiency exam scores, along with more qualitative interview, focus group and classroom observation data from the two school visits.

The *PL2000* study in particular benefited from information supplied by 'participants who know each other and have some familiarity with the researcher' (Duff 2002:1–10). This was exemplified in teacher interviews and focus groups, and in correspondence with some participants between study visits to the schools. It is difficult to conceive of an impact study that does *not* look for Duff's' 'recurrent patterns across data sets', but both the Cambridge ESOL studies described here certainly did so, often when seeking triangulation (Weiss's 'cross-check through different modes of inquiry', 1998:263). Examples of triangulation in the two studies include comparisons between:

- closed and open-ended student questionnaire responses
- teacher questionnaire responses and teacher interviews
- student and teacher questionnaire responses and classroom observation analyses.

The collection and analyses of these triangulated data are examined in some detail in Chapters 5–7.

Experimental research

It is because of the complexity of washback and impact (see Chapter 1), because of the consequent need to probe webs of inter-relationships between variables and processes, and to seek insights from a range of stakeholders, that the *experimental research* paradigm is rarely evident in impact research. It is significant that in *Washback in Language Testing: Research Contexts and Methods* (Cheng and Watanabe 2004) the sections on impact and washback methodology focus on qualitative approaches to the study of washback and impact, although with quantitative data analytical methods often used. The eight 'washback studies carried out in many different parts of the world' (2004:xiv) then described in the Cheng and Watanabe collection all use survey, interview and observation techniques rather than experimental research designs.

The experimental research approach is applicable where the researcher is in a position to control both independent variables being studied (for example, the teacher and the classroom materials) and the conditions under which a study takes place (for example, the class size and the target language proficiency level of the learners). Then, to satisfy the conditions of a truly experimental design, participants are assigned randomly to the experimental group (all of whom receive the programme or test under investigation) and the control group (selected randomly from the same population as the

experimental group, but not receiving the programme or test concerned). The differences in performance between the experimental and control groups can then be measured and inferences drawn about the effects of the programme or test concerned. Experimental studies with such parameters are relatively rare in educational research, especially where larger numbers of learners are involved, and data are collected from numerous and distant centres.

Watanabe indicates a further reason for the rarity of controlled experiments in impact studies, namely the inevitability of a role for the contexts in which a test is administered:

> Qualitative research also stresses gathering data in 'real', that is, non-experimental settings. The test always plays a certain role in a specific context, so even if it were found that a test has some impact on teaching and learning under controlled settings, it is likely that the result would not apply to situations where the teaching is actually being done for test preparation (2004:23).

Watanabe uses the term 'selection' rather than 'sampling' when discussing participants invited to take part in an impact study's quest 'to examine the validity of predictions' (2004:29). 'The selection', Watanabe continues 'is not to be made at random, but purposefully', often the case, it would seem, when 'selecting cases for study in depth' (Patton 1987:52). Watanabe also suggests, basing his judgement on the eight washback or impact studies reported in *Washback in Language Testing* (2004), that 'it is normal to select various groups of participants rather than one single population'. This is one way of facilitating the triangulation of data, of course, and the possible re-focusing of some research questions as new, potentially significant variables emerge.

In the case of the studies into IELTS impact, as we shall see in greater detail in Chapter 4, centres were selected for the invitation to participate in data provision according to the results of a pre-survey of more than 300 institutions and according to the nationality proportions in the overall IELTS candidate population. For the *PL2000* study, participating schools were selected (see also Chapter 4) according to strata such as school type (elementary, middle and high), geographical location (northern, central and southern Italy) and external exam certification (Cambridge ESOL exams at four levels). These school selections were confirmed or modified through contacts made during pre-study visits. On the key related issue of generalisation from qualitative data, Duff (2002) reminds us that qualitative researchers should acknowledge any limitations in their participant groups and the context-bound nature of their findings. Attempts are made in Chapters 4, 6, 7 and 8 to make such acknowledgements with regard to both studies.

Local contexts

Miller suggests that more ethnomethodological studies 'focus on the social and political contexts within which members use available interpretative methods to construct social realities' (in Silverman 2001:29). This appears to accord with Duff's further proposal for qualitative studies. They should, she suggests, seek to fulfil as far as possible their potential 'to yield interesting understanding of local contexts and to examine the interplay of variables and issues to a greater extent than quantitative research typically sets out to do; the latter, conversely, attempts to control as many extraneous variables as possible' (2002:6).

Weiss emphasises similar advantages in qualitative research approaches with reference to the concept of the case study, 'a research strategy that investigates a phenomenon in its natural setting using multiple sources of evidence' (1998:328). In the study of a high-stakes test such as IELTS, the receiving institutional context was important on key issues such as the IELTS band score cut-off points accepted by particular departments (see Chapter 6). There are findings, too, from the *PL2000* Impact Study (see Chapter 7) which relate interestingly to their particular contexts yet have potential generalisability, for example the influence of parental attitudes to foreign languages on student presence and progress on courses run under the *PL2000* initiative.

Survey research

Both our example impact studies employ what are usually termed s*urvey research* approaches especially, perhaps, in their use of questionnaires, interviews, and observation. Survey research methods are not confined to large sample studies and simple statistical analyses. They appear to suit impact studies, which tend (see Baker 1997:35) to be seeking probabilistic and interactive, rather than deterministic, relationships between individual and group characteristics such as: language background, attitudes, motivation; language learning approaches, strategies; styles of language teaching, and target language performance. Baker also, however, reminds us of problems sometimes associated with survey research. These can include: a lack of clear aims; implicit rather than explicit theoretical input; un-established causal relationships; inadequate sampling; instruments containing invalid items; lack of triangulation through other data collection methods; and interviewer or researcher effects. It is clear that such validity problems are not confined to survey research but apply to other methods of enquiry, including experimental research. Baker's warnings, however, are salutary.

Both the studies described in this book attempt to minimise survey-related problems. Both define and adhere to their objectives and research questions

(see Chapter 3). Both attempt to validate their data collection instruments (see this chapter and Chapter 4). Both, as already indicated, triangulate data and acknowledge the limitations of their samples.

Pre- and post-validation of instrumentation and data

At the time when Cambridge ESOL (UCLES EFL as it was then) was in discussion with Charles Alderson (consultant to the IELTS impact study project from 1995–1996), it was agreed that one of the starting points for such a project was an in-depth review of key constructs. The result was a paper entitled *Ideas for Research into Impact, Washback and IELTS*, submitted to UCLES by Alderson in June 1995. The paper (already cited in Chapter 1 and see also Chapter 4) reviews washback/impact definitions, reminds us of the complexity of relationships between both and any test, suggests that the primary goal of any washback/impact study is 'to describe', and stresses the importance in the conduct of such studies of prediction, baselines and comparisons, the collection of data on IELTS-related classroom activity and teaching/learning materials, and on stakeholder attitudes. A further example of the focus, throughout the study of IELTS impact, on the *validity* of data collected was the commissioned paper *How might impact study instruments be validated?* (Alderson and Banerjee, 1996).

Chapter 4 describes the development, applications and analyses of impact study instrumentation in detail, but it is relevant in this chapter to infer principles and practices which inform our discussion of impact study approaches. It is certainly clear that, whether a study is to the left or the right on the Lazaraton quantitative … qualitative research continua, the means of collecting data should be as valid and reliable as possible.

Key recommended actions and approaches emerging from the 1995 UCLES brief and the resultant papers were:

- a careful exploration of constructs and advance hypotheses by 'insiders' and 'outsiders' before the drafting of instruments
- checks that every item contributes to the corroboration or refutation of washback and impact hypotheses
- the *prediction* of participant responses to compare with actual responses at piloting and trialling stages, with adjustments to instruments made accordingly
- the use of expert and stakeholder judgements on draft instruments
- the comparison of draft instruments with other instruments
- the use of a combination of methods of data collection, approaching validation from a number of angles with data triangulation used as a check.

The studies of IELTS and *PL2000* impact, as will be seen in Chapters 3–6 below, included comprehensive instrument validation measures.

Strategies for data analysis

Validity is also, of course, an issue when it comes to data *analysis*, although sound work at the *pre*-validation stage should help reduce some of the problems that can be associated with the production of valid data.

Weiss points to differences in the analysis strategies applied to qualitative and quantitative data because of the 'nature of the data and the analytic intent' (1998:283). Quantitative research, seeking a degree of generalisability through larger sample size, will produce measures yielding numerical values or narrative data coded into numerical values, analysed and post-validated using statistical methods. As Weiss also notes, much of the quantitative researchers' data analysis work 'they have already done through the development and/or selection of measures and data collection'. Once the data are in, quantitative research will:

> … focus on locating significant relationships among them. Through statistical techniques, they identify associations among variables and the likelihood that associations are real and not mere chance fluctuations. Their aim is to model the system of cause and effect (1998:284).

Qualitative studies, on the other hand, seeking more holistic and dynamic information, may well produce more, and more varied, data, and from smaller numbers of participants. The data may be less objectively chronicled, as in the case of field notes, narrative accounts of conversations, recordings of interviews or group discussions, school documents, video-recorded classroom lessons and so on. Such data may be less susceptible to statistical analysis and thus require more qualitative interpretation and validation, often, however, through triangulation with more quantitative data on the same subject. After all, as we have suggested already, much social scientific research collects and analyses both qualitative and quantitative data. Both of the studies exemplified in this book certainly do so.

Davis suggests the following pattern of data analysis in applied linguistic research: a search for patterns of generalisation across multiple sources of data; and the establishment of credibility through systematic evidence from thick description (1995:446). Watanabe, drawing on experience specific to the test washback field, sees the researcher with an array of data to be analysed, including classroom materials, audio and/or video recordings, field notes, interview data, memos, e-mails and computer files. Throughout the data analysis, Watanabe reminds us, research verification must also be taken 'to establish reliability (or the consistency of data analysis) and validity (the relevance of the data) in the quantitative research tradition, and credibility,

transferability, dependability, and confirmability in the qualitative research tradition' (2004:34). His analysis steps include: reviewing the field notes; completing observation analyses; qualitative refinement of relevant categories; quantitative analyses of the extent of their relevance; interpretation through the interplay of data and theory; checking rival interpretations; drawing final inferences (2004:34–35). These data analysis and interpretation steps seem close to the 'basic analytic strategies' suggested by Weiss (1998:285) as used in both quantitative and qualitative data analysis, namely:

- descriptions
- counting
- factoring
- clustering
- comparing
- co-variation
- finding commonalities
- ruling out rival explanations
- modelling, and
- telling a story.

Descriptions from data analysed are used to describe a test such as the IELTS or a foreign language learning reform project such as the *PL2000.* The descriptions may be derived from quantitative analyses such as a bar chart of post-IELTS candidate closed-item responses to a questionnaire on factors most affecting their performance in the test in Chapter 6, or qualitative analyses such as those of teacher focus groups discussing the advantages and disadvantages of the *PL2000* in Chapter 7. In both cases, the presentation and interpretation of impact study findings require description.

Weiss's analytical category of *counting* is a part of the strategy of describing, explicitly or implicitly, most other data analysis operations. The counting involved in the establishment of population samples, and in response selection frequencies on key issues, informs most parts of the data analyses of both IELTS and *PL2000* studies. Wherever it appears necessary, actual counts of participants or responses are included to counteract tendencies to refer to 'more', 'a few' and so on when the scale and proportions of such references are not clear. Where participant opinions are cited in the *PL2000* study, for example, the numbers or prevalence of such opinions are also given.

Factoring in the sense of dividing aggregates into constituent parts, applies, as used in both our example studies, to the coding of narrative information into key words or categories. This technique is employed in the classroom observation and interview analyses to facilitate inferences about the content, frequency and co-variation of categories. *Factor analysis* is used in the confirmation of constructs. The language learning and test-taking

strategies in the Language Learning Questionnaire (see Chapters 1, 4 and elsewhere), for example, are categories confirmed through such *clustering* techniques. Qualitative data analysis also involves clustering, as exemplified, perhaps, in the close analysis of subjective data on learner motivation inferred from the *PL2000* impact study teacher, school head and parent interviews (Chapter 7).

Comparison represents a central operation in impact studies, which are by definition concerned with how things are, with and without, before, during and after, a programme or test. The *PL2000* Impact Study compares student baseline and post-*Progetto* course data. It also compares the performance of students on these courses with the performance of those on courses *not* run according to the principles of the Project. Analysis of variance (comparing independent and dependent variables across several groups to see if there is significant difference) and regression analysis (to explain and predict relationships between independent and dependent variables) are statistical techniques often used to investigate such differences (see references in the IELTS study in Chapter 4). Covariation is another key concept in impact studies as it refers to the relating of changes in one phenomenon to those in another. An example of this in both IELTS and the *PL2000* studies is the prediction that the test and the new curricula respectively would lead to different lesson content and teaching approaches (see Chapters 6 and 7). Covariance analysis, an extension of the analysis of variance, may be used to measure such phenomena quantitatively.

Davis's *search for patterns of generalisation* (1995) and Weiss's *finding commonalities,* are basic strategies of data analysis, entailing the identification of trends and significant effects. What are the most common preparation course experiences of IELTS candidates? What are the main impacts on head teachers of the *PL2000?* What are the common elements in learner and teacher reactions to both? The examination of 'deviant' cases is the corollary to finding commonalities, suggesting that it may not always be appropriate to drop such cases from one's data in favour of concentrating on the central tendency, since important impact information may 'lurk at the extreme ends of a distribution' (Weiss 1998:287). Analytic induction entails a focus on any evidence that challenges hypothesised constructs, and encourages researchers to keep modifying these constructs until they cover all the data; 'outlier' cases may indicate unexpected but nevertheless significant impacts. Interesting, possibly contradictory, responses on IELTS stress and motivation impacts from the teacher questionnaire used in this study were investigated (see Chapter 6). Differences in student and teacher perceptions of activities in the same lessons were found to be revealing in the *PL2000* study (see Chapter 7).

Watanabe's *'checking rival interpretations'* is clearly a crucial process in all kinds of research. In real-time educational research in particular, every

effort should be made to distinguish changes brought about by the programme or test concerned from changes caused by extraneous factors. But with the complexity of factors involved in learning, teaching and testing, it is not at all easy to rule out the possibility that factors or events other than the programme or test are responsible for observed washback or impacts. The data analyst must seek possible alternative explanations. To what extent is tension caused in candidates for the IELTS test to do with the test itself, to what extent by their dislike of *any* test? (See Chapter 6.) To what extent was the claimed target language improvement of students under the *PL2000* the result of the teaching approaches of the Project itself, and/or to what extent was the improvement motivated by the introduction of higher stakes external exams at the end of the course? (See Chapter 7.) The impact researcher can attempt to resolve such questions through closer examination of existing data, by collecting new data on the issue, or, at least, by acknowledging the query and analysing the possibilities.

Modelling and *telling the story* are the final data analysis strategies listed by Weiss, and would, of course, include Watanabe's *drawing final inferences,* a balanced final evaluation of which research hypotheses are supported, and which are not, to what extent, and why. Modelling, in this sense, means pulling together the main threads of the data to support descriptions and explanations of key processes and outcomes. The 'telling of the story' is the communication to readers of the findings of an impact study, clarifying results and inferences, re-emphasising contextual and generalisability factors and linking recommendations with future developments. A good story is most likely to be told when the researcher fully understands the purpose, processes and outcome of the study, and can present its conclusions with confidence, openness and well-supported argument. The 'story to be told' also needs to suit the book (or institutional context) in which it appears, to be accessible and to engage the readers (or stakeholders).

This chapter has attempted to summarise qualitative and quantitative data collection and analysis approaches in general, illustrating them with references to studies of IELTS and *PL2000* impact. We have examined, so far mainly at the level of principle rather than practice, key research issues such as the selection of participants, survey methods, pre- and post-data collection, the validation of instruments, generalising, data analytic patterns and strategies. In more human terms, these phases in the research process are intended to produce research that, as Weir suggests, should be *believable, logical, feasible*, have *value* and *interest.* Even better, Weir adds, if the research is 'important to the person doing it' (2004:221–222).

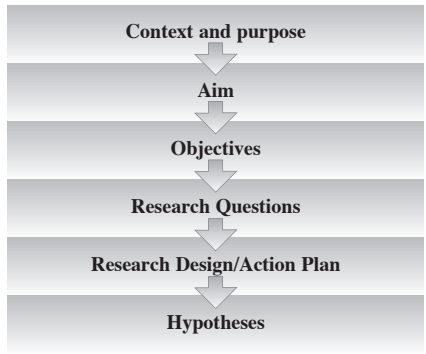
It would seem appropriate now to consider the specification of aims and research questions, and the development of research designs, in pursuit of answers. These are the themes of Chapter 3.

3 Impact study objectives, designs and research questions

Chapter 2 illustrated certain qualitative and quantitative data collection and analysis approaches with reference to the studies of IELTS and *PL2000* impacts. It is logical now to progress to the first *implementational* step in such studies, that is the specification of aims, conventionally translated into research questions and action plans. These are the themes of Chapter 3.

There would appear to be a hierarchy of preparation, planning and research design procedures for impact studies (and other kinds of project research) which it might be helpful to discuss here, before moving to the instrumentation and implementation of such studies in Chapters 4–7. A reasonable context from which to derive conceptual definitions and distinctions might be academic research, for which the conventional hierarchy of pre-implementational categories may be summarised as in Figure 3.1.

Figure 3.1 Hierarchy of research pre-implementation



The labels here are often flexibly defined and inconsistently used, as might be expected, but there is a conceptual pattern to the pre-implementation planning and development process of research projects (Bunton 1998). As is the case throughout this book, the relevant processes will be discussed here with reference to our two case studies.

Context and purpose

In the case of practical research such as an impact study, 'context' will have a definition close to the conventional academic research term 'scope', which is normally used to define the broad subject area of a programme of research and its relationship with other work in the field. In a sense, we have already established the context and purpose for our two studies. Their context is the *validation* (see Chapter 1) of high-stakes language tests through checks on their validity, reliability, impact and practicality (Cambridge ESOL's VRIP), in which the investigation of the effects of such tests on various stakeholders has the *purpose* of providing data and recommendations on test washback and impact.

Also to be included under the construct of context might be what Weiss calls 'evaluability' (1998:95), which entails making sure that a project, programme or test is amenable to washback and impact study. Both the IELTS test and the *Progetto Lingue* were agreed to be appropriate for impact research. The IELTS test, with its very high-stakes role and growing candidature worldwide, was already being evaluated and monitored through routine and project-based studies (see Chapter 1 above). What is more, a study into the test's washback and impact had been foreseen since the mid-1990s when 'it was agreed that procedures would be developed to monitor the impact of the test and to contribute to the next revision cycle' (Saville 2001:5).

As discussed in Chapter 1, the Italian Ministry of Education had included in its plan for the *PL2000* systematic monitoring at provincial, regional and national levels. Cambridge ESOL saw the *Progetto* as appropriate for impact study in a validation context because its own language tests had been approved for the external certification of the foreign language proficiency and progress of students on *PL2000* courses. Cambridge ESOL exams were selected for this purpose along with the tests of other international language test providers such as the Trinity International Examinations Board (also for students of English language), the Alliance Française (for French), the Goethe Institut (for German) and the Instituto Cervantes (for Spanish). A study of the processes and products of English language teaching, learning and assessment on *PL2000* English language courses would, it was felt, produce revealing data on Cambridge ESOL Main Suite exams, especially the Key English Test (KET), the Preliminary English Test (PET), the First Certificate in English (FCE) and the Certificate in Advanced English (CAE), which were being used for the external certification, respectively, of *PL2000* students at CEF levels A2, B1, B2 and C1.

Aims and objectives

Aims and objectives are often seen as representing different levels in the research project hierarchy, the aim representing the broad purpose of a project or study, the objectives defined as the desired or intended outcomes of the

project. In some project planning, however, aims and objectives are not hierarchically distinct. Characteristic, though flexible, patterns for the pre-implementation phase of academic research include the following, each with a differently named but similar *initial* step:

- aim > purposes > research objectives > research questions (for each objective)
- purpose/rationale/significance > research questions > hypotheses
- general research question > related research questions > hypotheses.

In a less academic context but nevertheless referring to educational research initiatives such as impact studies, Weiss specifies *setting the scene, purposes, understanding the program* and *specifying the key questions* as the four initial steps in a research project (1998:72).

The stated aim of the phases of IELTS impact study described in this book, in the context of Cambridge ESOL's continuous and iterative IELTS validation programme, was summarised by Saville as follows:

> In order to understand the test impact better and to conduct effective surveys to monitor it, it was decided that a range of standardised instruments and procedures should be developed to focus on the following aspects of the test:
>
> - *the content and nature of classroom activity in IELTS-related classes*
> - *the content and nature of IELTS teaching materials, including textbooks*
> - *the views and attitudes of user groups towards IELTS*
> - *the IELTS test taking population and the use of results* (2001:5).

This statement of aim is broad because it includes reference to the context and purpose of IELTS impact research likely to embrace more than one study. As we shall see in Chapters 4 and 5, the statement is in a sense a *meta*-aim, of which the phases described in this book are a rationalised part. The research questions derived from this broad IELTS impact study aim were, as we note in the next section of this chapter, framed to assist the achievement of objectives in the four research areas described by Saville above.

The *PL2000* itself was originally presented by the Italian Ministry of Education in terms of one aim and two objectives. Its *aim* was:

> To set up a continuous course of studies from elementary school to the second year of secondary school for the teaching/learning of foreign languages with a view to the acquisition of a pragmatic-communicative competence in accordance with the directions contained in the Common European Framework of Reference of the Council of Europe. By the end of the programme, pupils should have learned at least two foreign languages regardless of the type of school attended (Ministry of Education, Italy, 1999, *translated*).

The *Progetto*'s stated *objectives* were:

- to develop communicative competence in reading, written and oral interaction and production appropriate to the learners' age, level, domains and contexts of specific language use in the various school types
- to allow for the design of courses for specific purposes, including literary and professional language, as well as the development of partial competence in a single language skill (Ministry of Education, Italy, 1999, *translated*).

The language teaching and learning background to the *PL2000* is described vividly by Duguid in *Anatomy of a Context: English Language Teaching in Italy* (2001). She sees as a factor motivating the Project the poor performance in English of Italian students in a context where the language was becoming increasingly necessary to help Italians take advantage of study and work opportunities in the European Union and global market. Before the *PL2000*, foreign language teaching was dominated by grammar-translation, teacher-centred approaches. These were often in the hands of untrained teachers since qualifications were not a requirement until 1999. There was little or no formal inspection of teachers, who tended to write their own syllabuses and do their own language testing, often using discrete-item writing formats, and long oral tests. Given all this, and the lack of co-ordination between primary and secondary schools, students entered university with little ability to communicate in English.

The *PL2000* was thus conceived as a project to develop standards for language teaching in schools linked to the Common European Framework (CEF), subject to certification by selected external exams and providing improved professional development support for teachers.

The *PL2000* was first proposed as an appropriate focus for an impact study in August 2000 by Dr Peter Hargreaves, then Chief Executive Officer of Cambridge ESOL. In March 2001, following discussions involving the Italian Ministry of Education, senior Cambridge ESOL staff members and an external consultant, the following impact study *aims* were presented for a Cambridge ESOL impact study of the *PL2000*:

- to ascertain the effects of *PL2000* on English language performance through pedagogy, textbooks and other media, evaluation and assessment, and resource centres
- to explore the significance of *PL2000* to main stakeholders, including learners, teachers, teacher trainers, parents, evaluators, and educational managers
- to analyse the development of examinations for the certification of language levels as appropriate to the needs of the growing range of language learners in Italy

- to identify further areas of potential support for teacher in-service and test-awareness programmes, resource centre development, and textbook design
- to provide a focus for the presentation, to other EU countries, of the Italian model for state-sector adoption of the Common European Framework of Reference for language learning.

It might be argued that these aims could have been specified in terms of one aim, that is to ascertain 'the effects of *PL2000* on English language performance through pedagogy, textbooks and other media, evaluation and assessment, resource centres'. This aim could then have been specified in terms of four objectives. Meeting these objectives would entail seeking data on the effects of the Project on:

- a range of stakeholders in the language classroom and beyond
- their relationships with external language exams
- project support facilities, materials and training
- other potential foreign language programme reformers.

The *PL2000* Impact Study is in a way a *two-layered* study, investigating the washback and impact of the *PL2000,* and, within these, the impacts and washback of Cambridge ESOL tests used in the *PL2000* for the certification of English language learners' performance and progress.

There are, it would appear, a variety of ways of specifying impact study aims and objectives.

Research questions

Research questions appear consistently in the various research planning outlines mentioned so far. This is hardly surprising given the element of enquiry that is common to all forms of research. That research questions should be dictated by the context, purpose, aims and objectives of the study is also clear, which locates them after these levels in the sequence of inter-dependent research planning steps. However, experienced researchers warn of the dangers of asking too many questions, producing too many data to analyse, framing questions which stray from the research aims or objectives. Light, Singer and Willett advise that: 'Well-crafted questions guide the systematic planning of research. Formulating your questions precisely enables you to design a study with a good chance of answering the questions' (1990:13). Note here the implied chronological precedence of research questions over research design.

Weiss (1998:95) also emphasises the need for the researcher to frame the right research questions, that is questions that relate essentially to research focus on programme or process, outcomes, and the links between them. Decisions on research questions must also take account of practical matters

such as: access to participants and data; the interests of stakeholders; time, personnel and financial resources, including available research expertise and, in the case of data to be collected in working institutions, the need to minimise disruption to programmes and participants (see, especially, Chapter 5).

Ten research questions were framed to mediate the achievement of the objectives of Phase 3 of the study of IELTS impact in the four research areas cited above. The questions are presented here according to participant or activity focus, namely:

- IELTS pre- and post-test candidates, preparation courses, test impacts on candidates
- IELTS preparation course teachers, test impacts on these teachers
- IELTS preparation materials and test impacts on these materials
- IELTS preparation lessons
- IELTS impacts on receiving institution administrators.

The research questions are sequenced in the order in which they are raised in the five data collection instruments and corresponding face-to-face enquiry activities to be described in Chapter 5.

Focus on IELTS candidates:

- Research Question 1: What are the profiles of the candidates taking the IELTS test?
- Research Question 2: What is the washback of the IELTS test on courses preparing candidates to take it?
- Research Question 3: What are the profiles of the participants who have already taken the IELTS test?
- Research Question 4: What is the impact of IELTS on the participants who have taken the test?

Focus on IELTS preparation course teachers:

- Research Question 5: What are the profiles of the teachers preparing candidates to take the IELTS test?
- Research Question 6: What is the washback of the IELTS test on the teachers preparing candidates to take the test?

Focus on IELTS-related course materials:

- Research Question 7: Which textbooks and other materials are used on IELTS preparation courses?
- Research Question 8: What is the washback of IELTS on these preparation course materials?

Focus on IELTS preparation lesson observation:

- Research Question 9: What do IELTS-preparation lesson observations and analyses indicate about the washback of the IELTS test on the lessons?

Focus on receiving institutions:

- Research Question 10: What is the impact of the IELTS test on receiving institution administrators?

The *PL2000* Impact Study sought to identify, describe and explain the washback and impact of an important and radical educational reform programme, the *PL2000,* on a range of participants. Given the aims of the study as defined above, the main research questions which the study set out to answer were, according to the *PL2000* Impact Study Main Report*,* the following:

> - What washback is the *PL2000* having on the pedagogy, materials, and media for language teaching and learning?
> - What washback are changes in language teaching and learning pedagogy, materials, and media having on the performance and attitudes of the students and the teachers?
> - What washback is the *PL2000* having on language evaluation and assessment?
> - What impacts is the *PL2000* having on educational managers, including Heads of Schools?
> - What impacts is the *PL2000* having on support for teacher in-service and test-awareness programmes, resource centre development, and textbook design? (Hawkey, 2003:17)

Apart from being couched in question form, the research areas indicated by the *PL2000* study aims are rationalised in these questions into more discrete and consecutive elements. The rationalised questions imply that some of the answers might constitute a potential response to the fifth aim of the study, namely 'to provide a focus for the presentation, to other EU countries, of the Italian model for state-sector adoption of the ideas and proposals of the Common European Framework of Reference for Languages: Learning, teaching, assessment' (see above).

Chapter 4 will detail the long and sometimes tortuous route between the initial specification of research questions and their ultimate representation in the various impact study data collection instruments – but that belongs to the implementation rather than the planning phase of a study.

Research designs and action plans

Most of the models of pre-implementation research steps discussed include what may be called *research designs* and/or *action plans*, *work plans* or *project development plans*, depending on whether a study is conducted as academic or as project research. Sometimes, of course, it may be both. The research of Clapham (1996), for example, on the effect of background knowledge on reading comprehension, investigates English for specific purposes constructs underlying the ELTS and IELTS Reading modules, and is described as 'doctoral work carried out … over a number of years and supported by UCLES'.

An *action plan* is concerned with translating the purpose, aim and objectives of the study into a route to be followed, as far as possible, to reach the intended destination, that is the achievement of the study objectives. The action plan should demonstrate that the study team:

- specified the main phases and activities of the project (e.g. instrument development, participant identification and contacts, data collection design, data management systems etc.) which may actually be called the *research design*, that term representing one *part of* the action plan)
- defined the study management structure, staffing and responsibilities
- obtained approved finance for personnel, supplies and equipment and overhead resources
- agreed study monitoring and reporting systems.

Weiss defines a *research design* as 'the plan or structure a researcher develops to guide the study' (1998:330), which distinguishes it from an action plan, which is intended to ensure the framework and resources for the study are defined and available. For Weiss, a research design should specify target participant groups and numbers, selection criteria, data collection intervals, methods and means.

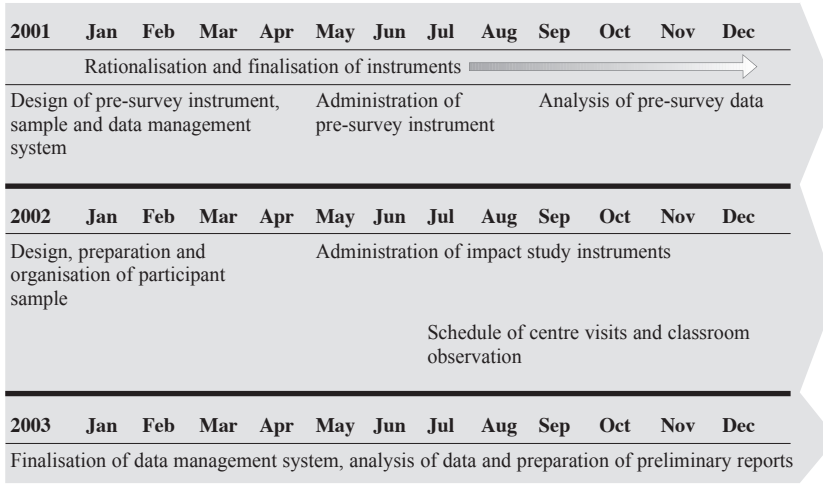
The study of IELTS impact described in this book, which, as we have seen, is a less one-off, short-duration project than the *PL2000* Impact Study, and part of a continuous, formative test consultation and validation programme, had a broad action plan envisaged as comprising three phases:

> Phase 1 for the identification of areas to be targeted and development of data collection instrumentation; Phase 2 for the validation and rationalisation of these instruments, and Phase 3 for the collection and analysis of impact data (Saville and Hawkey, 2004:76).

Each of these phases involved its own action plans (see Chapter 5 for plans guiding the instrument development phase).

The IELTS impact study research questions shown above were contextualised into the impact study implementation research design for Phases 2 and 3 of the project, as shown in Figure 3.2.

Figure 3.2 IELTS impact study Phase 3 research design of data collection chronology, types and inter-relationships



Notice the use of a pre-survey (of 300+ IELTS centres worldwide) to collect up-to-date preparatory data on centres currently administering IELTS, including: their key contacts and schedules; course numbers, durations; participants' backgrounds and numbers; teacher numbers; main IELTS textbooks. This early implementational step (see Chapter 5) also provided an opportunity for the centres concerned to indicate their willingness to participate in the subsequent data collection phase, which would involve some considerable time and effort on the part of their students, candidates and teachers.

Figure 3.3, which shows the very straightforward project planning *framework* agreed for the *PL2000* study, also illustrates the place of an action plan in the project development phase. Figure 3.4 (page 50) reproduces the *PL2000* Impact Study reseach implementation design.

Figure 3.3 *PL2000* Impact Study project plan model

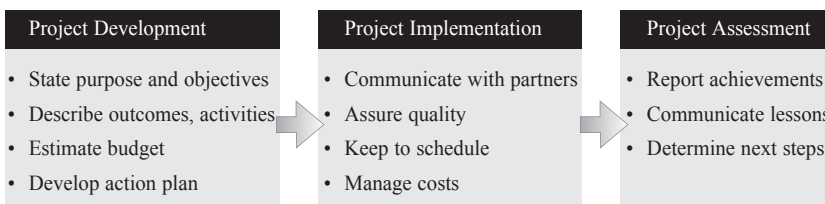
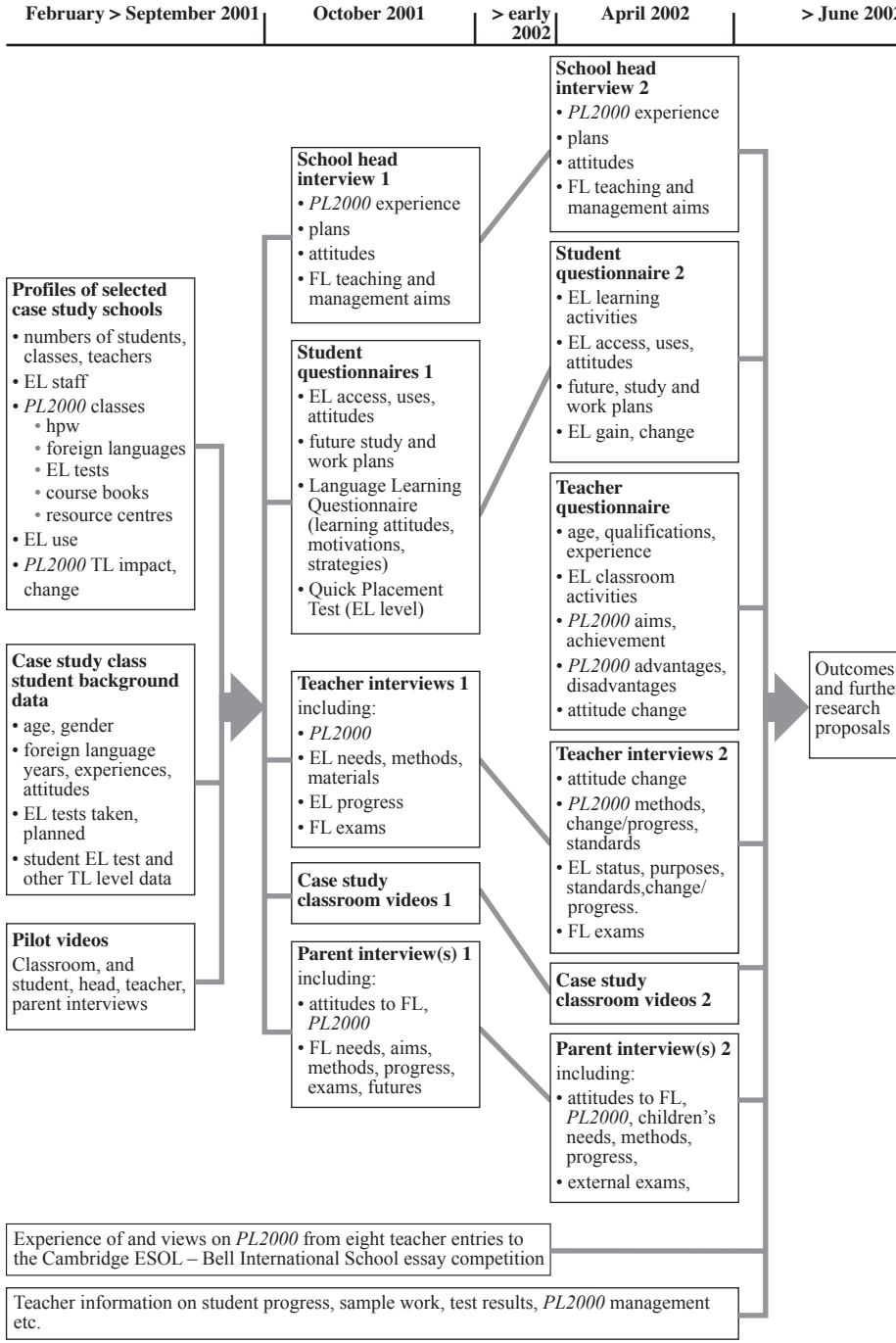


Figure 3.4 *PL2000* Impact Study research design of data collection chronology, types and inter-relationships



The development phase of the *PL2000* Impact Study, from the first proposal for a study in August 2000 to its approved action plan and agreement in November 2000, saw regular draft study proposals circulated, feedback reports exchanged, Cambridge ESOL team policy and planning meetings held at Cambridge and in Italy, and liaison and approval meetings with appropriate Ministry of Education officials. The study attempted and, in the main, adhered to the research implementation design specifying participants, types of data, and data collection methods under the longitudinal study timeline.

Hypotheses

In the models of research planning cited, *hypotheses* tended to come at the end of the planning steps; that is when the study really begins. In fact, though, research, undertaken as it is because there are questions that appear to require answers in some kind of empirical way, also *starts,* implicitly, from hypotheses. In the lay sense of the term, as an unproved theory or proposition tentatively accepted to explain certain facts or, as with a *working* hypothesis, to provide a basis for further investigation, both the IELTS and the *PL2000* studies began with hypotheses.

With the IELTS study, the research question: 'What is the washback of the IELTS test on courses preparing candidates to take it?' implies a working hypothesis:

- the IELTS test affects courses preparing candidates for the test.

When Cambridge ESOL felt that a 'study of the processes and products of English language teaching, learning and assessment on *PL2000* English language courses would produce revealing data on the VRIP of Cambridge Main Suite exams' (see previous), hypotheses were already implied, for example:

- the *PL2000* affects English language teaching on courses under the Project, or
- the use of a Cambridge ESOL exam at the end of an English course under the *PL2000* affects English language learning.

The definition and status of hypotheses in research depend on the research *approach*. Weiss, tending to espouse more qualitative approaches, defines a hypothesis as follows:

> The researcher's firm hunch, arising from theory or experience, that a particular predictor (in the evaluator's case a programme or an aspect of a programme) causes an outcome. Hypotheses are confirmed or denied using empirical analysis (1998:331).

In more quantitative research circles, the definitions are somewhat stricter. As in Snodgrass, for example:

> In hypothesis testing, we divide all the possible states of the world into two competing hypotheses such that if one is true the other is not, and vice versa. One of these hypotheses is called the null hypothesis, and the other is called the alternative hypothesis. The only hypothesis that is tested directly is the null hypothesis, and the investigator either rejects, or fails to reject, the null hypothesis (1977:187).

Restated as a null hypothesis, the IELTS study hypothesis above would be:

- the IELTS test does not affect courses preparing candidates for the test.

It is clear that null hypotheses and their testing are most closely associated with the experimental research paradigm (see previous), where the researcher controls the key variables and the conditions in which the research or hypothesis-testing study takes place. Given the more qualitative research paradigm adopted by both examples and by most impact studies, for the reasons given in Chapter 2, we tend to be dealing with research questions and working hypotheses rather than null and alternative hypotheses.

Green (2003), in his doctoral thesis, a comparative study in backwash between IELTS preparation and university pre-sessional courses, does present his research hypotheses in explicitly null terms, for example: 'IELTS preparation courses do not reflect the English for academic purposes (EAP) construct of academic writing'. The testing of this hypothesis involves the use of questionnaires to language teaching centres, repertory grid interviews (see Kelly, 1955), case study interviews and questionnaires, classroom observation and the analysis of classroom artefacts. Some of Green's hypothesis-testing measures do not differ greatly from those used to collect data in response to the research questions of the IELTS and *PL2000* studies, both of which use questionnaires, case study interviews and classroom observation and the analysis of 'classroom artefacts'. This will become clearer when we look more closely at the implementation of both studies, in Chapters 5–7.

This chapter has illustrated how, both in theory and in practice, and in both our impact studies, the constructs at the top of the research planning hierarchy, purpose, aims, objectives, though not always defined or used consistently, are as essential to the planning of impact studies as they are in other forms of research. There is clearly the need, in the pre-implementation stage of an impact study, to state what the study is setting out to do, and how it relates to its field. Research *aims* are analysed into research *questions,* specified rationally to take account not only of answers that the aims of the research imply are needed, but also of the practical constraints faced by the study. A research design specifies the target participant groups and numbers, selection criteria, methods and means for the collection of data in response to

the research questions. Good research project action plans attempt to provide the essential prompts and criteria for the resourcing and tracking of progress in an impact study. Hypotheses, whether of the null, alternative or working kind, tend to represent the final step in the research planning process, and signal the beginning of the research proper.

In Chapter 4, we shall look at study *implementation.* In particular we shall trace the development of instrumentation for the collection of data in answer to research questions and the support or refutation of hypotheses, all in the process of the fulfilment of research objectives for the achievement of research aims.

4 Impact study instrumentation

Chapter 3 has illustrated the importance of analysing impact research aims into *questions, research design*s, *action plans*, and *hypotheses*. In this chapter we shall report on the next steps in the implementation of our two studies, in particular the development and validation of data collection instrumentation.

The chapter will suggest examples of the conceptual and procedural differences between language tests and impact data collection instruments, then put into context the data collection instruments used in the sample studies. The development, validation and use of the instruments for the two studies will next be summarised. Throughout, there will be a focus on key principles and processes, including operationalising, prediction, brainstorming and review, piloting and trialling, post-trial validation, and, in the case of the IELTS impact, the rationalisation of instrument use.

Distinctions and similarities between data collection and test instruments and their validation

In the paper commissioned as part of the early research for the study of the impacts of IELTS, *How might impact study instruments be validated?* (see Chapter 2), Alderson and Banerjee express some surprise at the apparent lack of 'a well-developed literature on how the reliability and validity' of 'questionnaires, interviews, surveys and classroom observation schedules' should be established (1996:1). They then discuss some of the distinctions and similarities between tests and data collection instruments and in their validation. The insights provided by their paper and experience gained as research progressed on both the impact studies under discussion here may help to explain some of the relationships between the validation of language tests and the validation of data collection instruments. In Chapter 2, we have already seen, at the level mainly of principle rather than practice, how analytic strategies such as the ones described here are essential to *telling the story* of an impact study.

Distinctions

- In data collection instruments, for example in open-ended questionnaire items such as those in the student and teacher questionnaires for both our studies, it is the veracity of participant responses that is the focus of the analysis rather than, as in language tests, their linguistic accuracy or appropriacy. However, since responses to questionnaire items may not always, for a variety of reasons, reflect what participants actually think or feel, checks on veracity need to be made, where feasible, whether through related items or through triangulation from other data sources on the matters concerned.
- Item difficulty analysis, to ascertain 'the relative difficulty of a test item for a given group of test takers' (McNamara 2000:134), would not seem to apply, in its comparative, language proficiency level-setting sense, to questionnaires. The consistent ease of understanding of data-collecting items, from both a conceptual and language difficulty point of view, is, however, crucial, and was a constant concern throughout the development of IELTS and *PL2000* instrumentation and face-to-face data collection.
- Data collection questionnaire items may not be susceptible to certain other language test item analysis operations. Internal consistency measures, methods to estimate test reliability depending on the 'homogeneity of item variance as a reflection of consistency of scoring' (Henning 1987:193), may not be appropriate, for example. Questionnaires such as the IELTS or *PL2000* Impact Study student questionnaires normally seek a *range of information* from members of a participant group. They may well thus contain *single items on a particular autonomous matter* such as these from the *PL2000* study questionnaire, requiring responses on the frequency of activities in participants' foreign language learning classes:

	Frequently	Quite often	Sometimes	Never	Comments
grammar exercises					
discussions with the whole class					

  There is no necessary reason why students' responses on the two items should coincide across classes and teachers, or be 'consistent'.
- Item Response Theory (IRT), a form of statistical analysis that 'gives an estimate score on a test for each individual, basing this estimate on that individual's performance on each of the items on the test' (Hughes

2003:42), may be used to validate test items by ascertaining their match with a test-taker's 'estimated performance'. With questionnaires seeking a range of participant experiential or attitudinal matters, IRT may not be useful for instrument validation. An individual's 'unusual' responses may not be caused by flaws in the items but rather represent the actual value on the variable concerned.

Similarities

It was also becoming clear from our studies that the design of data collection questionnaires and language tests shared certain similarities, including the need for validation so that claims to be made on the basis of their findings would be well-founded. Questionnaires, interviews, surveys and observations are, as Alderson and Banerjee point out:

> … not dissimilar to tests in that they too seek information about a person, be it their spending behaviour, their attitudes towards corporal punishment or the way they learn a new language. So, as with tests, it would not be unreasonable to assume the importance of establishing the reliability and validity of data collection instruments i.e. whether they are capturing the same information each time they are administered and whether these instruments are actually capturing the information researchers expect them to capture (1996:7).

Some of the similarities between test and data collection instruments and their validation potential are now discussed.

- Data instruments measuring cognitive or affective traits by inviting participant responses to statements representing the trait concerned are similar to certain kinds of language tests, for example those using multiple choice items. The Language Learning Questionnaires (LLQ), one such example, were originally developed, in collaboration with UCLES EFL, by Bachman, Cushing and Purpura (1993) and finalised (Purpura 1999) in the form of questionnaires on socio-psychological factors, namely language learning attitudes, anxiety, motivation and effort. The LLQ which was used to collect data in both the IELTS (see Appendix A) and the *PL2000* studies uses sets of items representing particular traits (or abilities). It is thus more suited to some of the validation operations associated with language tests, for example:
  - *Descriptive analyses* e.g. totals; proportions; means; standard deviation: skew and kurtosis (both these latter terms referring to departures from the normal distribution of scores, skew so that 'the pattern of scores is not symmetrical around the mean', kurtosis so that the distribution is 'flatter or more peaked than normal', Henning

1987:183, 197); such analyses led, for example, to proposals during Phase 2 of the IELTS study (Gardiner 1999) for the modification of the 7-point scale on the Test-taker Background Questionnaire of socio-cognitive and cognitive, and test-taking variables to five points (see also below).

  - *Factor analysis*, which refers to 'a variety of multivariate, correlational statistical procedures used to aggregate data into non-overlapping or minimally overlapping categories or factors' (Henning 1987:192), is used in two ways. The first, exploratory factor analysis, examines patterns in the observed questionnaire data but assumes 'no a priori patterns in the data'. Confirmatory factor analysis is used 'to determine the extent to which items designed to measure a particular factor actually do so'. Both these definitions are given by Purpura himself (2004:100) writing on the statistical procedures considered and used in the validation of the Language Learning Questionnaires.

- Questionnaires, like multiple-item language tests, are amenable to validation through correlation analyses within instruments, for example *PL2000* student baseline data performance on the Oxford University Press Quick Placement Test (QPT) (see Chapter 5), and subsequent performance on their First Certificate in English (FCE) exams (both also see Chapter 7). Questionnaire item responses may also be checked for validity through triangulation across data sources, for example IELTS student questionnaire responses on test module difficulty compared with student interview and focus group responses on the same matter.
- Retrospective interviews, as in some of the instrument piloting for the study of IELTS impact referred to in this chapter, and participant think-aloud protocols are seen as possible validating tools for questionnaire instruments, as they are for some test tasks.
- Inter-evaluator (c.f. inter-rater) reliability may be checked, as it was in the trialling of the Instrument for the Analysis of Textbook Materials (IATM), (see Appendix C), in the study of IELTS impact. Comparisons were made between the responses of different evaluators on the same textbooks and the responses of the same evaluators on different textbooks, representing an attempt to establish convergent – divergent validity (Litwin 1995:44 cited in Alderson and Banerjee 1996:27).
- Test-retest reliability validation is an 'estimate of the likelihood that the test would rank test takers in the same order from one administration to another proximate one' (Henning 1987:196). The technique is mentioned as planned for the validation of the IELTS impact Test-taker Background Questionnaire (precursor to the LLQ), when trial questionnaires were sent to IELTS centres during 1997–8, with 'a number of candidates completing the questionnaire twice, so a test-retest reliability analysis

could be carried out' (Gardiner 1999:4). No test-retest data were, in the event, analysed from this trial, but there would appear to be questions over the use of the technique with instruments seeking attitudinal data, which are subject to fluctuation, although instruments such as the LLQ would seem more amenable to test-retest analyses since they are measuring socio-psychological traits of some permanency.

- Multi-trait, multi-method validation procedures are also referenced by Alderson and Banerjee, premised, as they are, on 'the idea that if a construct exists it should be possible to measure it in more than one way' in the expectation that 'the correlation among these different measures of the same thing should be strong and positive' Fitz-Gibbon and Morris (1987:15). Logically, this principle should apply both to tests and means of data collection. It also appears to overlap triangulation (see Chapter 2 and this chapter) quite strongly. In the first three phases of the study of IELTS impact study, attention was paid during instrument development and validation, through piloting, trialling and use, to the comparison of data across, for example:
  - closed and open-items on the same construct in the same instrument e.g. in the Teacher Questionnaire (see Appendix B):

    If an IELTS score had not been a requirement would you have prepared your students for their future studies in the same way? Yes/No

    Would your IELTS preparation course be a good way to learn English for someone going to university but who is not going to take IELTS? Why? Why not?
  - responses on the same construct across different data collection means, e.g. Teacher Questionnaire and teacher interviews/focus groups (see this chapter and Chapter 6).

Since, as Alderson and Banerjee (1996:7) suggest, impact study data collection instruments tend sometimes not to be validated at all and since, it would appear, no standard models of instrument validation for use in the study of impact appear to exist, it seems natural to look to language test validation principles and procedures for direction. As the examples above indicate, there is clearly no exact fit across the two fields, given key differences of topic, purpose and enquiry type. But there are lessons to be learned.

None of the validation techniques covered above will, of course, in themselves prevent the fundamental problems of *construct under-* or *over-representation,* that is, making sure that data collection instruments actually cover adequately the areas required by the research focus. It is only through comprehensive and thorough work on study aims and research questions, followed by their iterative verification using approaches such as those discussed below (e.g. brainstorming, expert opinion, prediction and review), that this will be achieved.

Contextualising the instruments

IELTS impact study

As discussed in Chapter 2, work on IELTS impact studies was to begin by re-examining fundamental impact and washback concepts. Thus when, in June 1995, in response to a request by Dr Mike Milanovic, then Director of the Validation Unit of UCLES EFL, Charles Alderson submitted the paper *Ideas for research into impact, washback and IELTS* (see Chapter 2), it was exploratory and explanatory on complex issues such as:

- the major descriptive function of washback/impact studies
- their role in test validation
- high-stakes test stakeholders
- the scope and limitations of impact studies.

The paper also covered 'specific ideas for research' such as:

- predicting impacts, then comparing predictions with response data
- the need for baseline data (that is 'measurements obtained to indicate levels of performance before the intervention of an educational or other programme against which to assess change' (Satterly 1989:336))
- the relevance of attitudinal data
- the need to distinguish between real impacts and coincidental policy changes
- the importance of systematising test impact research including 'specifically set-up studies' (1995:11) as a contribution to continuing test revisions.

The difficult problem of identifying with any certainty cases of classroom washback remained, of course. It is not easy to 'prove' that events in a classroom or performances beyond it are actually washback from tests or curricular innovation, rather than the result of other factors. These 'other factors' may include learner characteristics, attitudes, levels, preferences and so on. They may also include teachers' characteristics, attitudes, levels, preferences, approaches – or the changes in classroom processes and learner performance may be the result of extraneous influences not really to do with learners or teachers (see also Alderson and Hamp-Lyons 1996, Andrews 1995, Watanabe 1996). Given the complexity of washback relationships, washback studies can only seek to ask the right people the right questions, analyse the answers rigorously, present findings accurately and fairly, infer and explain insightfully on such matters.

Alderson's June 1995 paper ended with this 'final thought':

> One of the most significant ways in which UCLES can have impact on the profession – of teachers as well as testers – is by showing that it takes seriously its responsibility to have positive impact where at all possible and that it is willing to investigate and establish this. In doing this, UCLES will be taking a very important lead in a traditionally conservative area, and I pay tribute to UCLES' willingness to innovate, investigate, and hopefully report on its findings in the general area of impact and washback (1995:12).

A further obvious concern, from the outset of the IELTS impact study project, was the closest possible attention to the *validity* of data collected, the theme of the further paper commissioned for the study: *How might impact study instruments be validated?* (Alderson and Banerjee 1996), which proved seminal to the IELTS impact study (see above).

The historical context for such concerns is of relevance to this book in that it once again illustrates the systemic role of impact study in the validation of tests and programmes. The specifications document of the revised (1995) form of the IELTS test identifies the collection of test impact data as part of UCLES' continuing test validation programme. It was around this time, too, that UCLES was espousing the 'expanded' view of impact (see Chapter 1) with four maxims of test impact to 'achieve appropriate impact with international examinations', namely:

- *plan* (through a 'rational and explicit approach to test development')
- *support* (of 'stakeholders in the testing process')
- *communicate* (by providing 'comprehensive, useful and transparent information') and
- *monitor and evaluate* (by collecting 'all relevant data' and 'analysing' as required') (Milanovic and Saville 1996:10).

Alderson was the obvious choice for a key role in the development of instrumentation for studies of IELTS impact. He was co-editor, with Caroline Clapham his colleague at Lancaster, of IELTS Research Reports 2 and 3 in the English Language Testing Service (ELTS) revision project which led to the revised test, the IELTS, in 1989. Alderson also, of course, had experience of the practices as well as the theory of impact studies through the Sri Lankan Impact Study (Wall and Alderson 1993) and a study of the washback of the TOEFL test on preparation courses (Alderson and Hamp-Lyons 1996).

The 15 June 1995 commissioning letter from UCLES EFL requested Alderson and his team to 'carry out the next stage of activity', namely to specify the four main project areas as described by UCLES in terms of staffing and the delivery of a set of instruments for the collection of relevant impact data. The original wording of the four project area descriptions is significant in that it illustrates the breadth and continuity of the studies envisaged by UCLES EFL:

1. The design of a range of instruments and procedures for the collection of classroom data which will be applicable to IELTS and eventually to other UCLES EFL exams. This will be based on work done by Lancaster staff and will include the following: a) a questionnaire for teachers; b) a questionnaire for students; c) an observation schedule for classroom activity; d) a procedure for producing summaries of critical classroom activity; e) a procedure for setting up focus groups and recording the data (questions, audio recording etc.).
2. An instrument for materials analysis which will focus on IELTS materials which are currently available. This can be based on a number of sources including the Colorado experiment [reference to the Alderson and Hamp-Lyons 1996 study of the washback of the TOEFL test], work by post-graduate students at Lancaster and existing checklists (such as the ALTE checklists).
3. The design of four questionnaires to collect views and attitudes about IELTS from the 4 main user groups: a) test takers b) EFL teachers; c) administrators; d) lecturers in receiving institutions.
4. An instrument for collecting information about the test-taking population and the use of test results i.e. in addition to the standard demographic data collected using the [Cambridge ESOL] Candidate Information System. This instrument focusing on IELTS candidates can be adapted from a range of existing questionnaires (such as the UCLES/UCLA questionnaires [reference to Bachman, Cushing and Purpura (1993)].

Note then that an extensive range of instruments was foreseen for a series of IELTS-related impact studies, and with the possibility of the eventual use of the instruments beyond the IELTS study.

Contextualising the instruments: the *PL2000* Impact Study

The instrumentation for the *Progetto Lingue 2000* Impact Study was derived from the aims of the study as agreed between the Italian Ministry of Education, senior Cambridge ESOL staff members and an external consultant (see Chapter 3). The study was envisaged as seeking data on influences of *PL2000* on English language performance, pedagogy, textbooks and other media, evaluation and assessment, and resource centres. The data were to be collected from stakeholders such as learners, teachers, teacher trainers, parents, evaluators, and educational managers.

The data collection instruments designed for the study were:

1. A school profile *pro-forma* (see Appendix F).
2. A student questionnaire (in three parallel versions) (see Appendix G).

3. A teacher questionnaire (see Appendix H).
4. A classroom observation analysis form (see Appendix D).
5. Structured interview or focus group forms for use with teachers, school heads and parents (see Appendix I).

IELTS impact study instrument development and validation: principles and processes

At this stage of a chapter summarising the development and validation of the instrumentation used in our two studies of impact, it seems appropriate to trace how the instruments were developed and validated. Phase 1 of the study of IELTS impact would cover 'the identification of areas to be targeted and development of data collection instrumentation', and Phase 2 'the validation and rationalisation of these instruments'. Phase 3, as already noted, would cover 'the collection and analysis of impact data' (all from Saville and Hawkey 2004:76).

The instruments for use in the study of IELTS impact were characterised by the following *cycles of procedures*, all contributing, it was agreed, to the design of *valid* instruments:

- *operationalisation*
- *prediction*
- *brainstorming and review*
- *piloting and trialling*
- *qualitative and quantitative validation, and*
- *rationalisation.*

IELTS impact study: operationalisation

IELTS impact study aims (see Chapter 3) had been agreed and specified in advance, as had the scope, main research questions, and appropriate methodological options for the study. The *constructs,* that is the traits, characteristics or features hypothesised as relevant to the impact study research questions had, of course, to be *operationalised,* that is converted into traits that can be measured, 'relating constructs defined theoretically to our observations of behaviour' (Bachman 1990:42). The second stage of operationalisation involves including the constructs in an instrument or procedure for collecting data on them.

Figure 4.1 (page 67) will summarise the contents and item types of the data collection instruments to be used in IELTS impact study Phase 3. To illustrate

the processes entailed in developing, trialling and validating the instruments, experience with particular IELTS impact study instruments (student questionnaires, teacher questionnaires, textbook analysis instrument, classroom observation form) will be cited here, then summarised in the Figure.

IELTS impact study: prediction

As noted in Chapter 2, a prominent role in the instrument development and validation process was given to clarifying constructs through *predictions* of the washback/impact of the test on major test areas, e.g. reading, writing, listening, speaking. The study of IELTS impact also used response prediction at the questionnaire *item* level to help ensure the relevance and workability of items. The *Proposal for work to develop instruments to study the washback and impact of IELTS* (Alderson), notes that it 'is essential to gather *predictions* [my italics] as to the likely washback effects of IELTS and associated materials from item writers, EAP teachers, teacher trainers and language testers' (3 August 1995). The April 1996 *IELTS Impact Study Report* from Lancaster to UCLES notes that such predictions were collected from participants at a British Council Linguistics seminar (September 1995), which also included members of the Lancaster Testing Research Group (LTRG). Participants should, Alderson suggested, consider washback effects 'from the point of view of: topics, content, texts, skills, feedback types, test-taking and in-class use' (ibid:1).

A further Phase 1 Lancaster discussion meeting on 24 October 1995 illustrated the importance attached to predicting washback, and the fact that the team pursued such predictions assiduously. The report of the meeting notes that participants (13 researchers, an MA student and staff from the Institute of English Language Education and the Department of Linguistics and Modern English Language) were asked actually to *take* the IELTS sample Listening test and receive an orientation briefing on the Reading module, before brainstorming on: how they would expect students to prepare for the tests, what sort of test preparation techniques to expect, what sort of preparation would be appropriate, what methodology, and what they would consider to be negative washback in preparation practice. The usefulness of this exercise in the development of Phase 1 student and teacher questionnaire items and in the design of classroom observation instruments will be appreciated.

An example of the useful debate and thought encouraged by the use of response prediction was on the question of test fairness. Test taker perceptions of the fairnesss of the IELTS test seemed to various impact study data instrument developers relevant to a research question such as: What is the impact of IELTS on the participants who have taken the test? But predicting the response in the process of validating items brought a range of responses

suggesting the need for careful follow-up sub-items (for example on reasons for the response) and related items (for example on the test takers' typical and IELTS-specific views of their own test success).

IELTS impact study: brainstorming and review

Brainstorming, meaning discussion of evidence and the offering of different views to enable movement towards problem solution (May and Sigsworth 1983:257–58) also proved important in the development of IELTS impact study instruments. Brainstorming sessions like those mentioned in the previous paragraph would typically cover test effects on Alderson's (above) teaching/learning topics, texts, tasks, skills, feedback types, test-taking and in-class use. In Phase 1, such matters as these were *reviewed* regularly by project staff at Lancaster and the Project Co-ordinator (Alderson), and at impact study liaison meetings between Alderson and relevant UCLES EFL senior management and IELTS staff. Records of such meetings were circulated at UCLES and Lancaster for further comment. The *Notes on Meeting at UCLES, IELTS Washback Project, 1 November 1995,* exemplified the combination of prediction and brainstorming methods of instrument validation, as well as the fairly formalised dissemination and review procedures:

> The Lancaster Language Testing Research Group had already conducted a brainstorming on predicted washback of Listening and Speaking [*see preceding section*]. The report would be edited by the Lancaster team, and then sent to UCLES for comment and additions. A similar procedure would be followed for the results of today's meeting, with the Lancaster team commenting on the edited version of today. The two edited sets of predictions would then be combined into one document which would be sent to those who had not been able to be present at either meeting, as well as other interested and affected parties in the UK and [Australia]. A version of the predictions would be discussed at the [IELTS] Advisory Committee in February [1996], but instrument construction would have to commence using whatever sets of predictions were available when the work began (1 November 1995).

The May 1996 *IELTS Impact Study Report* from Lancaster to UCLES also underlines the importance in instrument development of brainstorming on item and scale-types. These sessions would attempt to ensure as far as possible that item wording enabled participants to understand the questions in the way they were intended (taking account of the fact that learner participants in the study would not, in the main, be first-language users of English) and to review draft instruments. At the Lancaster Language Testing Research Group (LTRG) meeting of 23 April 1996, for example, draft instruments had been reviewed by members with reference to four questions:

- is the instrument/procedure likely to measure IELTS washback adequately?
- have important aspects of washback/impact been overlooked?
- can the instrument/procedure be shortened?
- are there any practical points on instrument administration or data analysis?

Note the concern here for measurement efficiency, construct under-representation, economy of time, and practicality.

IELTS impact study: piloting and trialling draft instruments

Alderson and Banerjee suggest that 'pilot surveys involve the administration of the instrument on a small number of respondents with the intention of revealing problem areas in the instrument'(1996:7). They also feel that much of the relevant literature on *piloting* concentrates on procedures for eliciting information and views from participants that may be useful to 'check the adequacy of questions' rather than running 'statistical analyses on the pilot data' (op. cit). Presumably, however, the latter procedures would depend on the size of the pilot sample.

In the case of Phase 1 of the IELTS impact study, the piloting was carried out on small numbers of participants but with intensive analysis of their reactions to the draft instruments and items. Studies also involved subsequent 'wider scale international piloting' (IELTS impact study report (1996:3), often called *'trialling'* in project documentation to distinguish it from the small-scale Phase 1 piloting. The trials used larger sample participant groups to allow for statistical analyses and instrument modifications based on these. So, the piloting was carried out during the Phase 1 Lancaster development phase of the IELTS study, the trialling by UCLES in Phase 2.

IELTS impact study: post-trial validation

Examples will be seen (Figure 4.1 below) of a wide range of techniques, both quantitative and qualitative, used in the validation of draft data collection instruments for the study of IELTS impact. The Figure will also describe their modification into the forms in which they were used to collect impact data. Descriptive analyses (frequency statistics, mean, standard deviation, skew, kurtosis, see above) were run using an Access database, with the data imported into the Statistical Package for Social Science (SPSS), for example on the Student and Teacher Questionnaires). Factor analyses used the EQS for Windows programme, for example on the trial data from the Test Takers'

Background Questionnaire (TTBQ). This instrument also made use of quantitative and qualitative revision data in the form of reviews by UCLES EFL Validation Group staff, as well as outside consultants such as Hawkey and Green (2000), Kunnan (2000), Purpura (1996b). These reviews led to modifications to constructs, scope, scales and item formats. Convergent – divergent validation data on the draft Instrument for the Analysis of Textbook Materials (IATM) were derived from more than one evaluator analysing the same book, and the same evaluators analysing different books (see previous). Rater responses from the trial analyses of the draft IATM were consolidated on to a comparative item analysis form covering all draft items. The response aggregates were investigated in particular for inter-rater differences and inconsistencies. Qualitative validating data on the draft IATM were also collected at the intensive brainstorming sessions between evaluators and the questionnaire reviser. Brainstorming and review exercises took similar forms in the validation of all the IELTS Phase 3 impact study instruments. The full, finalised instruments appear at Appendices A–E.

Figure 4.1 (pages 67–72) now summarises main events in the operationalisation, piloting, trialling and validation of instruments during IELTS impact study Phases 1 and 2.

IELTS impact study: rationalisation

With phased projects such as the study of IELTS impact, which have an iterative design and include instrumentation not necessarily intended for immediate or all-at-once use, it is likely that the collection of data for any particular study phase will entail *rationalisation* of instrument use. The discussions of instrument use in Figure 4.1 will already have indicated that some instruments were not yet to be used in Phase 3 of the study, and that others were to be used in modified form.

Figure 4.1 Summary of main operationalisation, piloting, trialling and post-draft instrument validation exercises for IELTS impact study instruments

IELTS Study Instrument	Operationalisation	Piloting and Trialling	Validation
Student Questionnaire (SQ, *see Appendix A*)	1) Focus areas of Phase 3 version derived from Phase 1 draft versions of SQ i.e.: • Questionnaire 5 (Q5) for students preparing for IELTS • Q8 for students who have taken IELTS • Test Taker Background Q12 (TTBQ) and LLQ modified on basis of trial data from Phase 2. 2) Most focus areas retained, i.e.: Student: • L1, Target Language (TL) background • IELTS prep course type, activities, materials, views • for post-test candidates: IELTS outcomes, aims, pressure, fairness, difficulties, performance self-assessment, likes/dislikes, advice and general comment. 3) Validated TTBQ and LLQ items included in SQ. 4) Item response formats: open-ended, Yes/No, option selection (inventory -categorial, proportion, frequency/time, usefulness, intensity, Likert scales, rank order). 5) Phase 3 SQ includes items from Phase 1 TTBQ and Phase 2 LLQ development and trialling.	• Final version of Phase 1 draft SQ (without LLQ) after local piloting (Horak 1996), then trialled on 379 students in nine countries in Phase 2 • TTQB (Herington 1996) draft informed by LLQ , (Bachman, Cushing, Purpura 1993), trialled on 126 IELTS test takers at four UK test centres. • Phase 2 revised TTBQ, including LLQ items, trialled on 500 candidates before their IELTS (Gardiner, 1999).	• Trialled instrument data interpreted through descriptive analysis by SPSS: *value label, frequency, percentage, valid percentage,* and *cumulative percentage*. • Open-ended responses analysed qualitatively for additional insights into research questions. • Triangulation of quantitative and qualitative data, backing or rebutting related data. • Analyses included frequencies, percentages, reliability estimates (Kuder-Richardson 20 for internal consistency, acknowledged at the time as not entirely appropriate since constructs represented in the TTQB by only two items each). • Mean, standard deviation, skew, kurtosis, frequency and factor analyses on trial data, with recommendations for revision. • Revisions also suggested by UCLES EFL Validation Unit, Kunnan, Purpura, Hawkey and Green (see Hawkey and Green 2000, Kunnan 2000, Purpura 1996b). • Quantitative and qualitative analyses taken into account in further revisions, incl. changes to constructs, scope, scales and format. • Phase 3 SQ subjected to expert analysis (Green, 2003).

Figure 4.1 continued

IELTS Study Instrument	Operationalisation	Piloting and Trialling	Validation
Teacher Questionnaire (TQ, *see Appendix B*)	1) Constructs related to those in SQ by aims and research questions; some items identical e.g. on IELTS suitability, preparation course type, activities, test module difficulty, for candidate and teacher attitude comparisons. 2) TQ contents: Teacher • background and experience • students: ages, countries, edn., target countries, module, IELTS status • IELTS suitability, reliability, pressures, washback • IELTS preparation likes and dislikes • candidate success features • advice on IELTS preparation • preparation course success • test features • IELTS prep classes: type, length, course activities; test practice, materials; participant success characteristics, suitability • further comments. 3) Item response formats: open ended, Yes/No, option selection (inventory-categorial, proportion, frequency/time/proportion, usefulness, intensity) and rank order.	• Final version after local piloting is eighth Phase 1 edition; this (Horak, 1996) trialled in Phase 2 with 60 teachers in seven countries.	• Trialled instrument data interpreted through same validating processes as Student Questionnaire (SQ).
Instrument for the Analysis of Textbook Materials (IATM, *see Appendix C*)	1) IATM Constructs from study aims and research questions. 2) Main sources for lists of categories: ALTE Development and descriptive checklists for tasks and examinations (1995), Munby (1978)	• Phase 1 (Bonkowski, 1996) trialling through nine trial IATM textbook rater analyses by practising teachers. • Phase 2 analysis and revision, final draft trialled with two expert evaluators.	• Phase 1 convergent-divergent, multi-trait, multi-method validation data on draft IATM, from four raters evaluating IELTS-oriented textbook, two a TOEFL prep book, two an upper-intermediate book, one a different further IELTS prep book.

Figure 4.1 continued

IELTS Study Instrument	Operationalisation	Piloting and Trialling	Validation
	3) Teacher /evaluator/ textbook details • teacher evaluator background • textbook ID, type • book organisation units • language features • enabling skills • question/tasking techniques • communicative opportunities • text types • text topics • authenticity • open-ended comment: listening, reading, writing, speaking • open-ended comment on whole book • open-ended comment on the relationship between the book and test(s).		• Phase 1 brainstorming, expert opinion and review incl. by textbook writers. • inter-rater discussion of IATM findings included by Yue in thesis (1997). • Phase 2 quantitative analysis, rater responses given comparative item analysis, re. differences, inconsistencies, insights whether problems caused by interpretations of wording or raters' interpretation of textbook. • Saville, Hawkey (2004: 81) note five kinds of modifications to draft IATM: exclusion, items modified, items merged, items moved, items supplemented. • length of the Phase 1 version a major consideration.
Classroom Observation Instruments (*see Appendix D*)	• Phase 1 instrumentation: (Banerjee 1996), with insights from Sri Lankan Impact Study (Wall, Alderson 1993), Nicaraguan, Impact study (Luxon 1996), TOEFL washback study (Alderson, Hamp-Lyons 1996), IELTS Specifications. • pre-observation questionnaire re: observer, teacher, venue, class, course, IELTS relationship, students, materials, lesson focus • observation schedule: six-column matrix observation schedule for activities, timings, teacher and students' actions, interaction, skills; inventories for selection re. text related activities, homework • post-observation: teacher EL use, IELTS impact, lesson feedback.	• Phase 1 instruments piloted on a small scale with staff and students at Lancaster University, and modified (six drafts) in line with extensive feedback from individuals with a research interest in classroom observation. • Phase 2 comparative use by Green and Hawkey of COLT and simpler draft alternative classroom observation analysis model on the same two video-recorded lessons collected for the IELTS impact and Green's own studies.	• Phase 1 instrument used by Hayes and Read (2004: 97–111) in their own IELTS washback research. • Phase 3 analysis, by two expert judges, of 52 test-related classroom lessons, including IELTS and *PL2000* Impact Study lessons, in the course of the design of a Cambridge ESOL classroom video database (see Chapter 5). • Analyses in general agreement; two versions reconciled and observation analysis model agreed for use in Phase 3 of IELTS and *PL2000* studies.

Figure 4.1 continued

IELTS Study Instrument	Operationalisation	Piloting and Trialling	Validation
	• summary section • inter-observer discussion • teacher post-observation questionnaire: lesson objectives, success, materials; IELTS impact, features; skills activities, likes, dislikes, materials and handbook, activities, teacher background • student post-observation questionnaire: background, IELTS knowledge, lesson feedback. 2) Phase 3 instrument: more economic classroom observation to collect IELTS-related lessons for observation by video; categories: *timing, level, participants, macro-skills, mode, activity* and *comment* as key elements in classroom observation analysis for test washback.		
Interviews and Focus Groups	1) Close construct relationship with SQ, TQ; and receiving institution questionnaire (RIQ below). 2) Face-to-face data collection forms and methods informed by survey, formal/structured interviews, focus group literature (Hawkey 2000, 2001, Weiss 1998: 140–144) 3) *Informal, un- or semi-structured* interviews preferred (Farley Templeton 1996, Gunter 2001, Morrison 1998, Weiss 1998: 140–4).	• Phase 2: main interviewer/focus group facilitator attends British Council IELTS impact study survey sessions as observer	• Sessions at five of the seven impact study UK venues co-observed, at three video-recorded and double reported. • Phase 3 face-to-face data triangulated with relevant questionnaire data.

Figure 4.1 continued

IELTS Study Instrument	Operationalisation	Piloting and Trialling	Validation
	4) Interactions more like discussions; facilitator with list of key relevant topics but individual or group may react to these freely though with probes and follow-up questions to elicit further information and views or to keep interactions on track. 5) Weiss, on key face-to-face data collection points: • relevance and specificity of questions • correct level and clarity of language • adequate definition of terms • use of categorial rather than just yes/no responses where appropriate • avoidance of double negatives and double-barrel items • caution with second-hand information (1998: 140–142). 6) Probably easier to avoid or remedy such problems in less structured interviews or focus groups; problem reactions can be seen as they happen; in structured sessions responses may be too controlled by the format to reveal uncertainties or misunderstandings.		
Receiving Institution Questionnire (RIQ, *see Appendix E*)	1) Draft Phase 1 instruments foresee separate data collection for *IELTS administrators, admissions officers, receiving institution subject teachers.* 2) But early Phase 2 visits note staff often share *IELTS-related* responsibilities: IELTS prep course language teachers also IELTS administrators or admissions officers; some directors of language schools or centres are also teachers.	• Phase 2 trial data from 54 administrators in 21 countries. • Trial data and review of Phase 1 administrators, admissions officers, receiving institution subject teachers' questionnaires used to develop single RIQ for possible use in Phase 3 or later.	• Phase 3 data collected face-to-face and from secondary sources for impact study use and for possible modifications of RIQ for future use.

Figure 4.1 continued

IELTS Study Instrument	Operationalisation	Piloting and Trialling	Validation
	3) Phase 3 team also note valuable receiving institution data collected by Cambridge ESOL outside strict limits of the study. 4) Thus for Phase 3, *one* instrument developed, the Questionnaire for Receiving Institutions (RIQ). But main data collection means at this stage is *face-to-face* IELTS impact comment on IELTS administrative matters. 5) Focus areas in Phase 3: • IELTS general and skill module validity, reliability and bias • *use* of IELTS scores, cut-off band levels • administration and management issues (e.g. dates, venues, examiners, retakes, security, cost). 6) Test information and other data from receiving institution *subject teachers* deferred to a later iteration of the study.		

In fact, as is indicated in Figure 4.2, the original 12 instruments developed in Phase 1, were rationalised to five, for use in Phase 3 of IELTS impact study (2002–2003).

Figure 4.2 Rationalisation of original data collection instruments

Phase 1 draft instruments	
Instruments	**Target participants**
1	for classroom observers pre-, in- and post-observation
2	for teachers, post-classroom observation
3	for students, post-classroom observation
4	for teachers using student textbook materials
5	for students preparing for IELTS
6	for teachers preparing students for IELTS
7	for IELTS administrators
8	for students who have taken IELTS
9	for admissions officers in receiving institutions
10	for receiving institution subject teachers
11	for teachers preparing students for academic study post-IELTS
12	for test takers' background/characteristics

Phase 3 rationalised instruments	
Instruments	**Target participants**
SQ	test-taker students
TQ	IELTS prep teachers
IATM	IELTS prep teachers evaluating textbooks
CRO	classroom observation of IELTS prep class students and teachers
RIQ	receiving institution administrators

A further increasingly prominent consideration at the rationalisation stage of Phase 2 of the IELTS study was the scale of the demands to be made on participants. Though the instruments had, as a result of the trial and rationalisation process, tended to be shortened (see IATM), they were still extensive and demanding since so many dimensions needed to be included, as relevant to the washback and impact research questions.

A revised Student Questionnaire (SQ in Figures 4.1 and 4.2), combined items from three draft questionnaires from Phases 1 and 2 of the impact study: the original questionnaire for students preparing for IELTS (instrument 5 in 4.2), the one for students who have taken IELTS (8) and 12 on test takers' background and characteristics. The new, composite instrument incorporated Language Learning Questionnaire and Test Taker Background Questionnaire constructs and validated items relevant to the IELTS impact study research questions. This Student Questionnaire became the main student data collection

instrument for Phase 3 of the study. It was subject to a significant expert view validation session with Tony Green, himself involved (Green 2003) at the time in his related doctoral research (see above). The option of one student questionnaire instrument rather than three was appealing from a practical point of view, of course. Hawkey's August 2001 paper to UCLES EFL (summarised in Figure 4.1) thus proposed a student questionnaire instrument incorporating relevant coverage of language learner characteristics.

As Figures 4.1. and 4.2 indicate, the final version of the IELTS Teachers' Questionnaire instrument (see Appendix B) differs much less from its pilot version than does the Student Questionnaire.

The validation process for the Instrument for the Analysis of Textbook Materials (IATM) is also detailed in Figure 4.1, the need to rationalise from the extremely long Phase 1 version being a priority. Work on the IATM included rationalising to elicit summary, rather than over-specific, responses, reducing numbers of sub-sections, rationalising most checklists through merging and deletion, but seeking *more* information on the intention, focus and approaches of materials through enhanced open-ended comment sections. As with the student and teacher questionnaires, however, the 'uniformity of item format' suggested by Kunnan (2000) in his Phase 2 instrument validation report, was not attempted. Trialling and review had indicated that it was often appropriate to include open-ended as well as closed items on key issues because the open format was providing valuable elaborations of and triangulation checks on the more quantitative questionnaire data.

The rationalising decision to use a single, simplified *classroom observation instrument* (Q4) rather than the multi-part Phase 1 instrument was made for purpose-specific Phase 3 reasons and after related validation operations (see Figure 4.1). The original instrument proves insightful, well-constructed, comprehensive and very useful for intensive research use as by Hayes and Read (see Figure 4.1 above and Banerjee 1996). However, it would be less suitable for a study such as IELTS impact study Phase 3, conducted mainly at a distance from the participating centres. The purpose of the classroom observations carried out in Phase 3 was, crucially, to triangulate observed data on classroom activity and attitudes with learner and teacher *perceptions* of such activities and attitudes as analysed from the IELTS candidate and IELTS preparation class teacher questionnaires.

The rationalisation of Phase 1 instruments 7, 9 and 10 (for receiving institution administrators, admissions officers and subject teachers, respectively) was on practical grounds (see Figure 4.1) related to the decision to limit Phase 3 of the IELTS study to receiving institution views, collected mainly face-to-face but with the three instruments originating in Phase 1 ready for future use. A wide range of useful comment, spanning both washback and impact matters, had emerged on receiving institution visits made mainly to triangulate questionnaire data (see Figure 4.1 and

Chapter 6). It also became clear on our own early visits to IELTS-related institutions including universities and language schools, how valuable receiving institution data were being collected by Cambridge ESOL outside the strict limits of the study, for example by Lee Knapp, Cambridge ESOL UK Development Manager (Knapp 2003). The breadth and depth of such comment, on washback and impact matters, can be seen in Chapter 6, when the main messages from the face-to-face data are presented. So it was decided that, given the already broad scope of the study, the pursuit of data from receiving institution subject teachers (instrument 10 in Figure 4.2) would be deferred to a later iteration of the study (see Chapter 8).

The final rationalisations of IELTS study instruments for use in Phase 3 of the project were made by the now IELTS impact study Project Co-ordinator, Roger Hawkey, using all the trial data, proposals from a workshop at UCLES led by Antony Kunnan in spring 1999, additional feedback from researchers, including Green, working on related projects, and after consultations with UCLES senior management. It should be remembered that the original IELTS impact study design had always seen IELTS impact research as continuous and iterative, not necessarily involving the use of all instrumentation at once or indeed for the same project.

PL2000 Impact Study instrument development and validation: principles and processes

The *PL2000* Impact Study was, as we have noted, an altogether smaller and more compact study than Phase 3 of the study into IELTS impact. It nevertheless used a range of data collection instruments (see Figure 3.3) as well as collecting and recording face-to-face data through interviews, focus groups and classroom observations. Figure 4.3, which summarises the operationalisation, piloting and trialling and validation of the *PL2000* Impact Study instruments replicates Figure 4.1, which does this for the study of IELTS impact.

The Cambridge ESOL *PL2000* Impact Study did not have the same long research group development and pre-validation phase as the IELTS study. This was because the study was always recognised as more single-focus, national rather than international, and as a study of a particular, time-bound project rather than a high-stakes test with a long history and future. The main principles and prompts for the development and validation of *PL2000* Impact Study instrumentation were thus:

- the relevant literature
- *PL2000* documentation
- the impact study project aims, research questions, target participants, and

action plan as agreed by Cambridge ESOL management, headquarters and local, and endorsed by the Italian Ministry of Education
- case study school pilot visit interviews and observations (January to March 2001)
- regular draft study proposals, circulated among *PL2000* Impact Study project team members, with feedback reports exchanged
- Cambridge ESOL team policy and planning meetings held at Cambridge or in Italy, and liaison and approval meetings with appropriate Italian Ministry of Education officials, including the *PL2000* Director, Dr Raffaele Sanzo
- the use, where appropriate, of impact study approaches and data collection items validated and used in other studies, including the study of IELTS impact.

PL2000 Impact Study: operationalisation

The motivation for all the *PL2000* Impact Study data collection was response to our research questions (see Chapter 3). These, it will be recalled, concerned *PL2000* washback and impact on language teaching and learning: on students, teachers, educational managers and other stakeholders; on textbooks, language evaluation and assessment; on support for teacher in-service and test-awareness programmes and resource centre development.

The study of IELTS impact, which, as we have seen, has been less a one-off, short-duration project than the *PL2000* Impact Study, and more an element in a continuous, formative test consultation and validation programme, had a broad action plan envisaged as comprising three phases (see previous). Each of these phases involved its own action plan (see Chapter 5, for example, for plans guiding the Phase 2 instrument development).

The design and implementation of the *PL2000* Impact Study, on the other hand, were more concentrated and continuous, consisting of:

- a relatively short planning, development and pilot phase, from initial proposal in August 2000 to first case study school visits in October 2001
- a seven-month implementation phase (October 2001 to April 2002)
- a further seven months for data analysis and final report publication (in January 2003).

As with the aims of the study of IELTS impact, *PL2000* Impact Study aims had been agreed and specified in advance, as had the scope, main research questions, and appropriate methodological options (see Chapter 3). The study constructs had, of course, to be operationalised, that is made measurable and incorporated into data collection instrumentation.

Given the narrower scope and scale of the *PL2000* study, the instrumentation was relatively simpler. The data collection instruments

described in Figure 4.3 and beyond, with information on their design and validation were:

- the school profile form
- the student questionnaire (three versions)
- the teacher questionnaire
- the classroom observation analysis form
- the structured interview and focus group forms.

Since the design and timing of the *PL2000* Impact Study precluded the kind of research *group* development used especially in Phase 1 of the IELTS study, the 'literature and related-project review' (see above) was led by the study co-ordinator. He based the design of the attitudinal parts of the questionnaires (in this case, the student and the teacher questionnaires) on the following key principles (see Henerson, Lyons Morris, Fitz-Gibbon 1987):

- determining specific information needed according to agreed *PL2000* features, *PL2000* study aims, research questions and construct operationalisations
- using response formats which combine open and closed responses because of the potential advantage (revealed in IELTS Phase 2 instrument trials) of obtaining measurable data enhanced and triangulated by open comment
- using closed response formats with scale categories or choices that require relatively less *subjective* selections (e.g. *frequently, never* etc. as opposed to *very good, good* etc.)
- becoming familiar with the frame of reference of the participants, i.e. in the case of the *PL2000* study, children, some young, for whom English was a foreign language, and adjusting instrument item language accordingly and/or allowing an explanatory and support role for L1 speakers, including the students' teachers (see Chapters 5 and 7)
- modifying and pre-validating instrument sections and items through brainstorming, team liaison and review meetings, and pilot-visit data
- keeping the instruments concerned as short and straightforward as possible
- agreeing the appearance of questionnaires taking account of factors such as house style and user-friendliness in terms of:
  - appropriate introductory comments re. the purpose of, permissions for the study (see Chapter 5) and anticipated response areas
  - minimised length, completion time, item numbers
  - clear and logical section and item sequencing and demarcation
  - succinct and appropriate wording
  - transparent and convenient response entry, including appropriate space allocation to open-ended items.

PL2000 Impact Study prediction: brainstorming and review

As noted in Chapter 2, the prediction of the impacts of the *PL2000* on the selected stakeholders and activities was used to help ensure relevant and workable data collection means and items. Given that there was no permanent group of researchers working together on the development of the *PL2000* Impact Study instruments, the brainstorming and review processes were less intensive and regular than for the IELTS studies. However, brainstorming and review, in the sense of the discussion of study objectives and information and the exchange of views on priorities and action for their achievement, were attempted through regular Cambridge ESOL team meetings, with agenda and reports, and attended by the UCLES EFL CEO, the Head of the Research and Validation Group, representatives from UCLES EFL Projects, Marketing and Events departments and the co-ordinating consultant. All data collection instrumentation was also discussed with the then UCLES Development Manager, Italy. The *PL2000* Impact Study student, teacher, and school profile instruments were also shared with and discussed with the key contacts at the case study schools.

Figure 4.3 (pages 79–82) summarises main events in the operationalisation, piloting, trialling and validation of impact study instruments for the *PL2000* Impact Study.

The use, performance and findings of the *PL2000* Impact Study questionnaires are detailed in Chapters 5 and 7.

Other *PL2000* Impact Study data

The baseline data on the students in the case study classes at the seven schools were, in addition to their responses to the student questionnaire and their videoed classroom action, supplemented by the use of the *Oxford University Press Quick Placement Test (QPT*) (see Chapter 5), a test validated, before publication, from performances by more than 5,000 students in 20 countries. The test was used in all *PL2000* Impact Study case study classes except those at elementary level, where it was considered that the pupils were too young to participate in providing data for the impact study, except through the observation of their classes and interviews with their teachers and some parents.

The baseline data on the students in the middle and high school case study classes at the seven schools were also supplemented by the use of the Cambridge University Press *Language Learning Questionnaire* (LLQ) (see above). A selection of students completed the instrument in its full computerised format. This instrument sought information, in a limited pilot form, on their approach to learning English, profiled their language learning attitudes and motivations, the time and effort they were willing to give to

Figure 4.3 Summary of main operationalisation, piloting and trialling and post-draft validation exercises for *PL2000 Impact Study* instruments

PL2000 Impact Study Instrument	Operationalisation	Piloting and Trialling	Validation
School profile form (*see Appendix F*)	1) Given the case study sample, instrument designed to summarise the defining features of the seven case-study schools; information entered on school profile forms to serve as context to responses to research questions. 2) *PL2000 Impact Study* research questions indicated form should collect objective background and baseline information on: • numbers of students, classes • teachers, English language teaching staff • *PL2000* classes, hours per week • other foreign languages • English tests used • course books • resource centres.	• Pre-validation of data areas through, Ministry *PL2000* documentation and meetings, discussions. • Data also from Jan to March 2001 recorded pilot visit interviews with *PL2000*-participant school teachers, heads, students and parents, local *PL2000* officials; also *PL2000* related information from Cambridge ESOL Development Manager, Italy and contacts.	• Consultation process confirmed relevance of school profile information areas. • Process also suggested form should include two open-ended, more subjective questions: one on *PL* student use of English outside and beyond school, and one a general-opinion item on the differences the *PL2000* seemed to be making to the schools concerned. • School profile form, almost completely fact-seeking; no statistical validation required. • Single form to be completed one per school; completers invited to query or discuss items, but no problem encountered with completion or presentation of data from form.
Student Questionnaire (SQ, *see Appendix G*)	1) Impact study longitudinal, collecting data early on in and towards end of school year, so *three* versions: one for first-time students, parallel version for second-time end-of-year students, and third for (few) students present only for the second visit. 2) First student questionnaire mainly factual information i.e.: • background identity data • home and school languages • years of English, foreign languages	• Early 2001 pilot visits to schools confirm classroom activities on SQ (e.g. *discussions with whole class, using computers*) as relevant for enquiry by interviews with *PL2000* teachers, students and officials and from observation of *PL2000* classes.	• Activities in accordance with the 'acquisition of a pragmatic-communicative competence' as in Ministry *PL2000* aims (e.g. working *with a partner*). • Validated items from student / teacher questionnaires as used in • Green (2003) (e.g. *listening and taking notes*) • Banerjee (1996) draft IELTS impact study Phase 1 classroom observation instrument (e.g. *students read the questions and guess what listening passage is about*) and

Figure 4.3 continued

***PL2000 Impact Study* Instrument**	**Operationalisation**	**Piloting and Trialling**	**Validation**
	• exam history and intentions • overseas English language experience • English language self-access use • career intentions • self-access English language *use* item invites some quantification (*never, almost never, occasionally, often*) for each activity, e.g. *reading English language newspapers* • one open-ended question, i.e. *'What are your opinions about learning English?'* 3) 2 sections of second-time student questionnaire seek attitudinal data through systematic use of closed-item formats; section on student perceptions on relative frequencies of classroom activities e.g.: • *discussion with a partner; vocabulary exercises,* or *taking practice exams* • five multi-choice items on: *perceptions of target language improvement over the year (2 items), study time outside class, attitudes to external exams, main reason for learning English*. • these areas related directly to research questions (Chapter 3) and key to any evaluation of *PL2000* washback /impact.		• related items in validated IELTS impact study student, teacher and IATM questionnaires (see previous)
Teacher Questionnaire (TQ, *see Appendix H*)	1) Case study school English teachers, key stakeholders in *PL2000* and impact study, especially teachers of students responding to SQ (above)	• Early 2001 pilot visits to schools confirm classroom activities on TQ as relevant for enquiry through interviews with *PL2000* teachers, students and officials, and from observation of *PL2000* classes.	• SQ and TQ shared items facilitate triangulation; even if perceptions differed of what was happening on particular courses, similar overall patterns of activity should also be perceptible across learner and teacher data (see Chapter 7).

Figure 4.3 continued

PL2000 Impact Study Instrument	Operationalisation	Piloting and Trialling	Validation
	2) TQ seeks factual participant background, then scale responses on same list of classroom activity frequency as on SQ. This permits analyses of perceived *PL2000* lesson content and activities, and comparisons between learner and teacher perceptions of same lessons. 3) Two open-ended items allow further comparisons between student and teacher attitudes to *Progetto*; one seeking teacher views on how *PL2000* impacted on their students; second paralleling student questionnaire item on changes over year. 4) Second attitudinal section in TQ explores teachers' views of successful or otherwise achievement of *PL2000* stated objectives.	• All *PL2000* teachers completing TQ also interviewed and consulted on the form.	• The analyses of video-recordings of *PL2000* classroom lessons (see below) would complete the data triangle here.
Classroom Observation Instruments (*see Appendix D*)	1) *PL2000 Impact Study* classroom observation instrument shared objectives and development with similar IELTS study instrument (see 4.1); both seeking accurate analysis of what really happened in classroom for insights into washback from a language education reform project and high-stakes tests. 2) So, for *PL2000* study, use of same instrument to analyse 20 closely observed, video-recorded classroom language lessons in case study schools in terms of episodes, timings, activities, participation, materials and comment; these analyses offered most direct evidence on *Progetto* pedagogy and ethos, collected on two visits seven months apart.	• See 4.1, trial comparative use by Green and Hawkey of COLT and simpler draft alternative classroom observation model. • Decision to use, in both impact studies, version of Hawkey's classroom observation analysis framework (see Appendix D).	• Shared validation with IELTS study; analysis, by two expert judges, of 52 test-related classroom lessons, including from IELTS and *PL2000* studies. • Analyses in general agreement; final observation analysis model agreed (see Appendix D).

Figure 4.3 continued

PL2000 Impact Study Instrument	Operationalisation	Piloting and Trialling	Validation
Interviews and Focus Groups	1) Theory and practice of interviews and focus groups surveyed (Farley Templeton 1996, Gunter 2001, Morrison 1998); account taken of relevant experience in IELTS study. 2) As with IELTS study, key questions are: • generalisation • *social* nature of focus groups • facilitator role and actions. 3) Key focus group and interview data collection validation points noted for moderating/inter-viewing: • working mainly to formalised schedule of topic headings to be addressed by participants • keeping discussion on track but without inhibiting flow of ideas or emergence of unanticipated views • ensuring all group members contribute with no one participant's opinions dominating • avoiding overt support for or criticism of views • resolution of issues or achievement of goals • N.B. real aim of data-collecting interviews and focus groups being to gather as much valid information as possible.	• As for IELTS study: main interviewer/focus group facilitator attends British Council IELTS impact study survey sessions as observer.	• All sessions at all case study schools with observer present, often as interpreter. • All sessions but one video recorded; reports of analyses of recorded sessions verified by observer/interpreter. • Face-to-face data triangulated with relevant questionnaire data.

learning the language, the strategies they used and how they organised their learning (see Chapter 5 for further details of the management of the LLQ in the *PL2000* study).

Additional *PL2000* Impact Study insights on pedagogy, materials, and media were drawn from information provided by some of the case study school teachers writing direct to the study co-ordinator, and gained from UCLES EFL : Bell International *PL2000* teacher essay competition entries in response to the question: 'What the *Progetto Lingue 2000* means to me as a teacher'. These had always been seen, by Dr Peter Hargreaves, then CEO of Cambridge ESOL, as potentially useful triangulation data for the study.

It is possible that the *PL2000* Impact Study, in which a relatively small number of students and teachers were in direct contact with researchers, their video camcorders and microphones, computerised tests and their questionnaires, was vulnerable to the Hawthorne Effect. This is defined by Weiss as a 'change in the outcome measure that is the result of the attention that participants receive during intervention rather than a result of the program under study' (1998:331). The possibility of such an effect cannot, of course, be ruled out. However, participants in the *PL2000* Impact Study, like those in the study of IELTS impact, were not actually subject to a treatment, in the sense of a continuous programme that they were aware other similar participants were not receiving. Rather, they were subject to two or three data collection contacts during an academic year, none of which, with the possible exception of the QPT with its on-screen learner profile, actually left the students with anything their peers outside the impact study did not receive. Given the relatively fleeting involvements of *PL2000* participants (and the even fewer face-to-face contacts of most IELTS impact participants) Hawthorne Effects seem fairly unlikely to have played a part in the impact studies.

There may well have been Hawthorne Effects from the *PL2000* itself, of course, where students at the same schools could be, for a whole school year, attending courses under the project, when their friends were not. In the case of the *Progetto,* the Ministry's aim was, in a sense, to create a Hawthorne Effect of successful language learning and use.

The problem of ascertaining that classroom events and the English language performance of learners are actually evidence of the washback of tests or curricular innovation rather than the result of other factors remains, of course. It is the expectation, however, that research such as the *PL2000* Impact Study, with its inclusion of observational as well as survey data and its triangulation of evidence, will succeed in identifying likely signs of washback and impact (see Chapters 5 and 7). Chapter 5 will explain the approaches taken to try to collect valid data. Chapter 7 will summarise and interpret the main findings, and describe further attempts to come to valid impact conclusions.

Following the enquiry into impact research aims, questions, action plan, research designs, and hypotheses in Chapter 3, this chapter has described the second step in the two studies, of the IELTS and *PL2000,* with a particular focus on the development and validation of data collection instruments. General and instrument-specific points have been made about the main data collection tools of both studies, the validation of questionnaires and similar instruments emerging as just as important as, though in many cases procedurally different from, the validation of language tests.

The next logical step, for Chapter 5, is an investigation, exemplified as ever with reference to our two impact studies, of the project *implementation* steps. This will cover data collection, bases, analyses, and validation.

5 Collecting and managing the data

Chapter 4 has traced the development, validation and production of impact study instrumentation for studies of IELTS and *PL2000* impact, in their own right, and as examples of the processes that may be expected in the study of the language learning and testing aspects of educational impact. Chapter 5 now examines the collection, management and analysis of impact data. In Chapters 6 and 7, the main findings of the two studies will be presented, while Chapter 8 draws conclusions on the study of impact, seeks lessons to be learned, looks forward to future related research and speculates on future developments in models of impact research.

IELTS impact pre-survey

Following the development of pilot data collection instruments in Phase 1 and their validation and rationalisation in Phase 2, steps were taken to ensure that the implementation of Phase 3 of IELTS impact study collected data from an appropriate sample of candidates and preparation course teachers. To this end, it was felt necessary to update the information on IELTS centres. In May 2001, therefore, a *pre-survey* questionnaire was sent to over 300 University, British Council, IDP Education Australia and other test venues worldwide (see Figure 3.2 in Chapter 3 above). The survey, which achieved a high response rate of over 65% from centres in 41 countries, ascertained up-to-date objective baseline data such as the following:

- the language tests for which the centres were running courses
- the annual numbers, durations and dates of such courses
- the numbers and nationalities of students
- teacher strength
- the textbooks and other materials used.

All these data were used to help select IELTS centres for the main data-collecting phase of the study. The numbers foreseen for Phase 3 were a case study sample of around 40 centres, representing IELTS test taker nationality and language populations (see Chapter 6).

Points of interest from the pre-survey data, analysed using straightforward *descriptive* statistics (frequencies, percentages, means), are summarised here.

Thirty per cent of the pre-survey returns were from university centres, 19% from British Council and IDP Education Australia centres, and 51% from other types of language teaching institutions. IELTS was, unsurprisingly given the criteria for the selection of centres for the pre-survey and the objectives of the study, the test for which preparation courses were most frequently run at these centres (83% of the cases). However, courses were also being offered for the Test of English as a Foreign Language (TOEFL) at 34 of the 203 centres offering preparation for tests in addition to IELTS, and for the Test of English for International Communication (TOEIC) at 18. Forty-five per cent of the centres were providing preparation courses for one or more Cambridge ESOL Main Suite exams (see Chapter 3).

A further pre-survey question asked about the length of IELTS preparation courses at the centres concerned. Figure 5.1 summarises the responses, indicating the highest proportion of courses at between nine and 12 weeks in length, but with significant percentages half that length or less.

Figure 5.1 Lengths of IELTS preparation courses at IELTS impact study pre-survey centres

Weeks	Centres
1-2	27
3-5	35
6-8	26
9-12	73
13-16	10
Longer	22

From our contacts with the centres concerned we also gathered that the concept of 'IELTS' courses might need clarifying as the preparation for the test sometimes figured as *part of* other programmes, for example, 'General English' or English for Academic Purposes (EAP) courses. This clarification was made (see Chapter 6) in the questions asked of IELTS preparation students in the questionnaire for Phase 3 of the study itself.

Figure 5.2 indicates that at the pre-survey centres most IELTS preparation classes had relatively small student numbers, 6–10, and 11–15 being the most frequent class sizes reported.

Figure 5.2 Responses from 177 centres on approximate numbers of students on each IELTS preparation course

Number of students	Number of centres
1-5	21
6-10	65
11-15	47
16-20	23
More	21

Final key figures from our pre-survey data, suggesting that the student nationality groupings were fairly representative of the IELTS test taker population, are given in Figure 5.3.

Figure 5.3 Responses from 193 centres on the main nationality groups on their course(s)

Regions	Centres claiming as main nationality groups
East and South East Asia	125
Europe	65
Mid-East	23
Latin America	18
S Asia	10
Africa	3

Centres were, of course, quite likely to name more than one nationality group as 'main' participants on their courses.

The IELTS impact study pre-survey also provided a useful list of IELTS preparation course textbooks in use. This informed the study's investigation, through the Instrument for the Analysis of Textbook Materials (IATM, see Chapter 4), into teachers' choice and opinions of such books.

Above all, the contacts with centres made through the pre-survey helped the study team to select appropriate centres for Phase 3, based on the factual data provided (for the pre-survey was requesting only objective facts from the centres, not opinions or attitudes).

Instrument delivery and data return

There was, at this stage of Phase 3 of the IELTS study, considerable discussion of two important practical points in the management of impact research projects, the decisions on which would affect both our studies. The first discussion point was the *means of delivery* of the data collection instruments, the second the types of *direct contact needed with participants*.

Phase 3 of the IELTS impact study

There was extensive debate within the Phase 3 IELTS impact study team, with Cambridge ESOL senior management, with potential participants and with outside consultancy contacts on how the impact study questionnaires should be delivered, by what means they should be completed and how they should

be returned for analysis. The pre-survey was administered using a straightforward two-page questionnaire, sent by post to the 313 centres concerned, fronted by an explanatory invitation to participate, and pursued, if necessary, with e-mail reminders. 43% of the returns were received by fax, and the response rate, as stated above, was high, at 65%.

Frankfort-Nachmias and Nachmias (1996) refer to a 20–50% response rate as typical of conventional mail surveys. However, there are indications that electronic questionnaires, that is instruments sent and returned by e-mail and with responses entered electronically, can achieve higher response rates. Walsh et al (1992) claimed a 76% response rate with a randomly selected sample and 96% with a self-selected sample, although their survey was about computer networks, a fact that, 14 years ago at least, would appear likely to invite a high computerised response rate. Anderson and Gansneder (1995) report a 68% response to their e-mail survey, with 76% of responses returned via e-mail and 24% conventionally. Mehta and Sivadas (1995) compared a conventional mailed questionnaire with an e-mail equivalent, finding that the unsolicited mail questionnaire achieved a 45% rate of return, compared with the 40% response to its e-mail equivalent. The e-mail response rate increased to 63%, however, if an initial e-mail was sent requesting participation in the study. The approach to initial invitations to participate, reminders and thank you contacts in the studies of IELTS and *PL2000* impact are discussed in the next section of this chapter.

The deliberations on how to deliver IELTS impact study questionnaires and receive the return data raised a range of issues before decision. These included the following potential advantages and disadvantages.

The Internet's 'easy access to world-wide samples', and potential relative 'unobtrusiveness' to respondents are clear advantages (Selwyn and Robson 1998:1). However, the use of e-mail as a research tool may still be somewhat constrained by the bias and limits of its user population in terms of variables such as age, socio-economics or nationality. In 2001/2, of course, the period of IELTS impact study Phase 3, with e-mail use many times more common than it was in 1997, the danger was of *information overload* and that research via e-mail was already running the risk of becoming marginalised as a form of electronic 'junk mail'. In more recent research (Archer and Kendall 2003), questionnaire topic, length and sending institution were found to be the main factors that affected response rates. Furthermore, although 'almost all' schools in the NFER survey of head teachers' views on participating in research 'had internet access and email, the number indicating that they would prefer to complete electronic questionnaires rather than paper-based versions was considerably lower than this' (ibid:1).

However, we were, in the main data collection operation of Phase 3, contacting IELTS candidates or potential candidates *through* the centres

where they were studying. Since not all such centres provided e-mail addresses for their students, the IELTS impact study team was not in as advantageous a position as researchers contacting their sample populations through independent or official institutional e-mail addresses. If this had been the case, the convenience and reduced costs of e-mail contacts would have been a stronger factor in decisions on how to deliver and receive the questionnaires.

E-mail questionnaire designs also offer the advantage of data that are transferable in electronic form to the databases set up for the analyses required. Miller explains:

> Questionnaires may be posted online and submitted electronically. This can significantly reduce the time involved in administering and analysing questionnaires. Answers submitted on a spreadsheet can be entered on to a summary sheet and summary scores computed much more quickly when the spreadsheet is submitted electronically (2002:21).

We see below the actual methods used in the analysis of IELTS impact study data, applying spreadsheet operations but only *after* the data had been entered indirectly from the completed questionnaires.

The decision finally to use hard copy instruments was influenced by the following factors:

- the global coverage of Phase 3 of IELTS impact study using centres with varying IT facilities and equipment
- the comprehensive nature of the data collection instruments involved, with their significant number of open-ended items
- the success of hard copy instrumentation in both the Phase 2 trials and the pre-survey.

Further such studies are nevertheless likely to use electronic instrument delivery and collection modes, especially as the relevant technologies improve and spread so quickly.

The preliminary contact letters or e-mails to 72 centres selected from the pre-survey sample and others contacted since, offered the centre authorities the option of receiving, by post or special delivery, either the number of student, teacher and textbook assessment instruments they wished to use or single copies of the instruments concerned, which they could copy as required. In the event, the study obtained a satisfactory number of questionnaires returned by mail (see Chapter 6 for details of the 572 student, 83 teacher and 45 textbook evaluator participants), in proportions reasonably close to the IELTS test taker and preparation course teacher populations. The instruments were, in general, comprehensively and conscientiously completed.

Instrument delivery and data return

PL2000 Impact Study

The *PL2000* Impact Study, where similar discussions of the instrument delivery and return options took place, used a combination of electronic and hard copy data collection. The main instruments concerned are re-listed here, along with justification for decisions on how they should be administered:

1. School profile pro-forma (see Appendix F): This form, it will be recalled from Chapter 4, was to collect background information on student numbers, teachers, *PL2000* classes, external English language certification, course books, foreign language resource centres, *PL*2000 experience. It seemed clear that this form, one only *per* case study school, should be completed by someone in authority, with access to the information concerned. The data were mainly documentary, to be reproduced in the *PL2000* report as background information rather than analysable data. Hard copy format was the preferred option when the schools were consulted in advance. A single-page form was thus used and seemed to suit the school authorities concerned. The form was completed promptly, and received no negative comment.
2. Student questionnaires (in three parallel versions, see Figure 4.3 in Chapter 4 and Appendix G) were administered through the English language teachers of the case study classes in hard copy format. The completed forms were returned to Cambridge ESOL by hand during Impact Study venue visits, or later by post, especially in cases where class English teachers wished to arrange discussions on the questionnaire with their students (see Chapter 7).
3. *PL2000* Impact Study participant teachers completed the hard copy Questionnaires for Teachers (see Appendix H) during the second round of school visits in April 2002.
4. *PL2000* classroom video-recordings were made by camcorder and directional microphone on the February 2001 pilot visit by Cambridge ESOL Video Unit specialists, who trained the Cambridge ESOL country manager and the Impact Study co-ordinator to make the subsequent recordings on the October 2001 and March 2002 visits.
5. The Oxford University Press Quick Placement Test (QPT), designed for learners of all levels and all ages, was selected for use with the case study school students near the beginning of their courses as it would give an immediate broad assessment of their English language proficiency in the case study classes at the seven schools. The *computer-based adaptive* version of the test was selected. The computer estimates the level of the test taker on the basis of his/her responses and selects from its bank items of the appropriate difficulty for that level. Because of this, each response

contributes a maximum amount of information and the test can thus be shorter than equivalent paper and pen tests, without sacrificing reliability. QPT results are generated immediately. On the *PL2000* study, the test was sent to and administered by the case study schools on a networked system, following detailed instructions from Cambridge ESOL Projects Office.

6. The UCLES Questionnaire (LLQ, see Chapter 4) was administered in its Italian language version through CD-ROM delivered to selected case study schools with full instructions for installation on the school computer laboratory systems for student self-completion. Participants keyed in or ticked electronically information about their attitudes and approaches to learning English, and were able to view an immediate on-screen profile of their approach to learning a foreign language in terms of:
   - attitudes, concerns, motivations regarding English language
   - willingness to commit time and effort to learning the language
   - language learning strategies.

   The LLQ was intended to take no more than 40 minutes to complete and had an on-screen help option. Further details of arrangements and permissions for LLQ use appear in the next section.
7. As mentioned in Chapter 4, additional *PL2000* impact data were drawn from competitor submissions to the UCLES EFL : Bell International *PL2000* teacher essay competition, arranged at the same time as the *PL2000* Impact Study.

The *PL2000* Impact Study combined hard copy and electronic instrument delivery and data return. Decisions on which to use were made, as with the IELTS study, on practical grounds, after discussion by the project team in consultation with participating centres.

Permissions, confidentiality and accountability

Like all researchers seeking information from people, impact evaluators need to be aware throughout the research process of the *ethical* implications of their study for the participants. Tindall identifies the following 'closely intertwined ethical issues' (2001:153): *informed consent, protection of participants, confidentiality and anonymity,* and *accountability*. Weiss (1998:95) adds *reciprocity* as a further ethical issue.

Informed consent from participants presupposes, of course, that they are fully aware of the purposes and approaches of the research concerned and that their initial consent to participate includes the right to withdraw at any time, even retrospectively. Participant protection is achieved mainly through the confidentiality and anonymity of participant information.

IELTS impact study

The student questionnaire used in Phase 3 of the study of IELTS impact attempted to obtain the informed consent of test candidates by:

- telling them of the research purpose and accountability (in the sense of reference to the sponsors to whom the researchers are answerable, and the results they are expected to achieve)
- defining the context, nature and potential participants in the study
- assuring participants of the confidentiality of their responses
- thanking them for their time and co-operation
- inviting their consent to participate under conditions of:
  - anonymity
  - right to refuse or withdraw participation, and
  - guarantee of restriction to the research team of access to information given.

The full IELTS study Student Questionnaire introduction and permission slip was as follows (see also Appendix A):

> Dear Participant,
>
> As part of the continuing programme to update and refine its International English Language Testing System (IELTS), the University of Cambridge Local Examinations Syndicate (UCLES) is conducting a study of the impact of the test. We are contacting students, candidates, English teachers and university administrators for information and comment.
>
> Your responses to this questionnaire will be treated in confidence, and only used for the stated purposes of the study.
>
> Thank you very much for your time and co-operation. We should also be grateful if you would complete and sign the consent note below.
>
> Yours sincerely
> Nick Saville
>
> EFL Director
> Research and Validation
> University of Cambridge Local Examinations Syndicate
> English as a Foreign Language
> Cambridge
> England

Your consent to participate in the UCLES IELTS impact study

I understand that:

- the purpose of the study is to collect and analyse information from those familiar with the IELTS
- my name will not appear in any project publication
- the information I give, but not my name, may be quoted
- I am free to refuse to participate in the study and may withdraw at any time
- my completed questionnaire is for the study team only; it will not be shown to anyone not connected with the UCLES study.

Signature: Date:

The teacher, textbook evaluation and administrator questionnaire permissions took a similar form.

PL2000 Impact Study

The letter below to case study school heads from the *PL2000* Impact Study co-ordinator characterises the question of participant permissions for the study, which, unlike the study for IELTS, involved personal contact with the study team at all venues. It noted that the careful, though not necessarily formalised, approach to permissions (to enter schools, administer questionnaires etc.) would be maintained, with the Cambridge ESOL Development Manager to be responsible first-instance. The fact that the Impact Study was approved by the *PL2000* Co-ordinator at the Ministry of Education in Italy meant that general principles of consent, confidentiality and accountability had already been established. The following mid-project letter from the Cambridge ESOL *PL2000* Impact Study co-ordinator to the head of one of the case study schools exemplifies the way in which key project contacts were maintained, arrangements pursued, purposes confirmed, and permissions sought.

18 March 02

Dear Prof xxxxxxx

As co-ordinator of the Cambridge *Progetto Lingue 2000* Impact Study, I should like to request your approval for the follow-up visit to your school by the Cambridge ESOL Development Manager for Italy and myself, as discussed when we were at xxxxxx School last October. May I suggest Tuesday April 15, 2002 for our visit. We should be in [two case study school towns] for one or two days and intend, if the date is convenient, to be working in [town 1] on Monday April 15, and at the xxxxx school on Tuesday morning.

We should like to achieve the following objectives during the time we are with you:

1) the video-recording of :

- an interview with yourself on latest developments re the *PL2000*
- a lesson in a *PL2000* English class which would suitably illustrate some of the effects of the *Progetto*
- interviews before or after the lesson with teacher(s) and, perhaps, some of the students concerned
- a meeting of some of your English and other foreign language teachers.

We would welcome your assistance in ensuring that permissions to video and talk to students and staff have been obtained.

2) discussions with you and appropriate English language staff of the use at your school of the *Cambridge Language Learners Questionnaire (LLQ)* and the *Oxford Quick Placement Test (QPT).*

We would, of course, make every effort to minimise disruption during the visit.

I do hope the proposed visit on 15 April is acceptable to you. Your agreement represents a further step in the excellent support you have already given to the *PL2000* and the Cambridge Impact Study. The classroom video, interview and questionnaire data we collected at xxxx school last time have proved very useful indeed for our Study.

I trust that the academic year has been going well for your school and you, and look forward to hearing from you as soon as possible on our planned visit so that we can take the necessary action on arrangements.

Yours sincerely
Dr Roger Hawkey
Co-ordinator, *PL2000* Impact Study

When an independently produced data collection instrument is being used, as in the case, for example, of the LLQ in the *PL2000* Study, agreements would be signed covering both providers and users. Letters to *PL2000* Impact Study schools in January 2002, for example, included:

- a statement that the agreement covered terms and conditions of use of the LLQ computer-based software for the purposes of the Cambridge *PL2000* Impact Study in Italy
- Cambridge ESOL agreement to provide one set of LLQ materials to each of the Italian centres identified in the project (including LLQ CD-ROMS and Italian installation manuals)
- agreement that the school, as one of the schools participating in the Impact Study project, would install the LLQ software on its own equipment
- a statement that the licensed LLQ software and documentation contain confidential and proprietary information of the licensor and that all copyright trade marks and other intellectual property rights in the licensed software and the documentation are the exclusive property of Cambridge ESOL
- agreement that material would be used for the *PL2000* project alone and that it would remain the property of UCLES EFL throughout
- agreement that the LLQ would be completed by students in the selected classes in the *PL2000* Impact Study by the middle of March 2002, after which all material would be returned to Cambridge ESOL
- agreement that users of the software would not copy, modify or merge the whole or any part of the licensed software or documentation; nor use them on behalf of or make them available to any third party.

In cases such as this, permissions and confidentiality are clearly a two-way process. There is little doubt that issues of informed consent, participant protection, confidentiality, anonymity, and accountability in research will continue to grow in significance and require increased formalisation. In their template for agreements to permit the use of examination data for academic research purposes, for example, Cambridge ESOL Examinations seeks signature to a statement of the following:

- purpose, use, request context of data
- detailed description of data requested
- specification and limitations on use of the data provided
- permanent ownership of data
- date of return of data to owner
- assurance of safe-keeping of data on loan
- guarantee of restriction of use of data by signing party only unless with prior permission of owner

- agreement by temporary user of the data to offer to owner for prior comment all papers or forms of report using data, with owner right to refuse permission to publish
- data user undertaking to maintain anonymity and confidentiality of data participants unless agreed otherwise in writing with Cambridge ESOL.

Reciprocity

Both the IELTS and *PL2000* studies made significant demands on people's time and information. With the Italian project, the case study schools tended to be establishments where Cambridge ESOL had already made beneficial contacts, through, for example, teacher-support programmes. With Phase 3 of the IELTS impact study, however, carried out mainly at a greater distance, it was felt that the collaboration of participating centres should be *reciprocated* in some way. It was thus agreed that the letters inviting centres to participate would reiterate the data provision tasks they were being invited to perform, then offer the project sponsor's tokens of gratitude in the following words:

> We know that this is asking rather a lot. However, we intend to support your participation in the IELTS impact study as much as we can with any advice needed, by providing all the questionnaires you require and by sending you the results of the analysis of the data you submit. We are also offering a book award to every teacher completing the textbook evaluation questionnaire.

The promised data report was individualised where appropriate for each centre and sent by e-mail. The book award was in the form of an international book token sent by post. The IELTS impact study team was pleased to receive acknowledgements from centres for both tokens of gratitude.

Data analysis and management

IELTS

The means of analysis used as appropriate to the validation of Impact Study data collection instruments have been discussed in some detail in Chapter 4. The data *management and analysis system* for Phase 3 of the IELTS study was designed by Richard Turner of the Cambridge ESOL Research and Validation Group, in consultation with the co-ordinator of the study and author of this book, and with assistance from Research and Validation Group staff. The system covered the analysis of the data from the questionnaires, closed and open-ended items, the interviews and focus groups, and the classroom observations.

Since Phase 3 of the study of IELTS impact was a fairly long-term project, with data-collection through the questionnaires and face-to-face modes spanning several months, it was important to have prepared a data management system which would permit the entry and monitoring of data to identify early trends and possible problem areas. The *Microsoft Access 97 Database for Windows* was used for this purpose, with its system of:

- *tables,* for data storage
- *queries,* a function permitting the user to extract and draw together specific parts of the stored data (e.g. the number of female IELTS candidates with Cantonese as their first language, the highest frequency first language, the relationship between candidates' expected and required IELTS band scores)
- *forms*, for the viewing, editing, inputting and control of data.

To aid data evaluation and reporting, a link was developed between *Microsoft Access* and the *Microsoft Excel* spreadsheet-based programme with its strong chart-creation facilities. The link enabled a query posed in *Access* on the most recent data held in the database, to be pulled by *Excel* into a spreadsheet table for chart display. Once set up, the whole data action chain could be updated with a touch of the *Excel* 'refresh' button.

It was also possible to link the more powerful quantitative research package SPSS (Statistical Package for Social Sciences) once the limits of *Access* were reached. The SPSS package had already been used in Phase 2 of the IELTS study (see Chapter 4), although the data analysis operations remain, as may be seen from the data presentations in Chapter 6, mainly at the *descriptive* statistical level. It will be recalled from Chapter 2 that the IELTS study was categorised as a mainly descriptive study towards the qualitative side of the quantitative/qualitative continuum, the *inferential statistics* being reserved for the validation of some data collection instruments (see Chapter 4).

Open-ended data from the IELTS impact study questionnaires were also entered into the *Access* system and submitted to keyword analyses. Semi-structured interviews and focus groups were summarised through manual note-taking, in some cases double-checked across versions taken by two recorders.

To store and enable further analyses of the IELTS and *PL2000* video-recordings, a dynamic *Access*-platform video-data management system was designed by Richard Turner, with the assistance of Roger Hawkey and the Cambridge ESOL Video Unit (see Hawkey, Thompson and Turner forthcoming). This system permits the electronic retrieval of classroom video clips, which have been analysed according to 11 fields: *project, date, location, data source, time code, participation, level, mode, activities, quality* and *comment*. The now completed, but expandable, Impact Study research video

database covers 55 lessons, from the IELTS and *PL2000* studies and the Florence Language Learning Gain Project (see Chapter 8). Twenty videoed interviews and focus groups from the IELTS and *PL2000* studies are also in the database.

PL2000

The *PL2000* Impact Study was, as has already been indicated, smaller in scale and handling less complex data. The school profile form, student and teacher questionnaires and structured interviews and focus group data were analysed manually and summarised using *Microsoft Excel* data analysis and presentation operations (see Chapter 7). *PL2000* study videoed semi-structured interview and focus group data were summarised in note form, cross-checked with observer (or interpreter) and cited in the *PL2000* Impact Study report (see Chapter 7) with appropriate attribution.

This chapter has detailed some of the key features of IELTS and *PL2000* Impact Study data collection, management and analysis approaches. The two studies demonstrate both hard-copy and electronic data collection, manual and computer-programmed data analysis and management systems. The importance of introductory, follow-up, permission-seeking or granting and acknowledging correspondence with participants has also been recognised. Chapters 6 and 7 now present some of the main findings of the IELTS and *PL2000* studies respectively.

6 The IELTS impact study: main messages

In Chapters 6 and 7 we summarise some of the main findings of our two studies. This is intended both to present key messages for stakeholders and to illustrate how these messages emerged from the design, development, implementation and analysis stages of the two studies as described in Chapters 1–5. This chapter will focus on Phase 3 of the study of IELTS impact, Chapter 7 on the *PL2000* Impact Study.

Phase 3 of the IELTS study aimed, it will be recalled from Chapters 2 and 3, to respond to a series of washback and impact research questions seeking data on:

- profiles of IELTS test candidates and teachers preparing candidates to take the IELTS test
- the washback of the test on courses preparing candidates to take it
- the impact of IELTS on participants who have taken the test
- the impact of the IELTS test on institutions receiving IELTS candidates.

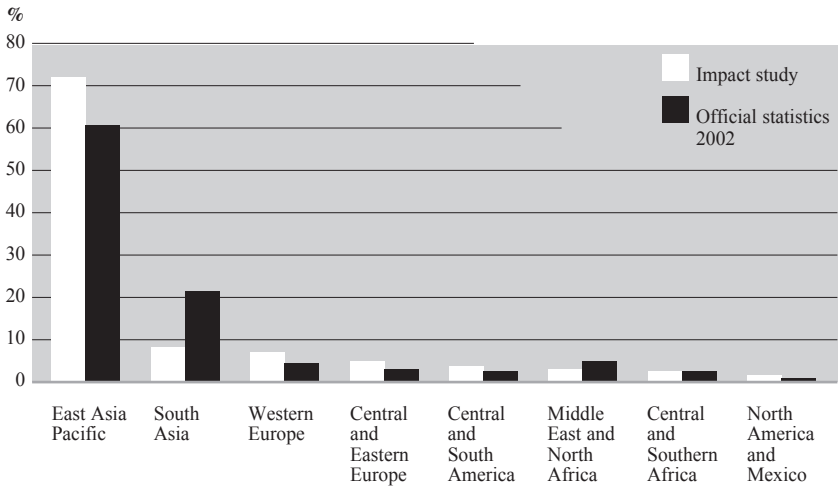
The chapter will first examine evidence from the study towards answers to the research questions on candidate participant profiles, then on the IELTS preparation courses and their teachers. Evidence will next be presented on participant perceptions of IELTS test fairness, likes and dislikes, pressures and motivation, test module difficulty and validity. Finally, mainly through an analysis of the study's face-to-face data, questions of receiving institution use of IELTS band scores and issues of test administration will be considered.

Impact study candidate profile

In pursuit of profiles of pre- and post-test IELTS candidates, a case study sample of IELTS centres representative of the IELTS nationality population was selected using data from the pre-survey (see Chapter 5). IELTS candidates and teachers at these centres were invited, through the individuals already nominated as key contacts on the basis of pre-survey responses, to respond using the appropriate data collection instruments (also see Chapter 5). The chart in Figure 6.1 summarises the IELTS impact study Phase 3 candidate population in terms of regional background compared with the full 2002 IELTS candidature, indicating a reasonable match between the two.

Of course, this candidature is changing constantly. The 2003 IELTS *Annual Review* notes, for example, a rise of 34%, from 355,000 to 475,000 candidates that year, with 'particularly impressive' growth 'in South Asia and the Middle East' (2003:1). As established in Chapter 1, a 'high-stakes, global English language test, accepted as fulfilling English language requirements for people whose first language is not English and who need to study, work or live where English is used as the language of communication' (www.ielts.org home page), the IELTS test is naturally subject to changes in policies and procedures both in the home and destination countries of candidates. There is therefore never likely to be a *perfect* sample of the test's candidate population.

Figure 6.1 IELTS impact study and IELTS test candidate regional backgrounds



The closer profile of the candidates participating in IELTS impact study Phase 3 is summarised in Figure 6.2. in terms of participant gender, age, English language background, academic level and IELTS test status. The figures are based on the full sample of 572 participants, percentages, however, calculated without missing entries, unless stated. Further baseline information on the test takers, including LLQ data, is available in the main report with the Cambridge ESOL Research and Validation Group.

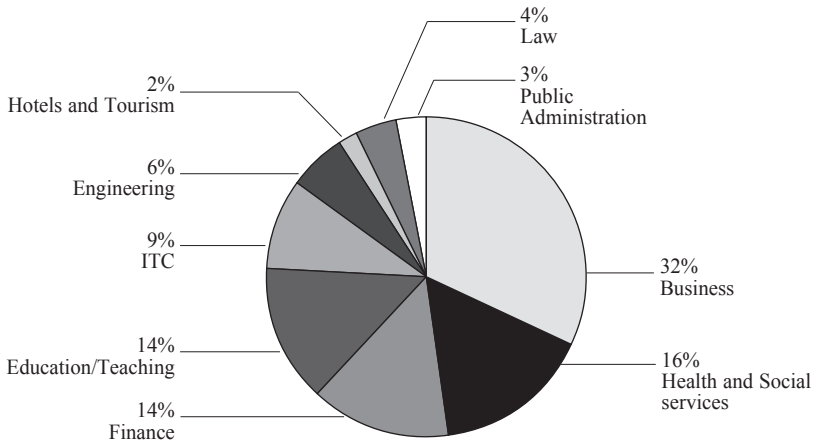
Figure 6.2 Profile of candidate participants (n=572)

Gender:	**No.**	**% (rounded)**
female	305	55
male	252	45
Age:		**% of stated**
15 – 20	181	34.7
21 – 25	192	36.8
26 – 30	84	16.1
31 – 35	33	6.3
36 – 40	23	4.4
41 – 45	6	1.2
46 – 50	2	0.4
Not stated	51	
First languages	32	
Countries of origin	34	
Years of English:		
kindergarten to college/university	51	9
primary school to college/university	90	16
secondary school to college/university	210	38
kindergarten to secondary	17	3
kindergarten only	3	1
primary only	28	5
primary + secondary only	28	5
secondary only	89	16
college/ university	26	5
outside classes only	8	1
above levels *with* extra language classes?	310	56
Educational levels*		
pre-university	7	6
undergraduate	56	46
post-graduate	59	48
IELTS test status		
Pre-IELTS	368	64
Post-IELTS	204	36
IELTS Module taken*		
academic	165	89
general training	20	11

** Only **post**-IELTS candidates asked*

The fields of study of the impact study IELTS participant candidates are summarised in Figure 6.3.

Figure 6.3 Participant candidate fields of study



Recent evidence from the Cambridge ESOL IELTS Candidate Information Sheets (CIS) suggests that this distribution of fields is fairly typical of current IELTS entries.

The profiling of the candidates responding to the Phase 3 Student Questionnaires serves two main purposes. It is necessary, first of all, as a data baseline against which to evaluate our later analyses of the impacts of the IELTS test on these candidates, and, secondly, it provides further evidence that the participants concerned are not untypical of the IELTS candidate population in general. When we begin to discuss *how* candidates are affected by the test, we shall be aware who the candidates are who are thus affected.

IELTS washback on preparation courses

The data analyses in this section are in response to the research question: What is the washback of the IELTS test on courses preparing candidates to take it?

The courses

96% of the IELTS impact study Phase 3 candidate participants were attending or had attended a preparation course for the test. According to the responses to the questionnaires completed by the 83 preparation course

teachers participating in the study, most of whom, of course, were teaching the participating IELTS candidates or their colleagues, the main *destination countries* for academic study or other entry purposes were the UK (36%), Australia (28%), Canada (12%), New Zealand (12%) and the USA (5%).

When asked what kind of IELTS preparation classes they were attending or had attended, the participating candidates responded as in the table in Figure 6.4.

Figure 6.4 Student experience of IELTS preparation course types (%)

Course categories	% of responses
Course with 'IELTS' in its title	40%
EAP/English language study skills course	42%
General English course with IELTS preparation elements	17%

These course categories (by no means mutually exclusive, naturally) are of particular interest in the context of an investigation of test washback and impact. A key issue is the extent to which students are participating in a course to prepare for a test or participating to improve their target language proficiency. The teachers in the study responded to a parallel item in their questionnaire, on the kinds of IELTS preparation classes they were currently teaching. According to the table in Figure 6.5 they, too, were involved with a similar variety of IELTS-related course types.

Figure 6.5 Teachers' preparation course types (%)

Course categories	% of responses
Course with 'IELTS' in its title	53%
EAP/English language study skills course	35%
General English course with IELTS preparation elements	12%

The table in Figure 6.6, using Teacher Questionnaire response data, indicates the class sizes in IELTS preparation courses, groups of 11–15, 16–20 and 6–10, in descending order, apparently the most common.

The figures from the impact study centres differ somewhat from the range and frequencies of the class sizes reported in the pre-survey (see Chapter 5). There, it may be recalled, class sizes of 6–10 were more common than 11–15. However, the three most common class sizes in both the pre-survey and the study are 6–10, 11–15, and 16–20, the three sizes accounting for 76% and 81% of the class sizes respectively.

Figure 6.6 Average numbers of students in IELTS-related classes

	Number	%
No answer	2	
1–5	4	5
6–10	15	19
11–15	34	42
16–20	16	20
More	12	14
Total (*without no answer*)	81	100

The IELTS preparation teachers

In response to research question 5: What are the profiles of the teachers preparing candidates to take the IELTS test? the objective data from the 83 IELTS preparation course teachers completing the teacher questionnaire give key baseline information (see Figure 6.7, page 105).

The teacher profile here rather suggests that IELTS preparation course teaching at the centres concerned was normally in qualified, though not always IELTS-experienced, hands. Like the candidate profile preceding it, the teacher profiling is useful as baseline data. It will be important as we report findings on IELTS impact on the teachers below, for example on their views of and attitudes to the test, to be aware of the fact that the reactions are coming from an apparently well-qualified group. The baseline data will also be informative, as IELTS impact studies progress through Phase 3 and beyond, on how typical the participants are of IELTS preparation course teachers in general.

Figure 6.7 Profile of IELTS preparation teachers

Gender:	**No.**	**%**
Female	48	58
Male	35	42
Age:		
Under 30	10	12
31–40	30	36
41–50	19	23
51–60	20	24
61+	4	5
Years teaching	**No.**	**%**
No answer	5	
1–5	11	14
6–10	21	27
11–15	14	18
16–20	5	6
20+	27	35
Total excluding no answer	78	100
Qualifications (more than one possible)	**No.**	**%**
No answer	7	
Teacher training college	1	1
Post-grad certificate	43	27
Diploma	40	25
Bachelor degree	36	23
Master degree	33	21
Doctorate/PhD	4	3
Total (excluding no answer)	157	100
Position at institution	**No.**	**%**
Teacher	64	80
Director	3	4
Manager	4	5
Administrator	2	3
Professor	2	3
Other	5	6
No answer	3	
Total (excluding no answer)	80	100
Trained as an IELTS examiner:		
Yes		52
No		48
Received training in preparation of students for IELTS:		
Yes		32
No		68

Teacher perceptions of the influence of IELTS on preparation course content, methods

Pursuing answers to the research question What is the washback of the IELTS test on courses preparing candidates to take it? we find, according to teacher questionnaire responses, that 90% of the participant teachers agreed that the test influences the content of their lessons, 63% that it also influences their methodology. (The two questions were: Does the IELTS test influence your choice of the content of your IELTS preparation lessons (i.e. what you teach), and Does the IELTS test influence your choice of methodology (i.e the way you teach) for IELTS preparation lessons?

Both these percentages are high, the second especially so given one of Alderson and Wall's findings in their 1993 study of washback from a revised O Level English exam in Sri Lanka. This was that 'the examination had had considerable impact on the content of English lessons and on the way English teachers designed their classroom tests (some of this was positive and some negative), but it had little or no impact on the methodology they used in the classroom …' (1993:348). Note that Alderson retains this view, presenting it now almost as a given: 'We now know, for instance, that tests will have more impact on the content of teaching and the materials that are used than they will on the teacher's methodology' (2004a:1). In the same piece, Alderson relates this phenomenon of impact (or washback) with 'the teacher factor in washback', adding that 'it is at least as much the teacher who brings about washback, be it positive or negative, as it is the test' (ibid). Other factors matter too, of course, as we inferred from the discussion of the *complexity* of washback and impact in Chapter 1. Note also how this focus on the teacher factor further emphasises the need for teacher baseline data in test impact studies.

Liying Cheng, reporting a study on the (intended) washback of a Hong Kong secondary school exam on teaching and learning, concludes that teachers' 'perceptions of the underlying teaching methodology associated with the new [Hong Kong Certificate Examinations in English] remained relatively unchanged' (2004:163).

Judging from teacher responses to the follow-up open-ended items in their questionnaire on IELTS washback on preparation courses, the distinction between course content and methodology is not always clearcut. Whereas there was only a single no-response to the question '… please note here how the [IELTS] test influences your decisions on lesson content (i.e what you teach)', 29 of the teachers gave no response to the parallel item on how the IELTS tests influences teacher choice of methodology for IELTS preparation lessons. The implication is that they were not easily distinguishing between content and methods. Three of the 79 teachers

responding actually wrote 'see 3.1', that is referring the question on methods to their response on content.

Both the perceived overlap between course content and methods and an apparently strong focus on the IELTS test itself are clear from the teacher responses. Fifty-five of 82 responses to the question on course content (67%) referred specifically to teaching for IELTS test elements, including:

- three references to 'teaching for the test'
- five to *focusing* on the test
- five to content *relevant* to the test
- three to content *related* (exactly, directly) to the test
- seven to content *dictated by, concentrating on, oriented to, influenced by, consistent with* the test
- four to work *specifically* for the test
- eight responses which referred to teaching with time pressures or time management in mind, reflecting that aspect of the IELTS test
- four stating that a test focus was *insisted on by the students*
- one claiming a 100% focus on the test, one to the test being *'all we do'*, one to it being 90%
- five mentioning *test practice*
- nine referring to test techniques/strategies
- four to predicting test items
- three to old/sample papers.

These responses in the context of questions on test *content* compare with easily the highest number of responses (23) referring to teaching IELTS formats, techniques, skills in response to the *methodology* question, and seven references to administering practice tests. Three teachers also referred, responding to the methodology question, to *student demands* for a close test focus on their preparation courses.

The message from these Teacher Questionnaire items is that there appears to be IELTS washback on the preparation courses in terms of content *and* methodology, a conclusion apparently differing from that drawn by Alderson and Hamp-Lyons (1996), Watanabe (1996) and also by Green (2003), whose observations of IELTS preparation and EAP classrooms indicate that 'course content was very clearly influenced by the test, but any influence on teaching and learning methods was less obvious and was mediated by participant beliefs' (2003:45). Green also reminds us, however, of the need to triangulate 'potentially misleading interview/questionnaire data'. 'Wall and Alderson' (1993:65), Green notes, 'could find no evidence of observable changes in instructional methodology, despite the teachers' assertions that the new test affected how they taught.' IELTS study Phase 3 classroom observation data are cited in the next section, suggesting again the importance of the teacher

factor and implying, like Green, a typical eclectic methodological approach: 'For the individual teachers, methodology did not emerge as a core issue, perhaps varying little across the two course types. Teachers did not tend to construe class activities in terms of methodologies…' (op. cit.:128–29). As the classroom observation data indicate, however, the eclectic methods tended to be influenced by the Communicative Approach to Language Teaching (CALT) (see Chapter 7).

In case the IELTS study Phase 3 teacher quotes above might suggest such relationships are too straightforward when we have noted above how complex they actually are, it should be noted that the teacher responses to the course content and methodology questions *also* featured the following views and practices:

- Six of the 83 teachers (7%) responding to the two items indicated that the test did *not* influence the choice of content of their IELTS preparation lessons.
- 27 (32%) of the teachers claimed that IELTS did not affect their methodology on such courses.
- Five responses emphasised the teaching of *general* English or improving students' general proficiency, two improving their study skills, two encouraging wide reading.
- Three responses emphasised the teachers' student-centred view, trying to challenge learners and to build their confidence.
- There were 10 references to a focus on communicative *micro-skills* (e.g. prediction, reading and listening for (key) information, recognising text types, organising text, expressing ideas).
- Nine teachers claimed an emphasis on grammar/structure, five on vocabulary.
- Five of the responses each referred to the courses being more structured because they were preparing for IELTS, and to their being task-based, three to a focus on topics.
- There were two references to pair work, two to group work.
- Five teachers described their approach as 'structured', three as 'intensive',' two as 'focused', one as 'serious' (without specific reference to IELTS as a cause of these features).
- Two teachers said there was more teacher talk and input than on other (non-IELTS) courses.
- Teachers referred once each to their IELTS prep courses as being less 'entertaining', 'dynamic', 'personal', 'reflective', and having less 'discovery', 'interaction', 'flexibility'.

The Phase 3 Teacher Questionnaire asked more *targeted* questions relating to lesson content and methodology in addition to the two general items on IELTS

preparation course content and methodology (see above). Figure 6.8 in the next section shows prominent IELTS-preparation course activities identified by teachers and students. These are taken from an inventory of 28 such activities developed and validated during Phases 1 and 2 for the original classroom observation analysis instrument (Banerjee 1996) and the student and teacher questionnaires (Horak 1996, Winetroube 1997). All four macro-skills are covered and the activities specified have a methodological tone (e.g. *listening and taking part in seminar/workshop activities; interpreting statistics/graphs/diagrams; short report writing; group discussions/debates*). These are similar to methodology-oriented questions to the teachers in Cheng's Hong Kong study (1998:170), for example*: explain the meaning of a text; organise group work or discussion.* Green refers helpfully to categories such as *'participant organisation, content control* or *student modality'* in describing the 'broad similarities in teaching methods across courses' (2003a: 339) revealed by his observation of IELTS and EAP preparation, his categories surely supporting the potential overlap between content and methodology. Cyril Weir (personal communication) wonders whether, by their very academic reading and writing emphasis, both course types are likely to be target domain and, thus, content influenced.

The richness of the Phase 3 data on course content and methodology so far certainly justifies the inclusion of open-ended questions following the Yes/No/Don't know items. As anticipated in Chapter 4, the open items served both a triangulation and a data-enhancing function.

Summary of preparation course contents, skills, activities

We now move from teachers' reactions to questions designed to encourage them to evaluate the influence of the IELTS test on their preparation courses, to attempts to discover what students and teachers perceived as the *actual content* of such courses. The table in Figure 6.8 (page 110) summarises activities selected by high percentages of both the teachers and the students as prominent in their IELTS preparation courses. The activities concerned have, it will be noted in connection with discussion in the previous section, both content and methodological implications.

The influences of the IELTS test may still be seen in these activities, for example in their relationships with the four modules (Reading, Listening, Writing, Speaking). The prominent items selected may also appear quite likely to make for interesting classroom lessons, including the use of a rather appropriate and lively set of communicative activities, relevant to students' needs.

As for the sources of IELTS preparation lesson content, a Teacher Questionnaire item asked what materials, other than a main textbook, the teachers used. The answer was additional textbooks (in 19 cases of the 62

teachers specifying materials in response to this question, 10 of these identifying books offering extra test hints or practice, the rest textbooks (20) targeted at specific language skills or components). However, supplementary materials from the press (11), TV and radio (7), video and audio (11) and the Internet (5) were also mentioned by the teachers, as were centre in-house or teachers' own materials (11), mostly not further specified.

Figure 6.8 Candidate and teacher perceptions of prominent IELTS-preparation course activities

Activities	Students %	Teachers %
Reading questions and predicting listening text and answer types	89	86
Listening to live or recorded talks and note-taking	83	63
Analysing text structure and organisation	74	90
Interpreting and describing statistics/graphs/diagrams	74	90
Learning quick and efficient ways of reading texts	73	93
Reading quickly to get main idea of text	77	96
Learning how to organise essays	82	99
Practising making a point and providing supporting examples	78	88
Group discussion/debates	83	76
Practising using words to organise a speech	74	83

There are related data of interest from the Instrument for the Analysis of Textbook Materials (IATM, see Chapter 4), from items seeking answers to Research Question 7: Which textbook and other materials are used on IELTS preparation courses? and Research Question 8: What is the washback of IELTS on these preparation course materials? Teacher textbook evaluators completing the instrument on IELTS-related textbooks, most of which (66%) were language teaching and test-preparation books combined, found that they covered micro-skills such as those listed, in rank order of selections by the 43 teacher-evaluators, in Figure 6.9.

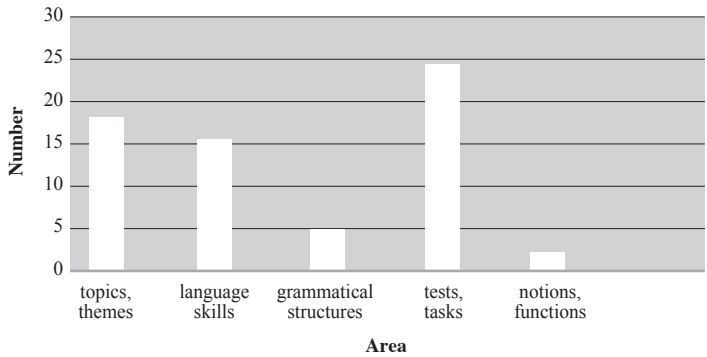
As the chart in Figure 6.10 indicates, too, there is a post-structural, communicative ethos to the organisational framework of most of the books analysed by the 43 teachers concerned.

IELTS coursebook activities considered by the teacher evaluators to encourage communicative opportunity were, in rank order: *essays, group work, pair work, reports, reviews, reading, listening, viewing for personal interest,* and *role plays*.

Figure 6.9 Perceived micro-skills coverage in IELTS preparation books

Micro-skills	No. of selections (N=43)
identifying main points	40
identifying overall meaning	38
predicting information	36
retrieving and stating factual information	34
planning and organising information	34
distinguishing fact from opinion	31
drawing conclusions	30
making inferences	29
evaluating evidence	27
identifying attitudes	23
understanding, conveying meaning through stress and intonation	15
recognising roles	12

Figure 6.10 Teachers' views of organisational units of their IELTS preparation coursebooks



As we have seen in Chapters 2, 3 and 4, it is part of the IELTS claim to theory-based validity that it reflects the language macro- and micro-skills needed by candidates to carry out their target communicative activities. The student and teacher data already presented in this chapter indicate that the preparation courses and their purpose-designed textbook materials are perceived as both influenced by IELTS and characterised by an emphasis on communicative activities and micro-skills. If this is definitely the case, it could again be evidence of Green's washback : construct overlap model (Figure 1.3 in

Chapter 1), where the target language domain and the language teaching, learning and use coincide, at least partly due to washback from the constructs of the test.

A useful check on whether the IELTS preparation courses really are as the students and teachers describe them in their questionnaire responses could be data from IELTS preparation lessons recorded and analysed using classroom observation instrumentation (see Chapter 4 and Appendix D). The Phase 3 instrument used categories such as *participations, timings, materials,* and *activities* to describe the lessons concerned. It was also felt appropriate to include a *student communicative opportunity time* criterion for listening, speaking, reading and writing activities in an attempt to detect significant differences in how much learning and communicating time individual students were actually getting.

The findings from the 10 IELTS preparation lessons analysed, supported by analyses of 20 recorded *PL2000* lessons, should, of course, be treated with caution, given the limited number of recordings and analyses so far in the video database concerned (see Chapter 5). This is already being expanded, however, through related impact study research (see Chapter 8).

Impressions from the IELTS preparation lessons recorded and analysed suggested that, while communicative activities across the four skills in line with the test tasks were evidence of CALT influences (see previous), the opportunities for learners to communicate on their own behalf seemed to vary considerably, with teacher approach and attitude a particular influence. This was further evidence, of course, of the influence of the teacher factor. Communicative opportunity for learning in the target language was enhanced where teachers were willing and able to allow learners the time and space to try to communicate, even when they were still uncertain of their ability to do so. The observed IELTS lessons included significant moments, when, thanks to a teacher's well-prepared but flexibly learner-centred lesson, a learner was able to strain towards adequate communication of a message with the help of a fellow-student partner also communicating for that purpose in the target language.

General impressions from the analyses of the IELTS preparation lessons may be summarised as follows:

- learners who are motivated, but sometimes to the extent of wanting, even demanding, a narrower IELTS focus than their teacher would otherwise tend to offer
- teachers' preference for task-based, often inter-related, macro-skills activities, involving micro-skills relevant to IELTS
- the use of materials from within and beyond the textbook
- multicultural learning and communicating between learners often one of the most engaging features in mixed nationality classes

- an ethos of focused activity within a coherent, often learner-centred, institutional approach to preparation for IELTS, but
- teacher willingness to try a range of teaching methods and approaches.

Student and teacher preparation course satisfaction

An open-ended item invited candidates already studying in an English-medium situation to comment on whether their IELTS preparation courses provided them with the language knowledge and skills they needed. There is a high 83% positive response from the 282 participants concerned. But this apparently positive view of preparation courses did not necessarily mean that candidates were satisfied with their own performance on the courses. When all (N=431) students attending preparation courses were asked whether they felt they were successful on the courses, the responses were less positive (see Figure 6.11), with only 49.5% of those who were sure of their answer saying yes.

Figure 6.11 Candidate views on their success on IELTS preparation courses

Do you think you were/are successful on the preparation course(s)?	
Yes	184
No	187
Unsure	39
Other	20

The main criteria for success mentioned by those responding positively to this item and adding their comments (130 students) were: *perceived improvement in English proficiency level (26 responses) or skills (14 responses)* and *increased familiarity with the test (33 positive responses)*. This result suggests an interesting balance in students' perceptions of success between improvement in their target language proficiency and increased test familiarity.

But (see the summary figures in Figure 6.12) a closer analysis of the reasons given by the 133 students elaborating their *negative* answers on preparation course success reveals the candidates focusing mainly on problems of their own rather than with the course itself.

Figure 6.12: Reasons given by candidates for lack of success on preparation courses

Reason	No.
Did not work hard enough	29
Target language (TL) problems: vocabulary (8), grammar (2), reading (2), listening (2), writing (4), speaking (4), spelling (1)	23
Did not have enough study time, had other demands on time	17
General TL level perceived as not good enough	16
Perceived personal characteristics: lack of concentration (4), lack of confidence (5), inconsistency (3), age (1), over use of L1 (1), poor problem-solving (1)	15
Not enough practice (8), test techniques (3) given on course	11
Course too short	10
Did not achieve required IELTS band	6
Test time pressure	5
Course too academic	1
Test topics	1
Bad test	1

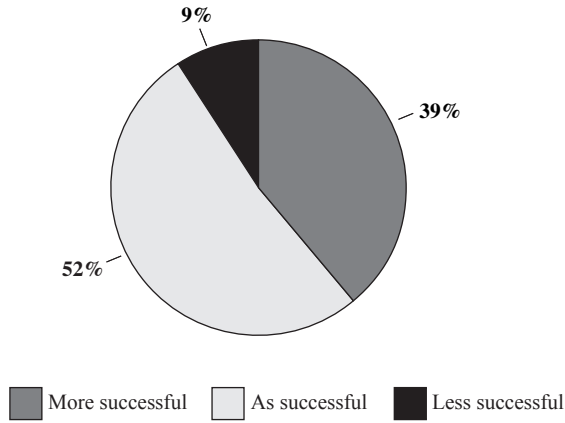
Significant among the explanations given by the participants was: the question of *time,* including the inability of a significant number to devote enough time to preparation for the test; the time span of the courses themselves; and the time pressures in the test itself, a factor we shall return to. Perceived English language proficiency problems are also a significant factor here. And it seems that, although participants are unlikely to see their preparation course as successful if they do not subsequently achieve their required IELTS band score, they tend not to blame the test itself for the situation. The next section of this chapter investigates further the impacts of the IELTS on students and teachers.

Figure 6.13 summarises preparation course *teachers'* perceptions of the success of their IELTS courses compared with other courses they teach.

While only 9% of the 83 teachers felt that their IELTS courses were less successful than other courses they taught, note the reasons given for their positive views:

- clear course goals and focus (21)
- high student motivation (16) (an important issue discussed further when we investigate IELTS impact on candidates)
- clearcut student achievement potential (12), and
- course validity (11) in terms of student target language needs, topics and skills.

Figure 6.13 Teacher perception of IELTS preparation course success



The negative points raised by the teachers concerning their IELTS preparation courses form an interesting counterpoint to the positive perceptions. Nineteen of the teachers, for example, were concerned about the narrowness of their students' focus while on IELTS preparation courses, six about pressure on the students, and six about the validity of the IELTS Writing test, also a point pursued further under IELTS washback.

The picture appears mixed on IELTS preparation courses. The teacher data so far on the influence of IELTS on course content and methods are interesting, emphasising once again the complexity of test washback, and the obvious intervening strength of teacher and teaching variables. The tendency towards eclectic methodological approaches to language teaching on IELTS preparation may again be emerging here (see above). We may recall one of the responses in the Phase 3 teacher profile at 6.7 above, where 68% of the participants claim that they have not received training in the preparation of students for IELTS. Five of the 33 teachers who responded to the final teacher questionnaire item offering the opportunity to 'note anything else you wish to say about your IELTS preparation course' refer to the need for more training in IELTS preparation teaching. In one case, supported by views expressed at interview by five face-to-face contacts (see below), training and technical support for IELTS preparation centres were seen as an antidote to centres that 'provide training just for money and focus only on the passing rate of the test takers'. The teacher concerned then suggested that, in order to 'realise the original concept of IELTS', key centres should be provided with 'more technical support so as to have them play an even more influential effect on

the others and lead the training on to the right track'. The next section of the chapter will pursue this point further, the teacher training variable being one of those identified by Alderson (e.g. 1993, 1995) and leading him to express the view that: 'in short, it is at least as much the teacher who brings about washback, be it positive or negative, as it is the test' (2004a:x).

The study's face-to-face contacts emphasise the need to improve IELTS results at centres running courses for the test, for example: by identifying and disseminating examples of good preparation course practice, through closer monitoring of IELTS centre preparation courses and by more seminars for IELTS teachers. At a language centre student focus group (August 2002) it was suggested that more self-study materials were needed for IELTS candidates. But a research student interview (June 2002) with a participant with first-hand knowledge of the increasing numbers of online IELTS preparation websites in China warned that these, though often apparently informed by specialists on the test, tended to focus on test tips and problem avoidance strategies rather than communicative language skills development.

Impact of IELTS on candidates and IELTS preparation teachers

Perceptions of test fairness

The focus now is on IELTS impact study Research Question 4: What is the impact of IELTS on the participants who have taken the test? This is in many ways a key question. It is central to the study of washback and impact, including as it is intended to not only language learning and teaching matters, but also matters pertaining to test fairness. The question is, as discussed in Chapters 1 and 4, crucial to the validation of high-stakes tests. In this section, therefore, we shall cover perceptions of test fairness, likes and dislikes, pressures, test difficulty and test performance levels. As in the previous sections, these factors will be investigated as they apply to teachers as well as test candidates, thus also responding to Research Question 6: What is the impact of the IELTS test on the teachers preparing candidates to take the test?

Phase 3 IELTS impact study candidate participants who had already taken IELTS were asked whether they thought IELTS was a fair way to test their proficiency in English. Figure 6.14 summarises the responses (of the 190 concerned) with the option to explain why/why not.

Figure 6.14 IELTS takers' perceptions of the fairness of the test

Do you think IELTS is a fair way to test your proficiency in English? (N=190)	
YES	72%
NO	28%
If No, why not?	
1 opposition to all tests 2 pressure, especially of time 3 topics 4 rating of writing and speaking 5 no grammar test	

The 72% : 28% split on perceived test fairness may be considered a rather positive response, especially when we predict people's *expected* response to a question on their perceptions of test fairness in general. Following through to the test takers' specific objections to the test, we find the most frequent of the 49 follow-up responders (some making more than one point) was opposition to *all* tests, an interesting finding and perhaps somewhat unexpected. Among the 25 comments indicating that tests *in general* were seen as unfair were the following:

- 'Any test is unfair as they're tested for a day while they have done a lot before.'
- 'It just depends on one test.'
- 'Because it is a test, it is different from doing it at home.'
- 'It is a test – some students can get a good mark in the test but are not able to use it in real life.'
- 'I usually cannot perform well on exam day.'
- 'Because sometimes it depends on your fate.'

But notice, too, how, in a way typical of open-ended responses, some students claim a feeling that all tests are unfair, yet for reasons which probably do not really apply to all tests:

- 'All tests are unfair because the candidate may not do well due to their fear, confusing questions, time pressure and health problems.'

Some of the responses seem to be focusing on the pressure of high-stakes tests per se, others on the time factor in IELTS specifically, for example:

- 'Performance is reduced under pressure conditions.'
- 'Not fair to test speaking proficiency in a 10-minute test.'

The three remaining rank-ordered participant categories of doubt over IELTS fairness in Figure 6.14, namely test topics, the rating of speaking and writing, and the absence of a grammar test, are language oriented and, like the first three objections, will be investigated further.

Teacher Questionnaire trialling had indicated that *teachers* found a straight question on IELTS fairness rather too broad. The related item on the Teacher Questionnaire thus invited teachers to compare their students' results on the IELTS test with their own assessment of the students' language proficiency across the four macro-skills. Figure 6.15 summarises the results.

Figure 6.15 Teachers' comparisons of their students' IELTS test results and their own assessment of their students' language ability

	Overall %	Reading %	Listening %	Writing %	Speaking %
Same	70	69	64	61	64
Some higher/some lower	23	15	16	23	16
Consistently lower	2	8	10	8	10
Consistently higher	5	8	10	8	10

The overall ratings in the table may be seen, in a way, as the teachers' estimate of test fairness. If so, they are close to the students' view as indicated in Figure 6.14. The least satisfactory match between teachers' and IELTS assessments appears to be for the writing module (see below).

Test taker IELTS likes and dislikes

In the Student Questionnaire, candidates who had already taken the IELTS test were asked for their *likes and dislikes* in the test, these items seen as related to the item on test fairness and supporting reasons (see above). The items were open-ended, and analysed, like other open-ended items, using key words (see Chapters 4 and 5). Figure 6.16 summarises the top likes and dislikes, the comparisons between the two being quite revealing.

Several of the positive perceptions of the test (including fairness, comprehensiveness) have been combined into a 'validity' category. Three of the test modules appear as both likes and dislikes, but Reading (referring to the IELTS Reading test module) is notably absent from the 'likes' column, only to appear as a major 'dislike'. This invites further investigation of the IELTS Reading module below. Yet again the time pressure aspect of the IELTS test re-appears as a major perceived problem.

Figure 6.16 Comparison of test taker IELTS likes and dislikes

	LIKES (132 post-IELTS students responding)	**n**		**DISLIKES** (138 post-IELTS students responding)	**n**
1	VALIDITY [fair (17), 4-skills/ comprehensiveness (15), recognition (7), language and study skills (2)]	41	1	TIME PRESSURE	50
2	SPEAKING	17	2	READING	41
3	STRUCTURE, ORGANISATION, FORMAT	16	3	LISTENING	18
4	WRITING	15	4	WRITING	16
5	INCENTIVE, CHALLENGE, INTEREST, VARIETY	14	5	'COMPLICATED QUESTIONS'	9
6	LISTENING	13	6	LANGUAGE DIFFICULTY	8
			7	SPEAKING	7

Test anxiety

There have been indications from the data already analysed on test fairness and test likes and dislikes that the IELTS test does indeed cause anxiety. That is also one of Alderson's givens:

> We know that high-stakes tests – tests that have important consequences for individuals and institutions – will have more impact than low-stakes tests, although it is not always clear how to identify and define the nature of those stakes, since what is a trivial consequence for one person may be an important matter for another (2004a:ix-x).

The test takers themselves responded to a Likert scale item 'Did you worry about taking the IELTS test?' as in Figure 6.17.

Figure 6.17 IELTS test taker anxiety

Did you worry about taking the IELTS test?		
	No.	%
Very much	78	41
	58	31
	36	19
Very little	18	9

It is very clear that IELTS causes anxiety, 72% of the 190 responding post-test participants claiming to have been *worried* or *very worried* by the test. Applying the response-prediction check to this matter, however, one might

expect that range of professed anxiety about most of the high-stakes tests we encounter in our lives. And the figures of 19% and 9% respectively of participants who claim to have been *not very* or *very little* worried actually seem, totalling as they do 28%, quite high.

Of course, in test impact and validation terms, the concern must be whether a test is causing levels of anxiety that could distort results, the validity of inferences made according to test scores and thus the test's fitness for use. To investigate this matter further from test taker point of view and to triangulate evidence, we can refer to the Language Learning Questionnaire items also included in the IELTS impact study Student Questionnaire (see Chapter 4). Of particular interest should be the section on test-taking attitudes and strategies, the final item of which states *After a test, I usually feel that I have done as well as my knowledge and ability deserve.* This is informative on candidates' test-taking in general, to set against the perceptions of their performance of IELTS candidates who had already taken the test. Figures 6.18 and 6.19 provide the required data analyses.

Figure 6.18 IELTS candidate perceptions of their test performance in general

'After a test, I feel that I have done as well as my knowledge and ability deserve'		
	No.	%
Always	75	14
Often	237	35
Rarely	181	34
Never	37	7

Figure 6.19 Post-IELTS test taker perception of IELTS performance (N=190)

Do you feel that you performed to the best of your ability in the [IELTS] test?	Yes 46%			No 54%
What affected your performance?				
	1 Time pressure (n)	2 Unfamiliarity of topics (n)	3 Difficulty of questions (n)	4 Fear of tests (n)	5 Difficulty of language (n)
1 A lot	91	48	33	26	20
2 Quite a lot	55	58	70	44	56
3 Not a lot	25	51	68	71	78
4 Not at all	10	22	10	32	27

The two tables suggest the following about IELTS test anxieties:

- The IELTS candidate population represents a mixture of optimistic and rather pessimistic views of own *general* test performance i.e. 49% positive, as in Figure 6.18.
- This compares with the fairly similar 46% of those who had already taken the IELTS test who felt that they had performed to the best of their ability in the test (see Figures 6.18 and 6.19).
- Time pressure (see above) figures once again as a major cause of IELTS test anxiety, 81% of the test takers rating it as affecting their performance *a lot* or *quite a lot* (Figure 6.19).
- Topic unfamiliarity again emerges as an issue of IELTS test concern for 59% of the IELTS test takers (Figure 6.19).
- *Difficulty of questions* affects the performance of 57% of the test takers, which, if it relates to the category 'complicated questions' mentioned as an IELTS 'dislike' in Figure 6.16, may be a matter warranting attention in the continuing validation processes for IELTS.
- General test anxiety (40% rating this as affecting performance *a lot* or *quite a lot*) and language level difficulty (42%) are significant but less dominant anxiety sources for the post-IELTS candidates.

The Phase 3 IELTS impact study teachers remind us of the frequently perceived relationship between anxiety and motivation. Figure 6.20 presents teacher responses to two items, the first asking their view on IELTS anxiety, the second on test motivation.

Figure 6.20 IELTS preparation course teacher (N=83) perceptions of IELTS impact on their students

Does the IELTS test cause stress for your students?	
Yes	44
No	27
Don't know	11
Does the IELTS test provide motivation for your students?	
Yes	70
No	8

It will be recalled from the discussion above of teachers' views on the success of their IELTS preparation courses that 'high student motivation' was considered a major positive factor in this. But, of course, the complexity and dynamism of both constructs is clear, as Green tellingly comments:

> If it can be agreed that raising the stakes for participants increases their extrinsic motivation, this is not regarded by all as a benefit for teaching and learning. As with other sources of motivation and anxiety (Alderson and Wall, 1993; Skehan, 1989; Spolsky, 1989), it seems likely that individuals will be differently affected and that benefits for one educational context may be deleterious to another. Indeed, the same effects that are judged to be beneficial by some observers may be judged to be damaging by others (2003:26).

It will be seen in Chapter 8 that the complex issue of motivation in language learning is being pursued in a further Cambridge ESOL impact study research project.

Test module difficulty and validity

As Figure 6.16 indicates, the fact that IELTS tests all four macro-skills, Listening, Reading, Writing and Speaking, is seen by candidates as a significant positive point. Figure 6.16 also shows that there is a difference in the perceived levels of difficulty across the skills modules. This is confirmed by both candidate and teacher questionnaire responses to closely-related items inviting them to rank the IELTS skill modules in order of difficulty. Figure 6.21 summarises these results, indicating that both candidates and IELTS preparation teachers have very similar perceptions on the relative difficulties of the IELTS skills modules.

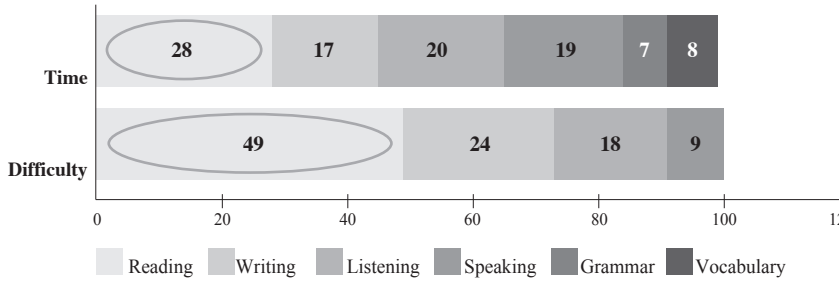
Table 6.21 Student and teacher perceptions of IELTS module difficulty

Most difficult IELTS Module? (%)		
	Students	**Teachers**
Reading	49	45
Writing	24	26
Listening	18	20
Speaking	9	9

The Reading module is seen as clearly the most difficult of the four IELTS test modules across our candidate and preparation course teacher participants. The point is pursued further, but it is of interest at this stage to note (from Figure 6.14 above) that the Reading test does not appear in the top five reasons given by the 28% of IELTS candidates who feel IELTS is unfair. This is despite the fact that the Reading module is the second most *disliked* element in IELTS, according to Figure 6.16.

This variation across the perceived difficulty of the skill modules may to some extent be reflected in the organisation of IELTS preparation lessons. Reading activities (see Figure 6.22) are considered by the candidates as occupying an average of 28% of course time, rather more than the almost equal proportions for Listening, Speaking and Writing.

Figure 6.22 Chart of students' perceived module difficulty and preparation course timings (%)



The observed lessons, however, suggested that it is not always straightforward to identify a classroom activity with a particular macro-skill. Integrated-skill activities seem common. Group oral brainstorming or video note-taking work as a preliminary to a writing topic, for example, may be categorised differently by different students (and teachers).

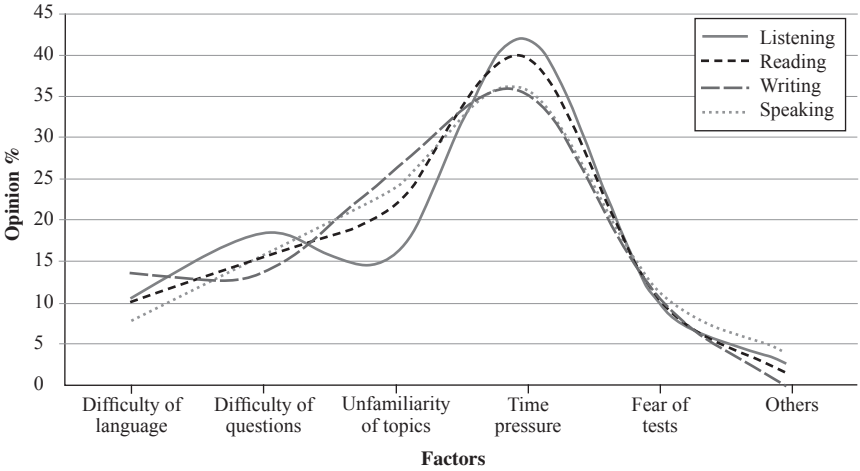
The inter-relationships between perceived difficulties emerging from the questionnaire data seemed to invite further investigation through second-level analysis (see Chapter 4). One such analysis was intended to discover whether there was a correlation between the perceived most difficult test skill, reading, and other factors perceived as affecting candidates' performance, in particular, *time*, which was also frequently mentioned (see previous) as a significant cause of worry for candidates. Figure 6.23 emphasises the dominance of the Reading test module as the most difficult according to test takers and of time pressure as the most prominent problem with the Reading test, as it is, relatively, with the other tests, especially with the Listening test.

Figure 6.23 Relationship between perceived skill difficulty and other factors perceived as affecting candidate test performance

	Difficulty of language	Difficulty of questions	Unfamiliarity of topics	Time pressure	Fear of tests	Others	Total
Listening	4	7	6	16	4	1	38
Reading	13	20	28	51	14	2	128
Writing	10	10	19	26	8	0	73
Speaking	2	4	6	9	3	1	25

Figure 6.24 represents an attempt at a further consolidation among candidates claiming to worry 'very much' about the test and suggesting that the Reading module was a factor in their worry (i.e. their most or second most difficult skill) and also considering that they did not perform to the best of their ability in the IELTS test. It was found that the perception of such candidates of factors that affected their performance *a lot* was again dominated by the time factor.

Figure 6.24 Consolidation between perception of skill difficulty and perception of other factors affecting candidate performance



The perception of the Reading module as more difficult than the other IELTS test modules, and of time pressure as affecting all modules, may be seen as test validity concerns of the theory-based kind. One might ask, for example, whether the reading expected of post-IELTS test takers in their target higher education domain is more 'difficult' than their Writing, Listening and Speaking activities in that domain. Such a question would be part of a set of questions about how the IELTS test modules compare with the skills, macro and micro, that they purport to be testing. Should the Reading module focus more on expeditious reading skills (e.g. skimming, search reading, scanning) rather than or as well as narrower, text part specific micro-skills? And is time such a pressure on those skills in the real higher educational life that candidates are preparing to face? Candidate responses in Figure 6.25 to the Student Questionnaire item on the occurrence of particular reading micro-skills in the IELTS preparation lessons (see also 6.8 above), suggest good coverage of expeditious reading skills.

Figure 6.25 IELTS students' views on occurrence of reading skills on their preparation courses

Reading skill	Yes	No	Not sure
Analysing text structure and organisation.	74	13	12
Interpreting statistics/graphs/diagrams.	74	15	12
Reading texts to predict test questions and tasks.	69	19	11
Learning quick and efficient ways of reading texts in English.	73	13	14
Reading articles, reports, books in your specialist subject area.	44	43	13
Using English-only dictionaries to complete reading tasks.	32	57	11
Reading quickly to get the main idea of a text.	77	13	10

It is interesting to note at this point that test users' perceptions of the impact of time pressure on their performance, especially in reading, are not necessarily supported by results from internal validation analyses which monitor the presence and effect of test speededness. Level of speededness is typically evaluated in the language testing profession on the basis of 'omitted responses', i.e. test items which candidates fail to answer. According to this:

1. 80% of candidates ought to be able to respond to almost every item in each section.
2. 75% of the items in a section ought to be completed by almost the entire group.

By these criteria, analysis of test response data shows that the IELTS Reading tests are not in fact speeded, even though candidates may perceive them as such. Relevant points from face-to-face impact study data were examined, collected, it will be remembered from Chapter 4, from a selection of the centres visited for Phase 3 of the study. In the section of the face-to-face data report headed *Skill module validity,* a range of related matters are raised.

A centre director and a teacher from a university English language centre (interviewed in June 2002) represented a frequently mentioned view, that it is, by definition, difficult to test reading and writing skills directly in ways that authentically relate them to real academic reading and writing activities. In the target academic study domain, these tend to involve long, multi-sourced texts, handled receptively and productively, and heavy in references and statistics. A focus group of 15 IELTS candidates from five countries at the same university questioned why reading and writing should be tested under time constraints when in real higher education practice, students were able, to a large extent, to make their own time. But the student group appeared less in favour of target domain authenticity when its members suggested that there should be choices of topics in the IELTS Writing test because a single topic might be 'unsuitable'.

Members of a teacher focus group at the language centre of a different university in November 2002 agreed that on balance the time pressure in the IELTS Reading module was still excessive, and suggested that the Reading test as it stands should be allocated 1 hour 20 minutes rather than 1 hour. The same group recommended 1½ hours rather than the current 1 hour for the Writing module on the grounds, again, that the timing was both too tight and unrepresentative of most of the writing students were called on to do in their actual academic life.

For some post-IELTS international students, the heavy (immediately-required) reading and writing load on entry to university or college proved a major difficulty. Teachers at the June 2002 interviews and in the November 2002 focus group agreed. That the IELTS Reading and Writing modules were perceived as the two most difficult was thus to be expected. Some of the teachers in the face-to-face sessions and the focus group, however, wondered why some very difficult tasks in the IELTS Reading module demanded careful reading and analysis *yet* under time pressure. Were these there, they asked, to discriminate for the 'Band 8/9 candidates'? Were they, perhaps, the kinds of items referred to above among IELTS candidate dislikes, as 'complicated questions'?

The complexity of the validity notion and the difficulty of designing and delivering reading tasks on which test taker scores could be interpreted as a valid reflection of their competence in the skills concerned (as they occur in the target language domain) appeared to be accepted in the face-to-face discussions. A November 2002 focus group questioned the authenticity of IELTS academic writing module task A, where candidates are asked to write a report 'based on materials found in a table or diagram' (*IELTS Handbook*, 2003). A teacher from another centre (September 2002) wondered whether this task was in fact 'interpretative' (as claimed in the *IELTS Handbook*) since candidates appeared mainly to be asked to *describe*. A teacher, at a third centre (also November 2002) asked why the data labels in the same tasks were often so obviously non-authentic. Some teachers (at September 2002 interviews at a major IELTS centre) were concerned that the Reading test tasks could become formulaic, predictable and thus vulnerable to slick test techniques training.

A call for *specialised subject modules* was made by two language school focus groups, one of 10 students from seven countries (July 2002) and one of 12 students from eight (November 2002). Six subject-specific Reading and Writing modules had been offered, of course, in the IELTS predecessor test, the ELTS (see Chapter 4) and three such discipline areas (physical sciences and technology; life and medical sciences; arts and social sciences) by the IELTS test itself from 1989 to 1995. However, the matching of subject-specific sub-tests to receiving institution courses was by no means a clear-cut matter (Charge and Taylor 1997:374). Research into the potential effectiveness of a one-module approach (see Clapham 1996) indicated that this would not

discriminate against candidates, whatever their field of study. Thus, in April 1995 the subject-specific IELTS modules were replaced by single academic Reading and Writing modules. Davies (forthcoming) discusses in some detail the issues and considerations involved in this decision.

Teachers from more than one centre (in September and November 2002 interviews) felt that the IELTS test attempts, rightly, to assess micro-skills, including those with a study-skill orientation, for example *evaluating evidence* (see Figure 6.9) as well as language performance. In fact, IELTS is not normally considered to test study skills in the conventional sense such as library use or referencing, but it is of interest that some IELTS preparation course teachers think that it does. Note also that 'study skills' in the Common European Framework interpretation mean 'ability to make effective use of the learning opportunities created by teaching situations' and include 'ability to use available materials for independent learning' (2001:107). Teachers at almost all centres visited referred to the need to help some pre-IELTS students with skills (which the teachers sometimes referred to as 'study skills') likely to enable them to cope with the more independent study modes of the target receiving institutions.

Test performance

IELTS impact study figures on band scores suggest that, when they were attending their preparation courses, participant candidates had not yet achieved the band scores required by their target institutions, and were not yet expecting to achieve them. The mode bands of the relevant sub-groups of the Phase 3 study's student populations were as follows:

- band 6 for the scores *achieved* by those who had already taken the test
- band 6.5 for the bands stated by the students as *required*, and
- band 6 as the band expected.

The table in Figure 6.26 compares the performance of impact study candidates who had taken the test with IELTS candidates in general.

Figure 6.26 IELTS impact study and IELTS general population band scores

	Overall Band score	Reading	Listening	Writing	Speaking
Phase 3 Study Mean	6.0	5.7	5.9	5.7	6.1
All IELTS Academic Module Candidates (2002)	5.86	5.79	5.79	5.64	5.94
All IELTS GT General Training Module Candidates (2002)	5.65	5.43	5.48	5.64	5.79

Main points emerging from the table are:

1. That the band scores of the impact study and the general IELTS populations are close, with the impact study participants scoring slightly higher across all modules apart from Reading (which, of course, they have identified as their most difficult module).
2. That impact study test taker candidate band scores for the Reading and Writing modules are slightly lower than for the other modules (again in line with their view of the difficulty of these two modules); their Reading and Writing scores are also lower than their overall band score, although all-IELTS mean band scores for Reading are equal third of the five Academic Module bands (but just the lowest for General Training Module candidates).

Phase 3 of the impact study did not, it will be recalled (see Chapter 3), have as one of its research questions the investigation of IELTS scores. Routine Cambridge ESOL analyses already include performance analyses with reference to Candidate Information Sheet (CIS) data. Given a system combining routine test validation operations such as those carried out by the Research and Validation Group (see Chapters 1 and 4) with validation-related research projects, such as IELTS impact studies and IELTS funded research (see also Chapter 1), it would seem more appropriate to cover whole-population studies of relationships between candidate baseline data and test performance through the routine Research and Validation Group systems (see Weir and Milanovic 2003 Chapter 2). Closer investigation of particular test taker characteristics or test washback effects and performance may well be pursued through narrower washback studies (see Chapter 8).

However, impact studies, with their broad coverage of research questions and multi-faceted data sources, would appear to add rich data to those used in routine research and validation operations in answer to such questions.

Impacts on receiving institution administrators

Although a receiving institution questionnaire (see Chapter 4 and Appendix E) had been developed, validated and was used at Phase 3 meetings with IELTS administrators, most of the administrative impact data for the study were collected through *interviews*. Fuller further impact inquiry is suggested for later phases in the study of IELTS impact (see Chapter 8).

The Phase 3 administrator interviews covered areas similar to those specified in the questionnaires, that is: participant background and institutional role(s); IELTS entry cut-off levels and decisions; perceptions of the test's validity (including bias) and reliability; the four modules; IELTS pressures and other impacts; candidate language difficulties; preparation

courses; fees; test information. It will be noted that many of the information areas in the receiving institution questionnaire are shared with the other IELTS impact study questionnaires, especially the student and teacher instruments.

A university IELTS administrator interviewed for the study noted that the broad range of responsibilities in her own brief included the following: test information, application processing, written, phone and online; test schedules, centres and rooms; test administration, invigilation and security; examiner contacts, local marking, fee payments; overseas student advisory support and certification; liaison with academic departments, and university administration and finance; co-ordination with Cambridge ESOL and British Council; test monitoring and staff training. It was clear that useful institutional views on such matters were held not only by full-time administrators but also by IELTS candidates and teachers.

Receiving institution IELTS cut-off points

This key IELTS issue is covered in depth by Banerjee (2003) in her doctoral research, which is a further example of the range of research approaches employed in the continuous validation and revision of the IELTS test.

In Phase 3 of the study of IELTS impact, interviews and focus groups were arranged with people who had administrative responsibilities for IELTS. The contacts took place between June and November 2002, at six IELTS centres in the UK, including two universities, and four language centres or schools. As already mentioned (see Chapter 4) IELTS administration data were also collected through external contacts made by the Cambridge ESOL UK Development Manager.

Administrators and teachers at a UK university (June 2002) noted variable IELTS cut-offs across departments and under-graduate vs graduate course levels. The following pattern was typical:

- arts and humanities IELTS cut-off band score of 6 at under-graduate level vs 6–7 for post-graduate programmes
- 6.5 for both levels in the sciences and technology
- social sciences: 6.5 to 7
- business: 7 (postgraduate only).

The group commented on tensions between a university's pressure for higher cut-off scores to maintain or raise standards and departmental requests to lower cut-off band scores in order to admit students with particularly strong non-linguistic strengths.

A focus group at a second university (November 2002) identified 6.5 as the university cut-off norm but with departments ultimately deciding their own IELTS cut-offs, with the flexibility to select according to the claims of

individual applicants. Teachers in the focus group felt that post-graduate courses would be particularly difficult for students who had not achieved at least an IELTS 6 band score. There were reports from our interviews with university administrators, however, that some institutions were willing to accept students with bands below six. The IELTS band 5 *'Modest user'* descriptor refers to a 'partial command of the language, coping with overall meaning in most situations, though is likely to make many mistakes. Should be able to handle basic communication in own field'. The band 4, *'Limited user',* is described as having a basic competence which is 'limited to familiar situations', as having 'frequent problems in understanding and expression' and being unable 'to use complex language'.

The mode required band of the IELTS impact study Phase 3 student population, it will be recalled from the previous section, was also 6.5 (compared with expected and already achieved bands of 6). At an IELTS managers' and administrators' meeting (January 2002) the question was raised whether the test discriminated proficiency levels finely enough at bands 6–7, given the prominence of receiving institution admission cut-off points around these levels. The November 2002 university focus group, however, felt that the IELTS appeared already to have been 'designed for' students at bands 5–7.

Administrators, as well as language teachers, tended to link the question of IELTS cut-off points with IELTS preparation course provision. The feeling from teacher and administrator interviews at a language school (September 2002) was that students should not enter IELTS preparation courses unless they were already at least IELTS band 6 level. It was difficult, according to teachers at a university language centre (November 2002), to provide for students at bands 4 and 6 on the same course; students with lower proficiency levels should, the teachers felt, enter 'general English' courses. Some of their IELTS candidates were assessed as needing intensive remedial grammar and vocabulary courses before their IELTS course, although such students were sometimes unwilling to take these ('non-IELTS') courses even though they were offered by the university.

At the other end of the proficiency scale were reports, from two of the language schools visited, of first-language or ambilingual speakers of English required, unnecessarily it was perceived, to take the IELTS test. The preparation of such candidates for the test thus often consisted merely of familiarising themselves with its format through IELTS specimen materials packs. All the language centres contacted face-to-face referred to an increasing demand for IELTS preparation courses as candidacies for both academic and general training modules rose. Given the significant managerial and financial implications of official IELTS centre status, the opening of new IELTS centres to meet growing demand, it was emphasised, must be carried out in a fair and systematic manner.

The IELTS partners now publish *IELTS Band Score Requirements (*the

2004 edition was compiled by IELTS Australia) providing 'a sample of organisations recognising IELTS and their required band scores' (2004:2). This names institutions, course types and minimum band score requirements, although 'as a guide only' (op. cit:1).

IELTS administration

The question of IELTS test availability was raised at several of the IELTS impact study Phase 3 interviews (in July, September and November 2002), the general belief at that time being that students should be able to take the test when they wished. A change from IELTS tests arranged by centres on demand to fixed-date scheduling was implemented in September 2003 to enhance security, with test papers distributed nearer to test dates and with the use of unique test versions. This change appeared to have caused much discussion among IELTS test centres because of the various impacts, for example, room bookings, administrative and examiner staff commitments, the reduction at some centres from five to four sittings per month. Also related to test scheduling was the two-week results turnaround for the IELTS, considered by one administrator interviewed as an important justification for the IELTS fee.

Some IELTS impact study candidate participants disliked the compulsory three-month gap before they were permitted to re-take the test. An administrators' meeting (January 2002), presumably reflecting the urgency for many candidates of institutional qualifying deadlines, suggested that candidates should be able to accumulate band scores on the test skill modules over a specified period of time rather than having to take all modules at the same sitting. This would offer the opportunity for those who did not achieve the required band score in one module to focus on that skill in a retaken test three months later.

A fairly general administrative point raised at several of the face-to-face meetings with IELTS candidates, teachers and administrators was that IELTS information was not as transparent or as user friendly as it might have been. It seems clear that user expectations for more in the way of information, not only on practical matters but also on research and validation processes and outcomes, are rising, in tune, perhaps, with the times (see also Chapter 8). IELTS handbooks, annual reports and websites (also Chapter 8) might thus be expected to become more comprehensive and open. This is indeed what has been happening over recent years as IELTS has grown in use and recognition.

Conclusions

This chapter has presented key messages from IELTS impact study Phase 3 data, collected using the instruments described in Chapter 4, through the

systems outlined in Chapter 5 and attempting to answer the research questions presented in Chapter 3. The focus here has been on: the profile of the impact study candidate and teacher participants, IELTS washback on preparation content, methods, skills, activities and materials; perceptions of IELTS test fairness, pressures, likes and dislikes; a summary of impact study face-to-face data on test module difficulty and validity; and receiving institution views on various administrative impacts of the IELTS test.

The analyses above of impact study data from candidates, teachers and administrators suggest that, in general, the IELTS test is perceived as:

- a reasonable direct communicative performance test of all four macro-skills
- hard but mainly fair, considered appropriate for use with the candidates in the study who were heading for under-graduate and post-graduate studies
- using authentic and mainly appropriate content, on topics most of which seem relevant to the target communicative activities, covering a suitable range of micro-skills
- having Reading and Writing modules generally considered the more difficult of the four modules, partly because of time constraints often perceived as severe and not necessarily authentic to the target domain activities
- a high-stakes test, causing anxiety though not to an out-of-the ordinary extent, motivating both candidates and teachers, although sometimes to the extent of demands for narrower, test-based preparation courses
- encouraging some of these courses to motivate and satisfy both students and teachers because of their focus and the range of relevant activities, skills and materials inspired by the content of the IELTS test itself.

These are the main conclusions from the data and findings summarised in this chapter.

The chapter has covered selected key areas of inquiry in Phase 3 of the IELTS study. Cambridge ESOL will now decide on measures to take account of the findings of the study in the continuing validation of the IELTS test. It is likely that the Research and Validation Group would advise the Examinations and Assessment Group, Operations Group, Business Support Group, appropriate senior management members and IELTS partners of matters requiring discussion and decision in the interests of continuing IELTS validation work. The role of the impact study is to describe and try to explain rather than to recommend. Chapter 8 continues the discussion of the implications of the IELTS Phase 3 study within its Cambridge ESOL context.

7 The *PL2000* Impact Study: main messages

Following the presentation of main messages from the study of IELTS impact in Chapter 6, Chapter 7 now summarises some of the key findings from the *PL2000* Impact Study, its main messages for stakeholders emerging from responses to the research questions defined in Chapter 3:

- What washback is the *PL2000* having on the pedagogy, materials, and media for language teaching and learning?
- What washback are changes in language teaching and learning pedagogy, materials, and media having on the performance and attitudes of the students and the teachers?
- What washback is the *PL2000* having on language evaluation and assessment?
- What impacts is the *PL2000* having on educational managers, including heads of schools?
- What impacts is the *PL2000* having on support for teacher in-service and test-awareness programmes, and resource centre development?

PL2000 Impact Study profile

Having agreed the purposes, approaches, outcomes and resources needed, and defined the research questions and action plan for the *PL2000* Impact Study (see Chapter 3), then agreed data collection instrumentation and approaches (see Chapter 4) and management (Chapter 5), an Impact Study population was decided, as follows:

- an opportunity sample of seven case study schools with region and school type strata, in the north, central region, and south of Italy and representing middle and high school levels, with some contact, also, at elementary level
- to be seen on two visits (not including the January to March 2001 pilot visits) to each of the case study schools, at the beginning and towards the end of the school year.

The table in Figure 7.1 summarises *PL2000* Impact Study participation data collection on school pilot visits.

Figure 7. 1 Participation on *PL2000* Impact Study pilot visits

Participation on pilot visits to case-study schools (January – March 2001)	Class videos	Head interviews	Teacher interviews	Parent interviews
Technical institute	1	1	3	3
High school	–	1	3	–
Middle school	2*	–	–	–

* sound inadequate for analysis

The table in Figure 7.2 indicates the data collection, participation and scheduling on the main visits to the case study schools.

Figure 7.2 Data collection and participation on *PL2000* Impact Study October 2001 and April 2002 visits

Schools	Class videos		Interviews, focus groups, written contact						Questionnaires		
			Heads		Teachers		Parents		Students		Teachers
Types / Dates	10/01	4/02	10/01	4/02	10/01	3-4/02	10/01	4/02	10/01	4/02	4/02
Elementary	2	2	1	–	–	1	2	–	**	**	2
Comprehensive	1	1	1	–	1	3	2	–	23	24	2
Middle (1)	2	2	–	–	2*	2	–	–	43	42	2
Middle (2)	1	1	1	1	–	–	6	2	13	13****	1
High school (1)	1	1	–	–	–	4	–	–	18	5	1
High school (2)	1	2	1	1	2	7	2	2	14	16	2
High school (3)	–	–	–	–	3(1)***	3	–	–	–	–	
Technical Institute	1	1	1	1	2	1	1	1	7	10	1
	9	**10**	**5**	**3**	**11**	**21**	**13**	**5**	**118**	**110**	**11**

* session not videoed
** questionnaires not administered because pupils considered too young
*** private teacher interviewed in addition to three high school teachers
**** student questionnaires for the two visits were from different students, though from the same classes

Figure 7.3 summarises participations in the *PL2000* Impact Study.

7.3 *PL2000* Impact Study total participations

	Lessons videoed	Head interviews	Teachers interviewed etc*	Parent interviews	Student questionnaires	Teacher questionnaires
TOTALS	**20**	**10**	**38**	**21**	**228**	**11**

* these included some teachers of other *PL2000* foreign languages (French, German, Spanish)

In the structure of the *PL2000 Impact Study Main Report* (Hawkey 2003:2–3), answers to the research questions are sought in the data analysis chapter, which is structured as follows:

- washback on the *PL2000* classroom
  - communicative approach to language teaching (CALT)
  - pedagogy and ethos
  - lesson planning and management
  - function and form
  - media and modes
  - books
- impacts on attitudes
  - from the students
  - from the teachers
  - from the heads
  - from the parents
- washback on assessment
  - from the students, the teachers and the classroom
  - from the school heads and parents
  - Cambridge exams
- washback on performance
  - *PL2000* impacts on exam candidacies and performance
- impacts on management and support
  - management
  - teacher professional support.

The sections below will follow the pattern of the main report in summarising the main messages of the Cambridge ESOL *PL2000* Impact Study.

Washback on the *PL2000* classroom

The Communicative Approach to Language Teaching

The first *PL2000* Impact Study research area was washback from the project on foreign language learning, pedagogy, materials, and media.

The study elicited relevant data from several sources:

- its classroom video analyses (see Appendix D)
- the Student and Teacher Questionnaires (Appendices G and H respectively)
- teacher interviews (Appendix I)
- school head and parent interviews
- information which some of the case study school teachers provided in writing direct to the Impact Study co-ordinator

- UCLES EFL : Bell International *PL2000* teacher essay competition entries.

The first point emerging from such data is that the *PL2000* appeared to have transmitted successfully the message of its aim and objectives in favour of the communicative approach to language teaching (CALT). The *PL2000*, it will be recalled, targeted 'the acquisition of a pragmatic-communicative competence according to the guidelines of the material contained in the Common European Framework of Reference for Languages', and 'communicative competence in reading, written and oral interaction and production' *(*see Chapter 3). We note immediately that the *PL2000* did not offer new curricula but rather referred to the Common European Framework for Languages and its levels as targets for the Italian school cycles (A1 Breakthrough level for primary school; A2 Waystage for middle school; and B1 Threshold for high school exit).

The case study teachers talked freely and knowledgeably about the CALT. Typical views, that is views *exemplifying* comments on the CALT *and*, in combination, encapsulating the overall views of the case study teachers on the communicative approach, are cited here. They are taken from the questionnaires completed by the 11 teachers of our case study classes and their colleagues, and of the 37 teachers participating in teacher interviews, focus groups or written contacts. The words given in single quotation marks are direct quotations from the teacher interviews or questionnaires; those without are taken from the teacher interview summaries:

- Teachers are now more aware of the CALT and teach with more confidence.
- English teaching now concentrates on students' communication needs; new students speak the language much better and are more confident doing so (both from a comprehensive school teacher, October 2001).
- *PL2000* changes approaches to emphasise communication, listening and speaking in particular.
- 'My attitude has changed a lot; I speak English in class more often than before. I cover all the four language abilities, listening included, and I tend to use more enjoyable activities' (both from a middle school teacher, April 2002).
- Italian language teachers favour the CALT; previous FL teaching aims and approaches emphasised gap-filling and reading, not listening or speaking (two native English speaker high school teachers, October 2001).
- Teaching approaches are communicative, so the students communicate in English, not 'in an Italian way', which is noticeably different from non-*PL2000* students (high school teacher, April 2002).
- 'CLT (Communicative Language Teaching) is the best way but it is hard', especially as teachers should adapt approaches to the needs of different

students, motivated as well as unmotivated; CALT theory and practice differ (three high school teachers April 2002).

- 'New emphasis on communication, on students doing things with the TL, and learning how to learn, through which foreign language doesn't matter' (high school teacher of German, April 2002).

A further sign of positive *PL2000* impact was the agreement of all 11 teachers from the case study schools who completed the teacher questionnaire in April 2002 that *'communication skills relevant to students' language needs'* was a Project objective which had been achieved '*very well*' or *'well'* in their schools.

The teachers' analysis, in their questionnaires, of activities that were prominent in their lessons suggested a fairly balanced coverage of the four skills. The following student activities were all selected by nine or 10 of the 11 case study teachers as '*very frequently'* or '*quite often'* occurring in their lessons:

- pair discussions
- listening to the teacher talking to the whole class
- listening and note-taking
- reading books or other materials
- writing notes, letters
- doing practice exams.

This compares interestingly with the case study *student* views on their class activities. The activities perceived by the students as occurring 'frequently' or 'quite often' in their classes were, in order of frequency of selection as indicated in Figure 7.4:

Figure 7.4 Student perceptions of frequency of activities in their English classes at school (N=118)

Activities in class	Frequently	Quite Often	Sometimes	Never
Individual students				
1. listening to the teacher talking to the whole class	35	20	9	2
2 reading texts (from books or other materials)	25	24	17	1
3. writing notes, letters or compositions	18	27	20	2
4. reading then writing answers to questions	22	23	20	2
5. discussions with whole class	20	24	17	5
Two classes (total 42 students) negotiating block votes*				
1. listening to the teacher talking to the whole class	√	√		
2 reading texts (from books or other materials)		√√		
3. writing notes, letters or compositions	√	√		
4. reading then writing answers to questions			√√	
5. discussions with whole class	√	√		

* in two of the case-study classes, the teacher and students worked together to produce joint responses to some of the student questionnaire items.

Note that the students appeared to regard the *speaking* activities as somewhat less prominent in their *PL2000* classes than their teachers did. There may be a message here (see Hawkey, forthcoming) on the difference in perceptions between learners and teacher of the shape and elements of a classroom lesson. But the students also felt that their English speaking skills were the most *improved* over the school year. Of the 161 responses to the questionnaire item concerned in April 2002 (some students specified more than one 'most improved' skill) the following selections were made:

Figure 7.5 Student questionnaire responses on the most improved skill(s) over the 2001/2002 school year

Reading	Listening	Writing	Speaking
29	25	51	56

A significant majority of students perceived their productive skills as improving more than their receptive skills. Yet both the case study teachers and, even more strongly, the students, saw '*the teacher talking to the whole class'* as the dominant language classroom activity (see further below).

Cross-checking with the entries to the 2001 UCLES EFL : Bell International *PL2000* teacher essay competition entries, on the topic 'What the *Progetto Lingue 2000* means to me as a teacher', we find that the competitors appear to support both the content and tone of our other data on the CALT. One candidate calls her chosen methodology *'functional-notional-communicative'* and enjoys her students' very motivating realisation '*that they could express their ideas and wishes in English'*. Another sees the *PL2000* approach as offering '*a risk-free environment'* and *'improved teacher attitude and interesting materials'*. The same teacher, however, warns against the *'students-need-to-talk syndrome',* that is *'an excessive emphasis on oral-aural skills realised in pair or group work with the exclusion of individual tasks'*.

Pedagogy and ethos

Here, we refer primarily to the ways in which the teachers (and the learners) sought to achieve the communicative objectives signposted by the *PL2000*.

The most direct evidence on *PL2000* classroom pedagogy was gleaned from the 20 sessions of closely observed and video-recorded classroom language lessons in our case study schools, analysed (see Chapter 5) using the classroom observation analysis instrument (Appendix D) shared with Phase 3 of the IELTS impact study. The *PL2000 Impact Study Main Report* suggests that 'there is good evidence in almost all the lessons analysed' for the study that the CALT did indeed motivate 'teachers to seek teaching/learning

activities that were relevant to the communicative needs of their students' (2003:32). Examples of these from the video-recorded lessons, in rising order of the ages of the students concerned, were:

- middle school, Common European Framework (CEF) Level A2 students, invited by their teacher not only to exchange personal information with their friends, in pairs, groups and in informal presentations to the whole class, but also to amuse their colleagues (and themselves), as 13-year-olds often like to do, with flights of imagination and humour, including, in the observed lesson concerned, stories of intelligent, tennis-playing tropical fish, and living as tramps
- high school CEF Level B2 students (16–18 years old) working in pairs to compose e-mails to their English teacher suggesting the kinds of English language learning activities they considered would meet their needs on their course
- CEF Level C1 (18-year-old) students, working in small groups, discussing and writing an action plan for the multimedia promotion of a local product.

Yet, as suggested in Chapter 6, the idea of increased communicative opportunities for learners may appear to be contradicted by the fairly frequent reference in the classroom observation video analyses to teacher communicative dominance. In 11 of the 20 case study lessons analysed and observed in October 2001 and in April 2002, there is reference to teachers interrupting or supplying words and ideas to fill silences, limiting time and opportunity for learners to learn and feel their way towards coping with communication in a foreign language, thereby perhaps reducing learning impetus.

The data under our *'pedagogy and ethos'* heading here are clearly relevant to both the first and second research questions of the Impact Study, namely:

What washback is the *PL2000* having on the pedagogy, materials, and media for language teaching and learning?

What washback are changes in language teaching and learning pedagogy, materials, and media having on the performance and attitudes of the students and the teachers?

The data discussed here also connect with the next issue, that of the planning and facilitation of lessons.

Lesson planning and management

There are comments in the classroom analyses which suggest that restrictions on learner communicating time may be the result of lesson planning and management problems as well as teacher reluctance not to intervene. Stated or

implied in comments on six of the 20 analysed lessons was the suggestion that their perception of the CALT might have induced in some teachers a changed view of the importance of lesson planning. In observed lessons where there were signs of somewhat unproductive disorganisation, it appeared that the teacher might have had an appropriate *main idea* for a lesson (as in examples such as a relevant role play or simulation, a spidergram on a currrent topic, a survey questionnaire) but without necessarily considering the procedures, tasks or activities that might best ensure that learners (and teacher) were able to make the most communicatively of the lesson planned round the main idea.

Where planning for useful communication by learners appeared to be lacking, problems such as the following tended to arise:

- teacher dominating talking time, thus too little opportunity for students to create their own interactions (in 10 of the lessons)
- students left inactive and, possibly, bored (9)
- lesson fragmentation into too many short episodes caused by changes of pedagogic direction in a sometimes vain attempt to find paths (unmapped in the lesson plan) to more effective communication by learners, and/or, relatedly, stop-start episodes, repetition of activities and frequent resort to fairly random question and answer sessions (6)
- an absence of learning round-ups, reinforcement, consolidation (4)
- reduced educational value added (4)
- long and sometimes confusing task organisation explanations (3).

The penultimate problem above, perhaps less self-explanatory than the others in the list, seemed sometimes to emerge in the form of mundane and predictable comment and topic references inserted by teachers to or about the students, when communicative lessons had lost their way through lack of adequate planning.

In eight of the 20 *PL2000* Impact Study language teacher interviews and focus group discussions, it was agreed that lesson planning, far from being *less* important once teaching approaches moved towards an emphasis on learner-centred communicative opportunity, was still crucial or had become even more so. An experienced English teacher responsible for the co-ordination of *PL2000* activities at her high school, summed up this view with her remark that, given the communicative construct and the aim of facilitating activities in line with students' communicative needs, particularly careful lesson planning is needed to create the conditions for real communication. Related points made by other teachers were:

- Classroom and lesson management are harder and more time-consuming with the CALT. With the emphasis on speaking, it is easy to lose control of the lesson (A2 teacher, April 2002).
- 'We're not really scientific in planning our lessons; we're not exact in our

lesson planning or our evaluation'. Lesson preparation is about 'what we do and say' (middle school teacher of A2 level students, April 2002).

- Team teaching is encouraged but successful teamwork and sharing of responsibility needs careful planning (two English mother tongue teachers of B2 level students, October 2001).

One of the essay competition candidates suggested that one washback of the *PL2000* on teachers was '[f]orcing us to view the planning and assessment stages of the process as extremely important parts of our teaching, as steps which require far more serious consideration than they have ever had'.

There was strong evidence from the Impact Study all round, in fact, that lesson planning and management were a crucial aspect of the implementation of the communicative teaching and learning approaches advocated by the *PL2000*. Both lesson planning and lesson management thus appeared to invite prominent attention on language teacher professional programmes set up in connection with the *PL2000* and related initiatives.

Function and form

Also relevant to the research question on the washback of the *PL2000* on language teaching and learning, were data on the balance between language *functions*, the communication of meaning, fluency and appropriacy, and language *forms,* correct language usage, and accuracy. These, of course, represent continua conventionally referred to when the communicative and the structural approaches to language teaching are compared (see, for example, Cook, 2003:31–35).

A tendency was noted in the observed *PL2000* Impact Study lessons for teachers to correct some pronunciation, grammar and vocabulary errors (though not others) while students were in the process of communicating. This may be acceptable, of course, in the interests of maximising communicative opportunity, as discussed above. But inconsistency in error correction was noted with regard to the communicative impact of errors, the stage of communication, and the level and confidence of performance of the students concerned. There was also little evidence of systematic consolidation, although this may have come in subsequent lessons. The occasional and fleeting treatment of error does, of course, carry the risk that students' oral communication may become inaccurate even to the point of impeding meaning. There is comment in 10 of the 20 *PL2000* Impact Study classroom lesson analyses on the apparent lack of a consistent error correction approach or on the prominence of unexpected errors.

Yet Figure 7.6, summarising teacher and student questionnaire data for April 2002, indicates that grammar and vocabulary practice activities, normally associated with a focus on form and accuracy, as well as function,

figured quite prominently in the students' perceptions of their English language lessons.

Figure 7.6 Perceived prominence of grammar and vocabulary exercises in the case study school lessons

Individual students							
Grammar exercises				**Vocabulary exercises**			
Frequently	Quite often	Sometimes	Never	Frequently	Quite often	Sometimes	Never
24	19	20	2	9	24	22	9
Two classes (42 students) negotiating block votes							
Frequently		Quite often		Frequently		Quite often	
√√						√√	
Case study teachers							
Grammar exercises				**Vocabulary exercises**			
Frequently	Quite often	Sometimes	Never	Frequently	Quite often	Sometimes	Never
3	3	5	0	3	6	2	0

The teacher interviews suggest conscious attempts to balance fluency and accuracy work at all levels, 10 of the sessions including comment on the role of grammar in the teachers' classes. A B2 level case study teacher who appeared very 'communicative' in the theory and practice of the pedagogy in his observed lesson, felt nevertheless that, on his and other English courses at his school, '[t]he students are getting quality instruction in language grammar, syntax and structure'. A group of high school teachers agreed: 'CLT (communicative language teaching) does not mean no attention to grammar; all approaches are acceptable as long as they improve students' communicative competence'. The same teachers six months later confirmed that 'CLT has to be combined with more structural approaches'. A middle school teacher summed up the form/function relationship as follows:

> There is a need for a balance between allowing students to make mistakes in English communication, but not accepting everything; some correction of grammar and lexis errors is necessary; if you don't know the code, you're in trouble.

Once again the message from the essay competition entries confirms our observation on the accuracy/fluency question. One candidate noted the relatively poor level of listening comprehension of students who had been learning their English through traditional methods without really *using* the language, adding that they might nevertheless have only a basic grammatical competence, and a very limited vocabulary. Another candidate suggested the brainstorming of lexis as a remedy to vocabulary problems, a third the use of CD-ROM stories to improve weak understanding and pronunciation.

The whole accuracy/fluency question remained crucial to the achievement of the *PL2000*'s aim to enable students to meet their needs as specified by the appropriate CEF proficiency level statements. These, remember, include A1–C2 level can-do statements for both *grammatical accuracy* and *vocabulary control* (ibid. 2001:112–114).

Materials, media, modes

The *PL2000* project descriptions referred to the integration of information technology into the teaching and the '*use of diversified materials in place of textbooks*'. Eight of the 23 English language teaching textbooks specified by the case study schools on the school profile form had the name of an external exam in their titles (e.g. *FCE Gold*, *PET Practice Plus)*, such tests being among those approved by the Ministry for the external certification of students on *PL2000* courses. The remaining textbook selections in the Impact Study data indicated a preference for books taking a fairly communicative approach. However, teachers tended to use additional materials from a variety of other sources rather than a single coursebook, for example cut-out photographs, self-designed spider-grams, information-gap hand-outs, audio-cassettes, wall charts and, in four of the 20 lessons, external exam practice cassettes, papers or photographs.

The following are the comments on textbooks from the teacher interviews. Being teachers, our case study interviewees and focus group members tended to talk about language teaching books in terms of their pedagogical context of use:

- New and improved textbook materials are better designed for communicative activities such as letter writing, written dialogues … A textbook may be used as the main guide to a lesson plan, with the teacher adding material (often from other textbooks), usually because one textbook does not contain enough practice materials relevant to particular students' needs for the exams (from an interview with two middle school teachers of A2 level students, April 2002).
- Textbooks are lacking in up-to-date ideas re students' actual target language activities; 'so I use videos, BBC, photocopies, examples of Australian and US culture (not only UK), pop song lyrics' (middle school A2 class teacher, April 2002).
- The good new CALT textbooks can be useful in providing ideas, starting points and material; teachers must pick, from a range of books, what is suitable for their students. … Old coursebooks were artificial (culturally for example); new books use everyday language and real situations (C1 technical institute teacher, April 2002).
- CALT coursebooks and workbooks have useful exercises and activities, including attention to grammar (*PL2000* German teacher, April 2002).

- Past exam papers, videos and selected coursebooks are used on the external certification courses … Past papers are useful but, because they are international, lack a national (Italian) cultural relevance (case study high school teacher and *PL2000* course organiser, April 2002).

Though clearly an element in a teacher's pedagogical resources, information technology appeared as a separate project recommendation in the *PL2000* guideline document, which referred to *'the integration of information technology into the teaching and use of diversified materials in place of textbooks'*. Of the 20 lessons video-recorded, observed and analysed for the *PL* Impact Study, five, from two of the case study schools, both in B2 classes, used information technological support, including the following:

- personal and school computers used for student >< student and student >< teacher e-mails
- computer laboratory use for word-processing; Internet picture-viewing and discussion; video viewing and discussion; videocam/video-conferencing language practice; students' own test construction.

More traditional audio-visual support materials in the observed classes included:

- sample listening test cassettes for external exams (seen in use at A2 level)
- jigsaw listening practice cassettes (A2)
- cassettes of chants and stories (professionally recorded as support materials for a textbook) designed for use at elementary level (A1).

In the teacher interview reports, references to IT use included both conventional and newer technologies, and often, as the *PL2000* guidelines suggested, made the connection with other language learning and teaching means, for example:

- 'The computer is a perfect lingua vortex. With a few modifications and administrative systematising, [our computer lab] could become an international model cyber-seminar room' (B2 level teacher at a high school, October 2001).
- 'Setting Internet search tasks and creating CDs with students are useful communicative activities' (teacher at a high school focus group, April 2002).
- 'We try for the optimal use of multimedia, e.g. computers, video, the Internet' (B2 level teacher, April 2002).

Like other language teaching media, of course, the use of information technology needs the appropriate rationale, planning and management. Certain of the uses of computers in the observed lessons seemed somewhat less authentic and practical than the same activity might have been without their use. Nevertheless, most of the IT examples in the observed lessons

suggested positive washback from this area of *PL2000* recommendation and support.

PL2000 impacts and washback on attitudes

The quest for information on the impacts of the *PL2000* on the attitudes of key participants in the project sought answers to one research question in particular, namely: What washback are changes in language teaching and learning pedagogy, materials, and media having on the performance and attitudes of the students and the teachers?

From the students

The nature of our case study students' motivation for learning English was investigated in both the October 2001 and the April 2002 student questionnaires. In October, the students (n=118) responding to an open-ended item 'What are your opinions about learning English?' focused mainly on their *reasons* for learning the language, with:

- 51 references to the status of English as an international language
- 48 to its use in work
- 36 to their 'future' (which, one would assume, may well also be connected with their professions)
- fewer than 10 mentions each for: travel, university study, and a liking for the language or its culture.

Figure 7.7 summarises responses to the April 2002 student questionnaire item, on students' 'main reason for learning English'.

Figure 7.7 Students' main reasons for studying English (from Student Questionnaire, April 2002, n=110)

Reason selected*	Numbers of selections
To get a better job	49
To communicate with more people	48
For travel overseas	37
Because I like it	25

* Some students selected more than one of the reasons given; all selections were counted.

These figures again indicate the perceived importance of the English language as a means of international communication and of enhancing professional opportunity. There may also be an indication of rather more integrative motivations, in the *'because I like it'* and in the overseas travel figures collected on the April 2002 case study school visits.

In their October and April questionnaires, the students were also asked to rate the frequency of their activities in English outside class. Figure 7.8 summarises the responses.

Table 7.8 Summary of Student Questionnaire responses on the use of English outside class.

Out of class Activities	Responses in October 2001				Responses in April 2002			
	Never	Almost never	Occasionally	Often	Never	Almost never	Occasionally	Often
Reading books	43	16	24	13	28	16	55	8
Reading newspapers	38	30	22	2	–	–	–	–
Reading magazines	33	31	14	11	–	–	–	–
Reading newspapers, magazines	–	–	–	–	52	28	21	7
Writing letters	32	36	12	14	11	19	41	27
Radio	38	23	15	20	35	22	23	19
TV	13	26	36	20	17	19	38	16
Movies	40	26	22	8	29	41	26	13
Going to shows	37	27	15	3	57	18	8	3
Using email	22	34	16	18	46	21	22	20
Using the Internet	30	19	13	35	20	25	21	44
Talking with visitors	30	20	27	8	30	30	39	8

Apart from the general picture of relatively limited activity in English outside class for most students, there are some indications from the questionnaire responses on the October 2001 and April 2002 visits, that students were increasingly often participating in English in:

- reading books
- using the Internet, and
- writing letters.

These tentative indications recall a comment by a teacher at interview in April 2002: ‘Students read and write less nowadays in their real lives; language teaching should reflect their actual reading and writing activities’.

More inferences will be made about the students’ motivation and attitudes below, when their perceptions of their progress in English over the 2001/2002 school year are analysed. As suggested in Chapter 1, effects of a programme or test on student (and teacher) attitudes appear to span the washback/impact divide. Perhaps effects on attitudes in learning and teaching contexts belong to washback, while broader attitude changes might

more appropriately be attributed to impact. But the distinctions in this area are by no means certain.

From the teachers

In general, the above case study student data on motivations and attitudes in relation to the *PL2000* were supported by the perceptions of the 11 case study teachers, as expressed in their April 2002 questionnaires, for example:

- More motivation for students and teachers (elementary school French teacher).
- 'More interesting: children and teachers improve their use of the foreign language' (elementary school English teacher).
- 'More motivation, more involved in the lessons, more variety of lexis' (middle school teacher).
- 'We prepare students, in a limited number, of course, good students. They are more stimulated, they are more involved in their activities' (middle school teacher).
- 'Students are gaining motivation, extra hours, interest for languages' (high school teacher).
- 'Students' performance improves especially orally; students show more interest in different languages' (high school teacher).
- 'Students are extremely enthusiastic; their speaking/listening/reading abilities have greatly improved' (middle school teacher).

The case study teachers also appeared positive on the impact of the *PL2000* on themselves. In addition to the apparently enthusiastic ways in which they responded to the methodological approaches implied by the project (see above), the following points from the teacher interviews suggest that *PL2000* had a healthy impact on teacher as well as student attitudes. Note the relationships inferred by the teachers between motivation, clearcut aims and new approaches:

- *PL2000* means that 'teachers know where they are going'; previously FL teaching objectives were vague, the emphasis on the academic aspects of literature, with the basics of language forgotten.
- Teaching is now more rewarding because you can see the results.
- *PL2000* emphasis on communication is appropriate and 'more challenging for the teachers' (all from the high school teachers' focus group, April 2002).
- *PL2000* appeals even to language teachers not involved in it; some teachers talk of backwash from *PL2000* to non-*PL2000* courses (middle school English teachers, October 2001).

There were, of course, extra pressures from the *PL2000,* as from any such radical educational reform. The following three less positive comments from the teacher interviews are indications of this, again to be taken up below.

- Language teachers are sometimes de-motivated by the extra administrative work entailed by *PL2000*.
- There may be resistance from other subject teachers to the scheduling of *PL2000* afternoon classes (high school teachers' focus group, October 2001).
- Difficult to motivate all the students in a mixed-ability class; 'It is important for a teacher to stay with the same students over more than a year' (middle school teacher, April 2002).

The overall impression conveyed by the *PL2000* teachers contacted through our study was that they were well motivated by the *Progetto* and keen to pursue its aims and processes in a professional way. It should be mentioned here that the attitudes of the case study teachers to the *PL2000* Impact Study itself was unfailingly positive and helpful.

From the heads and the parents

This section moves the data analysis and findings on to the next research question, namely: What impacts is the *PL2000* having on educational managers, including heads of schools? Given that case study school heads referred regularly to the attitudes of parents to the *Progetto*, these are also covered here. Note also the evidence of developing views as the interviews enable the comparison of opinions over the longitudinal span of the *PL* Impact Study, across the March 2001 (pilot visit), October 2001 and April 2002 interviews.

The heads of the seven *PL2000* Impact Study case study schools took an appropriately broad view of the significance of the *PL2000,* with a focus on the project's impact as well as its washback, to use the terms as defined in Chapter 1. The summaries of typical comments here are taken from 10 interviews with the case study school heads, conducted in Italian, with simultaneous interpretation in English and subsequent notes made using the video-recordings and discussions with a bilingual interpreter.

- Very positive [impact] since the project highlights the EU and the need for foreign languages for communication, and an increase in students' enthusiasm, pride in their foreign languages, willingness to interact in them, and understanding of their importance (elementary school head, October 2001).
- Foreign languages are important as borders and barriers have fallen and communication is facilitated between European and non-European

countries. The students have more opportunities of communicating; a knowledge of foreign languages is fundamentally important for students completing secondary education as well as for continuing education at higher levels.

- The impact of *PL2000* has been extremely positive, participation in it being necessary to reinforce the position of the school locally and more widely.
- *PL2000* also helps the school to achieve a greater degree of openness.
- The parents seem happy and understand the objectives of *PL2000* language teaching (all four from the head of a comprehensive middle school, October 2001).
- *PL2000* is in line with pressure from parents who want to enrich the students' educational experience.
- *PL2000* impact is positive in terms of motivation for the students. Thus it is good for the school, but time is needed before final judgement can be made (from the head of a second middle school, October 2001).
- There are now more students (and parents) wanting to participate in *PL2000* courses, numbers rising from 31 to 50 over the year (same head, April 2002).
- Teachers are enthusiastic, prepared to work beyond the objectives the school sets; the school's and *PL2000*'s objectives are the same! (head of a technical institute, March 2001).
- The training has improved teachers' English, given them different kinds of English (for different purposes), brought conferences, films, discussions and debates in English, made them better teachers. So the *PL* has given the teachers the opportunity to broaden their knowledge, use better materials, made them more aware as teachers (same head April, 2002).
- *PL* is accepted at all school levels in this region. The students are now used to *PL* methods, so progress quickly (high school head, October 2001).
- The *PL* is great from the teacher's point of view given the training and satisfaction they get out of this kind of project. School will have conferences and meetings open to locals, to discuss projects of this kind and get suggestions from them (same head, April 2002).

The heads also refer to some typical difficulties with the *PL2000*, as an educational reform:

- Good parent/school relations, but, of course, it is the most motivated families who send students to the afternoon language learning activities, about 10–15%. Not necessarily the best students, but the mid to high ability ones. The students really needing help do not come; their families tend not to trust the schools (head, middle school, October 2001).

- The school is determined to expand and enrich the *PL2000* programmes in English and French to attract students (and parents) who are not so highly motivated as the initial groups (same head, April 2002).

The *PL2000* Impact Study held interviews or focus groups involving 21 parents with children at elementary, middle and high school levels. The parents concerned spoke both of the impacts of the *PL2000* on their children and on themselves. Their comments were overwhelmingly positive, which might, of course, be expected since it was more likely that parents with positive views would be called upon and accept invitations from the school heads to be interviewed for the Impact Study. Parents' comments focused mainly on the way English is or should be taught and learned, or on the broad values of competence in the language and on English as a tool for children's academic or working future.

PL2000 impacts on assessment

From the students, the teachers and the classroom

This section responds to research question 3, namely, What washback is the *PL2000* having on language evaluation and assessment?

One of the key *PL2000* innovations was that '[i]n addition to the internal evaluation which is an obligatory part of the school system, the project will permit the introduction of external certification by virtue of a protocol of intent ratified by the Ministry and examination boards recognised internationally in the four languages studied (English, French, German and Spanish)'. The rationale for this radical change was that '[e]xternal certification represents an added value in quality of 'transparent credit' which may facilitate the re-orientation of students in their educational programmes (the transition from one academic level to another or from a formal system of instruction to professional education) constituting a negotiable certification in the professional world and across national borders'. (Ministry of Education, Italy, 1999, translated.)

The *PL2000* Impact Study collected evidence and views on the impact of external (and internal) exams, and teaching/testing relationships.

The Student Questionnaire returns in April 2002 are fairly clearcut on how students felt about external exams, as Figure 7.9 indicates, a strong majority of the case study class students suggesting that the prospect of external exams at the end of their *PL2000* courses motivated them to work harder.

Figure 7.9 Summary Student Questionnaire responses on attitudes to external exams

*My main feeling about taking **external** exams in English is they*			
are more difficult than school exams	help me to work harder	are for listening and speaking	are necessary to enter university
12	54	13	16

Responding to their questionnaire item on the achievement at their schools of the *PL2000* objective of the 'use of external language exams at recognised Council of Europe levels', 10 of the 11 case study teachers rated this objective as *'very well'* (5) or '*well achieved'* (5).

The case study class observations raise the question of the relationship between external exams and the teaching/learning activities that occurred in courses at the end of which an external language was to be taken, in other words external exam washback (see Chapter 1 and the discussion of the issue in connection with the IELTS test in Chapter 6).

The case study classroom observations may help us answer this complex question. The communicative tasks and activities in a lesson taken by a *PL2000* class of A2 level middle school students in April 2002 are summarised in Figure 7.10 (page 152). The lesson described here using the form of classroom observation analysis discussed in Chapter 4, is typical of *PL2000* lessons implemented with a clear awareness of the aims of both the *PL2000* and the target external exam, yet with some resistance to too direct a relationship between the test and the learning.

The target exam of the course was the Cambridge ESOL Key English Test (KET) a Common European Framework (CEF) level A2 exam, at the upper of the two 'Basic User' levels. In the CEF global scales, A2 describes a user who can 'communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters' (2001:24). The KET exam consists of three papers: Reading and Writing; Listening; and Speaking. The KET Speaking test (the closest to the activities analysed in the observed lesson summarised in Figure 7.10) assesses candidates' ability to interact in conversational English with an examiner and with another candidate. Students need to be able to answer and ask questions about themselves and talk freely about their likes and dislikes. The Speaking test is normally taken in pairs (www.cambridgeesol.org/exams).

Figure 7.10 *PL2000* Impact Study lesson taken by A2 students in April 2002; observation analysis

No.	Episode	Timing (mins)	Activity, participation, materials	Comment
1	Questionnaire completion practice: task preparation	7	Pre-prepared worksheets given out. Teacher (T) explains the questionnaire form and format, and how to use for pair 'survey' work. Examples given of A and B student sheets, A a questionnaire sheet on reading habits, B a self-evaluation questionnaire on the same topic.	T uses target language (TL) for classroom language, with a little L1. T encourages students to use TL.
		3	Students study sheets silently. T monitors and helps.	
2	Pair interviews	10	Students work in pairs, A students asking Qs and completing questionnaires, B students answering questions on their reading habits.	Students use L1 a lot in pairs. Quite a lot of silences. T uses TL mainly when monitoring. Some good pair interactions.
3	Report-writing	2	T explains report-writing task: A students to write report of questionnaire findings, B students to write a self-evaluation of own reading habits.	T has to explain quite basic words e.g. 'good'.
		8	Report-writing begins (completing pre-prepared information frames). T monitors.	
4	Report back	9	A students read their reports aloud to class. B students read out their self-evaluations. T adds questions to individual reporting students.	6 students get to present.
5	T >< student Oral question and answer	4	T asks questions of individual students re. the questionnaires and self-evaluations.	Most students questioned.
Students seem rather more confident than in October 2001? T has positive and sympathetic attitude. Perhaps helps and intervenes too much. Not much intensive student communication?				

The relationship between the CEF level, the KET exam and the A2 class lesson would seem *positive* here. The learners were participating in several communicative activities, listening, speaking, writing and reading, which they most probably considered relevant to their own communicative needs and life aspirations. These activities also coincided with their CEF target level and the requirements of the external exam due at the end of their course. But the exam was not mentioned, and the materials for the lesson were by no means just practice tests.

Yet *PL2000* Impact Study data do also indicate a tendency for external exams to be regarded as, in a sense, the syllabus and the target criterion for a *PL2000* course. It was commonplace during the study to hear school heads, teachers and students refer to 'the PET (KET or FCE) course'; or to see curriculum specifications expressed in exam preparation terms. Making sure that students are prepared for the formats and conventions of external exams is, of course, an important and natural part of a course leading to an exam, as discussed with reference to IELTS above. School language courses that are simply exam cramming, however, causing 'curricular alignment', that is a narrowing of curricula to fit closely and exclusively with the demands of a test (see Hamp-Lyons in Chapter 1), would not be in the spirit of the needs and task-based communicative approaches suggested by the *PL2000*. It was encouraging in this respect that 17 out of 20 of our observed lessons were not explicitly exam-oriented, but appeared rather to be attempting to involve the learners in communicative activities appropriate to their English language levels, needs and wants. A middle school teacher noted in his April 2002 interview that the washback of a test is positive if the test and the teaching have common aims. But the overlap relationship can be complex, we already know. Teachers' perceptions of the route to the achievement of teaching aims may vary.

Balancing the demands of an exam with the communicative needs of learners remains a key washback issue, one on which the data from our teachers and head of school interviews revealed further insights, many of which could inform teacher professional support programmes.

At the teacher interviews and focus groups, the issue of external exams was raised frequently, often with regard to the status and effect of external and other exams, and to the teaching/testing relationship. The responses typified here reveal a mature and generally positive view of external foreign language exams:

- 'Teachers were surprised at first, at the idea of external certification. They didn't think it suited their teaching. But external certification is important as it means that students can be checked for their level of competence at any stage, from A1 to C2. Teachers are used to the idea now.' (Teacher of German and *PL2000* official, February 2001)

From the teacher and her English first language colleague teaching at C1 level towards the Cambridge ESOL Certificate in Advanced English (CAE) exam, comments (in October 2001) bring a useful perspective to the teaching/exam relationship question:

- Exams lead in the right direction, use lively, up-to-date everyday language, in line with *PL2000*.
- 'External certification increases student motivation.'

But, at the same time cautioning:

- Course targets and lengths are set according to the final, external test.
- 'Students know the exam aims better than *PL2000* aims.'

A case study teacher at a middle school, interviewed twice, expressed the following enlightened views on external exam impact and the teaching/testing relationship:

- There is a close teaching/testing relationship but teaching should never be 'just for the test'.
- External exams are an important motivation through their international credibility but should not be emphasised too much.
- A test must not be a threat to the students (October 2001).

At the end of the school year, in April 2001, the same teacher had this to say:

- Don't teach just for a test; get 'the test objectives and the student learning objectives to work together' in a programme; an external test should be 'motivation but not a threat'.
- Where a test represents real-life communication, it is good; this is not always the case, though, e.g. tests with tasks from contexts (e.g. road signs) students are not familiar with … Target language letter-writing may not be a realistic communicative activity for students who write only notes in real life.
- Exam preparation is important but teachers are given the freedom to approach this in their own ways.

Then, accepting, perhaps, the added validity and reliability that should be expected from external certification, the teacher added:

- It is harder for teachers to evaluate students 'scientifically' with large classes, mixed abilities and some less able students.

A case study high school teacher responsible for organising the *PL2000* programme at her school wrote, in March 2002, of the *instrumental* motivation for external exams that had to be taken into account:

- 'Parents encourage students to take external certification, as giving useful credits at the end of the year and facilitating university entrance or English course exemption.'

In April 2002, the same teacher, in her interview, raised the important question of the *other* tests on *PL2000* courses with an external final exam:

- Placement, progress and mock tests are important on a course leading to an external exam.
- Next year it is planned to make the external tests less of an end in themselves; 'remember that it takes time for the levels of the CEF to have their full influence on planning and teaching'.

A teacher at the same school agreed, in March 2002, with the use of external exams:

- 'Earning a benchmark of proficiency gives students something to strive for and motivates language acquisition.'

From the classroom observations, the students and the teachers, we gained important insights into the influences of external exams on learning and teaching. The message so far appeared to be that external certification was welcomed by the learners and their teachers, and that much thought was going into ensuring that the washback was positive because the right teaching/testing relationships were being developed. Implied in the comments cited above, as in nearly all the discussions of *PL2000* aims and their achievement, was a good fit between *PL2000* English language communicative targets, the Common European Framework levels and the external exams. *PL2000* levels were, it will be recalled, expressed in explicitly CEF level terms. And Cambridge ESOL would claim both theory-based and empirical evidence for 'a coherent link' (Taylor 2004:2) between their Main Suite of exams and the CEF, as established through the ALTE Can Do Project (Jones and Hirtzel 2001).

Growing interest in 2002 in the European Language Portfolio (ELP), which, with its 'format in which language learning and intercultural experience of the most diverse kinds can be recorded and formally recognised' (Council of Europe 2001:5) was consolidating the links between *PL2000*/CEF levels and the recognised external exams. Chapter 8 reports further on this in the aftermath of the *PL2000.*

From the school heads and the parents

For the Heads of the Impact Study schools, the selection of exams from the approved external examination boards had implications, both educational and managerial, of course. From our ten interviews with the heads, we gained insights into both.

On the March 2001 *PL2000* Impact Study pilot visit to his high school, one head provided the study with a thought-provoking view of testing. He was, he said, 'trying to teach students to view the exams as a kind of research into their own knowledge and ability, a sort of self-evaluation rather than a check linked to promotion or other rewards'. The case study school heads also commented on the positive effect of good Cambridge ESOL exam results on performance in state exams, the apparent improved student confidence and motivation resulting from their taking external certification through *PL2000*.

The focus of almost all the comments on external exams from our 11 interview and focus groups with 21 *parents* was on the greater international recognition of such exams. These were seen as '*good for their work in Italy and abroad',* or offering *'wider work opportunities in the EU and beyond'*, to cite typical parent interview responses.

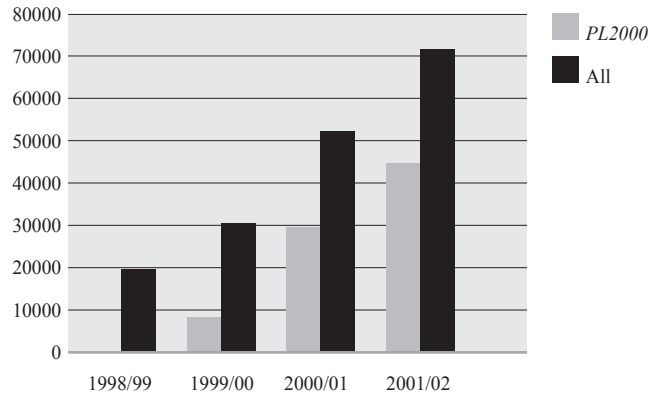
PL2000 washback on exam candidacies and performance

We now return to the research question: What washback are changes in language teaching and learning pedagogy, materials, and media having on the performance and attitudes of the students and the teachers? The focus here is evidence of impact on performance from both a broad and a case study specific perspective.

Exam candidacies

It is clear that the *PL2000,* with its advocacy of the use of the tests of approved examination boards to provide the added value of internationally recognised certification to students' foreign language qualifications, was likely to increase the number of external foreign language exams taken in Italy. Figure 7.11, giving official Cambridge ESOL *PL2000* exam and total candidate numbers from 1998/99 to 2001/02, showed this indeed to be the case.

Figure 7.11 *PL2000*: all Cambridge exam entries 1998/2002



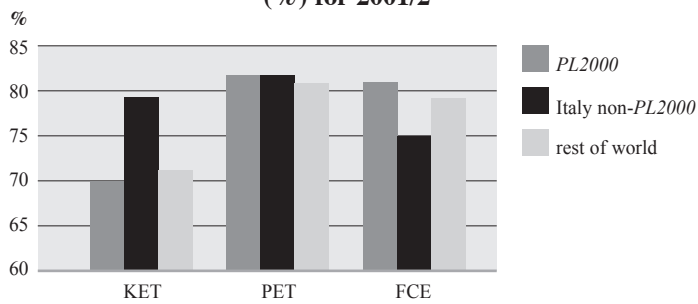
Cambridge exam candidacies for Young Learners English Tests (YLE), KET, PET, FCE, CAE and CPE, rose by 54% in the 1999/2000 academic year, by 71% in 2000/01 and by 36% in 2001/02, to a 2002 total of 71,953. In 2002, *PL2000* entries accounted for 68% of entries at A1 level, 81% at A2, 64% at B1, the level with the most Cambridge exam candidates, and 49% at B2 level. The *PL2000* candidatures at C1 and C2 levels were still few, presumably because most school age students did not take exams at that high CEF level

(although several students in our case study group did take the CAE exam at C1 level).

Exam performances

It should eventually be possible to see how learning English on *PL2000* courses actually affected student exam results in the long term. The *PL2000*, which ended as a formal Ministry project in 2003, certainly provided interesting insights into learning processes and attitudes as students moved from one level of foreign language proficiency towards another. In Figure 7.12, the pass rates at national level are compared across the main Cambridge exams for *PL2000*, Italy non-*PL2000* candidates, and for candidates from the rest of the world, that is countries other than Italy where the Cambridge exams are taken. The figures indicate the kinds of patterns of change that might be expected at the early stages of educational reform programmes.

Figure 7.12 *PL2000*, Italy non-*PL2000* and rest of the world pass rates (%) for 2001/2



At A2 level, the pass rate for *PL2000* candidates in the 2001/2002 academic year was 70%, just below the rest of the world average (71%), but not yet as high as the A2 Italy non-*PL2000* figure of 79%. This latter figure may be high because it embraces a wider A2 level age range and includes more learners taking private tuition. Checking the *PL* vs. non-*PL* pass rates for the two preceding years (1999/2000 and 2000/01) we note that the pattern was the same, with *PL2000* KET (A2) pass rates below both the Italy non-*PL* and rest of the world figures. In fact, the 2001/02 rates were a little more encouraging as the gap between *PL2000* and the rest of the world had narrowed to only 1%.

At B1 level, the *PL2000* pass rate was 82%, a little higher than the world average (81%) and the same as Italy non-*PL*, thus an improvement on 2000/01. The FCE results for 2001/02 showed a higher pass rate for *PL2000*

students than non-*PL* or the rest of the world, this for the third year running. The CAE (*PL2000* 126, all Italy 1686) and CPE (*PL2000* 3, all Italy 741) results are not included in Figure 7.13 because the *PL2000* numbers are too low for valid national or international comparisons to be made.

The table in Figure 7.13 presents performance data for the 60 students in our case study group who had already taken an external Cambridge exam at the end of the academic year during which we had become acquainted with them. Thirty-one from the case study schools, had moved on from A1 to A2 level to take KET; 25 had moved from B1 level (PET) to B2, to take FCE, and six had advanced from B2 level to take CAE at C1 level. The pass rate patterns were encouraging, although it is no surprise at all to note that they vary according to school. QPT data (see Chapters 3,4,5 above) tended to confirm Impact Study students at the CEF levels at which their schools had placed them.

Figure 7.13 Levels and external exam performance in the *PL* Impact Study case study group

ID	School cycle	Start level	End exam level	Grades				
				Numbers of candidates				
				Merit	Pass	Narrow Fail	Fail	
A	Middle	YLE (A1)	KET (A2)	3	14	2	4	
B	Middle	YLE (A1)	KET (A2)	1	1	2	3	
				A	B	C	D	E
				Pass			Fail	
C	High	PET (B1)	FCE (B2)	6	11	–	–	–
D	High	PET (B1)	FCE (B2)	–	–	7	–	1
E	High	FCE (B2)	CAE (C1)	–	–	2	1	2

PL2000 impacts on management and support

This section covers two further *PL* Impact Study research questions, namely: What impacts is the *PL2000* having on educational managers, including heads of schools? and What impacts is the *PL2000* having on support for teacher in-service and test-awareness programmes, resource centre development, and textbook design?

We have heard already, in the quotations from the school head interviews, of several of the *PL2000* impacts on management. These included (see previous): the added motivation the Project appeared to bring to students, teachers and parents; the additional professional training and resource support it provided for language teachers; and the increased importance it brought to some schools as regional language teaching resource centres.

One of the principles of the *PL2000* which was mainly administrative, and to which we have not yet referred was ‘the division of classes into learning

groups of 12–15 students homogeneous in terms of initial competence, interests or shared specific educational needs' (Ministry of Education, Italy 1999). Most of the classes we saw which were specifically under the *PL2000*, and thus, as was often the case, held in the afternoons, were of groups of around the indicated number of students. We saw such classes in one of the middle schools, both the high schools and at the technical institute. There was evidence, too, that the schools were taking advantage of the *PL2000*'s 'specification of a number of hours over several years for every school cycle, thus a choice of level by individual schools of the number of programme hours per year and of the modality of such programmes (extensive, intensive or mixed mode)' (ibid).

As far as the homogeneity of the groups was concerned, management and teachers were aware of the relevant *PL2000* objective on group size. However, this was sometimes irrelevant, given the relatively small overall numbers involved (30 students *altogethe*r requesting *PL2000* afternoon classes at one of the middle schools, 18 at the beginning of the year at a high school, dwindling to fewer than 10 by the end). Yet most of the *PL2000* groups we saw seemed reasonably homogeneous in terms of English language competence level, and attempts were being made to sort and place students at the beginning of the school year.

One of the case study high school teachers, who was also the school's *PL2000* course organiser, described her school's *PL2000* course structure thus:

> … short early afternoon courses (20–30 hrs.) to offer classes to a larger number of students. More or less homogenous groups of about 10–15 students per class, according to CEF parameters.

The same teacher, like two others in the case study group, referred to the possible conflict between the idea of groups selected for homogeneity of level of competence and the comprehensive school principle:

> I had also thought of organising curricular classes [i.e non-*PL2000*] dividing students according to their initial language competence (this seems to be one of the aims of *PL2000*), but this seems very difficult to do in Italian schools, where class groups are based by law on the union of mixed ability students. When you cannot always have homogeneous groups, you have to plan to cope with mixed abilities and give extra teaching for the weaker students.

The additional work load of teachers organising *PL2000* programmes could clearly be quite heavy. One of our case study teachers suggests that a 'major defect of *PL2000* courses is both administrative organization and coordination which all rely on one person, already engaged in full time work'. She described graphically the work of a *PL2000* official:

> The responsibilities of *PL2000 referenti* [officials] are considerable, including: the administration of *PL* course schedules, placements and certification; the collection and analyses of performance data; the development and use of resource centres.

One of the reasons why such efforts were worthwhile, of course, was the Ministry funding available for the running of *PL2000* courses, for teacher professional support, and for student exam enrolments.

Teacher professsional support

Returning to our research questions, we are reminded that impact on the various possible forms of support for *PL2000* teachers was included, in the question: What washback and impact is the *PL2000* having on support for teacher in-service and test-awareness programmes, resource centre development, and textbook design?

A good range of information and views on professional support emerge from the *PL2000* Impact Study data, first of all from the teacher interviews and focus groups:

- The *PL2000* regional resource centre does a lot of planning, holds teachers' meetings, distributes materials, liaises with other centres, takes part in Europrojects, and has a website (*PL2000* official and teacher of German at a technical institute, March 2001).
- Teacher meetings help through mutual advice, problem-solving as a team, the selection and use of equipment, books, test development and use. There is no peer observation, though.
- A young teacher learns little by little, should steal ideas from more experienced colleagues, and make use of all facilities and support (both from a new middle school teacher, April 2002).
- *PL2000* has led to improvements in IT facilities and more, though not enough, in-service training for teachers (two high school teachers, October 2001).
- Preparing tests is time consuming and needs special professional competence; Italian teachers are always preparing tests, but should undergo more training (high school teacher, written communication, March 2002).
- The project is useful for teachers because it offers the chance of extra professional training and qualifications (technical institute head, April 2002).
- The teachers, and Cambridge, have worked hard for the *PL*. Some new teachers, weaker in English language level and culture resisted a bit, but

> once they were trained and once they understood why the school was supporting the *PL*, they were more positive (high school head, October 2001).

The additional need for professional support for teachers required by the *PL2000* also had an impact on the selected examination boards. In the UCLES *Progetto Lingue Full Report* for January 2000–2001, the establishment of the *Progetto* is described as having led to a change in both the focus and the scale of the Cambridge seminar programme to *'reflect the needs of state-school teachers'*. To meet the increased demand for seminars in support of the *PL2000*, Cambridge ESOL set up a team of 25 experienced and trained presenters across Italy. Print and web format Cambridge support materials have also been developed in Italian for teachers, parents and students. Cambridge ESOL also appointed a part time *Progetto Lingue 2000* Co-ordinator, based in Rome.

A fuller account of the *PL2000* Impact Study data and analyses appears in Hawkey (2003a).

Chapters 6 and 7 have presented the main findings of the IELTS and *PL2000* impact studies. The two chapters should serve both to exemplify the implementation and outcomes of such studies, and to convey their main impact messages on the IELTS test and the *PL2000.*

8 Related developments, further research and lessons learned

Chapters 6 and 7 have presented key messages from IELTS and *PL2000* impact study data. These messages represented some of the answers to the research questions presented in Chapter 3 above, using the instruments described in Chapter 4, delivered through the systems outlined in Chapter 5. Chapters 3–7 all deal with the constructs and approaches to their investigation which were discussed in Chapters 1 and 2.

Now in Chapter 8, the intention is to reach some kind of closure by:

- tracing further developments, research and planning related to the two studies
- suggesting some lessons to be learned, in terms of impact study development, validation, implementation and models.

IELTS impact study developments, research and planning

Research implied in previous IELTS impact study phases

IELTS impact study Phases 1, 2 and 3 as described in this book were neither conceived nor implemented as an autonomous, monolithic inquiry, but rather as part of a system of continuing and iterative research intended to help ensure the test's continuing validity. The study investigated the washback of the test on learning and learners, teaching and teachers, and the impact of the test on stakeholders beyond the teaching/learning context including educational administrators, candidate selectors, testers, writers, decision-makers. As the analysis in Chapter 6 indicates, data from this phase of IELTS impact study help us to understand and consider means of improving language teaching, learning, methods, materials, activities and attitudes with regard to IELTS. In addition to insights into improved test washback elements such as these, our study of test impact has helped us to understand reactions to and recommendations on IELTS test administration and candidate selection.

True to the acknowledged iterative nature of high-stakes impact research, and as indicated in the original project design for IELTS impact studies

described in Chapter 4, there were certain focus areas in the original long term research design (see Chapters 3 and 4) which are still to be covered:

- pursuit through receiving institute questionnaire of IELTS impact on *administrators*; this area of inquiry has so far been handled mainly through face-to-face rather than questionnaire data
- additional inquiry, perhaps using the trial instrumentation already developed, into receiving institution *subject lecturer* views on IELTS
- iterative research into the existing IELTS impact study research questions but with *General Training Module* takers; the Phase 3 data were primarily from Academic Module test takers.

Other key targets for new research could be IELTS-related studies linking test impact and test performance, taking as criteria for test washback measurement matters raised by the findings of the impact study phases described in this book (see Chapters 3, 4 and 6), for example:

- given criticisms in the study of the limited time available for the IELTS Reading module, exploring, through sample test-task scores under different timings, relationships between performance and test time
- given evidence of the need for further investigation of the validity of IELTS reading and writing tasks, research involving observation and/or the views of receiving institution subject lecturers, into the nature of reading and writing activities on a range of higher education courses. This could be linked, perhaps, to
  - an investigation of candidate reading and writing performances on different reading and writing test tasks
  - further investigation of candidate test and target domain performance in relationship to their test-preparation courses (see Green, 2003 on writing) across the four skills.

Role of the Research and Validation Group

Such further research under the IELTS impact study would feed into the routine monitoring and evaluation of the IELTS undertaken by the Cambridge ESOL Research and Validation Group, whose systems cover analyses and research projects for all Cambridge ESOL exams, including IELTS. It will be recalled, from Chapter 4, that it was from this group that initiatives for the IELTS impact studies first came. The group's role is, after all, to provide quality assurance for Cambridge ESOL exams at every stage of the assessment process through routine operational analyses of the administration cycle of all exams, i.e. exam production, exam conduct, marking/grading, and post-exam evaluation (see Chapter 1). The Group continues to support instrumental

research through small-scale projects designed to inform operational activities and through longer-term language assessment research such as the phased IELTS impact study, to meet validation targets relevant to Cambridge ESOL's objectives and future developments. Detailed research and analysis of both test material and test takers' performance are carried out to ensure that not only does IELTS provide accurate information for the institutions that recognise it, but that the tests are fair to test takers whatever their nationality, first language and gender.

The IELTS funded-research programme and *Research Notes*

Areas of interest for external research purposes under the joint-funding and support of IELTS Australia and the British Council, and co-ordinated with the Cambridge ESOL framework for research and validation were identified by the IELTS Research Committee and summarised in *Research Notes* 16 (May 2004) as follows:

- the use of IELTS and IELTS scores
- IELTS and its impact on the teaching/learning environment
- perceptions of and attitudes to IELTS among stakeholder groups
- individuals and small groups taking/using IELTS (2004:22).

Chapter 1 referred to IELTS-related research activities rounds 1–7 of the funded IELTS research agenda between 1995 and 2001 (see *Research Notes* 8, May 2002), a significant proportion of which covered test impact issues. The research topics *since* rounds 1–7 of the funded research programme, indicate a continuing washback and impact interest. Recent topics with this focus, and indeed a shared focus on areas also investigated to a greater or lesser extent in the study analysed in this volume, have included (see rounds 8–10, 2002–2004):

- candidate identity, learning and performance, with specific reference to the affective and academic impact of IELTS on successful IELTS students (Rea-Dickins, Kiely and Yu)
- the relationships between IELTS and preparation courses for it, including ethnographic study of classroom instruction (Mickan), and the relationship between teacher background and classroom instruction on an IELTS preparation programme (Swan and Gibson)
- the impact of IELTS on receiving institutions, for example, tertiary decision-maker attitudes to English language tests (Smith and Haslett); the use of IELTS for university selection (O'Loughlin), and IELTS as a predictor of academic language performance (Ingram, Bayliss and Paul)
- perceptions on the IELTS skills modules, for example the Speaking test (Brown; Hampton and Huang Chung; O'Sullivan, Weir and Horai;

O'Sullivan and Yang Lu; Read and Nation; Seedhouse and Egbert) and the Writing test, features of written language production (Franceschina and Banerjee), impact of computer vs pen-and-paper versions (Weir and O'Sullivan).

All these research topics focus on the effects of the tests on stakeholders, on their attitudes or on their actions, or on the complex of variables (for example, here, the backgrounds of preparation course teachers) intervening between an aspect of the test and one of its impacts (see Chapter 1).

It is the aim of the UCLES/Cambridge University Press *Studies in Language Testing* (SILT) series, to which this volume also belongs, to disseminate knowledge, research and experience with regard to test development, validation and use. SILT volume 19 *IELTS Collected Papers* (Taylor and Falvey, forthcoming) gathers together a set of 10 IELTS-funded research projects on the Speaking and the Writing tests. As well as evaluating the wide range of methodologies used in these projects, the volume reviews the importance of their research findings on the Speaking and Writing modules of the IELTS test in the past five years. In Part 1 of the volume (on the IELTS Speaking module) the research studies investigate impact relationships such as interviewer styles and gender, with candidate performance (Brown; Brown and Hill; Merrylees; O'Loughlin). The Part 2 projects investigate relationships between target domain and IELTS academic writing tasks (Moore and Morton); training and rater judgments (Rignall and Furneaux); quantity and presentation and performance (O'Loughlin and Wigglesworth), bias analysis feedback and rater performance (O'Sullivan and Rignall) and analyses of linguistic output (Kennedy and Thorp; Mayor, Hewings and Swann).

The messages from such a volume emphasise:

- the importance of disseminating impact-related studies on a high-stakes test such as IELTS in the context of continuing and iterative research
- the close relationships of broader, multi-faceted research such as that described in this volume, with smaller-scale studies such as those supported by the IELTS funded-research initiative, and the need to integrate the two as far as possible
- the superordinate role of the routine test validation work carried out by Cambridge ESOL with its focus on test taker *performance* across the whole IELTS population, and the desirability of relating all other IELTS impact study systematically to this.

We can already see, from the references here (and in preceding chapters) the important role played by Cambridge ESOL's *Research Notes* (www.cambridgeesol.org/rs_notes) in the dissemination of test development,

administration and validation information. The November 2004 issue of the journal, for example, which was devoted to IELTS, articles covered: IELTS : CEF level comparability; paper- and computer-based versions; the suitability of the General Training (GT) modules for 16–17-year-olds; Writing test revision Phase 4 (see Chapter 4); frequently asked questions, and test performance data for 2003.

Several points of relevance to the discussions in this book arise from the selection of articles in the issue. The paper on the IELTS GT module is headed *IELTS Impact: a study on the accessibility of IELTS GT Modules to 16–17 year old candidates* (Smith 2004). The study 'required a comparison between materials used in classes preparing students for entry to upper secondary and those used in the General Training Modules' (2004:6). The instrument used to pursue this purpose was the Instrument for the Analysis of Textbook Materials (IATM, see Chapters 3, 4 and 6), developed for and used in IELTS impact study Phase 3. It will be also recalled from Chapter 4 that the instrumentation developed in Phases 1 and 2 of the study, was intended for use in as many future studies as they were felt appropriate for.

The *Research Notes* 18, 2004 paper on IELTS performance data (Blackhurst 2004:18–21) presents mean module score data for the whole 2003 IELTS population and for a range of sub-populations (for example by Academic and GT module, frequent language backgrounds, nationalities). This is precisely the kind of routine validation-oriented performance analysis that is best handled by the Research and Validation Group rather than by smaller scale studies. This point was made in Chapter 6, it will be recalled, to explain why the various aspects of IELTS washback on candidate *performance* was not a major objective of Phase 3 of the IELTS study, with its test taker sample of 572 participants. But medium and smaller-scale impact studies may be useful in following up particular leads provided by the large-scale data. Green, for example, reports in *Research Notes* 16 (2004) on studies of score gains on the IELTS Writing test (see also below).

IELTS website

The IELTS website (www.ielts.org) contains details of and links to both current and completed IELTS research awards, and summaries of the most recent candidate and test data. The stated aim of the IELTS partners, in keeping with the high-stakes nature of the test, is 'to provide appropriate public access to both comprehensive test data and research findings'.

The very significant increase in IELTS candidature over the past few years (see Chapter 6) reflects the increase in recognition of IELTS as 'a benchmarking tool for English language proficiency' (www.ielts.org:1). To date, education institutions, professional bodies, government departments and

employers in approximately 40 countries use the test as one of their entry requirements. The last few years have seen a parallel increase in demand for IELTS preparation services. As a result of this development, the IELTS test partners, the British Council, IDP: IELTS Australia and the University of Cambridge ESOL Examinations, have introduced a Website designed for teachers of IELTS preparation classes. This includes the following features:

- sample tasks
- tips for teachers and students
- dos and don'ts
- ready-made activities for use in preparation classes
- skills' development activities.

The contents and approaches of this site may be seen as a further example of the strength of test washback. The test is influencing the website directly and the website is likely to have washback on its users, mainly teachers and test takers, and on IELTS preparation courses.

The findings of Phase 3 of the IELTS impact study have already been disseminated at various conferences and seminars. These include IATEFL annual conferences, York (March 2002) and Liverpool (April 2004); a Cambridge ESOL staff seminar (July 2004); the EFL Testing and Evaluation Conference, Athens (October 2004); the 18th IDP International Conference in Sydney (October 2004); the British Council *Going Global* Conference, Edinburgh, (December 2004), and the ALTE second International Conference, Berlin, (May 2005). Publications have included Saville and Hawkey (2004), Hawkey (2001, 2003b).

PL2000: further impact research and related developments

Research pursuing issues from the *PL2000* Impact Study

The *PL2000* Impact Study described had always been envisaged as part of continuing research into the effects of changes in foreign language teaching, learning and testing in Italy, with particular reference to Cambridge ESOL exams. As with research into the impacts of the IELTS test, impact is seen here as part of Cambridge ESOL's four-element test development focus, to optimise, continuously and iteratively, the validity, reliability, impact and practicality (VRIP), in fact the usefulness, of its tests. The *PL2000* Impact Study, like IELTS impact study Phase 3, has thus led into *related* studies.

The Florence Language Learning Gain Project

In order to attempt to investigate further the vital yet elusive matter of high-stakes test washback on language *performance,* language learning *growth and change*, Cambridge ESOL, in co-operation with the British Institute of Florence, initiated the Florence Language Learning Gain Project (FLLGP). This second-phase study took as its sample populations English language learners in young learner and adult classes at the British Institute. Some of the participants had the Cambridge ESOL Preliminary English Test (PET) at B1 level, others the First Certificate in English (FCE) at B2 level as their target exams. Some had no immediate external examination target.

Participant learner groups and individuals would be compared, in terms of variables such as competence level, age, stage, motivation, high-stakes or internal final exams, and language gain. All groups were pre-tested using the *Communicat* English language placement test*,* a computer adaptive language test designed to provide prompt and economical assessment across language skill areas, and available in six languages. The participant learners were also pre-profiled through the Language Learning Questionnaire (LLQ), 'Can Do' statements in Italian, and Student and Teacher Questionnaires adapted from the parallel instruments used in the *PL2000* Impact Study. Teacher interview and log data, and classroom video analyses at the beginning, in the middle and at the end of the study year (2002–2003), samples of English language work throughout the year, and final exam (external and internal) performance data, are being analysed for insights into factors affecting language learning and performance. The issue of learner motivations, in their multi-faceted and dynamic forms, are a key focus of the research. The data analyses are being conducted in collaboration with Professor Jim Purpura and doctoral student Michael Perrone of the Teachers College, Columbia University, New York.

The Siena University–Cambridge ESOL collaboration project

The *PL2000* did not, of course, cover higher educational institutions, nor, therefore, did the Impact Study. However, it is intended that students who studied some of their English under the *PL2000* should be a focus of collaborative research between Cambridge ESOL and the University of Siena Language Centre (*Centro Linguistico dell'Ateneo (CLA))*. The focus of such research would again be learning gain and performance data collection, related to the question of which CEF level and which exams are appropriate for students learning and using English as a foreign language during their university studies. Areas for possible research questions in the collaborative study, especially in the light of the increasing acknowledgement of the impact of the *Common European Framework of Reference for Languages: learning, teaching and assessment* (2001) (see previous), could be:

- How do CEF aims compare with school and university views of foreign language needs and target levels? Are foreign language target proficiency levels changing for various learner groups and cohorts?
- Is CEF Level B1 an appropriate achievement level for students by the time they leave school or enter university? Should universities turn away those who have not reached B1? What is the right entry level for foreign language *specialists*?
- Can the CEF help ensure high validity high stakes exams? Is the use of university-developed exams a viable alternative to external exams?

The collaboration will use insights from Alison Duguid's *Anatomy of a Context: English Language Teaching in Italy* (2001), which describes a university response to higher education reforms in Italy, in particular the international certification (*idoneità internazionale)* project at Siena University. This introduced the Cambridge ESOL PET exam as a B1 English language qualification for university degree programmes.

The recent Campus 1 project, which involves 71 Italian universities, means that over half of all Italian universities are enrolling students for Cambridge exams either directly (through 26 university centres) or indirectly. The Cambridge ESOL–University of Siena *Centro Linguistico dell'Ateneo* (CLA) collaboration would make use of Cambridge ESOL impact study experience and instrumentation to enhance CLA language course teacher data collection, including further classroom video recording and analyses.

Related *PL2000* developments

A report in the *EL Gazette* (Vint, June 2004; 1) noted significant impacts from the *PL2000.* The 'overall consensus' was that the Project achieved 'a great deal of success in moving language learning in state schools towards a more language-oriented approach, with syllabuses less dominated by literature or grammar and lessons increasingly conducted in the L2'. External examinations such as those of Cambridge ESOL encouraged schools to set targets according to CEF language 'Can Do' levels (A1 for primary, A2 for middle schools, B1 for high school). More than 400,000 state school students were reported to have taken external certificates in one of the four foreign languages since January 2000, 350,000 or more of these in English. The *EL Gazette* also reported that most certificates were at levels A2 and B1, but that a large proportion also took oral exams only at below A2 in primary and middle schools, rather than exams covering all the four macro-skills, this in line perhaps with the concept of 'partial skills' permitted by the *Progetto*. An increasing number of candidates was reported to be taking four-skill examinations at B2 level (FCE).

The Italian Ministry of Education had always stressed that external exam boards were not brought into the *PL2000* in order to evaluate the success of

Italian language teachers. Regional inspectors, however, are said to have pointed out that the pass rates for *PL2000* course students in four-skill exams, particularly at B1 and B2 levels, suggest teacher success in helping Italian students towards appropriate CEF levels.

The *PL2000* officially terminated in December 2002 but the number of state school students taking external exam certificates has continued to increase, rising to 140,000 students in the academic year 2002/03 with indications of further increases in 2003/04. The *EL Gazette* articles suggested that the reason for this was probably that 'parents are prepared to pay for the exams if school funds are not available, and that most exam boards have continued with special prices and support activity' (2004:1). Figure 8.1 shows the trend.

Figure 8.1 Total Cambridge ESOL exam enrolments in Italy 1999–2003

ITALY	1999/00	2000/01	% Incr.	2001/02	% Incr.	2002/03	% Incr.
Non-*PL2000*	23,978	25,251	*5*	28,594	*13*	33,850	*18*
Progetto Lingue	8,058	29,647	*268*	45,491	*53*	52,465	*15*
Total	32,036	54,898	*71*	74,085	*35*	86,315	*17*

Post-*PL2000*, the Ministry of Education is supporting the preparation of state-school students for the required exit levels as part of the normal curriculum, 'taking greater stock of the CEF objectives' (ibid). From September 2004, regional local education authorities were able to offer language certification to primary school teachers as part of a general teacher training programme.

As noted in the *PL2000* Impact Study report (and see Chapters 1 and 7), impact has tended to be *two-way*, the external exams approved under the project influencing teaching, learning and performance, and the project influencing the external exams. A number of changes were reported in the exams on offer since the start of *PL2000.* Trinity International offers more skills with its International Skills exam, and Cambridge-ESOL *modular* certificates through its Certificates in English Language Skills exam (CELS). Primary school teachers would also have the option of taking an exam based on teaching methodology at B1 level (the new Cambridge ESOL Teaching Knowledge Test [TKT]), as well as more traditional teaching certificates (from Trinity International, UCLES and City and Guilds).

The *PL2000* had emphasised the importance of ensuring that external examinations are selected according to how well they assess agreed foreign language targets or syllabus. The *Progetto* also advised that courses should be

seen as helping learners towards a CEF level rather than a particular exam (see, for example, evidence from the Impact Study in Chapter 7). In this connection, the Council of Europe has produced a draft manual *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment – Manual* (Council of Europe, 2003). This should also help schools and teachers respond to the key message of the *PL2000* Impact Study, to think first about a syllabus in line with the CEF and its exemplification in the European Language Portfolio (ELP) 'to help them monitor progress and understand competency objectives' (Vint 2004). Decisions can then be taken on appropriate language learning methods, materials and activities, before the selection of an appropriate exam for external certification.

Teacher support seminars run by the external exam providers (see Chapter 7) have continued after *PL2000*, now with a stronger ELP focus (there are now at least 11 different versions of the portfolio available in Italy). And on the performance front, it is noted (Vint 2004) that A2 (KET) pass rates have improved considerably since the *PL2000*, perhaps 'as a result of increased support attention' (for example, teachers' seminars and special guides in Italian).

Information dissemination

Cambridge ESOL has disseminated information on the *PL2000* Impact Study through its formal reports to the Ministry of Education and through presentations at conferences and seminars. In Italy, for example, these have included British Council conferences in Venice (March 2001), and Florence, (February–March 2002); and TESOL Italy conferences in Rome, (November–December 2001 and November 2002). Details of the Impact Study were also presented outside Italy, for example at the Association of Language Testers in Europe (ALTE) European Year of Languages Conference, Barcelona (July 2001), at the ALTE meeting in Salamanca, Spain (November 2002), and at the 36th Annual Conference of the International Association of Teachers of English as a Foreign Language (IATEFL) in York (March 2002). *PL2000* Impact Study publications include a chapter in the volume *European Languages in a Global Context* (2004) in the UCLES/Cambridge University Press *Studies in Language Testing* series, and in the *EL Gazette* (Vint 2004). This dissemination process fulfils one of the original aims of the *PL2000* Impact Study, to 'provide a focus for the presentation, to other EU countries, of the Italian model for state-sector adoption of the Common European Framework of Reference for language' *(*see Chapters 3 and 6).

Impact studies: lessons to be learned

The design and implementation of the two studies of impact presented in this book have suggested both theoretical and practical lessons which could inform future similar studies. Most of these lessons will already have been suggested in preceding chapters, but they can usefully be summarised here.

Define and adhere to the purpose in context

It is important to have established, agreed and to reiterate regularly the purpose, aims and institutional context of an impact study. Phases 1 to 3 of the IELTS impact study, and the *PL2000* Impact Study were fundamentally Cambridge ESOL exam validation exercises. They were to be carried out within continuing and iterative Cambridge ESOL systems of test validation, pursuing data and findings on the impact (including the washback) of Cambridge language tests, impact being a fundamental quality of a test or examination, along with its validity, reliability and practicality. Impact study implementers should never lose sight of the wood of the impact study purpose as they pick their way through the trees of study data collection and analysis.

The ownership of the two studies covered in this book was clear, their purposes agreed at senior management level. It was nevertheless always important, both at project design and at implementation stages, to recall and restate the objectives of and target stakeholders for all instruments, interviews, focus groups and data analyses. This helped ensure that they remained on target and that the resulting data were constantly checked for their relevance to impact study purpose. The research questions should guide all data collection. If instrumentation development or data analysis ever appear to lose direction, it will normally clarify matters to re-pose the relevant research question.

Define and maintain the research structure

Impact studies should be based on sound principles of research design and project management. In our two studies, the use of an academic research pattern of organisation was found useful.

First a study was envisaged in terms of context and purpose, aim and objectives, research questions, research design/action plan and hypotheses. This helped to focus the pre-implementation stage, both in terms of theory-based validity in the research design, and an action plan, which was a practical structure assigning roles and responsibilities, including the allocation of resources.

Then, for the implementation phase, the choice of appropriate research *type* is made, leading inevitably to the use of both quantitative *and* qualitative

approaches to data collection. The techniques of wide-scale surveys and limited case studies can both provide valuable evidence in answer to a study's research questions. The research should try for symmetry and control but accepting that studies carried out in real time and places have their own value, despite possible inconvenience and the possibility of some patchy data. It seems crucial for an impact study to include observation as well as stakeholder response data. Since impact studies are, by definition, making comparisons, it is important that baseline data collection is included in the research design and implementation.

Timescale is an important dimension in impact studies. It may be necessary to conduct a study project in phases over a significant period of time. Like other aspects of test validation, these phases should form part of a cyclical and iterative approach, where evidence collected along the way feeds into the planning of subsequent phases.

Validate data collection instruments through agreed steps and appropriate language test validation operations

It is important to start from and sustain the view that impact data collection approaches are, in many cases, like language tests and must be validated. The techniques for the design and development of data collection means should be agreed in advance (though, of course, subject to modification and addition where events really dictate this). Approaches such as brainstorming, response prediction, expert opinion, report and review, revision, piloting, trialling and systematic revision and rationalisation should be planned into the development stage. Language test validation techniques used in the process of data collection instrument validation could include, depending on instrument format, descriptive analyses; factor analysis; convergent–divergent, multi-trait, multi-method validation; triangulation within and across instruments and data collection modes.

Plan and implement the most efficient applicable data management systems

The evidence from an impact study, probably in the form of various types of data, will need to be adequately stored, for retrieval, analysis and reporting purposes. Implications for the type and quantity of analyses to be carried out need to be carefully considered from the start of the impact study design process and in formulating the action plan. These considerations can affect the design of the instruments. Data collection and management systems should take account of the practical constraints of the implementers of and the participants in a study. But the most advanced possible technology should be favoured within those constraints.

Establish and sustain good practice

The conduct of an impact study should always adhere to good practice, in terms, for example, of: informed consent, protection of participants, confidentiality and anonymity, accountability, reciprocity and intellectual property and data protection legislation. These are matters where up-to-date official sources of information need to be consulted regularly.

Use ideas and instrumentation, adapted, across research projects

Impact studies tend to be demanding both of time and effort. It would be rare for a study implementation team, however well developed its action plan, to consider itself adequately staffed. Instrumentation and items were shared, modified as required, by the IELTS Phase 3 and the *PL2000* studies. We have also seen that related projects, for example in the IELTS funded research programme, have used Phase 3 instruments. This policy of instrument sharing is to be recommended wherever feasible.

Towards revised models for test validation and the place of impact study in them

Certain constant themes will have been noted in this book on the matter of systematic approaches to impact studies within an institutional context. These have included:

- the view of impact study as part of routine test validation research
- the study of impact, like the test validation process, as continuing and iterative
- an emphasis on good practice in the planning and conduct of impact studies
- the need to seek ever-greater coherence and rigour in models and systems for test specification, design, development, validation, production administration and use.

Significant developments and achievements in Cambridge ESOL systems and models for test development, validation and use have been traced in preceding chapters, leading to a model of test research, development, implemention and revision based on the principles of validity, reliability, impact and practicality (VRIP). These principles and the systems to implement them were arrived at through internal and external collaboration, including:

- long-term theoretical and empirical investigation:
  - by Cambridge ESOL's Research and Validation Group, in co-operation with the Examinations and Assessment Group, responsible for

professional aspects of the development, production and on-going servicing of all Cambridge ESOL exams, and the Operations Group which handles, among other matters, exam materials, commissioning and pretesting, candidacy entry and results processing, marking, and script management.

- collaboration with partners:
  - in the case of the IELTS test, the British Council and IDP Education: IELTS Australia, both fully informed of Cambridge ESOL test development and validation systems and models and very much involved in IELTS validation research and dissemination, in particular through the funded-research programme
  - the Association of Language Testers in Europe (ALTE) in the development and finalisation of the Code of Practice, which, it will be remembered, lays a strong emphasis on provider responsibility for exam development, scoring, result interpretation, fairness and transparency
- consultation with key outside language testing figures including:
  - Bachman, in particular on the relationship between VRIP and Bachman and Palmer's 1996 six test usefulness qualities, reliability, construct validity, authenticity, interactiveness, impact and practicality, and on test comparability (Bachman et al 1995)
  - Weir (2004), with his important socio-cognitive framework for validating tests, with test validity the superordinate category to theory-based, context, scoring, concurrent and consequential validities
  - Alderson, as in Phase 1 of the study of IELTS impact and at the ALTE Barcelona Conference (Alderson in Milanovic and Weir (eds) 2004.
  - Kunnan (2000) and Purpura (1996b, 2001), as in their contributions to the study of IELTS impact.

Such research, practical investigation, collaboration and consultation has led to the development of a range of frameworks and models designed to inform the test development, validation and management process. Among the frameworks and models encountered in this book are:

- the taxonomy of Cambridge ESOL stakeholders as they relate to test constructs, formats, conditions and assessment criteria, see Figure 1.4 (and Saville 2003, Saville and Hawkey 2004, Taylor 1999)
- the standard current Cambridge ESOL model of the test development or revision process (Figure 1.5), accounting for phases in the test cycle from perceived need for a new or revised test, through planning, design, development (trialling, analysis, evaluation and review), operation, further monitoring, review and evaluation through to further test revision or renewal
- representations used above (e.g. Figures 1.1, 1.2) to indicate the place and role of impact study in test development and validation.

This book should already have shown the important role played by the study of impact in test development, validation, production and administration models and systems.

Models and systems themselves, of course, are also subject to monitoring, revision and refinement. Recent discussion at Cambridge ESOL indicates trends towards renewed models. The direction of these trends appears to be influenced by the growing expectation, noted (for example in Chapter 1, see Saville and Hawkey, 2004) of good value, good practice and accountability. Test validity is expected to be strongly theory- and evidence-based (see the Weir 2004 framework) and there appears to be a stronger than ever expectation of tightly logical argument in support of validity claims. Relevant recent influences in this direction appear to be:

- the concept of interpretive argument proposed by Kane (1992) and Kane, Crooks and Cohen (1999), emphasising clarity of argumentation, coherence of argument, and the plausibility of assumptions
- the 'standards approach to validation' of Chapelle et al (2004), which 'directs validation researchers to gather kinds of evidence' which 'are needed to evaluate the intended interpretation of test scores' to be guided by 'developing a set of propositions that support the proposed interpretations'
- the work of Mislevy et al (2002, 2003) on evidence-centred assessment design (ECD), like the 'evidence-based' label hardly revolutionary, but as Saville (June 2004) says, 'a way of introducing greater systematicity in the design and development' of language tests
- the method of Toulmin, favouring analyses as a process of breaking an argument into its different parts to make judgements on how well the different parts work together; for example in terms of *claim* ('a conclusion whose merits we are seeking to establish'); *data* ('information on which the claim is based'), a *warrant* (a statement which 'provides legitimacy of the step in the argument and refers to the claim'); the *backing* (evidence from theory, research, evidence from validation operations (2003:90-96), 'without which the warrants themselves would possess neither authority nor currency'), and *rebuttals* ('exceptional conditions which might be capable of defeating or rebutting the warranted conclusion')
- Bachman (2004), with his characteristic emphasis on assessment *use* as well as assessment validation argument, suggesting a model where a *claim,* that is an inference that we make according to data from test taker responses, is justified on the basis of a warrant, where the backing is positive evidence (from theory, research, evidence from validation operations) or, refuted on the basis of a rebuttal if such evidence is significantly negative.

Saville (2004) claims that the Cambridge ESOL research and validation framework 'is being recast' taking these features into account. The intention is to adopt a process model and theoretical framework approach 'to develop a more coherent programme of test development, validation and research to support claims about the validity of the interpretation of our examinations results and their uses'. This, Saville adds, would enable ESOL 'to set out our CLAIMS relating to the usefulness of the test for its intended purpose, explain why each claim is appropriate by giving REASONS and justifications; provide adequate EVIDENCE to support the claims and the reasoning'.

It should be noted that, as this book has confirmed, Cambridge ESOL already employ models and systems making reasoned claims on test usefulness and validity based on evidence from the multi-faceted sources some of which are listed above. The continuing and iterative nature of Cambridge ESOL models for the key steps in the test cycle, however, requires that such models and systems are themselves revisited regularly.

Conclusion

In Chapter 8 further research and other post-impact study developments have been discussed, lessons to be learned for the design and implementation of impact studies suggested, based on the theory and the case studies described in the book. In both the current and possible future frameworks and models of test validation noted, there is clearly a place for the study of test impact.

Chapter 8 brings closure to the logic of the book. It has travelled a route from the constructs of impact and related concepts in Chapter 1, approaches to their investigation in Chapter 2, through impact study research design in Chapter 3, the histories of the development and validation of the data collection instruments described in Chapter 4 and the practicalities of their implementation in Chapter 5. Chapters 6 and 7 presented key messages from the IELTS and *PL2000* impact studies.

Appendices

APPENDIX A:
Student Questionnaire

STUDY OF THE IMPACT OF THE INTERNATIONAL ENGLISH LANGUAGE TESTING SYSTEM (IELTS)

STUDENT QUESTIONNAIRE

Dear Participant,

As part of the continuing programme to update and refine its International English Language Testing System (IELTS), the University of Cambridge Local Examinations Syndicate (UCLES) is conducting a study of the impact of the test. We are contacting students, candidates, English teachers, and university administrators for information and comment.

Your responses to this questionnaire will be treated in confidence, and only used for the stated purposes of the study.

Thank you very much for your time and co-operation. We should also be grateful if you would complete and sign the consent note below.

Yours sincerely

Nick Saville
EFL Director
Research and Validation
University of Cambridge Local Examinations Syndicate
English as a Foreign Language
Cambridge
England

Your consent to participate in the UCLES IELTS Impact Study

I understand that:

- the purpose of the study is to collect and analyse information from those familiar with the IELTS;
- my name will not appear in any project publication;
- the information I give, but not my name, may be quoted;
- I am free to refuse to participate in the study and may withdraw at any time;
- my completed questionnaire is for the study team only; it will not be shown to anyone not connected with the UCLES study.

Signature: ..... Date: .....

Please give the information about yourself requested below in the relevant spaces and by ticking (✓) the appropriate ☐.

Full name and current address		Home country	
		Nationality	
		Field of study / work	
Male ☐	Female ☐	Age	

There are 4 Parts in this questionnaire:

- **PART 1:** for all students and candidates, on participants' language learning background.
- **PART 2:** again for everyone to complete, about the ways people learn, study and take tests.

- **PART 3:** only for students who are taking or have taken a course to prepare them for the IELTS, on their experiences on the course.
- **PART 4:** only for students who have taken the IELTS, asking them about the test.

You will be advised on which Parts you should complete.

Please now complete PART 1 of the questionnaire, which is about your language learning background.

PART 1: ENGLISH LANGUAGE BACKGROUND

1. What language(s) do you usually speak at home when in your country?	

	Yes	No
2. As a child, did you live with a parent, guardian or other close relation who was a native speaker of English?	☐	☐

3. Did you study English in:	Yes	No
3.1 kindergarten (age 3-6)?	☐	☐
3.2 primary school (age 7-11)?	☐	☐
3.3 secondary school (age 12-17)?	☐	☐
3.4 college / university?	☐	☐
3.5 extra language classes?	☐	☐

4. How many of your English lessons in your country were taught or assisted by native speakers of English?	More than half	About half	Less than half	A few	None
	☐	☐	☐	☐	☐

5. How much of the time in class did your teachers of English speak to you in English?	All the time	More than half	About half	Less than half	Never
	☐	☐	☐	☐	☐

6. In your home country, how often:	Often	Some-times	Rarely	Never
6.1 do you use English for socialising (talking or writing to friends)?	☐	☐	☐	☐
6.2 are you in English contact with friends overseas?	☐	☐	☐	☐
6.3 are you exposed to English in the media?	☐	☐	☐	☐
6.4 do you use English for work?	☐	☐	☐	☐
6.5 do you read texts in English in your specialist subject?	☐	☐	☐	☐
6.6 do you write in English in a day?	☐	☐	☐	☐

7. Have you studied in an English-speaking country? Where, when and for how long?	

	Very often	Quite often	Once/twice only	Never
8. How often have you stayed in English-speaking countries for a week or more?	☐	☐	☐	☐

Thank you for completing Part 1.

Please now complete PART 2 of the questionnaire, which is about the ways you learn and study languages.

PART 2: LEARNING, STUDY AND TEST-TAKING APPROACHES

A.	***Please put a tick (✓) in the column that best describes your language learning or study habits.***	Strongly disagree	Moderately disagree	Slightly disagree	Slightly agree	Moderately agree	Strongly agree
1.	When I begin learning a new language, I think about the level to which I want to learn it.	☐	☐	☐	☐	☐	☐
2.	I set specific goals for myself in language learning.	☐	☐	☐	☐	☐	☐
3.	I plan how I am going to learn so that I can use my time effectively.	☐	☐	☐	☐	☐	☐
4.	I think about how I learn languages best.	☐	☐	☐	☐	☐	☐
5.	I decide in advance to pay special attention to particular skills in English so I can learn them best.	☐	☐	☐	☐	☐	☐
6.	I learn best when I am taught language rules.	☐	☐	☐	☐	☐	☐
7.	I make notes of the mistakes I make in English so that I can learn from them.	☐	☐	☐	☐	☐	☐
8.	I make efforts to improve my ability in English by spending time with native speakers of this language.	☐	☐	☐	☐	☐	☐
9.	I try to improve my ability in English by asking other people to tell me if I have understood or said something correctly.	☐	☐	☐	☐	☐	☐
10.	I repeat new words to make sure I have understood them correctly.	☐	☐	☐	☐	☐	☐
11.	I make charts, diagrams or tables to organise what I have learned.	☐	☐	☐	☐	☐	☐
12.	I try to make sure I remember new words by using them in new situations.	☐	☐	☐	☐	☐	☐
13.	I learn new words in English by thinking of words that I know that sound like the new word.	☐	☐	☐	☐	☐	☐
14.	I learn new words in English by dividing them into parts that I understand so I can figure out what they mean.	☐	☐	☐	☐	☐	☐
15.	I learn new words in English by remembering where the new word was on the page, or where I first saw or heard it.	☐	☐	☐	☐	☐	☐
16.	I learn grammar in English by comparing the rules in my language with grammar rules in English.	☐	☐	☐	☐	☐	☐
17.	I learn grammar in English by memorising the rules and applying them to new situations.	☐	☐	☐	☐	☐	☐
18.	When I speak or write in English I know when I make grammar mistakes.	☐	☐	☐	☐	☐	☐
19.	When I listen to or read incorrect English I recognise grammar mistakes.	☐	☐	☐	☐	☐	☐
20.	I make special efforts to improve my English by listening to programmes in English on the radio.	☐	☐	☐	☐	☐	☐

21. I try to improve my English by watching television programmes or films in English.	☐	☐	☐	☐	☐	☐
22. I try to improve my listening and reading in English by guessing the meanings of new words from the context.	☐	☐	☐	☐	☐	☐
23. I try to improve my reading in English by making notes to help me remember what I have read.	☐	☐	☐	☐	☐	☐
24. I try to improve my English by summarising new information I hear or read.	☐	☐	☐	☐	☐	☐
25. I try to improve my writing in English by analysing how writers organise their paragraphs.	☐	☐	☐	☐	☐	☐
26. I try to improve my writing in English by showing my writing to another person.	☐	☐	☐	☐	☐	☐
27. I try to improve my writing or speaking by including expressions I have read or heard.	☐	☐	☐	☐	☐	☐
28. I try to improve my writing in English by putting words and phrases that I meet into new practice sentences.	☐	☐	☐	☐	☐	☐
29. I try to improve my speaking in English by repeating sentences in English until I can say them easily.	☐	☐	☐	☐	☐	☐
30. I try to improve my speaking and writing in English by using my knowledge of grammar rules to help me form new sentences.	☐	☐	☐	☐	☐	☐
31. I try to improve my speaking in English by repeating what I hear native speakers say.	☐	☐	☐	☐	☐	☐
32. When I speak English I know when I have mispronounced something.	☐	☐	☐	☐	☐	☐
33. After I finish a conversation in English, I think about how I could have said things better.	☐	☐	☐	☐	☐	☐
34. After I have said something in English, I check whether the person I am talking to has really understood what I meant.	☐	☐	☐	☐	☐	☐
35. I want to improve my general ability to use English.	☐	☐	☐	☐	☐	☐
36. I encourage myself to use English even when I am afraid of making mistakes.	☐	☐	☐	☐	☐	☐
37. I read for pleasure in English.	☐	☐	☐	☐	☐	☐
38. I have several close friends who are native speakers of English.	☐	☐	☐	☐	☐	☐
39. I try to learn about the culture of English speakers.	☐	☐	☐	☐	☐	☐
40. I learn more by reading textbooks than by listening to lectures.	☐	☐	☐	☐	☐	☐
41. I enjoy working on an assignment with two or three classmates.	☐	☐	☐	☐	☐	☐
42. I learn better in class when the teacher gives a lecture.	☐	☐	☐	☐	☐	☐
43. I learn better working alone than in a group.	☐	☐	☐	☐	☐	☐
44. I enjoy learning in class by doing experiments.	☐	☐	☐	☐	☐	☐
45. I remember things I have heard in class better than things I have read.	☐	☐	☐	☐	☐	☐

B.	***Please tick (✓) the appropriate ☐ for each item on the ways you do tests.***	Never	Rarely	Often	Always
1.	I decide which parts of the test are the most important before I start.	☐	☐	☐	☐
2.	I glance through all the questions in the test before I start doing them.	☐	☐	☐	☐
3.	I plan so that I have enough time to answer each question.	☐	☐	☐	☐
4.	I read and think about the instructions in detail before I try to answer the questions.	☐	☐	☐	☐
5.	I highlight (or underline) key words in the instructions and keep them in mind while completing the task.	☐	☐	☐	☐
6.	I read the whole text before I start to answer the questions.	☐	☐	☐	☐
7.	I use the words in the questions to find the sentence in the text which contains the answer.	☐	☐	☐	☐
8.	I first skim a text to look for the main ideas.	☐	☐	☐	☐
9.	Before I listen to something, I try to guess what information is coming.	☐	☐	☐	☐
10.	I try to predict the questions when listening.	☐	☐	☐	☐
11.	I expect to hear the information for the answers in the same order as the questions.	☐	☐	☐	☐
12.	I make a plan of my whole answer before I write.	☐	☐	☐	☐
13.	I write a draft of my whole answer.	☐	☐	☐	☐
14.	I do not write much more than the minimum word requirement.	☐	☐	☐	☐
15.	I try to include words or phrases to organise my speaking /writing (e.g. *firstly, furthermore, secondly, I have two points....*)	☐	☐	☐	☐
16.	I practise using set conversational phrases (e.g. *let me see now...; what shall I say... etc.*) to fill in silences.	☐	☐	☐	☐
17.	I am prepared to speak first in a conversation test.	☐	☐	☐	☐
18.	I don't wait to be asked before speaking.	☐	☐	☐	☐
19.	I check my answers before I leave the exam.	☐	☐	☐	☐
20.	After a test, I usually feel that I have done as well as my knowledge and ability deserve.	☐	☐	☐	☐

Thank you for completing Part 2.

If you attended / are attending IELTS preparation classes, please now complete PART 3 of the questionnaire.

PART 3: FOR THOSE ATTENDING IELTS PREPARATION CLASSES

1.1 Are you still attending IELTS preparation classes?		**Yes** ☐	**No** ☐
If yes, where and who is organising?			
1.2 Did you attend such classes in the past?		**Yes** ☐	**No** ☐
If yes, where and who organised?			

2. What kind of IELTS preparation classes and for <u>how long?</u>	**Yes**	**No**	**Weeks**	**Hours /week**
2.1 A course with "IELTS" in its title?	☐	☐		
2.2 Part of a general English course?	☐	☐		
2.3 Part of an English language study skills course?	☐	☐		
2.4 Part of an English for Academic Purposes (EAP) course?	☐	☐		
2.5 Part of an English for Occupational Purposes (EOP) course?	☐	☐		

3. What proportion of the time on your IELTS-preparation course is / was spent working on the following, and how useful do you believe they are for the IELTS test?	**% of time**	**Very useful**	**Quite useful**	**Not very useful**
		for IELTS		
3.1 Reading		☐	☐	☐
3.2 Writing		☐	☐	☐
3.3 Listening		☐	☐	☐
3.4 Speaking		☐	☐	☐
3.5 Vocabulary		☐	☐	☐
3.6 Grammar		☐	☐	☐
3.7 Other (please specify)		☐	☐	☐

4.	Did / Do any of the following happen in your IELTS preparation classes?	Yes	No	Not sure
	Listening:			
4L.1	Reading the questions and predicting what listening passages would be about.	☐	☐	☐
4L.2	Listening to live, taped or video talks / lectures and taking notes.	☐	☐	☐
4L.3	Listening and taking part in seminar / workshop activities.	☐	☐	☐
4L.4	Using information from a lecture or talk to write reports.	☐	☐	☐
4L.5	Reading questions and guessing the type of answer required.	☐	☐	☐
4L.6	Practice in recognising previous information repeated in different words.	☐	☐	☐
	Reading:			
4R.1	Analysing text structure and organisation.	☐	☐	☐
4R.2	Interpreting statistics / graphs / diagrams.	☐	☐	☐
4R.3	Reading texts to predict test questions and tasks.	☐	☐	☐
4R.4	Learning quick and efficient ways of reading texts in English.	☐	☐	☐
4R.5	Reading articles, reports, books in your specialist subject area.	☐	☐	☐
4R.6	Using English-only dictionaries to complete reading tasks.	☐	☐	☐
4R.7	Reading quickly to get the main idea of a text.	☐	☐	☐
	Writing:			
4W.1	Copying out good paragraphs and model answers.	☐	☐	☐
4W.2	Describing a graph / a process / statistical data.	☐	☐	☐
4W.3	Learning how to organise essays.	☐	☐	☐
4W.4	Practising using words or phrases to organise a written text (e.g. *firstly, furthermore, secondly, therefore*)	☐	☐	☐
4W.5	Learning how to write in different styles.	☐	☐	☐
4W.6	Short report writing.	☐	☐	☐
4W.7	Planning written answers to test questions.	☐	☐	☐
4W.8	Editing written work.	☐	☐	☐
4W.9	Writing parts of test answers.	☐	☐	☐
4W.10	Writing long essays, reports (i.e. over 1000 words).	☐	☐	☐
	Speaking:			
4S.1	Practising making a point and providing supporting examples.	☐	☐	☐
4S.2	Planning and delivering oral presentations.	☐	☐	☐
4S.3	Group discussions / debates.	☐	☐	☐
4S.4	Practising using filler words to cover silences in your speech (e.g. *well...you see....*)	☐	☐	☐
4S.5	Practising using words or phrases to organise your speech (e.g. *firstly, furthermore, secondly, I have two points*)	☐	☐	☐

5. How much of the following kinds of specific exam practice do / did you do on your preparation course (as approximate percentages (%) of the course)?	
Specific IELTS Exam Practice	**Approx % of course time**
5.1 Information about the contents and format of the test	
5.2 Looking at past papers	
5.3 Taking practice tests	
5.4 Receiving feedback in the form of IELTS band scores	
5.5 Techniques for taking the test *(please specify)*	
5.6 Others *(please specify)*	

6. Did you use (a) textbook(s) on your IELTS preparation course? If so, what was / were the title(s) (approximately if you cannot remember exactly), and what did it / they contain?

7. If you did use (an) English textbook(s), please give your opinions of the good and less good points here.	
Good Points	**Not so good points**

8. What other materials did you use on the preparation course? How were they (e.g. useful / not useful; difficult / easy; enjoyable / not enjoyable)?

9. What did / does a good / successful student do on the IELTS preparation course that an unsuccessful one did / does not?

10. Do you think you were / are successful on the preparation course(s)? Why? Why not?

11. If an IELTS score had *not* been a requirement, would you have prepared for your studies abroad in the same way? Yes ☐ No ☐
Why / Why not?

12. If you are already studying / working in an English-medium situation now, do you find that your IELTS preparation course provided you with the language knowledge and skills you need? Please comment.

13. Would your IELTS preparation course be a good way to learn English for someone who is not going to take IELTS? Why? / Why not?

14. Would the IELTS preparation course be useful for someone who is not going to university? Why / Why not?

15. Please note here anything else you wish to say about your IELTS preparation course.

Thank you for completing Part 3.

If you have already taken the IELTS, please now complete PART 4 of the questionnaire.

PART 4: FOR THOSE WHO HAVE TAKEN IELTS

1.1 Are you a student at present?	**Yes ☐**	**No ☐**
1.2 If so, at what level are you studying?	**Under-graduate ☐**	**Post-graduate ☐**
	Other:	
1.3 What degree are you studying for (e.g. BA, MSc., PhD)?		
1.4 In which country are you studying now?		
1.5 If you are not a student at present, please describe your current status (e.g. *waiting to enter University, already graduated, teaching)* :		

2. When did you take the IELTS test? *(Please include all dates if you have taken the test more than once.)*	

3. Why did you take IELTS?	

4. Have you taken other similar tests (TOEFL; CAE/CPE/FCE; Michigan, others?) Why? / Why not?	

5. For the IELTS reading and writing section did you take the:	**General Training (GT) module? ☐**	**Academic (A) module? ☐**

6. If you know your IELTS band score, (range 1 to 9):	
6.1 which band score(s) did you get?	
6.2 which band score(s) did you need for your studies or work?	
6.3 which band score(s) did you expect?	

7. Did you worry about taking the IELTS test?	**Very much 1**	**2**	**3**	**Very little 4**
	☐	☐	☐	☐

8. What worried you most about the test?	

9. Please rank the sections of the IELTS test according to how difficult you found them

	The most difficult 1	**Second most difficult 2**	**Third most difficult 3**	**The easiest 4**
9.1 Listening	☐	☐	☐	☐
9.2 Reading	☐	☐	☐	☐
9.3 Writing	☐	☐	☐	☐
9.4 Speaking	☐	☐	☐	☐

10. Did you feel that you performed to the best of your ability in the test?	Yes ☐	No ☐

11. What affected your performance?	A lot	Quite a lot	Not a lot	Not at all
11.1 Difficulty of language	☐	☐	☐	☐
11.2 Difficulty of questions	☐	☐	☐	☐
11.3 Unfamiliarity of topics	☐	☐	☐	☐
11.4 Time pressure	☐	☐	☐	☐
11.5 My fear of tests	☐	☐	☐	☐
11.6 Others *(please specify)*				

12. Do you think IELTS is a fair way to test your proficiency in English?	Yes ☐	No ☐
Why (not)?		

13. What knowledge or skills (if any), apart from language ability, are needed for a good IELTS score?

14. In your opinion, is the IELTS exam appropriate for the following groups?	Yes	No	No opinion
14.1 all ages 15+	☐	☐	☐
14.2 post-graduates	☐	☐	☐
14.3 under-graduates	☐	☐	☐
14.4 professional people	☐	☐	☐
14.5 all nationalities/cultures	☐	☐	☐
14.6 students in all subject areas	☐	☐	☐

15. What did you *like* about the IELTS test?	16. What did you *dislike* about the IELTS test?

17. What advice would you give someone who is going to take the IELTS test?

18. Please write here any comments about IELTS, which are not covered in the above items:

That is the end of the questionnaire. Thank you very much for responding to the items and answering the questions.

APPENDIX B:
Teacher Questionnaire (for teachers preparing students for IELTS)

Dear Colleague

As part of the continuing programme to update and refine its International English Language Testing System (IELTS), the University of Cambridge Local Examinations Syndicate (UCLES) is conducting a study of the impact of the test. In the course of the study, we are contacting students, candidates, English teachers, and university administrators for information and comment.

As a teacher of English to IELTS students, your responses to this questionnaire will be treated in confidence, and used only for the stated purpose of the study.

We should be grateful if you would first sign the consent slip and answer the short background questionnaire below.

Thank you very much for your time and co-operation.

Yours sincerely

Nick Saville
EFL Director
Validation and Research
University of Cambridge Local Examinations Syndicate
English as a Foreign Language
Cambridge
England

Your consent to participate in the UCLES Impact Study

I understand that:

- the purpose of the study is to collect and analyse information from those familiar with international English language tests;
- my name will not appear in any project publication;
- the information I give, but not my name, may be quoted;
- I am free to refuse to participate in the study and may withdraw at any time;
- my completed questionnaire is for the study team only; it will not be shown to anyone not connected with the UCLES study.

Name: .....

Signature: ..... Date: .....

Thank you.

QUESTIONNAIRE FOR TEACHERS PREPARING STUDENTS FOR IELTS

Please tick (✓) the boxes or write in your responses as appropriate.

SECTION A: About you

1.	Your full name:								
2.	Form of address:	Miss ☐	Mrs ☐	Mr ☐	Dr ☐	Other, *please specify*.			
3.	Your age, please:	below 30 ☐	31-40 ☐	41-50 ☐	51-60 ☐	61+ ☐			
4.	Country you now work in:								
5.	Name, address and email of institution where you work:								
6.	Your position there:								
7.	Number of years you have been teaching English:		8. Your qualifications:						
9.	Have you been trained as an examiner for IELTS or other international proficiency test(s)?						Yes ☐	No ☐	
	If, yes, please explain a little:								
10.	Have you received any training in how to prepare students for IELTS?						Yes ☐	No ☐	
	If yes, please describe briefly:								

SECTION B: About your students

11.	Ages of your IELTS students?		Their country(ies) of origin:			
12.	Your IELTS students' level(s) of education?	secondary up to 16 years ☐	secondary 17-19 years ☐	degree or equivalent ☐	post-graduate ☐	
13.	Applying for which country(ies)?					
14.	Taking which IELTS modules?		All	Most	Half	A few
		Academic module	☐	☐	☐	☐
		General Training module	☐	☐	☐	☐
15.	Proportion of students who have already taken IELTS?	All	Most	About half	A few	None
		☐	☐	☐	☐	☐

If you have taught students who have already taken IELTS, please answer items 16-20. If not please go to Section C.

16-20 Comparing your students' results in the IELTS test with your own assessment of their language ability:

Compared with their IELTS results, I consider	higher	about the same	not consistently related	lower
16. my students' actual general English proficiency level is	☐	☐	☐	☐
17. my students' actual level in reading is	☐	☐	☐	☐
18. my students' actual level in writing is	☐	☐	☐	☐
19. my students' actual level in listening is	☐	☐	☐	☐
20. my students' actual level in speaking is	☐	☐	☐	☐

SECTION C: About the IELTS

21-26. Do you consider the IELTS test appropriate to candidates' future English language needs:

		Yes	No	Not sure
21.	at undergraduate level?	☐	☐	☐
22.	at postgraduate level?	☐	☐	☐
23.	at pre-university level?	☐	☐	☐
24.	for vocational studies?	☐	☐	☐
25.	in their professional work?	☐	☐	☐
26.	for immigration purposes?	☐	☐	☐

27.	How does a current IELTS Band 6 compare with a band 6 in previous years?	Higher ☐	Lower ☐	Unchanged ☐	Not sure ☐

		Yes	No	Don't know
28.	Does the IELTS test provide positive motivation for your students?	☐	☐	☐
29.	Does the test cause unhelpful stress for your students?	☐	☐	☐
30.	Does the IELTS test influence your choice of the content of your IELTS preparation lessons (i.e. what you teach)?	☐	☐	☐
31.	If yes, please note here how the test influences your decisions on lesson content:			
32.	Does the IELTS test influence your choice of methodology (i.e. the way you teach) for IELTS preparation lessons?	☐	☐	☐
33.	If yes, please note here how the IELTS test influences the way you teach:			

Please complete the following statements:

34. What I like about teaching for IELTS is .....

.....

35. What I dislike about teaching for IELTS is .....

.....

	What knowledge or skills other than proficiency in the English language do you think helps students achieve a good IELTS grade?			
37.	What advice would you give to a colleague who was about to prepare students for IELTS for the first time?			
38.	Compared with other English language classes you have taught, do you think your IELTS preparation classes are:	more successful? ☐	as successful? ☐	less successful? ☐
39.	Why?			

			Rank order
40.	Please rank the IELTS test sections in order of difficulty for most of your students (1 = most difficult ...4 = least difficult etc.):	40.1 reading	
		40.2 writing	
		40.3 listening	
		40.4 speaking	

41- 48 Are the following statements about the IELTS test correct?

	Yes	No	Not sure
41. The IELTS test includes a section testing grammar.	☐	☐	☐
42. In the speaking module, candidates have to both ask and answer questions.	☐	☐	☐
43. Reading and writing together carry more than half of the marks.	☐	☐	☐
44. Candidates have two opportunities to hear the cassette in the listening module.	☐	☐	☐
45. Candidates have to write at least 150 words for the first task in the writing module.	☐	☐	☐
46. Candidates often need to refer to the reading texts when they do the writing module.	☐	☐	☐
47. The reading module has three sections.	☐	☐	☐
48. In the listening module, candidates may have to label a diagram.	☐	☐	☐

SECTION D: About IELTS preparation classes

What kind(s) of IELTS preparation classes are you teaching?

	Yes	No	Weeks long	Hours/week
49. a course with "IELTS" in its title	☐	☐	☐	☐
50. part of a general English course	☐	☐	☐	☐
51. part of an English language study skills course	☐	☐	☐	☐
52. part of an English for academic purposes (EAP) course	☐	☐	☐	☐
53. part of an English for occupational purposes (EOP) course	☐	☐	☐	☐
54. other (please specify)				

55. How many students on average attend the IELTS class(es) you teach?	1-5 ☐	6-10 ☐	11-15 ☐	16-20 ☐	21-25 ☐	26+ ☐

56. Are the IELTS courses normally taught by one, or more than one teacher? Please explain:

57. What proportion of the time on your IELTS-preparation course is normally spent working on the following, and how useful do you believe they are for the IELTS test?

Skill, component etc	% of time	For IELTS		
		very useful	quite useful	not very useful
57.1 Reading		☐	☐	☐
57.2 Writing		☐	☐	☐
57.3 Listening		☐	☐	☐
57.4 Speaking		☐	☐	☐
57.5 Vocabulary		☐	☐	☐
57.6 Grammar		☐	☐	☐
57.7 Others (please specify)		☐	☐	☐

58. Which of the following activities take place in your normal IELTS-preparation class?			
	Yes	No	Not Sure
Listening:	☐	☐	☐
58L.1 reading questions and predicting what listening texts will be about	☐	☐	☐
58L.2 listening to live, cassette or video talks / lectures and taking notes	☐	☐	☐
58L.3 listening and taking part in seminar / workshop activities	☐	☐	☐
58L.4 using information from a lecture or talk in written reports	☐	☐	☐
58L.5 reading questions and guessing the types of answer required	☐	☐	☐
58L.6 practice in recognising previous information repeated in different words	☐	☐	☐
Reading:	☐	☐	☐
58R.1 analysing text structure and organisation	☐	☐	☐
58R.2 interpreting statistics / graphs / diagrams	☐	☐	☐
58R.3 reading texts to predict test questions and tasks	☐	☐	☐
58R.4 learning quick and efficient ways of reading texts in English	☐	☐	☐
58R.5 reading articles, reports, books in specialist subject area(s)	☐	☐	☐
58R.6 using monolingual dictionaries to complete reading tasks	☐	☐	☐
58R.7 reading quickly to get the main idea of a text	☐	☐	☐
Writing:	☐	☐	☐
58W.1 copying out good paragraphs and model answers	☐	☐	☐
58W.2 describing graph / process / statistical data	☐	☐	☐
58W.3 learning how to organise essays	☐	☐	☐
58W.4 practising using words or phrases to organise a written text (e.g. *firstly, furthermore, secondly, therefore*)	☐	☐	☐
58W.5 learning how to write in different styles	☐	☐	☐
58W.6 short report writing	☐	☐	☐
58W.7 planning written answers to test questions	☐	☐	☐
58W.8 editing written work	☐	☐	☐
58W.9 writing parts of test answers	☐	☐	☐
58W.10 writing long essays, reports (i.e. over 1000 words).	☐	☐	☐
Speaking:	☐	☐	☐
58S.1 practising making a point and providing supporting examples	☐	☐	☐
58S.2 planning and delivering oral presentations	☐	☐	☐
58S.3 group discussions / debates	☐	☐	☐
58S.4 practising using filler words to cover gaps in speech (e.g *well..., you see...*)	☐	☐	☐
58S.5 practising using words or phrases to organise a speech (e.g. *firstly, furthermore, secondly, I have two points*)	☐	☐	☐

59. How much, if any, of the following kinds of specific exam practice do you give on your preparation course (as approximate percentages (%) of the course):	
	approx % of course time
59.1 information about contents and format of the test	
59.2 looking at past papers	
59.3 taking practice tests	
59.4 marking and giving feedback in the form of IELTS band scores	
59.5 techniques for taking the test *(please specify)*	
59.6 others (please specify)	

60. Do you use (a) textbook(s) on your IELTS preparation course. If so, what is / are the title(s) *(approximately if you cannot remember exactly)*	

61. If you do/did use (a) textbook(s), please give your opinions of their good and less good points here:

Good points:

Not so good points:

62. What other teaching materials do you use on your IELTS preparation course(s) and why?

IELTS prep-course teaching materials	Reason

63. What does a good/ successful student do on the IELTS preparation course that an unsuccessful one does not?

64. If an IELTS score had *not* been a requirement would you have prepared your students for their future studies abroad in the same way?	Yes ☐	No ☐

65. Would your IELTS preparation course be a good way to learn English for someone going to university but who is not going to take IELTS? Why? / Why not?

66. Would the IELTS preparation course be useful for someone who is not going to university? Why? Why not?

67. Please note here anything else you wish to say about your IELTS preparation course.

Thank you very much for your help with the UCLES impact study.

APPENDIX C:
Instrument for the Analysis of Textbook Materials (IATM)

Dear Participant

As part of the continuing programme to update and refine its International English Language Testing System (IELTS), the University of Cambridge Local Examinations Syndicate (UCLES) is conducting a study of the impact of the test. In the course of the study, we are contacting students, candidates, English teachers, admissions officers and other university administrators for information and comment.

Your responses to this questionnaire will be treated in confidence, and used only for the stated purpose of the study.

We should be grateful if you would first sign the consent slip and answer the short background questionnaire below.

Thank you very much for your time and co-operation.

Yours sincerely

Nick Saville
EFL Director
Validation and Research
University of Cambridge Local Examinations Syndicate
English as a Foreign Language
Cambridge
England

Your consent to participate in the UCLES Impact Study

I understand that:

- the purpose of the study is to collect and analyse information from those familiar with international English language tests;
- my name will not appear in any project publication;
- the information I give, but not my name, may be quoted;
- I am free to refuse to participate in the study and may withdraw at any time;
- my completed questionnaire is for the study team only; it will not be shown to anyone not connected with the UCLES study.

Name: .....

Signature: ..... Date .....

Thank you.

A little about you *(please write, type, tick (✓) boxes ☐, leave blank as appropriate)*						
Your full name						
Form of address	Miss ☐	Mrs ☐	Mr ☐	Dr ☐	Other *(please specify)*	
Country where you were born						
Your first language						
Name and address of institution where you work						
Your position						
Your academic / professional qualifications						
International English language test(s) with which you are familiar						
Your experience with this / these test(s) (*teaching test preparation courses, test administration, being trained as examiner etc*), if any	Test(s)		Experience			
Your brief opinion of this test / these tests (e.g. any comments on: *test components, levels, topics, skills, format, scoring, administration, reliability in predicting a student's English language competence and performance?)*	Test(s)		Comment(s)			

Notes on the use of the Instrument for the Analysis of Textbook Materials (IATM)

1. This questionnaire seeks your analysis and evaluation of the content, level and approaches of a textbook and its support materials.
2. We are especially interested in the relationships between textbooks and international tests for the certification of language performance.
3. The questionnaire invites:
   - objective analytic responses on features of the book (using a tick (✓) in the appropriate boxes ☐)
   - additional comment on most items and on relevant matters not covered, to be written in the appropriate spaces
   - your evaluation of the book's coverage of the four skills, and of the book as a whole, to be written in the spaces provided.

We are grateful for your help and look forward to reading your responses. Thank you.

INSTRUMENT FOR THE ANALYSIS OF TEXBOOK MATERIALS (IATM)

0. The textbook being analysed:

Title:			
Author(s)			
Publishers			
Place of Publication		Year of publication	
What materials in addition to this book, if any, do you use when teaching students?			
Which students are you teaching using this book?			

1. What kind of book would you say this is?
(Please tick (✓) the box ☐ where appropriate).

1.1 a language teaching book with no specific reference to international tests ☐	1.2 a book of practice tests only ☐	1.3 a language teaching book and an international test preparation book combined ☐
1.4 If it is a test-related book, for which test(s)?		
1.5 Any other comment on the type of book this is?		

- ***If the book is a book of practice tests only, please go to Question 4.***
- ***If the book contains teaching material as well as practice tests, please go to Question 2, about the way the book is organised.***

2. The book's units / chapters etc. seem to be organised mainly according to:
(Please tick (✓) the box(es) ☐ where appropriate, more than one possible).

2.1 topics, themes ☐	2.2. language skills ☐	2.3 grammatical structures ☐	2.4 tests, tasks ☐	2.5 notions, functions ☐
2.6 other (please specify)				
2.7 Any further comment on the organisation of the book?				

3. Your analysis of the book's explicit practice of <u>language features.</u>
(Please tick (✓) appropriate boxes ☐.)

	A lot	A little	None		A lot	A little	None		A lot	A little	None
3.1 recognition of sounds	☐	☐	☐	3.2 pronunciation of sounds	☐	☐	☐	3.3 stress and intonation	☐	☐	☐
3.4 grammar	☐	☐	☐	3.5 sentence patterns	☐	☐	☐	3.6 notions, functions	☐	☐	☐
3.7 word formation	☐	☐	☐	3.8 connotation	☐	☐	☐	3.9 collocation	☐	☐	☐
3.10 idioms	☐	☐	☐	3.11 linking words, expressions	☐	☐	☐	3.12 punctuation	☐	☐	☐
3.13 Other language components or features given explicit practice:											
3.14 Related comments on how the book treats language features:											

4. <u>Enabling skills</u> you think are covered in the book:
(Please tick (✓) appropriate boxes ☐.)

4.1 understanding and conveying meaning through stress and intonation ☐	4.2 retrieving and stating factual information ☐	4.3 identifying main points ☐	4.4 drawing conclusions ☐
4.5 identifying overall meaning ☐	4.6 predicting information ☐	4.7 making inferences ☐	4.8 evaluating evidence ☐
4.9 distinguishing fact from opinion ☐	4.10 recognising roles ☐	4.11 identifying attitudes ☐	4.12 planning and organising information ☐
4.13 Other skills covered by the book (please specify):			
4.14 Further comment on skills covered or not covered by the book:			

5. Your summary of the use of question / tasking techniques in the book:
(Please tick (✓) appropriate boxes ☐.)

	Frequent	A little	None		Frequent	A little	None
5.1 multiple / dual choice	☐	☐	☐	5.2 conversion	☐	☐	☐
5.3 true / false	☐	☐	☐	5.4 sequencing	☐	☐	☐
5.5 matching	☐	☐	☐	5.6 paraphrasing	☐	☐	☐
5.7 substitution	☐	☐	☐	5.8 open-ended questions	☐	☐	☐
5.9 linking / joining	☐	☐	☐	5.10 note taking / making	☐	☐	☐
5.11 expansion	☐	☐	☐	5.12 correcting / editing	☐	☐	☐
5.13 gap filling / completion	☐	☐	☐	5.14 summarising	☐	☐	☐
5.15 Other techniques (please specify)							
5.16 Further comment on question and task techniques covered or not covered by the book							

6. Your evaluation of the extent to which the materials provide / encourage the following kinds of communicative opportunities. (Please tick (✓) appropriate boxes ☐.)

	A lot	Quite a lot	Very little	None		A lot	Quite a lot	Very little	None
6.1 pair communication	☐	☐	☐	☐	6.2 group discussions and debates	☐	☐	☐	☐
6.3 games, puzzles, quizzes	☐	☐	☐	☐	6.4 role play, simulations	☐	☐	☐	☐
6.5 surveys, other project work	☐	☐	☐	☐	6.6 report writing	☐	☐	☐	☐
6.7 review writing	☐	☐	☐	☐	6.8 essay writing	☐	☐	☐	☐
6.9 creative writing	☐	☐	☐	☐	6.10 IT e.g. telephone, fax, letters, email, web	☐	☐	☐	☐
6.11 listening, reading, viewing for personal interest	☐	☐	☐	☐	6.12 other communicative opportunities *(please specify)*:				
6.13 Further comment on the communicative opportunities offered by the book:									

7. How would you categorise the text types (heard, spoken, read, written) in the book?
(Please tick (✓) appropriate boxes ☐.)

7.1 public announcement ☐	7.2 lecture/ talk ☐	7.3 press report ☐	7.4 textbook/ journal article ☐
7.5 correspondence ☐	7.6 fiction ☐	7.7 discussion ☐	7.8 face-to-face conversation ☐
7.9 radio/TV report ☐	7.10 manual / brochure ☐	7.11 advertising ☐	7.12 maps, charts, tables and graphs ☐
7.13 interview ☐	7.14 telephone ☐	7.15 email ☐	7.16 internet ☐
7.17 other text type(s)? *(please specify)*			

8. And the book's text topics (heard, spoken, read, written)?
(Please tick (✓) appropriate boxes ☐.)

8.1 accommodation ☐	8.2 health ☐	8.3 physical environment ☐	8.4 leisure and sports ☐
8.5 daily routines ☐	8.6 education, training ☐	8.7 science ☐	8.8 travel ☐
8.9 shopping ☐	8.10 world of work ☐	8.11 arts ☐	8.12 current affairs ☐
8.13 food and drink ☐	8.14 social environment ☐	8.15 customs ☐	8.16 moral issues ☐
8.17 Other topics: *(please specify)*:			
8.18 Any inappropriate topics: *(please exemplify and explain)*:			

- ***If the book has no recorded texts, please go to Question 10.***

- Questions 9-11 seek information on the authenticity of the listening and reading texts and tasks.

9. Authenticity of listening texts and tasks:
(Please tick (✓) appropriate boxes ☐.)

9.1 Do the listening text(s) appear:	scripted? ☐	authentic? ☐	some of each? ☐
9.2 Do the recorded texts include redundancies such as:	repetition? ☐	rephrasing? ☐	hesitation? ☐
9.3 Please comment on the authenticity or realism of the listening tasks:			

10. Authenticity of reading texts and tasks: (please tick (✓) the appropriate boxes ☐)

10.1 Do the reading texts seem:	adapted or written for the book? ☐	authentic? ☐	some of each? ☐
10.2 Please comment on the authenticity or realism of the reading tasks:			

Most of the information you have been asked to provide so far has been relatively objective. Questions 11 and 12, here are very important as they request you to give your evaluation of how the book treats the main language skill areas, and of the book as a whole.

11. Please give your comments on the book's treatment of the four language skills:

11.1 Listening	
11.2 Reading	
11.3 Writing	
11.4 Speaking	

12. Please now evaluate the whole textbook, preferably in terms of:

• type • level • contents • pedagogical approach • interest

13. Finally please evaluate the book's relationship with the international test for which you use it to prepare your students.

How does the book help your students to cope with the international test you are preparing them for?

That is the end of the questionnaire. Thank you very much for answering the questions.

APPENDIX D:
Impact Study Lesson Observation Analysis Form

LESSON OBSERVATION ANALYSIS FORM			
Institution ID: **Type:**	**Lesson date, time, venue:**	**Teacher: background, qualifications, experience, other:**	**Materials, references, facilities, equipment etc:**
Course type, level, stage, main book(s), target exam:	**Lesson aim, type, etc:**		
Class number, m / f:	**Students present, m/f:**		
Nationalities, ages, TL level, etc:	**Recorded / not recorded**		
Other:	**Other:**		

Episode		Timing (mins)	Activity, participation, materials	Comment
1				
2				
3				
4				
5				
6				
7				
8				
9				

General Comments:

Student Communicative Opportunity Analysis data			
	Total time	**Time range across learners**	**Learner time average**
Class listening to T			
Individual/ pair/ gp / (other) listening to T			
Individual listening in pair / group / (other)			
Individual speaking in pair/ group / (other)			
Individual speaking to class			
Writing			
Reading			
Other:			

- *total times greater than lesson length because of some simultaneous communicative activities*
- *? denotes video does not reveal*

APPENDIX E:
Receiving Institution Questionnaire

Dear Colleague

As part of the continuing programme to update and refine its International English Language Testing System (IELTS), the University of Cambridge Local Examinations Syndicate (UCLES) is conducting a study of the impact of the test. In the course of the study, we are contacting students, candidates, English teachers, and university staff for information and comment.

As a member of an institution receiving students preparing for, or having taken the IELTS, your responses to this questionnaire will be valuable to our study. They will be treated in confidence and used only for its stated purpose.

We should be grateful if you would first sign the consent slip and answer the short background questionnaire below.

Thank you very much for your time and co-operation.

Yours sincerely

Nick Saville
EFL Director
Validation and Research
University of Cambridge Local Examinations Syndicate
English as a Foreign Language
Cambridge
England

Your consent to participate in the UCLES Impact Study

I understand that:

- the purpose of the study is to collect and analyse information from those familiar with international English language tests
- my name will not appear in any project publication
- the information I give, but not my name, may be quoted
- I am free to refuse to participate in the study and may withdraw at any time
- my completed questionnaire is for the study team only; it will not be shown to anyone not connected with the UCLES study.

Name: .....

Signature: ..... Date: .....

Thank you.

University of Cambridge Local Examinations Syndicate (UCLES) RIQ

QUESTIONNAIRE FOR RECEIVING INSTITUTIONS

Please write your response in the spaces provided, or tick (✓) the boxes as appropriate.

SECTION A: About you and any duties connected with IELTS

1.	Your full name and form of address (Miss, Mrs, Mr, Dr etc)	
2.	Name, address and email of institution where you work	
3.	Your position there, and main duties	
4.	If you teach / lecture at your institution, what subject(s)?	

In what way(s) have you been involved with IELTS tests? Please tick (✓) the appropriate box(es).

5.1	Advising international students, including on language proficiency matters	
5.2	Administration of university admissions	
5.3	Disseminating information, dealing with queries about language test(s)	
5.4	Test administration/ entry processing	
5.5	Test invigilating (proctoring/supervising))	
5.6	Administration of test preparation or other university entrance preparation courses	
5.7	Other (please specify)	

SECTION B: Your experience of and views on IELTS and other English language tests

- ***We know that you may not be in a position to answer all the questions, so there is a "don't know" option, which you can tick where appropriate.***
- ***Many of the questions also give space for you to comment. Your comments will be very helpful for our study.***

B1 : The use of IELTS and/or similar tests at you institution

		Yes	No	Don't know
6.1	Does your university / department use IELTS scores as an entry requirement?			
6.2	If so, what is / are the minimum IELTS level(s) required for the university / for your department(s)?			
6.3	Do you think that this / these requirements are fair?			
6.4	Who decides on the English language test(s) and the entry levels to be used by your institution?			
6.5	Any other comment on the use of IELTS results?			
7.1	Does your university / department use any other English language test scores as an entry requirement?			
7.2	If so, which test(s) and what are the minimum level(s) required for the university / for your department?			

7.3	Do you think that this / these requirements are fair?			
7.4	Any other comment on the use of other English language test results?			
8.1	Is/Are there other English language test(s) which you would like to see used as an entry requirement?			
8.2	If so what is it / are they?			
8.3	Please explain your answers:			
9.1	Have the English language entrance requirements at your institution changed recently?			
9.2	If yes, how and why?			

B2: The Validity and Reliability of IELTS and other tests

		Yes	No	Don't know
10.	Do you think that the IELTS test is appropriate for candidates who will study at university level?			
11.	Do you think that IELTS measures students' language proficiency effectively? (i.e. do only the students with adequate language abilities gain high scores and those without the necessary language ability gain low scores)?			
12.	Do you think that the IELTS score is a good predictor of academic success?			
13.	How do your students' results in the IELTS test compare with your own assessment of their language ability?			
14.1	Currently, IELTS candidates, regardless of their intended study areas, take the same IELTS language test. Do you think that this is reasonable?			
14.2	Is/Are there any subject area module(s) that you think should be added to the IELTS test?			
15.	Does a current IELTS Band 6 compare with a band 6 in previous years?			
16.1	Do you think that the IELTS test scores are biased against any of the nationality groups which you have taught?			
16.2	If so, please explain your answer.			
17.	If IELTS were not an entry requirement would you recommend students to take it anyway (*Please ignore the matter of the cost of taking the test for the purpose of this question*)?			
18.1	If you had the choice, would you include the IELTS or similar test score as a requirement for overseas applicants wishing to study in English at your institution?			
18.2	If not, what would be your alternative, and why?			
19.1	In the last two years have any of the international students you know about dropped out of their course because of language difficulties?			
19.2	If so, had they taken the IELTS test or similar tests?			
19.3	If you know what score(s) they had, what was it / were they?			
19.4	Your comments on international students' language difficulties, please:			
20.1	Do you feel that IELTS and/or similar language proficiency tests are an unreasonable barrier to overseas students wishing to study in an English-speaking environment?			
20.2	Please explain your answer:			
21.	Please now give your comments on the suitability of IELTS for your international students (compared with other English language proficiency tests if appropriate):			

B3: Your views on the impact of the IELTS				
22.	What impact do you think IELTS has? On:			
22.1	the students who take it:			
22.2	the teachers who prepare them for it:			
22.3	the subject lecturers who teach the students who have taken the IELTS:			
22.4	the university administration:			
22.5	Others *(please specify and comment)*:			
		Yes	**No**	**Don't know**
23.	Do think candidates need to prepare for IELTS using specifically designed materials?			
24.	Are candidates generally satisfied with their IELTS results?			
25.	Do you think that students who have taken the IELTS test have better study skills than those who haven't?			
26.	Is the standard of English of the IELTS candidates improving?			
27.	Do candidates have a better chance of getting good scores on IELTS if they have attended a preparation class?			
28.	Are there any particular areas of language knowledge or skills in which you feel that students who have taken the IELTS test need more proficiency?			
29.	Are teachers generally satisfied with their students' results on IELTS?			
30.	Do candidates feel that achieving a certain score on IELTS is merely an administrative obstacle to overcome?			
31.	Do IELTS candidates appear generally to suffer undue pressure from the IELTS test?			
32.	Does the registration fee for IELTS test act as a financial barrier to some candidates?			
33.	Do you have adequate access to information about the IELTS?			
34.	Do you have sufficient information to interpret IELTS scores?			
35.	Is the IELTS handbook available to you?			
36.	Please add your comments about the impacts of the IELTS test, compared, if appropriate, with other English proficiency tests?			

Many thanks for completing this questionnaire!

APPENDIX F:

Background information form for *PL2000* Impact Study schools

Full name of school							
Postal address phone, fax, email							
School Head							
Number of students in the school		Number of classes		Number of teachers		Number of English Teachers/ assistants	/
Number of *PL 2000* classes		Number of *PL2000* English classes		Hours a week for *PL* English classes			
Which foreign Languages are taught at the school?							
Which English Tests are taken in which classes / years?							
Which English language teaching textbooks are used?							
Does the school have access to a resource centre? Y/N ☐		Where is it?					
		What does it contain? How is it used?					
How do your *PL2000* students use their English while at, and after they leave, the school?							
What difference has the *PL2000* made to your school?							
This form completed by *(name, position, please)*							

Thank you very much for your help!

PLIS Q2 (10/01,04/02)

APPENDIX G:
PL2000 Impact Study questionnaire for students

About you and your language learning

Full name				School		Class	
Age		Male/ Female		Language(s) you study at school?		How many years have you studied English?	
Any extra English course(s) this school year? *(Yes/No)*		If yes, what course(s)? where, when?					

Please answer these questions about the use of English outside your school classes	Yes/No	If yes, where? when and how long?
Have you ever studied English abroad?		
Have you used English as a tourist?		
Have you ever used English in a job?		
Which English tests or examinations have you taken?		
Which English tests or examinations will you take in the future?		
What job / career are you aiming at for your future?		
Do you intend to study at University? if so, what subject(s)?		

Please put a tick (✓) in the boxes which show how often you do the following activities in your English classes at school.		Frequently	Quite Often	Sometimes	Never	Comments
1.	discussions in small groups					
2.	discussions with a partner					
3.	writing notes, letters or compositions					
4.	grammar exercises					
5.	discussions with the whole class					
6.	vocabulary exercises					
7.	listening to the teacher talking to the whole class					
8.	taking practice exams					
9	listening and taking notes					
10.	reading then writing answers to questions					
11.	reading texts (from books or other materials)					
12.	listening and choosing answers to questions					
13.	discussing the external exam					
14 +	Other activities in your English classes:					

Please put a tick (✓) in the column (a, b, c, or d) for the choice that best describes your response to each of the statements below.		a	b	c	d	Comments
1.	My English has improved this school year. a. A lot. b. Quite a lot. c. A little. d. Not at all.					
2.	My English language skill that has improved most is a. reading. b. listening. c. writing. d. speaking.					
3.	Most school days I spend .....studying English outside class. a. less than one hour b. one to two hours c. more than two hours d. no time					
4.	My main feeling about taking external exams in English is they a. are more difficult than school exams. b. help me to work harder. c. are for listening and speaking. d. are necessary to enter university.					
5.	My main reason for learning English is a. to communicate with more people b. to get a better job. c. for travel overseas. d. because I like it.					

Please put numbers in the boxes to show how often you do the following activities in English outside your school English classes: *0 = never; 1= almost never; 2 = occasionally; 3 = often*)

reading books		reading newspapers, magazines		writing letters		using email		watching TV	
listening to the radio		using the Internet		watching movies		going to shows		talking with visitors	
Other activities using English? *please specify*									

Please write here how you think your English language learning and attitudes have changed over the past year:

Thank you! PLIS SQ1A04/02

APPENDIX H:
PL2000 Impact Study questionnaire for teachers

About you

Your full name:															
Form of address:	Miss		Mrs		Mr		Dr		Other						
Your age, please:	20-30		31-40		41-50		51-60		61+						
Institution where you work							Your position there								
Your qualifications			Number of years teaching English				Types of schools								

About PL2000, your students and you

Please put a tick (✓) in the boxes which show how often you implement the following student activities in your *PL2000* English classes at school.		Frequently	Quite Often	Occasionally	Never	Comments
1.	discussions in small groups					
2.	pair discussions					
3.	writing notes, letters or compositions					
4.	doing grammar exercises					
5.	discussions involving the whole class					
6.	doing vocabulary exercises					
7.	listening to the teacher talking to the whole class					
8.	taking practice exams					
9	listening to recordings and taking notes					
10.	reading then writing answers to questions					
11.	reading texts (from books or other materials)					
12.	listening and choosing answers to questions					
13.	discussing the external exam					
14 +	Other activities in your English classes:					

Please put a tick (✓) in the boxes which show how well the following objectives of the *PL2000* have been achieved in your school:		Very well	Well	Not very well	Hardly at all	Comments
1.	modernising the teaching and learning of foreign languages					
2.	language teaching through modules with clear objectives, numbers of hours, content, expected outcomes					
3.	small homogeneous language learner groups					
4.	communication skills relevant to students' language needs					
5.	more self-access language learning					
6.	optimal use of modern technology in language learning					
7.	the use of external language exams at recognised Council of Europe levels					
8.	improved teacher support such as resource centres and teacher professional development					

Please list here and comment on what you see as the advantages and disadvantages of the *PL2000* for your students and you.

Please write here how you think your English language teaching and attitudes have changed over the past year:

PLIS TQ

Thank you very much!

APPENDIX I:

Semi-structured interview/focus group topics for *PL2000* Impact Study Italy visit 14–21 October 2001

0. Opening
   - intros and identities
   - *PL2000* Key points
   - *PL2000* Impact Study purpose, *bona fide*/ confidentiality, media, timing (ie max 30 mins)
   - gratitude

1. Knowledge of *PL2000*, how they know about it?

2. *PL2000* focus is language learning:
   - LL nowadays: teachers, learners
   - trends, reasons
   - methods
   - change / progress
   - standards

3. What about English?
   - ELL nowadays: teachers, learners
   - importance, reasons
   - purposes - global, local
   - change / progress
   - standards

4. English and other foreign language assessment?
   - *PL2000* stance on it
   - change, progress
   - external exams and exam routes
   - purposes and values

5. Other Views

6. Round-up
   - acknowledgements
   - next steps
   - thanks

PLIS Q3 RAH 10/01

References

Agar, M (1986) *Speaking of Ethnography*, Beverly Hills: Sage Publications.

Airasian, P (1988) *Measurement driven instruction: A closer look. Socio-cultural Issues in English for Academic Purposes*, London: Macmillan.

Alderson, J (1986) Innovations in language testing, in Portal, M (ed.) *Innovations in Language Testing*, Windsor: NFER Nelson, 93–105.

Alderson, J (1995) Ideas for research into impact, washback and IELTS, paper submitted to the University of Cambridge Local Examinations Syndicate.

Alderson, J (2004a) Introduction, in Cheng, L and Watanabe, Y (eds).

Alderson, J (2004b) The shape of things to come: will it be the normal distribution? in Milanovic, M and Weir, C (eds).

Alderson, J and Banerjee, J (1996) How might impact study instruments be validated? Paper commissioned by the University of Cambridge Local Examinations Syndicate (UCLES) as part of the IELTS impact study.

Alderson, J and Buck, G (1993) Standards in Testing: a study of the practice of UK examination boards in EFL/ESL testing, *Language Testing* 10, 1–26.

Alderson, J and Clapham, C (1993) *Examining the ELTS Test: an Account of the First Stage of the ELTS Revision Project*, IELTS Research Report 2, London, Cambridge and Canberra: British Council, UCLES, IDP Australia.

Alderson, J and Hamp-Lyons, L (1996) TOEFL preparation courses: a study of washback, *Language Testing* 13 (3), 280–297.

Alderson, J and Hughes, A (eds) (1981) *Issues in Language Testing*, ELT Documents 111, Oxford: Pergamon Press.

Alderson, J and North, B (eds) (1991) *Language Testing in the 1990s*, London: Macmillan.

Alderson, J and Wall, D (1993) Does washback exist? *Applied Linguistics* 14, 115–129.

Anderson, S and Gansneder, B (1995) Using electronic mail surveys and computer-monitored data for studying computer-mediated communication systems, *Social Sciences Computer Review*, 13 (1).

Andrews, S (1995) Washback or washout? The relationship between examination reform and curriculum innovation, in Nunan, D, Berry, V and Berry, R (eds) *Bringing about Change in Language Education*, Hong Kong: University of Hong Kong.

Andrews, S (2004) Washback and curriculum innovation, in Cheng, L and Watanabe, Y (eds).

Archer, T and Kendall, L (2003) Headteachers' main concerns, *Annual Trends in Education Digest* 12, Slough: NFER News, online.

Association of Language Testers in Europe (1994) *Code of Practice*, Cambridge: UCLES.

Association of Language Testers in Europe (1995) Development and descriptive checklists for tasks and examinations. Members of ALTE.

Association of Language Testers in Europe (1998) *ALTE Handbook of Language Examinations and Examination Systems*, Cambridge: UCLES.

Bachman, L (1990) *Fundamental Considerations in Language Testing*, Oxford: Oxford University Press.

Bachman, L (2005) Building and supporting a case for test use, *Language Assessment Quarterly* 2(1), 1–34.

Bachman, L, Cushing, S and Purpura, J (1993) Development of a questionnaire item bank to explore test-taker characteristics, Cambridge: UCLES interim internal report.

Bachman, L, Davidson, F, Ryan, K and Choi, I-C (1995) *An investigation into the comparability of two tests of English as a foreign language*, Cambridge: Cambridge University Press/UCLES.

Bachman, L and Palmer, A (1996) *Language Testing in Practice*, Oxford: Oxford University Press.

Baker, C (1997) Survey Methods in Researching Language and Education, in Hornberger and Corson (eds) *Encyclopedia of Language and Education, Volume 8: Research Methods in Language and Education*, Dordrecht: Kluwer Academic Publishers.

Banerjee, J (2003) Interpreting and using proficiency test scores, unpublished PhD thesis, University of Lancaster.

Banerjee, J (1996) The design of the classroom observation instrument, Cambridge: UCLES report.

Blackhurst, A (2004) IELTS test performance data 2003, *Research Notes* 18, 18–21.

Bogdan, R and Biklen, S (1982) *Qualitative Research for Education: An introduction to theory and methods*, Boston: Allyn and Bacon.

Bonkowski, F (1996) Instrument for the assessment of teaching materials, unpublished manuscript, Lancaster University.

Bracey, G (1987) Measurement-driven instruction: catch phrase, dangerous practice, *Phi Delta Kappa* 68 (9).

Brown, A (2002) An examination of candidate discourse in the revised IELTS speaking test, British Council/IDP Australia IELTS Research Report.

Brown, A (forthcoming) An investigation of raters' orientation in awarding scores in the IELTS oral interview, in Taylor, L and Falvey, P (eds).

Brown, A and Hill, K (forthcoming) Interviewer style and candidate performance in the IELTS oral interview, in Taylor, L and Falvey, P (eds).

Bunton, D (1998) Linguistic and textual problems in Ph.D and M.Phil theses: an analysis of genre moves and metatext, unpublished PhD thesis, University of Hong Kong.

Chapelle, C, Enright, M and Jamieson, J (2004) Issues in developing a TOEFL validity argument, draft paper presented at LTRC, California, March 2004.

Chapman, D and Snyder, C (2000) Can high stakes national testing improve instruction? Reexamining conventional wisdom, *International Journal of Educational Development* 20 (6), 457–474.

Charge, N and Taylor, L (1997) Recent Developments in IELTS, *ELT Journal* 51 (4), 374–380.

Cheng, L (1997) How does washback influence teaching? Implications for Hong Kong, *Language and Education* 11, 38–54.

Cheng, L (1998) The washback effect of public examination change on classroom

teaching. An impact study of the 1996 Hong Kong Certificate of Education in English on classroom teaching of English in Hong Kong secondary schools, unpublished PhD, University of Hong Kong.

Cheng, L and Watanabe, Y (eds) (2004) *Washback in Language Testing: Research Contexts and Methods*, New Jersey: Lawrence Erlbaum Associates.

Cheng, L and Curtis, A (2004) Washback or Backwash: A review of the impact of testing on teaching and learning, in Cheng and Watanabe (eds) (2004).

Clapham, C (1996) *The development of IELTS: A study of the effect of background knowledge on reading comprehension*, Studies in Language Testing 4, Cambridge: UCLES, Cambridge University Press.

Connell, J, Kubisch, A, Schorr, L and Weiss, C (eds) (1995) *New Approaches to Evaluating Community Initiatives: Concepts, methods, and contexts*, Washington DC: The Aspen Institute.

Cook, T and Reichardt, C (1979) *Qualitative and quantitative methods in evaluation research*, Beverly Hills: Sage Publications.

Cook, V (2003) Introduction: the changing L1 in the L2 user's mind, in Cook, V (ed.) *Effects of the Second Language on the First*, Clevedon: Multilingual Matters.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, teaching, assessment*, Cambridge: Cambridge University Press.

Council of Europe (2003) Relating language Examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment. Manual. Draft.

Criper, C and Davies, A (1988) ELTS Validation Project Report, ELTS Research Report 1 (i).

Cronbach, L (1982) *Designing Evaluations of Educational and Social Programs*, San Francisco: Jossey-Bass.

Davies, A (1996) Ironising the myth of linguicism, *Journal of Multilingual and Multicultural Development* 17 (6), 485–96.

Davies, A (1997) Demands of being professional in language testing, *Language Testing* 14 (3) 328–339.

Davies, A (forthcoming) *Assessing Academic English: testing English proficiency 1950–2005 – the IELTS solution,* Studies in Language Testing 23, Cambridge: UCLES, Cambridge University Press.

Davis, K (1995) Qualitative theory and methods in applied linguistics research, *TESOL Quarterly* 29, 3.

Davies, A, Brown, A, Elder, C, Hill, K, Lumley, T and McNamara, T (1999) *Dictionary of language testing*, Cambridge: Cambridge University Press/UCLES.

Denzin, N (1978) *Research Act: Theoretical Introduction to Sociological Methods*, New York: McGraw-Hill.

Department for International Development (DFID) (1998) Glossary of terms, internal DFID document.

Duff, P (2002) Beyond generalizability: contextualization, complexity and credibility in applied linguistics research, presentation to the AAL / ILTA colloquium, Salt Lake City, April 6, 2002.

Duguid, A (2001) *Anatomy of a Context : English Language Teaching in Italy*, London: Granville.

Farley Templeton, J (1996) *The Focus Group: A Strategic Guide to Organizing, Conducting and Analyzing the Focus Group Interview*, New York: McGraw-Hill.
Fitz-Gibbon, C and Morris, L (1987) *How to Analyse Data*, Newbury Park: Sage Publications.
Franceschina, F and Banerjee, J (2004) Documenting features of written language production typical at different IELTS bands levels, British Council/ IDP Australia IELTS Research Report.
Frankfort-Nachmias, C and Nachmias, D (1996) *Research Methods in the Social Sciences*, London: Arnold.
Gardiner, K (1999) Analysis of IELTS impact test-taker background questionnaire, Cambridge: UCLES IELTS EFL validation report.
Glaser, B and Strauss, A (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Chicago: Aldine Publishers.
Gunter, B (2001) *Media Research Methods: measuring audiences, reactions, impact*, London: Sage Publications.
Green, A (2003a) Test impact and English for academic purposes: a comparative study in backwash between IELTS preparation and university pre-sessional courses, unpublished PhD thesis, Centre for Research in Testing, Evaluation and Curriculum in ELT, University of Surrey, Roehampton.
Green, A (2003b) Monitoring score gain on the IELTS Academic Writing module in EAP programmes of varying duration: Phase 2 report, Cambridge: UCLES internal report.
Green, A (2004) Making the grade: score gains on the IELTS Writing test, *Research Notes* 16, 9–13.
Green, A and Hawkey, R (2004) Test Washback and Impact, *Modern English Teacher* 13 (4), 66–70.
Hamp-Lyons, L (1997) Washback, impact and validity: ethical concerns, *Language Testing* 14 (3), 295–303.
Hamp-Lyons, L (1998) Ethical test preparation practice: the case of the TOEFL, *TESOL Quarterly* 32 (2), 329–337.
Hamp-Lyons, L (2000) Social, professional and individual responsibility in language testing, *System* 28, 579–591.
Hampton, C and Huang Chung (2003) What makes a good IELTS speaking test? Perceptions of candidates and examiners, British Council/ IDP Australia IELTS Research Report.
Hawkey, R (2000) Draft Survey Questions and Recommendations on Specifications for the Institutional Survey Instrument, consultancy report to the British Council.
Hawkey, R (2001) Phase 2 Focus Groups: Analysis relating Participant Comment to other Current IELTS Research Projects, consultancy report to the British Council.
Hawkey, R (2003a) The Cambridge ESOL *Progetto Lingue 2000* Impact Study: Full report to UCLES.
Hawkey, R (2003b) IELTS Impact Study: Phase 3: Report to UCLES.
Hawkey, R (2005) *A Modular Approach to Testing English Language Skills,* Cambridge: Cambridge University Press/UCLES.
Hawkey, R (2006) Teacher and learner perceptions of language learning activity, *English Language Teaching Journal* 60 (3).
Hawkey, R and Green, A (2000) Comments on the Test takers' Background Questionnaires, Cambridge: Reports to UCLES.

Hawkey, R, Thompson, S and Turner, R (forthcoming) The development of a video database for language educational projects.

Hayes, B and Read, J (2004) IELTS test preparation in New Zealand: preparing students for the IELTS academic module, in Cheng, L and Watanabe, Y (eds).

Heath, S (1982) Ethnography in education: defining the essentials, in Gilmore, P and Glatthorn, A (eds) *Children in and out of School: Ethnography and education*, Washington DC: Centre for Applied Linguistics.

Henerson, M, Lyons Morris, L and Fitz-Gibbon, C (1987) *How to Measure Attitudes,* Newbury Park: Sage Publications.

Henning, G (1987) *A Guide to Language Testing*, Boston, Massachusetts: Heinle and Heinle.

Herington, R (1996) Test-takers' background questionnaire design characteristics: Report to UCLES.

Horak, T (1996) IELTS Impact Study Project, unpublished MA assignment, Lancaster University.

Hopkins, D (1985) *A teacher's guide to classroom research*, Philadelphia: Open University Press.

Hughes, A (2003) *Testing for Language Teachers*, Cambridge: Cambridge University Press.

Ingram, D, Bayliss, A and Paul, A (2004) IELTS as a predictor of academic language performance, British Council/ IDP Australia IELTS Research Report.

International English Language Testing System, Australia (2004) *IELTS Band Score Requirements,* IELTS Australia.

International Language Testing Association (ILTA) (1997) Draft code of practice for foreign/second language testing, Lancaster: ILTA.

Jones, N (2001) Reliability in UCLES Examinations, Cambridge: UCLES internal report.

Jones, N and Hirtzel, M (2001) Appendix D: The ALTE 'Can do' statements, in *Council of Europe Framework of Reference for Languages*.

Kane, M (1992) An argument-based approach to validity, *Psychological Bulletin* 112, 527–535

Kane, M, Crooks, T and Cohen, A (1999) Validating measures of performance, *Educational Measurement: Issues and Practices* 18 (2), 5–17.

Kelly, G (1955) *The Psychology of Personal Constructs*, New York: Norton.

Kennedy, C and Thorp, D (forthcoming) Investigation of linguistic output of Academic Writing Task 2, in Taylor, L and Falvey, P (eds).

Knapp, L (2003) Summary comments from UK centres, Cambridge: UCLES internal report.

Kunnan, A (2000) Reviewing the instruments (scales) for impact of the IELTS study, Cambridge: Report to UCLES.

Larsen-Freeman, D and Long, M (1991) *An Introduction to Second Language Acquisition Research*, London: Longman.

Lazaraton, A (1995) Qualitative research in applied linguistics: a progress report, *TESOL Quarterly* 29, 3.

Lazaraton, A (2004) Quantitative research methods in language test development and validation, in Milanovic, M and Weir, C (eds).

Levine, T (2002) Stability and change in curriculum evaluation, *Studies in Educational Evaluation* 28, 1–33.

Light, R, Singer, J and Willett, J (1990) *By Design: Planning research on higher education*, Cambridge Mass: Harvard University Press.

Litwin, M (1995) *How to Measure Survey Reliability and Validity*, Thousand Oaks: Sage Publications.

Luxon, A (1996) ODA ELT Project: Nicaragua – A Case Study in Project Management, in *Encuentro*, University of Alcala de Henares.

MacDonald, R (1974) Evaluation and the control of education, in MacDonald, B and Walker, R (eds) *Innovation, Evaluation research and the Problem of Control*, CARE, University of East Anglia.

May, N and Sigsworth, A (1983) Teacher-outsider partnerships in the observation of classrooms, in Murphy, R and Torrance, H (eds).

Mayor, B, Hewings, A and Swann, J (2000) An analysis of the linguistic features of output from IELTS academic writing tasks 1 and 2, British Council/ IDP Australia IELTS Research Report.

McNamara, T (1996) *Measuring Second Language Performance*, Harlow: Longman.

McNamara, T (2000) *Language Testing*, Oxford: Oxford University Press.

Mehrens, W (1998) Consequences of Assessment: What is the Evidence? Education Policy Analysis Archives 6 Number 13: http://epaa.asu.edu/epaa/v6n13.html

Mehta, R and Sivadas, E (1995) Comparing response rates and response content in mail versus electronic mail surveys, *Journal of the Market Research Society* 37 (4) 429–439.

Merrylees, B (1997) A monitoring program of examiner performance in IELTS Australia centres, British Council/ IDP Australia IELTS Research Report.

Merrylees, B (1998) An impact study of two IELTS user groups: immigration and secondary, British Council/ IDP Australia IELTS Research Report.

Messick, M (1996) Validity and washback in language testing, *Language Testing* 13 (3), 241–256.

Messick, S (1989) Validity, in Linn (ed.) *Educational Measurement*, New York: ACE/Macmillan.

Mickan, P (1998) A study of the response validity of the IELTS writing test – Stage 2, British Council/ IDP Australia IELTS Research Report.

Mickan, P (2004) An ethnographic study of classroom instruction in an IELTS preparation program, British Council/ IDP Australia IELTS Research Report.

Milanovic, M and Saville, N (1992) Principles of Good Practice for UCLES Examinations: a discussion paper, Cambridge: UCLES internal report.

Milanovic, M and Saville, N (1996) Considering the impact of Cambridge EFL examinations, Cambridge: UCLES internal report.

Milanovic, M and Weir, C (eds) (2004) *European Language Testing in a Global Context: Proceedings of the ALTE Barcelona Conference July 2001*, Cambridge: UCLES, Cambridge University Press.

Miller, G (1997) Building Bridges: the possibility of analytic dialogue between ethnography, conversation analysis and Foucault, in Silverman (ed.) (2001).

Miller, N (2002) Questionnaires, in *The Handbook for Economic Lecturers*, University of Bristol: The Higher Education Academy.

Ministry of Education, Italy (1999) Progetto Lingue in Comunicazione di Servizio, 23 settembre 1999, under *Circolari ministeriali e comunicazione di servizio* at www.istruzione.it/argomenti/autonomia/documenti

Mislevy, R, Steinberg, L and Almond, R (2002) Design & analysis in task-based language assessment, *Language Assessment* 19, 477–496.

Mislevy, R, Steinberg, L and Almond, R (2003) On the structure of educational assessments, *Measurement: Interdisciplinary Research & Perspectives* 1(1), 3–62.

Moore, T and Morton, J (forthcoming) Authenticity in the IELTS Academic Writing module: a comparative study of Task 2 items and university assignments, in Taylor, L and Falvey, P (eds).

Morrison, D (1998) *The Search for a Method: Focus Groups and the Development of Mass Communication,* Luton: University of Luton Press.

Munby, J (1978) *Communicative Syllabus Design*, Cambridge: Cambridge University Press.

Murphy, R and Torrance, H (eds) (1987) *Evaluating Education: Issues and methods*, London: Paul Chapman.

O'Dwyer, J (forthcoming) The management of a process of curriculum development: a longitudinal case study of the use of formative evaluation for curriculum renewal and development in a large English language teaching institution, unpublished PhD thesis, University of Roehampton.

O'Loughlin, K (forthcoming) The impact of gender in the IELTS oral interview, in Taylor, L and Falvey, P (eds).

O'Loughlin, K and Wigglesworth, G (forthcoming) Task design in Academic Writing Task 1: the effect of quantity and manner of presentation of information on candidate writing, in Taylor, L and Falvey, P (eds).

O'Sullivan, B, Weir, C and Horai, T (2003) Exploring difficulty in speaking tasks: an intra-task perspective, British Council/ IDP Australia IELTS Research Report.

O'Sullivan, B and Rignall, M (2000) Assessing the value of bias analysis feedback to raters for the IELTS Writing module, British Council/ IDP Australia IELTS Research Report.

O'Sullivan, B and Yu, L (2002) An empirical study on examiner deviation from the set interlocutor frames in the IELTS Speaking Test, British Council/ IDP Australia IELTS Research Report.

Oxford, R (1990) *Language Learning Strategies*, New York: Newbury House.

Parlett, M and Hamilton, D (1972) Evaluation as Illumination: a new approach to the study of innovatory programmes, in Glass, G (ed.) *Evaluation Studies Review Annual*, Vol. 1, Beverly Hills: Sage Publications.

Pavians, M, Bertocchi, D, Hofmannova, M and Kazianka, M (2001) *Introducing CLIL Content and Language Integrated Learning. Guides for teachers and schools to use foreign languages in content teaching*, edited in five languages by Gisella Langé, Milan: Ministry of Education Regional Direction Lombardy.

Patton, M (1987) *How to use Qualitative Methods in Evaluation*, London: Sage Publications.

Popham, W (1987) Two-plus decades of educational objectives, *International Journal of Educational Research* 11 (1), 31–41.

Popham, W (1993) Measurement-driven instruction as a 'quick-fix' reform stratagy, *Measurement and Evaluation in Counseling and Development* 26, 31–34.

Purpura, J (1994) The development and construct validation of an instrument designed to investigate selected cognitive background characteristics of test takers, unpublished doctoral qualifying paper, Dept of TESL/Applied Linguistics, UCLA.

Purpura, J (1996a) Modeling the relationships between test takers' reported cognitive

and meta-cognitive strategy use and performance on language tests, unpublished doctoral thesis, UCLA.
Purpura, J (1996b) What is the relationship between students' learning strategies and their performance on language tests? unpublished paper given at UCLES seminar, Cambridge, June 1996.
Purpura, J (1999) *Learner Strategy Use and Performance on Language Tests: A structural equation modelling approach*, Cambridge: UCLES, Cambridge University Press.
Purpura, J (2001) A proposal for continued research on the UCLES Language Learning Questionnaire item bank, Cambridge: paper to UCLES.
Purpura, J (2004) Validating questionnaires to examine personal factors in L2 test performance, in Milanovic, M and Weir, C (eds).
Ramirez, A (1999) Assessment-Driven Reform: The Emperor still has no clothes, *Phi Delta Kappan* 81 (3), 204–208.
Rea-Dickins, P (1997) So why do we need relationships with stakeholders in language testing? A view from the UK, *Language Testing* 14 (3), 304–314.
Rea-Dickins, P, Kiely, R and Yu (2003) Student identity, learning and progression: with specific reference to the affective and academic impact of 'successful' IELTS students, British Council/ IDP Australia IELTS Research Report.
Read, J and Nation, P (2002) An investigation of the lexical dimension of the IELTS speaking test, British Council/ IDP Australia IELTS Research Report.
Reichardt, C and Cook, T (eds) (1979) *Qualitative and Quantitative Methods in Evaluation Research,* Beverly Hills: Sage Publications.
Rignall, M and Furneaux, C (1997) *Speaking*, London: Longman.
Roberts, J (1993) Evaluating the impacts of teacher research, *System* 21 (1), 1–9.
Roy, S (1998) A general framework for evaluating educational programmes, in McKay, V and Treffgarne, C (eds.) *Evaluating Impact*, London: Department for International Development, 69–74.
Satterly, D (1989) *Assessment in Schools*, Oxford: Basil Blackwell.
Saville, N (2001) Investigating the impact of international language examinations. *Research Notes* 2.
Saville, N (2003) The process of test development and revision within UCLES EFL, in Weir, C and Milanovic, M (eds).
Saville, N (2004) The ESOL test development and validation strategy, Cambridge: Cambridge ESOL internal report.
Saville, N and Hawkey, R (2004) The IELTS Impact Study: Investigating washback on teaching materials, in Cheng, L and Watanabe, Y (eds).
Saxe, L and Fine, M (1979) Expanding our view of control groups in evaluations, in Datta, L and Perfoff, R (eds) *Improving Evaluations*, Beverly Hills: Sage Publications.
Scriven, M (1967) The methodology of evaluation, in Tyler, R, Gagne, M and Scriven, M (eds) *Perspectives of Curriculum Evaluation*, AERA Monograph Series on Curriculum Evaluation, 1, 39–83, Chicago: Rand McNally.
Seedhouse, P and Egbert, M (2004) The international organisation of the speaking test, British Council/ IDP Australia IELTS Research Report.
Selwyn, N and Robson, K (1998) Using e-mail as a research tool, *Social Research Update* 21, Department of Sociology: University of Surrey.
Shaw, S (2002) IELTS Writing: revising assesment criteria, *Research Notes* 9.

Shaw, S (2004) IELTS Writing: revising assessment criteria and scales (Phase 3), *Research Notes* 16, 3–7.
Shohamy, E (1997) Testing methods, testing consequences: Are they ethical? Are they fair? *Language Testing* 14 (3), 340–349.
Shohamy, E (1999) Language Testing: Impact, in Spolsky, B (ed.).
Shohamy, E (2001) *The Power of Tests: A Critical Perspective on the Use of Language Tests*, Harlow: Pearson Education.
Silverman, D (ed.) (2001) *Qualitative Research: Theory, Method and Practice*, London: Sage Publications.
Skehan, P (1989) Individual Differences in Second Language Learning, London: Edward Arnold.
Smith, H and Haslett, S (2003) Attitudes of tertiary key decision makers towards English tests: A New Zealand case study, British Council/ IDP Australia IELTS Research Report.
Smith, J (2004) IELTS Impact: a study on the accessibility of IELTS GT Modules to 16–17 year old candidates, *Research Notes* 18, 6–8.
Snodgrass, J (1977) *The Numbers Game*, Oxford: Oxford University Press.
Spolsky, B (1989) *Conditions for second language learning: Introduction to a general theory*, Oxford: Oxford University Press.
Spolsky, B (1995) *Measured Words: The Development of Objective Language Testing*, Oxford: Oxford University Press.
Spolsky, B (ed.) (1999) *Concise Encyclopedia of Educational Linguistics*, Oxford: Pergamon.
Swan, A and Gibson, C (2004) The significance of socio-linguistic backgrounds of teachers of IELTS preparation courses in selected Malaysian institutions, British Council/ IDP Australia IELTS Research Report.
Taylor, L (1999) Constituency matters: responsibilities and relationships in our testing community, paper delivered to the Language Testing Forum, University of Edinburgh, 19–21 November 1999.
Taylor, L (2001) Revising the IELTS Speaking Test, *Research Notes* 4.
Taylor, L (2001) Revising the IELTS Speaking Test: developments in test format and task design, *Research Notes* 5.
Taylor, L (2001) Revising the IELTS Speaking Test: retraining IELTS examiners worldwide, *Research Notes* 6.
Taylor, L (2004) IELTS, Cambridge ESOL examinations and the Common European Framework, *Research Notes* 18.
Taylor, L and Falvey, P (eds) (forthcoming) *IELTS Collected Papers: Research in speaking and writing assessment*, Cambridge: UCLES, Cambridge University Press.
Templeton, J Farley (1996) *The Focus Group: A Strategic Guide to Organizing, Conducting and Analyzing the Focus Group Interview*, New York: McGraw-Hill.
Tindall, D (2001) Review of Giugni, M, McAdam, A and Tilly, C (eds) *How Social Movements Matter,* Minnesota: University of Minnesota Press, *Canadian Journal of Sociology* 26 (4), 678–682
Toulmin, P (2003) *The uses of argument*, Cambridge: Cambridge University Press.
Turner, R, Hawkey, R and Thompson, S (forthcoming) Developing a video database for classroom observation research.

University of Cambridge ESOL Examinations (2005) *Teaching Knowledge Test Handbook.*

University of Cambridge ESOL Examinations, British Council, IDP Education in Australia and IELTS Australia (2003) *IELTS Annual Review*.

University of Cambridge Local Examinations Syndicate (1999) *The IELTS Handbook,* UCLES, The British Council, IDP Education, IELTS Australia.

Varghese, N (1998) Evaluation vs Impact studies, in McKay, V and Treffgarne, C (eds) *Evaluating Impact*, London: Department for International Development, 47–54.

Vint, L (2004) *PL2000*, an Italian government initiative to improve language learning brings home the bacon: examining the results, *EL Gazette*, 293.

Wall, D (1997) Impact and Washback in Language Testing, in Clapham, C and Corson, D (eds) *Encyclopaedia of Language and Education, Volume 7: Language Testing and Assessment*, Dordrecht: Kluwer Academic Publishers, 334–343.

Wall, D and Alderson, J (1993) Examining Washback: the Sri Lankan Impact Study, *Language Testing* 10, 41–70.

Watanabe, Y (1996) Investigating washback in Japanese EFL classrooms: Problems of methodology, in Wigglesworth, G and Elder, C (eds) *The language testing circle: From inception to washback,* Melbourne: Applied Linguistics Association of Australia.

Watanabe, Y (2000) Washback Effects of the English Section of Japanese University Entrance Examinations on Instruction in Pre-College Level EFL, *Language Testing Update* 27, 42–27.

Watanabe, Y (2004) Methodology in washback studies, in Cheng, L and Watanabe, Y (eds).

Weir, C (2004) *Language Testing and Validity Evidence*, Basingstoke: Palgrave.

Weir, C and Milanovic, M (eds) (2003) *Continuity and innovation: Revising the Cambridge Proficiency in English Examination 1913–2002*, Cambridge: UCLES, Cambridge University Press.

Weir, C and Roberts, J (1994) *Evaluation in ELT*, Oxford: Blackwell.

Weir, C and O'Sullivan, B (2002) Does the computer make a difference? Reactions of candidates to a CBT versus traditional hand-written form of IELTS Writing component: effects and impact, British Council/ IDP Australia IELTS Research Report.

Weiss, C (1998) *Evaluation*, New Jersey: Prentice Hall.

Winetroube, S (1997) The design of the teachers' attitude questionnaires, Cambridge: UCLES internal report.

Yue, W (1997) An investigation of textbook materials designed to prepare students for the IELTS test: A study of washback, unpublished Master thesis, Lancaster University.

Index

C

D

E

F

G

H

I

Q

R

S

T

Y