Research and Practice in Assessing Academic Reading: The Case of IELTS

Also in this series:

IELTS Washback in Context: Preparation for academic writing in higher education Anthony Green

Examining Writing: Research and practice in assessing second language writing Stuart D Shaw and Cyril J Weir

Multilingualism and Assessment: Achieving transparency, assuring quality, sustaining diversity – Proceedings of the ALTE Berlin Conference, May 2005 Edited by Lynda Taylor and Cyril J Weir

Language Testing Matters: Investigating the wider social and educational impact of assessment – Proceedings of the ALTE Cambridge Conference, April 2008 Edited by Lynda Taylor and Cyril J Weir

Components of L2 Reading: Linguistic and processing factors in the reading test performances of Japanese EFL learners *Toshihiko Shiotsu*

Aligning Tests with the CEFR: Reflections on using the Council of Europe's draft Manual Edited by Waldemar Martyniuk

Examining Reading: Research and practice in assessing second language reading Hanan Khalifa and Cyril J Weir

Examining Speaking: Research and practice in assessing second language speaking *Edited by Lynda Taylor*

IELTS Collected Papers 2: Research in reading and listening assessment Edited by Lynda Taylor and Cyril J Weir

Examining Listening: Research and practice in assessing second language listening *Edited by Ardeshir Geranpayeh and Lynda Taylor*

Exploring Language Frameworks: Proceedings of the ALTE Kraków Conference, July 2011 *Edited by Evelina D Galaczi and Cyril J Weir*

Measured Constructs: A history of Cambridge English language examinations 1913–2012 Cyril J Weir, Ivana Vidaković, Evelina D Galaczi

Cambridge English Exams – The First Hundred Years: A history of English language assessment from the University of Cambridge 1913–2013 *Roger Hawkey and Michael Milanovic* Testing Reading Through Summary: Investigating summary completion tasks for assessing reading comprehension ability Lynda Taylor

Multilingual Frameworks: The construction and use of multilingual proficiency frameworks Neil Jones

Validating Second Language Reading Examinations: Establishing the validity of the GEPT through alignment with the Common European Framework of Reference Rachel Yi-fen Wu

Assessing Language Teachers' Professional Skills and Knowledge Edited by Rosemary Wilson and Monica Poulter

Second Language Assessment and Mixed Methods Research Edited by Aleidine J Moeller, John W Creswell and Nick Saville

Language Assessment for Multilingualism: Proceedings of the ALTE Paris Conference, April 2014

Edited by Coreen Docherty and Fiona Barker

Learning Oriented Assessment: A systemic approach Neil Jones and Nick Saville

Advancing the Field of Language Assessment: Papers from TIRF doctoral dissertation grantees Edited by MaryAnn Christison and Nick Saville

Examining Young Learners: Research and practice in assessing the English of school-age learners

Szilvia Papp and Shelagh Rixon

Second Language Assessment and Action Research

Edited by Anne Burns and Hanan Khalifa

Applying the socio-cognitive model to the BioMedical Admissions Test

Edited by Kevin Y F Cheung, Sarah McElwee and Joanne Emery

Research and Practice in Assessing Academic Reading: The Case of IELTS

Cyril J Weir

Powdrill Professor in English Language Acquisition, University of Bedfordshire

and

Sathena Chan CRELLA, University of Bedfordshire

With contributions from Lynda Taylor CRELLA, University of Bedfordshire



CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom One Liberty Plaza, 20th Floor, New York, NY 10006, USA 477 Williamstown Road, Port Melbourne, VIC 3207, Australia 314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India 79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org Information on this title: www.cambridge.org/9781108733618

© Cambridge University Press 2019

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2019

20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Printed in

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

ISBN 978-1-108-73361-8

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate. Information regarding prices, travel timetables, and other factual information given in this work is correct at the time of first printing but Cambridge University Press does not guarantee the accuracy of such information thereafter.

Contents

Acl Ser Lis	Acknowledgements Series Editors' note List of abbreviations			
Section 1 The nature of academic reading				
1	Introduction: Some general considerations	3		
2	Establishing the cognitive and contextual parameters in academic reading: Insights from empirical research	22		
3	Modelling the construct of academic reading: Insights from	22		
	theoretical research	58		
Section 2 Academic reading tests in practice: The case of the IELTS Academic Reading Module				
4	Cognitive validity parameters	107		
5	Contextual validity parameters	149		
6 7	Consequential validity parameters	180		
Section 3 Enhancing validity in academic reading assessment				
8 9	The use of reading-into-writing tasks to assess academic reading Assessing academic reading in a digital age	221 242		
References and further reading Author index Subject index				

Acknowledgements

In bringing this volume to fruition, we are deeply indebted to a number of colleagues. Their expertise, sound advice and collaboration have undoubtedly contributed to the quality of this volume. Special thanks must go to Dr John Field (CRELLA, University of Bedfordshire) who reviewed Chapters 2, 3 and 4 and was a constant source of encouragement and insight, and also to Professor Barry O'Sullivan (British Council) who provided thoughtful comments on the whole volume. Professor Lynda Taylor (CRELLA, University of Bedfordshire) and Dr Nathaniel Owen (The Open University) served as external reviewers once the complete manuscript had been assembled. The discerning comments and constructive criticism of these reviewers enabled us to revise and refine the manuscript at various points before it went into production.

We are extremely grateful to a large number of other people from both academic and practitioner communities within and beyond Cambridge Assessment English who also provided valuable input to or feedback on the individual sections and chapters. Shigeko Weir and Nicola Latimer were a great help in tracking down information electronically and in compiling the rather extensive set of references.

We also acknowledge the support and encouragement of Dr Nick Saville (Director of Research and Thought Leadership at Cambridge Assessment English) who has maintained a close involvement with the volume throughout its development. Grateful thanks must also go to John Savage, Publications Assistant at Cambridge Assessment English, for his careful and supportive management of the publication project from initial manuscript through to printed volume.

Cambridge Assessment English funded an initial position paper on academic literacy that forms the basis for Chapter 1. Cambridge Assessment English and the British Council funded the research on IELTS that much of this book is based on and for which we are very grateful. We are grateful to the Steel Charitable Trust for their funding for Cyril Weir's Powdrill Professorship and their continuing support for our research on academic literacy assessment.

Thus the expertise, sound advice and collaboration of a large number of people have contributed significantly to the quality of this volume. To all of these and to any others we have failed to mention, we extend our sincere thanks and appreciation.

Series Editors' note

Introduction

This volume is aimed primarily at those working professionally in the field of language testing such as key personnel in examination/testing agencies and those with an interest in language testing and/or English for Academic Purposes (EAP). It is intended to provide a coherent account of the theoretical construct on which academic reading tests should be based and of the rigorous procedures that need to be followed to provide evidence concerning the various aspect of a test's validity which, when taken together, offer an argument for it being a sound measure of the academic reading construct. As such it is hoped that it will offer other institutions a useful framework for reviewing their own academic reading examinations/tests.

It stands as the fifth in the SiLT series of construct volumes designed to explore the theoretical underpinnings of the testing of adult English language skills (see Shaw and Weir 2007 on Examining Writing, Khalifa and Weir 2009 on Examining Reading, Taylor (Ed) 2011 on Examining Speaking and Geranpayeh and Taylor (Eds) 2013 on *Examining Listening*). The specific focus of the present volume is on the testing of second language academic reading ability with particular reference to the IELTS Academic Reading Module. It slightly modifies the sociocognitive validation framework employed in those earlier volumes for general English testing, in that it deals with the specific case of academic reading and reading-into-writing examinations. In addition, the volume uniquely examines the operationalisation of the framework by way of a critical evaluation of the IELTS test instead of the Cambridge English Qualifications that were the focus of the earlier volumes. This evaluation provides the context for the framework's exegesis, as in Chapters 4-7 IELTS practice in the Academic Reading Module is reviewed in terms of the individual aspects of the framework, the results of which, when taken together, constitute the validity argument for the test. Messick (1995:747) points out in relation to current orthodoxy on establishing validity for a test:

... what differentiates the approach to the old opportunistic nature of evidence collection was the need to touch all of the relevant bases by explicitly including a reference to all six aspects in an integrated validity argument ...

The unitary nature of the current consensus view of validity is discussed further in the next section.

Validity and validation

Language testing is faced with increasing demands for accountability in respect of all examinations/tests offered to the public. Examination boards are increasingly being required by their own governments and by other international authorities to demonstrate that the examinations they offer are well grounded in the language ability constructs they are attempting to measure. An explicit test validation framework is required which enables test providers to furnish comprehensive evidence in support of any claims about the validity of their tests. Examination boards and other institutions offering high-stakes tests need to demonstrate how they are seeking to meet the demands of all aspects of validity in their tests. This volume develops a theoretical framework for validating tests of second language academic reading ability, which then informs an evaluation of the IELTS approach to assessment in the skill area of academic reading.

The authors explore how the sociocognitive validity framework first described in Weir's *Language Testing and Validation: An Evidence-based Approach* (2005b) might contribute to an enhanced validation framework for use with academic reading tests. Weir's approach attempts to reconfigure validity to show how its various aspects (contextual parameters, cognitive processing, scoring validity, criterion-related validity and consequential validity) might interact with one other. Academic reading, the construct of interest in this volume, is viewed as not just the underlying latent trait of reading ability but as the result of the constructed triangle of trait, context and score (including its interpretation). The approach adopted in this volume is therefore effectively an interactionalist position, which sees the academic reading construct as residing in the interactions between the underlying cognitive ability, the context of use and the process of scoring.

Weir (2005b:85) stresses that:

... approximation to the construct in a measurement instrument is essentially the result of the interactions between its context and [cognitive]-based elements ... Establishing the nature of these interactions is what will take forward our understanding of language testing and the constructs it attempts to measure.

Khalifa and Weir (2009:8) suggest that:

Undoubtedly a close relationship exists between these elements, for example between context validity and cognitive validity, which together with scoring validity constitute for us what is frequently referred to as construct validity. Decisions taken with regard to parameters in terms of task context will impact on the processing that takes place in task completion. The interactions between, and especially within, these aspects of validity may well eventually offer further insights into a closer definition of different levels of task difficulty.

Dunlea (2015:48) observes that:

The components of the model are likely to interact and overlap in many dynamic ways. However, as Weir [2005b] and O'Sullivan and Weir (2011) note, for ease and clarity of description and in order to tease out the impact on each of these components at the point of interaction between test taker, test task, and the context of use in the testing situation, distinguishing these aspects is useful.

For the purposes of this volume, the distinctions between the various aspects of validity are maintained since they offer the reader a helpful descriptive route through the sociocognitive validation framework and, more importantly, a clear and systematic perspective on the literature which informs it. Construct validity, though, is seen as a superordinate, unifying concept to which the collection and evaluation of multiple forms of evidence would contribute.

This is very much in keeping with the Messickian tradition (1989, 1995). The categories in the sociocognitive framework in large part overlap with the six aspects of validity Messick himself identified (content relevance and representativeness; substantive theoretical rationales including process models of task performance; scoring structure; generalisability of score properties and interpretations; external aspects including convergent and discriminant evidence; and consequential aspects). Dunlea (2015:29) notes:

Messick (1995, p. 744) highlighted six "distinguishable aspects of construct validity", noting that the distinctions do not detract from the unified nature of validity but rather "provide a means of addressing functional aspects of validity that help disentangle some of the complexities inherent in appraising the appropriateness, meaningfulness, and usefulness of score inferences." Messick (1989, 1995) suggested that addressing these six aspects was crucial for validating score-based inferences and test use within a unified approach. Perhaps more importantly, he also suggested that centering validation around these six aspects of validity evidence would in fact be sufficient for doing so, noting the six aspects were applicable "to all educational and psychological measurement", and provided "a way of addressing the multiple and inter-related validity questions that need to be answered in justifying score interpretation and use" (Messick, 1995, p. 746). These six aspects of evidence collection and appraisal would thus ensure that "the theoretical rationale or persuasive argument linking the evidence to the inferences drawn touches the important bases" (Messick, 1995, p. 747)... Messick stressed that these distinctions did not allow a return to the selective use of one or more types of evidence as convenient to the researcher: "Evidence pertinent to all of these aspects needs to be integrated into an overall validity judgment to sustain score inferences and their action implications ... which is what is meant by validity as a unified concept" (Messick, 1995, p. 747).

Focuses of the volume

The focus for attention in this volume is the reading component of the IELTS Academic module, which is principally used for admissions purposes into tertiary-level institutions throughout the world (see Davies 2008 for a detailed history of the developments in EAP testing leading up to the current IELTS). According to the official website (www.cambridgeenglish.org/exams-andtests/ielts/test-format/), there are three reading passages in the Academic Reading Module with a total of c.2,150-2,750 words. Individual tasks are not timed. Texts are taken from journals, magazines, books, and newspapers. All the topics are of general interest and the texts have been written for a non-specialist audience. The readings are intended to be about issues that are appropriate to candidates who will enter postgraduate or undergraduate courses. At least one text will contain detailed logical argument. One of the texts may contain non-verbal materials such as graphs, illustrations or diagrams. If there are technical terms, which candidates may not know in the text, then a glossary is provided. The texts and questions become more difficult through the paper.

A number of specific critical questions are addressed in applying the sociocognitive validation framework to the IELTS Academic Reading Module:

- Are the cognitive processes required to complete the IELTS Reading test tasks appropriate and adequate in their coverage? (Focus on cognitive validity in Chapter 4.)
- Are the contextual characteristics of the test tasks and their administration appropriate and fair to the candidates who are taking them? (Focus on context validity in Chapter 5.)
- What effects do the test and test scores have on various stakeholders? (Focus on consequential validity in Chapter 6.)
- What external evidence is there that the test is fair? (Focus on criterion-related validity in Chapter 7.)

These are the types of critical questions that anyone intending to take a particular test or to use scores from that test would be advised to ask of the test developers in order to be confident that the nature and quality of the test matches their requirements. Cognitive validity is established by *a priori* evidence on the cognitive processing activated by the test task before the live test event (e.g. through verbal reports from test-takers), as well as through the more traditional *a posteriori* evidence on constructs measured by statistical analysis of scores following test administration. Language test constructors need to be aware of the established theory relating to the cognitive processing that underpins equivalent operations in real-life language use (see Chapters 2 and 3 for what the research literature says about this and Chapter 4 for further explication in relation to IELTS).

The term content validity was traditionally used to refer to the content coverage of the task. Context validity is preferred here as the more inclusive superordinate which signals the need to consider not just linguistic content parameters, but also the social and cultural contexts in which the task is performed (see Chapters 2, 3 and 5 for detail). Context validity for a reading task thus addresses particular performance conditions such as response method, time available, source and text length, as well as the linguistic demands inherent in the successful comprehension of the source text including lexical and syntactic complexity (Weir 2005b).

Scoring validity accounts for the extent to which test scores are arrived at through appropriate criteria in constructed response tasks and exhibit consensual agreement in their marking, are as free as possible from measurement error, stable over time, appropriate in terms of their content sampling and engender confidence as reliable decision-making indicators.

Messick (1989) argued the case for also considering consequential validity in judging the validity of scores on a test. From this point of view, it is necessary in validity studies to ascertain whether the social consequences of test interpretation support the intended testing purpose(s) and are consistent with other social values (see Chapter 6 for detail). There is also a concern here with the washback of the test on the learning and teaching that precedes it as well as with its impact on institutions and society more broadly.

Criterion-related validity is a predominantly quantitative and *a posteriori* concept, concerned with the extent to which test scores correlate with a suitable external criterion of performance with established properties (see Chapter 7 for detail). Evidence of criterion-related validity can come in three forms: firstly, if a relationship can be demonstrated between test scores and an external criterion-related validity takes two forms: concurrent and predictive. Concurrent validity seeks an external indicator that has a proven track record of measuring the ability being tested (Bachman 1990:248). It involves the comparison of the test scores with this other measure for the same candidates taken at roughly the same time as the test. This other measure may consist of scores from some other reading tests, or ratings of the candidate by teachers, subject specialists, or other informants (Alderson, Clapham and Wall 1995).

Predictive validity entails the comparison of test scores with another measure of the ability of interest for the same candidates taken some time after the test has been given (Alderson et al 1995). Demonstration of the qualitative and quantitative equivalence of different versions of the same test is a second source of evidence. A third source of evidence results from linking a test to an external standard such as the Common European Framework of Reference for Languages (CEFR, Council of Europe 2001) through the comprehensive and rigorous procedures of familiarisation, specification, standardisation and empirical validation (Council of Europe 2003).

Authenticity

As a general principle, language tests should, as far as is practicable, place the same requirements on test-takers as are involved in communicative settings in non-test 'real-life' situations. This approach requires particular attention to both cognitive and social dimensions of communication. A major focus of this volume is IELTS's concern with authenticity, which has been a dominant theme for adherents of the communicative testing approach as they attempt to develop tests that approximate to the 'reality' of non-test language use (real-life performance) (see Weir 1983, 1990, 1993, 2005b). The 'real-life' approach (Bachman 1990:41), though initially the subject of much criticism in the USA, has proved useful as a means of guiding practical test development. It is particularly useful in situations in which the domain of language use is relatively homogeneous and identifiable.

Authenticity is considered to have two characteristics. Firstly, interactional authenticity, which is a feature of the cognitive activities of the test-taker in performing the test task (see Chapters 2, 3 and 4 on cognitive validity), and secondly, situational authenticity, which attempts to take into account the contextual requirements of the tasks (see Chapters 2, 3 and 5 on context validity and see Bachman and Palmer 1996 for discussion of these concepts). Though full authenticity may be unattainable in the testing situation, as far as is possible, attempts should be made to use situations and tasks which are likely to be familiar and relevant to the intended test-taker (see Chapters 2 and 3 for an account of these based on the research literature on academic reading). The concern with situational authenticity requires readers to respond to contexts which simulate 'real life' in terms of criterial contextual parameters without necessarily replicating it exactly. In this paradigm, tests should be as direct as possible and, by employing tasks which activate the type of processing that characterise reading in the real-life target situation, interactional authenticity is enhanced. The more features of real-life use of language, in this case of academic reading, that can be built into test tasks, the greater the potential for positive washback on the learning that precedes the test-taking experience and the easier it will be to extrapolate from the test to make statements about what students can or cannot do in real-life reading situations. If the purpose is to measure academic reading ability, examination boards should be employing reading tasks that encourage teachers to equip candidates with the reading abilities they will need for performing in a realworld academic context.

Structure of the volume

The volume is structured according to three main sections.

Section 1 offers an introductory overview in Chapters 1–3 which address the nature of academic reading and how the construct of academic reading has come to be understood in light of insights from empirical and theoretical research. Parts of Chapters 2 and 3 that deal with cognitive validity review the available research literature on the processing involved in real-life reading, in preparation for Chapter 4's detailed examination of the cognitive processing involved in the IELTS Academic Reading Module. Parts of Chapters 2 and 3 dealing with context validity review the research literature on the impact of contextual variables on reading performance, and the findings of available research in this area relating to the contextual parameters of the IELTS Academic Reading Module are later explored in Chapter 5.

Section 2 (Chapters 4–7) explores the actual practice of assessing academic reading through IELTS in relation to four main areas of validity: cognitive validity; context validity; consequential validity; and criterion-related validity. There is a separate chapter on each of these areas. Again, IELTS Academic Reading tests and published research are the basis for the investigation. Scoring validity does not have its own chapter as the objective format of the current IELTS test largely precludes reliability concerns and the research literature that informs this volume has not concerned itself with this aspect of validity in relation to IELTS. However, the discussion of test formats in Chapter 2 covers the scoring validity of those formats currently employed in IELTS and there is also an extended discussion on the internal consistency of current IELTS test items.

Section 3 moves beyond the present towards consideration of potential task formats and the criterial features that future tests of academic reading should exhibit. Chapter 8 explores how reading-into-writing tasks can help to maximise the authenticity of an academic reading task, while Chapter 9 considers the growing contribution and benefit in a digital age of new technologies for the valid assessment of academic reading ability.

Conclusion

This Series Editors' note has supported the authors of this volume in their view that academic reading test developers should provide a clear definition

of the ability constructs which underpin the tests they offer in the public domain. Such an explication is increasingly necessary if claims about the validity of test score interpretation and use are to be supported both logically and with empirical evidence.

Weir and Chan propose a test validation framework that adopts a sociocognitive perspective in terms of its underlying theory and which conceptualises validity as a unitary concept; at the same time the framework embraces core aspects of validity, which reflect the practical nature and quality of an actual testing event. An understanding of the framework plus its various aspects and their application to academic reading tests can assist test developers in operationalising their tests in a more valid fashion and thereby provide a more accurate measurement of the construct of interest.

> Cyril J Weir Nick Saville Lynda Taylor June 2019

Postscript

The Series Editors' note for this volume was drafted in July 2018 by Cyril Weir and myself. At that time Cyril was about to have a major operation and we had just signed off the text a few days before he went into hospital. About a month later when he was recovering from his surgery, we met again at his house to discuss the final manuscript itself. In light of comments from the two external reviewers, we decided to make some significant changes to the design of the volume that also involved the rewriting of a number of sections. Given the circumstances, we decided to invite Lynda Taylor (one of the reviewers) to assist the authors in making the agreed changes.

The revisions were already underway by September 2018 but unfortunately Cyril passed away on September 28 2018, and before the manuscript could be finalised.

I am very grateful to Lynda and Sathena for their ongoing commitment to the project and their collaboration in bringing this volume to press in line with Cyril's wishes and intentions.

I believe this volume now makes a fitting tribute to Cyril's extensive contribution to the assessment of academic reading over many years.

Nick Saville July 2019

List of abbreviations

AERT	Advanced English Reading Test
AI	Artificial Intelligence
ALT	Academic Literacy Test
ALTE	Association of Language Testers in Europe
AWL	Academic Word List
BNC	British National Corpus
CAE	Certificate in Advanced English
CASCCC	Committee of the Academic Senates of the California
	Community Colleges
CEFR	Common European Framework of Reference for Languages
CLB	Canadian Language Benchmarks
Cohm	Coh-Metrix
CPE	Certificate of Proficiency in English
CRELLA	Centre for Research in English Language Learning and
	Assessment
CUP	Cambridge University Press
EAL	English as an Additional/Alternative Language
EAP	English for Academic Purposes
EAS	Explicit Across Sentences
ECPE	Examination for the Certificate in Proficiency in English
EEG	Electroencephalography
EFL	English as a Foreign Language
EL1	English as a First Language
ELBA	English Language Battery
ELTS	English Language Testing Service
ELTS-GA/1	English Language Testing Service – General Academic
	Module
ELTS-SS	English Language Testing Service – Social Science Module
ELTS-T	English Language Testing Service – Technology Module
EPTB	English Proficiency Test Battery
ESL	English as a Second Language
ESOL	English for Speakers of Other Languages
ESP	English for Specific Purposes
ESRC	Economic and Social Research Council
ETS	Educational Testing Service
EWS	Explicit Within Sentence

Research and Practice in Assessing Academic Reading

EXMI	Understanding Explicitly Stated Ideas
FCE	First Certificate in English
GCE	General Certificate of Education
GenZ	Generation Z
GEPT	General English Proficiency Test
GMC	General Medical Council
GSTEP	Georgia State Test of English Proficiency
HESA	Higher Education Statistics Agency
IATM	Instrument for the Analysis of Textbook Materials
IBS	Implicit Between Sentences
IDP	International Development Program
IELTS	International English Language Testing System
IESOL	City & Guilds International ESOL
IIS	IELTS Impact Study
ILA	International Language Assessment
IRT	Item Response Theory
ISE	Trinity College London's Integrated Skills of English
ISESOL	International Spoken English for Speakers of Other
	Languages
IWS	Implicit Within Sentence
L1	First Language
L2	Second Language
LIBS	Local Item Banking System
LTTC	Language Training and Testing Centre
MCQ	Multiple-Choice Question
MOOC	Massive Open Online Course
NP	Noun Phrase
OET	Occupational English Test
PET	Preliminary English Test
PTE	Pearson Test of English Academic
QMS	Quality Management System
QPP	Question Paper Production
RIELTS	Revised IELTS
SAQ	Short-Answer Question
SD	Standard Deviation
SiLT	Studies in Language Testing
SJTU	Shanghai Jiao Tong University
SKM	Surveying to Obtain the Gist
SM	Situation Model
SRI	Stimulated Recall Interview
TEAP	Test of English for Academic Purposes
TEEP	Test of English for Educational Purposes
TEFL	Teaching English as a Foreign Language

TEM	Test for English Majors
TESOL	Teaching English to Speakers of Other Languages
T/F/NG	True/False/Not Given
TL	Target Language
ТМ	Text Model
TOEFL	Test of English as a Foreign Language
TOEFL iBT	Internet-based Test of English as a Foreign Language
TTR	Type-token Ratio
UCLA	University of California, Los Angeles
UCLES	University of Cambridge Local Examinations Syndicate
UETESOL	University Entrance Test in English for Speakers of Other
	Languages
UKVI	UK Visas and Immigration
UoB	University of Bedfordshire
WPM	Words Per Minute
WS	Writing Summaries

Section 1 The nature of academic reading

This section of the volume maps out the essential cognitive and contextual parameters of academic reading ability.

Chapter 1 offers a general overview of how we might approach academic reading according to the existing literature and examines what universities themselves consider academic reading to be. The chapter unpacks the three main approaches that have emerged over the past 40 years as the theoretical basis for understanding the nature of academic reading: *generic study skills*; *genre-based*; and *socially situated discourse*. The chapter then examines how these three contemporary approaches might inform the development of a test of academic reading for screening students wishing to embark upon Englishmedium study at undergraduate and postgraduate level.

Chapter 2 revisits three major research studies that have been carried out over the past three decades to define the nature of academic reading at tertiary level. Though this information is for the most part already in the public domain, we saw an advantage in assembling conveniently in one place the available theory and major empirical research outcomes that are directly relevant to a description of academic reading (and writing) and the practice and experience of assessing those skills. This strategy enabled us to show the degree of agreement on both the tertiary-level reading activities and the attendant performance conditions for carrying them out across the three major studies. It also provided an outline framework for looking at the remaining literature in the field to fill out the detail of the descriptive parameters involved.

Chapter 3 then has recourse to the wider literature and builds a theoretical validation framework for the academic reading construct. This framework will be used in Section 2 of this volume to explore the extent to which the current IELTS Academic Reading test remains construct valid or might merit attention as part of any future revision project.

Introduction: Some general considerations

I found that I was fitted for nothing so well as for the study of Truth; as having a mind nimble and versatile enough to catch the resemblances of things (which is the chief point), and at the same time steady enough to fix and distinguish their subtler differences; as being gifted by nature with desire to seek, patience to doubt, fondness to meditate, slowness to assert, readiness to consider, carefulness to dispose and set in order; and as being a man that neither affects what is new nor admires what is old, and that hates every kind of imposture. So I thought my nature had a kind of familiarity and relationship with Truth. Sir Francis Bacon (1603)

Three approaches to defining academic reading

A recent survey of the literature indicates that three approaches have been put forward as the basis for teaching and testing academic reading (see Weir 2013b). We first examine these three contemporary approaches, which, contingent on their *practical usefulness*, might in turn inform the development of a reading test for screening students with regard to their ability to cope with tertiary-level English-medium academic study, i.e. higher education at both undergraduate and postgraduate level within a university or college context:

- The first focuses on the *generic study skills* that students need post entry to tertiary-level English-medium education, i.e. the traditional building blocks of academic literacy familiar to English for Academic Purposes (EAP) teachers throughout the world since the 1970s (Weir 1983).
- More recently, researchers employing corpus linguistics have suggested that a *genre-based* approach better prepares students for the discipline-specific demands of academic study (Nesi and Gardner 2006).
- Critical literacy adherents argue that we must go even further and consider academic discourse as *socially situated*, involving issues of power and authority (Lea and Street 1998, Murray 2016).

The generic study skills approach

Academic language proficiency is . . . the literacy of the educated, based on the construct of there being a general language factor relevant to all

Research and Practice in Assessing Academic Reading

those entering higher education, whatever specialist subject(s) they will study (Davies 2008:113).

Davies (2008) addresses the question 'what is academic language proficiency?' at the end of his book on testing EAP. He neatly encapsulates the argument for a single test of academic English based on cross-cutting EAP skills (2008:113). He argues (2008:113) that academic proficiency involves *performing the appropriate discourse*, which is in his view generalisable across all disciplines. He identifies the components of this skilled literacy of the educated as argument, logic, implication, analysis, explanation, and reporting, and makes the case for a single general approach relevant to all those entering higher education, whatever their area of specialism. Davies also offers a complementary definition of communicative language testing further supportive of the argument against discipline-specific modularity. He argues that what should be tested are the skills and features underpinning communicative behaviour, i.e. the abilities rather than the behaviour itself.

In the development phase of the Associated Examining Board's Test of English for Academic Purposes (TEAP), Weir (1983) surveyed the various academic activities in tertiary-medium education in the UK and the problems students encountered in coping with the language-related study skills involved in these activities. Responses were received from 940 overseas students, 530 British students and 560 academic staff, in respect of 43 post-graduate courses, 61 undergraduate courses and 39 A Level courses (i.e. the advanced-level courses taken by secondary school students and used for admission to university). Weir found a good deal of variety, but also considerable overlap in reading activities (1983:198–226) and writing activities (1983:226–242) across academic levels and across subject boundaries. The data generated by Weir's study encouraged the Associated Examining Board to follow a generic language-based study skills approach in developing TEAP.

For Weir (1983), reading in the academic context involves: reading carefully for comprehension of all the information in relevant written materials; reading to get a general idea of the main information about a topic, e.g. general background reading, as follow-up to lectures or in preparation for seminars; reading to check sources of new information, such as articles in recent journals, and new books to see how useful they might be to a course of study; skimming quickly to establish how useful it would be to study a particular text more intensively; search reading to get information specifically required for particular written assignments, e.g. for homework tasks and project work; critical reading to establish and evaluate the author's position on a particular topic (done at all levels across subjects but more often at postgraduate level and across the social sciences); and making notes, particularly in social sciences but less so in science and engineering. Weir (1983:332) summarised the data on which his academic reading test would be based:

According to our survey of the reading activities in which students were involved, it is possible to distinguish two different kinds of complementary reading activities to which students are exposed, namely extensive reading and intensive reading. As part of their preparation for written work and seminar discussions, students often have to search-read to get information specifically required for assignments. This requires the ability to read quickly and with ease, selecting salient features from paragraphs and longer units of prose. It further involves the skills of surveying, i.e. skimming through a text in order to become familiar with the gist of the content, and scanning, which refers to the skills used when reading quickly through a text in order to locate specific pieces of information. It also requires the ability to separate the essential from the non-essential in a text and presupposes understanding of explicitly stated information. Some of the reading material which students encounter will require more intensive study to understand all the information contained therein. In such cases they will need to examine the text as a unit in closer detail and understand how the various parts are related to each other.

The nature of the generic reading skills in an academic context is explored more fully in Chapters 2 and 3 where, based on the reading research literature, we develop a model of academic reading which takes into account: types of academic reading; cognitive processing levels in academic reading; and the contextual parameters encountered in academic reading. The cognitive and contextual parameters of the current IELTS Academic Reading test are then explored in Chapters 4 and 5.

Weir (1983) argues that writing in coursework and examinations in an academic context is usually based on prior reading of appropriate source texts; he found this was common to all disciplines, at all levels (with the exception of mathematics undergraduate students). Horowitz (1986a, 1986b) similarly found many of the same broad types of undergraduate writing task recurring in different fields, for example, 'synthesis of multiple sources', which he describes as a sort of 'essay', were set in hard, soft, pure and applied disciplines.

Transforming knowledge from information extracted from the reading of various source texts into new written texts is one form of mediation that is relevant to general communication, but is an absolute essential in academic life. The new *Companion Volume* to the CEFR (Council of Europe 2018) provides an extensive treatment of mediation, missing from the original Common European Framework of Reference for Languages (CEFR) specification (Council of Europe 2001), which opens up new avenues for assessment tasks that can be integrated in innovative ways. We will look more closely at

the benefits of mediation in approaches to academic reading-into-writing in Chapter 8.

According to Weir (1983), there was a good deal of homogeneity in how student writing was assessed by the 560 staff who completed a questionnaire for his research study (246–269, 390–392). For subject tutors, the relevance and adequacy of the subject matter in students' writing (a clear connection with prior reading here), the clarity of expression, and the arrangement and development of written work were paramount. Mechanical accuracy and grammar were considered of lesser importance, not least because overseas students were generally considered better with regard to these than home students. The clear emphasis on content and how it is organised again suggests the importance of prior reading in any consideration of students' writing performance in an academic context.

The percentage of academic staff who considered each criterion important in the assessment of written production ranked in order of preference:

The subject matter (1)	91.8
Expressing what you want to say clearly (2)	90.9
Arranging and developing written work (3)	82.1
Using appropriate vocabulary (4)	69.6
Tidiness (5)	62.8
Writing grammatically correct sentences (6)	46.9
Handwriting (7)	44.6
Using appropriate grammatical structures (8)	43.3
Spelling (9)	42.3
Using a wide and varied range of vocabulary (10)	41.3
Punctuation (11)	39.3
Using a variety of grammatical structures (12)	22.2

Weir (1983:391)

It is important to make it clear from the outset that exhibiting adequate generic study skills for academic success is not just a problem for overseas students. Weir's (1983) research clearly indicates the extent and gravity of the language-based study skills problems of a number of *home* students at both undergraduate and postgraduate levels across disciplines. Chapter 2, Research Study 3, nearly 30 years later, provides continuing evidence of the academic literacy problems experienced by home students in a post-1992 British university. With widening participation in university education now a societal goal in the UK, the problems are unlikely to ameliorate. There is clear evidence here of the need to establish the academic literacy of both home and overseas students on entry to a university so any necessary action can be taken to help enhance the student experience. A generic study skills approach offers the universities a feasible, and perhaps the only, option of doing this through a single test.

According to this first approach, we should conceive of academic literacy as comprising a set of generic study skills needed by both home and overseas students at university level with regard to reading and reading-into-writing. A key feature and potential weakness, however, is that they are not contextualised within a more closely defined institutional setting, in terms of level (e.g. undergraduate/postgraduate), disciplinary area (e.g. business and administration/STEM/humanities subjects) or academic social practices, and this leads us to consider the *genre-based* approach and then *socially-situated* discourse practices.

The genre-based approach

Lillis (2003:194) characterises the genre-based approach as academic socialisation, with two dimensions:

- Language as discourse practices which learners will/must gradually come to learn implicitly. Socialisation (1) teaching as (implicit) induction into established discourse practices.
- Language as genres which are characterised by specific clusters of linguistic features. Socialisation (2) explicit teaching of features of academic genres.

Lea and Street (1998:164) view the approach as involving 'how to write specific, course-based knowledge for a particular tutor or field of study'. They believe problems in students' assessments lie with a lack of familiarity with the subject matter of a particular discipline and how to write knowledge in that discipline, 'in particular the need to abstract theory rather than attend to factual detail as evidence . . . deeper epistemological issues associated with knowledge in different disciplines' (1998:164–166).

The Economic and Social Research Council (ESRC) study *An Investigation* of Genres of Assessed Writing in British Higher Education (Project No. RES-000-23-0800) by Hilary Nesi, Sheena Gardner, Paul Thompson, Paul Wickens et al is a seminal study which provides a useful starting point for a comprehensive description of the genres to be encountered in academic study. Nesi et al (2008) extend our knowledge of the genres of assessed student writing in British higher education and their work represents the first large-scale attempt to identify and describe the range of writing produced by university students for assessment purposes. They find that students in British universities are required to produce a range of different genres of assessed written work, reflecting a range of rhetorical structures, communicative purposes and audiences (see also Nesi and Gardner 2006:103–105 for details).

Nesi and Gardner (2006:99) show that there are clear cultural differences in academic literacy between discipline areas and Gardner and Nesi (2013:47) suggest the existence of at least 13 genre families in academic writing each with their own stages of production. Nesi and Gardner (2006:102) refer to an even greater diversity beyond that of genre families:

... undergraduate student writing is clearly complex, with many variations in practice dependent not only on discipline, level of study and educational approach, but also on the nature of the higher education institution, the particular focus of the department within that institution, and the idiosyncrasies of the lecturers who assign written work.

However, in the midst of all this diversity, Nesi et al (2008) also establish that the vast majority of students in the arts and humanities and in social sciences have to produce an *argumentative essay*. Nesi and Gardner (2006:106–107) find the discursive essay to be the most common form of writing:

The prototypical pedagogic genre is the traditional student essay. It is used by all departments in our sample with the exception of Physics, which has only recently abandoned it. When defined by tutors, it is taken to be discursive prose. Length and frequency varies: some tutors expect short essays every two weeks, others require a 3000-word essay per module per term, and possibly one longer "essay" of 8,000–10,000 words in the final year . . .

Essays have a basic, generally three-part structure:

- Introduction, body, conclusion (Biological Sciences)
- Introduction, logical sequence of argument, conclusion (Medicine)
- Argument, counter-argument, conclusion (Hospitality and Tourism)

Nesi and Gardner (2006:101–102), echoing Weir (1983), point out a further commonality in the criteria applied by academics to student assignments:

Nevertheless there is also evidence that academics in different disciplines value many of the same qualities in the written assignments their students produce . . . In interviews with academics in the humanities, sciences and social sciences, Lea and Street (2000) found that "structure", "argument" and "clarity" were commonly identified as crucial to student writing success (although their informants had difficulty in explaining what a well-developed argument actually looks like in a written assignment).

Nesi and Gardner (2006:113) describe the attributes looked for in student writing:

When we asked tutors about desirable characteristics of student writing, there was remarkable consistency within the group, and indeed with the literature. Economics tutors mentioned critical analysis and logical development, History tutors clarity of argument, taking the reader on "a journey through conflicting ideas". Tutors in Sociology and Medicine valued "a clearly stated argument". Engineering tutors liked succinct and well-structured writing, while Philosophy tutors liked clarity and clear signalling... Next to coherent structure, the most frequently stated desirable quality was originality or creativity, and we have seen how this interacts with logical thinking... Given the differences that have emerged surrounding writing purpose, audience and rhetorical structure, it is perhaps surprising that there are nevertheless shared qualities valued across the university.

One additional finding of interest was the change in the types of writing that university students are required to produce as they progress through their courses of study, at least within the UK university context. As they progress, students are increasingly expected to conform to the norms of favoured genres, and may also be given generically different writing tasks at different stages of study (see also Ganobcsik-Williams 2001, 2004, Hewings and Hewings 2001, Sommers 2000).

- First year student writing tends to be descriptive: sources are used to affirm students' own ideas and to demonstrate student comprehension.
- Second year students begin to develop a more 'questioning disposition' toward sources.
- There is an increase in confidence and critical evaluation of source materials between first year writing and the writing of degree finalists.

Nesi and Gardner note (2006:107):

... Essays were also thought to involve critical thinking. Particularly in essays, progression is marked by an increasingly critical and original response:

- "We'd expect much more of a critique of their work from a third year ... [student] than we would from a first year" (Computing)
- students become "more critical in the final stages" (Hospitality and Tourism)
- first year writing should be accurate, concise, explicit, but by the third year "originality should be added to the mixture" (Psychology)
- good students "develop a genuine personal voice" (Theatre Studies).

According to this view, therefore, academic literacy involves more than just acquiring a set of basic, generic study skills; it involves awareness of and increasing familiarisation with the genres and modes of discourse that characterise academic study within a particular field or discipline, in terms of their discoursal and surface language features. Paradoxically, their research also establishes that in the midst of such diversity and differentiation, commonalities nevertheless exist as can be seen in the pervasiveness of the argumentative essay across disciplines and a good deal of agreement on the assessment criteria that should be used to mark it.

Socially-situated discourse practice

Lea and Street (1998:170) regard the generic study skills approach as:

... focused on the student and suggests that students lack a set of basic skills that can be dealt with primarily in a remedial study skills or learning support unit.

However, they are critical that:

This takes no account of the interaction of the student with institutional practices and is based on the underlying principle that knowledge is transferred rather than mediated or constructed through writing practices . . . the implicit models that have generally been used to understand student writing do not adequately take account of the importance of issues of identity and the institutional relationships of power and authority that surround, and are embedded within, diverse student writing practices across the university . . .

Lillis (2003:195) argues for considering a model based on socially-situated discourse practices as well as the generic and disciplinary specific approaches:

The academic literacies frame has helped to foreground many dimensions to student academic writing which had previously remained invisible or had been ignored; these include the impact of power relations on student writing, the centrality of identity in academic writing, academic writing as ideologically inscribed knowledge construction (see e.g. Jones et al., 1999; Lea & Stierer, 1999; Lea & Street, 1998).

Lea and Street characterise language as socially-situated discourse practices, which are ideologically inscribed. Lea and Street (1998:161–163) observe:

... many of the difficulties [students] experienced with writing arose from the conflicting and contrasting requirements for writing on different courses and from the fact that these requirements were frequently left implicit ... Students described taking "ways of knowing" (Baker et al., 1995) and of writing from one course into another only to find that their attempt to do this was unsuccessful and met with negative feedback. They were consciously aware of switching between diverse writing requirements and knew that their task was to unpack what kind of writing any particular assignment might require. This was at a more complex level than genre, such as the "essay" or "report", but lay more deeply at the level of writing particular knowledge in a specific academic setting.

Lea and Street (1998:158) argue that the academic literacies model 'incorporates both of the other models into a more encompassing understanding

of the nature of student writing within institutional practices, power relations and identities'. According to this third approach, academic literacy thus embraces an additional dimension on top of generic and discipline-specific study skills: that of the socially-situated nature of academic interaction.

Towards a viable academic reading test

Clearly in Weir's (1983) extensive dataset on academic reading practices, one can find differences between academic levels and subject boundaries, but if one wishes to teach or test EAP beyond the individual or particular course of study, there would seem to be no real alternative other than focusing on important commonalities rather than individual differences. A critical *sine qua non* for language testing is that it must be *practical* which involves due consideration of logistics and cost (*Principles of Good Practice*, Cambridge Assessment English 2016). This would seem to preclude anything other than a generic study skills approach, which has been the approach of choice to date in most EAP pre-sessional courses, tests and materials (see Research Study 2 in Chapter 2 which investigated a wide range of EAP tests and coursebooks).

The history of major UK EAP testing projects like English Proficiency Test Battery (EPTB, 1965–1980) (see Davies 2008:71, Weir and O'Sullivan 2017: Chapter 3), English Language Testing Service (ELTS, 1975-1989) (see Davies 2008:Chapters 2 and 3, Weir and O'Sullivan 2017:Chapter 4), International English Language Testing System (IELTS, 1989-2018) (see Alderson and Clapham 1992, Davies 2008:Chapters 4 and 5, Weir and O'Sullivan 2017: Chapter 5) and TEAP (1980-2018) (Weir 1983) indicates that attempts to provide discipline-specific options within a testing system are faced with major problems in terms of content specificity and subject boundaries; for example, which of the 35 types of engineering should determine the content of a potential engineering module? Furthermore, such English for Specific Purposes (ESP) tests are difficult to market effectively as it is sometimes unclear which version is most suitable for any given individual. How do students choose an appropriate module where their courses cut across disciplinary boundaries? Even within a single course, students might have to cope with differing genres. Choosing the appropriate genre version of a test proved problematic for both students and receiving institutions even when there were only six modules to choose from in ELTS (Weir and O'Sullivan 2017:Chapter 4).

Furthermore, catering for discipline-specific tests proved to be administratively unwieldy, potentially unreliable in terms of parallelism of test forms and economically unviable. Their development and the need for iterative production of multiple subject-oriented versions of a test create enormous logistical problems for the test providers. When we consider that a very large number of different versions of the current general module of the IELTS Academic Reading test need to be created for security reasons, the consequences of going down the specific-module route again are obvious. The critical issue of comparability between all the different versions would be a very difficult hurdle to overcome. (For further discussion of these problems see Alderson and Clapham 1992, Alderson and Urquhart 1985, Charge and Taylor 1997, Clapham 1996a, Davidson 1998, Davies 2008, Henning 1988, Read 2015, Weir 1983, Weir and O'Sullivan 2017).

The current position of examining boards appears to be that, if an academic literacy test is to be viable, it must operate at a suitable level of generality, be administratively feasible and cater for as wide a spread of students as possible. The shortcomings of the genre and critical approaches for testing purposes discussed above are that they are both directed at finding what is **distinctive** rather than what is **common** across subject areas.

The *genre-based approach* is certainly more delicate and granular than the generic study skills approach to academic discourse definition, but it would create serious practical problems if it was used as the basis for creating a suite of reading/writing tests covering all 13 of the major genres identified by Gardner and Nesi (2013). The ESP approach, though, in principle, potentially more construct valid, has proven in the past to be simply not feasible for almost all industrial-scale global tests. See, however, further discussion of this issue in the section 'Can one size fit all?' in Chapter 5 on context validity, where research supporting the need to revisit these assumptions is considered.

The *critical linguists' perspective*, with its concern for a range of relationships (social and political as well as academic), at an individual as well as an institutional level, is attractive in terms of highly specific relevance to the discoursal demands made on students by their courses. It is, nevertheless, impossible to cater for this in anything but a highly specific and limited way, perhaps even down to the level of an individual student's context, as indicated by the case studies Lea and Street (1998) provide. For examination boards, this approach is even further down the road to impracticality than the genre-based approach and has never been a serious option as the basis for test design.

Murray (2016:107–108), though clearly favouring the *socially-situated approach* above all others for evaluating students' academic English, concedes that it is some time after entry to university that students will be inducted into the domain-specific requirements of socially situated discourse. They do not have this knowledge/awareness on entry to the university, which makes it rather pointless testing for it at that stage (though a different case could perhaps be made for admission to postgraduate study).

Although the idea of students taking English language tests based on the discipline area in which they intend to study and tailored accordingly might appear a logical option, in practice it makes little sense if (a) we cannot assume that students will come equipped with adequate conversancy in the literacy practices of their future disciplines, as a result of diverse educational experiences, and (b) those literacy practices will therefore need to be taught to them anyway, embedded in the curriculum.

This really only leaves industrial-scale academic English test developers, who wish to evaluate students on entry to university, with the generic academic study skills approach as a viable option. This clearly remains open to the criticisms that it does not cater specifically for differences in terms of:

- the varying genre(s) a student might have to deal with even on the same course (Gardner and Nesi 2013)
- the different epistemologies of individual academic staff for whom a student might have to write in their course of study (Lea and Street 1998)
- the contrasting official and unofficial discourse practices in the institutions where students find themselves (Lea and Street 1998).

Nevertheless, a generic study skills approach offers receiving institutions a feasible and plausible benchmark for deciding whether students possess the *baseline language skills* necessary for comprehending and collating key information from a variety of written academic sources and for reformulating these ideas in writing in a relevant, adequate, clear and coherent manner at the start of a course of study (see Weir 1983).

Before looking in detail at the published research in this area in Chapters 2 and 3, we felt it might be illuminating to provide a brief snapshot of what universities, as the principal end users of test results, consider to be the core generic language-related study skills their students require on entry. We looked at all those universities that provided details of academic reading activities on their websites. This will help provide further situational context for the research that is reported in later chapters (see Khalifa and Weir 2009). It is interesting to note that there is only limited reference to the genre-based or socially-situated, critical discourse approaches to defining skills requirements on any of the university websites we consulted. A summary of the impressive *Statement of Competencies* produced by academic institutions in California is provided first, followed by a synthesis of common descriptions by UK universities.

How universities themselves define academic reading and writing needs

Of particular note in attempts by the universities themselves to specify academic literacy requirements is *Academic Literacy: A Statement of Competencies Expected of Students Entering California's Public Colleges and Universities* (2002), developed by the Intersegmental Committee of the Academic Senates of the California Community Colleges (CASCCC), the California State University, and the University of California. There are clear resonances throughout this report with Weir's earlier (1983) findings for students at universities in the UK context. Even though Weir's study is now over 30 years old, and clearly a number of new skills are now relevant to university study (use of multimedia and management of a wider range of information sources to name but two), many of the core skills clearly remain the same.

A selection of the *Statement*'s target strategies and processes (CASCCC 2002:2–6) relevant to our defining an academic English test specification are reported below, in relation to three key aspects:

- critical thinking processes
- the reading and writing connection
- reading-into-writing.

Critical thinking processes

Critical thinking generally refers to a set of cognitive habits and processes. Thus, critical thinkers recursively engage in probative questioning, rigorous analyzing, imaginative synthesizing, and evaluating of ideas (CASCCC 2002:14).

The *Statement* also argues that the 'following intellectual habits of mind are important for students' success . . . College and university students should be able to engage in the following broad intellectual practices' (CASCCC 2002:13):

- compare and contrast own ideas with others'
- challenge their own beliefs
- generate hypotheses
- see and respect other viewpoints
- respect principles as well as observations and experiences
- read with awareness of self and others
- read sceptically
- respect facts and information in situations where feelings and intuitions often prevail
- postpone judgement and tolerate ambiguity
- synthesise multiple ideas into a theory
- sustain and support arguments with evidence.

(CASCCC 2002:38, 46)

The reading and writing connection

According to the *Statement* (CASCCC 2002:4), '83% of faculty say that the lack of analytical reading skills contributes to students' lack of success in a course'. Entering students thus need to have the following higher-level reading skills (CASCCC 2002:16, 39, 40):

- use the title of the article/essay/text as an indication of what will come next
- predict the intention of the author from extra-textual cues
- retain versatility in reading various forms of organisation, both essay and paragraph
- · read texts of complexity without instruction and guidance
- read a variety of texts, including news articles, textbooks, essays, research of others, internet resources
- retain the information read
- decipher the meaning of vocabulary from the context
- summarise information
- synthesise information from reading and incorporate it into a writing assignment
- understand separate ideas and then be able to see how these ideas form a whole
- relate prior knowledge and experience to new information
- make connections to related topics or information
- determine major and subordinate ideas in passages
- identify the evidence which supports, confutes, or contradicts a thesis
- identify key examples that attempt to prove the thesis
- anticipate the direction of the argument or narrative
- retain information while searching for answers to self-generated questions.

The *Statement* (CASCCC 2002) reported that many students appeared intimidated by these especially in tasks where both reading and writing were involved. Fewer than 50% could provide brief summaries of readings or analyse information based on their reading. Reading was seen to be *the most significant factor in the success of students in academia*.

Reading into writing

According to the report (CASCCC 2002:21), 'In writing for university courses, faculty in our study indicated that students will be asked to write papers that require them to do the following':

- generate ideas for writing by using texts in addition to past experience or observations
- represent the ideas of others responsibly
- critically analyse the ideas or arguments of others
- · summarise ideas and/or information contained in a text
- synthesise ideas from several sources
- report facts or narrate events
- develop main point or thesis
- develop thesis convincingly with well-chosen examples, reasons, and logic
- structure writing so that it is clearly organised, logically developed, and coherent
- correctly document research materials to avoid plagiarism.

(CASCCC 2002:4, 21-22, 40)

Academic literacy defined by other universities

To provide a sense of what other universities thought important in terms of academic literacy, we also examined the available websites of those responsible for providing assistance with academic English in British universities and synthesised views on academic reading and writing from the more insightful of these (viz The Open University at www.open.ac.uk/skillsforstudy; Dundee University at www.dundee.ac.uk/academic-skills; Monash University at www.monash.edu.au/lls/OffCampus/Improve/3.5.html, and University of Kent at www.kent.ac.uk/careers/postgradmenu.html).

We collate below some of the more informative findings from these sites, which collectively bear close comparison with those skills identified in the *Statement* we have quoted from above and for the most part accord with Weir's (1983) survey of study skills and habits of students in the UK. Taken together they offer the reader further evidence of the nature of the consensus on the academic skills required at university, with regard to:

- academic reading
- academic writing
- postgraduate research.

Academic reading

Identifying sources to read

This approach helps you to establish how useful a book will be for your purposes. You need to:

- understand how text is constructed by establishing and evaluating the value of the book through considering its aims, and examining the introduction, contents and index
- · identify passages that are important to your aims
- read the chosen sections in depth.

Reading efficiently

Once you have identified the sources to read, go through them quickly to make sure that they are indeed relevant to your studies and provide you with the information you need. Reading irrelevant material carefully is a waste of time. Adopting strategies so that you read quickly with understanding is imperative for reading productively and absorbing content. Skimming is the process of deciding whether or not a particular piece is worth reading more thoroughly. You might read:

- the abstract, summary or overview to get an insight into the purpose and the findings of the document
- the headings and the subheadings to see how the information is organised
- the conclusions to ensure that the outputs from the piece can deepen your understanding of a topic
- the first paragraph of each section, or the first couple of lines of each paragraph to decide whether or not to read the entire text.

Reading carefully

This approach helps you to:

- read the chosen sections in depth
- draw your own conclusions about the content and use this understanding to help you make notes from printed source material that are succinct and not simply a 'copy' of the original text.

Reading critically

The higher grades at every level of university study require some critical analysis. 'Critical' means discerning the strengths and the limitations of the work you are studying; applying reasoned and disciplined thinking to a subject. You might ask:

- What ideas and information are presented and how were they obtained?
- Are there unsupported assertions?
- Are relevant reasons or evidence provided?
- Is the method used to find the evidence sound?
- Is the evidence sound?

- What assumptions have been made?
- What is fact and what is opinion?
- Are there unreasonable generalisations?
- What has been omitted?
- How was the conclusion reached?
- Is the conclusion reasonable?
- What other perspectives or points of view could there be?

Academic writing

The following are key requirements of academic writing:

- Must address the key issues in the question in a manner that demonstrates a thorough understanding of the theories and the concepts studied.
- Present personal views and findings, based on well-reasoned views and judgements.
- Indicate understanding and evidence that you are capable of independent, well-reasoned thought, application and reflection.
- Translate the language of other authors into a form that represents your own perspectives and is comprehensible to your readers.
- Construct the arguments and counterarguments required in a balanced rather than biased manner; questions which ask for critical evaluation of something require a balanced treatment which considers both the positive and negative aspects of the argument. Only presenting one side of the argument is merely a value judgement whereas considering both sides helps develop and refine analytical skills which help in problem solving and decision making. Students in this way are obliged to think about things as reflective practice is thereby encouraged and they become receptive to new ideas and approaches.
- Back up own position with academic opinion, facts, examples and statistics, rather than mere personal opinion. Demonstrate a recognition of the difference between fact and conjecture.
- Cite sources in your text and include the publication details in a reference list or bibliography at the end of your work completely and accurately so the reader is able to access your sources and consider your treatment of them and avoid any charge of plagiarism. Such referencing helps the reader validate your argument.
Postgraduate research students

Clearly research skills (for example carrying out an extensive literature review, formulating research hypotheses, employing quantitative, qualitative or mixed method design methodologies, generating empirical data, carrying out data analysis) are more necessary at postgraduate level than at undergraduate level, but these are usually viewed as part of the apprenticeship of postgraduate study, i.e. performance-based add-ons to well-developed academic literacy.

Under the research skills construct, Wisker (2007), in her *Postgraduate Research Handbook*, attempts to distinguish research skills from the study skills construct. The research skills she identifies include starting, choosing, proposal writing, managing supervisors, timing and tasks, learning approaches and styles, research cultures, research questions, literature review, methods, data collection and analysis, thesis writing and examination. These add-on research skills are often covered by additional courses provided by the institution and, in any case, do not easily lend themselves to assessment in an exam format.

The Monash University website (www.monash.edu.au/lls/OffCampus/ Improve/3.5.html) details some further differences between undergraduate and postgraduate studies (emphasis added by authors):

There are obvious differences between undergraduate and postgraduate study: for example, you are usually expected to produce a thesis (up to 15,000 words for a shorter thesis and up to 80,000 words for a longer one), which is a longer piece of writing than is generally expected at the undergraduate level. Postgraduate study also typically involves the selection of a research topic that produces some kind of new insight, unique outcome or new knowledge. This isn't necessarily expected at the undergraduate level... It is a little more difficult to define the depth and breadth of understanding you are expected to demonstrate in postgraduate studies, and this also varies with disciplinary expectations. Breadth of research supports the comprehensive coverage of the relevant issues; depth is indicated in the level of **critical** engagement, analysis and application of theory. The importance of these two features of postgraduate assessment is often reflected in the proportion of marks allocated to the relevant criteria...

The website also emphasises the centrality of critical thinking for postgraduate students and provides the following useful discussion of it:

[Critical thinking] refers to a way of approaching the ideas of other scholars and researchers – questioning, evaluating, and checking the accuracy and validity of ideas and information. In writing, this often means evaluating the strengths and weaknesses of an idea, a response and/or a theory. We cannot assume that everything that is written, researched, or published is accurate, correct, justifiable or even worthwhile!

... In order to determine how robust a published piece of research is, look critically at how the research has been conducted, how data has been gathered, analysed and interpreted, to assess for yourself how valid and reliable you believe it to be. Are the writer's claims adequately supported by the research findings? Identify problems or inconsistencies. Read critically so that you can analyse and evaluate a work in terms of what it purports to achieve and also as it relates to other published work in the discipline ...

From our sample of the information available on current university websites, there appears to be substantial overlap between the academic study skills required of undergraduates and postgraduates. Weir (1983) had found that though some differences occurred in the demands made on, and the problems encountered by, students at undergraduate and postgraduate level, these were not considered significant enough to warrant a separate test for the two academic levels by the Associated Examining Board in their TEAP. The similarity between the study skills and language problems faced by postgraduates as against undergraduates was confirmed by Hawkey and Weir in their 2008 report *Language, study skill and related issues facing international students in the first year of their MPhil/PhD studies: Relevance to institutional language, academic and other support services.*

According to Hawkey and Weir (2008), differences can occur where postgraduates are expected to perform at a higher level in terms of the complexity and range of content they are exposed to, e.g. following complex arguments in difficult subject matter, and a greater degree of reading critically to evaluate an author's position is expected from them. Research design, data collection and analysis skills are also a more prominent feature of the postgraduate toolkit. However, most universities see this additional requirement of research skills as part of the added value their postgraduate courses supply, and accordingly clear provision is usually made for helping students with these. In any case the nature of these research skills would seem to preclude them from being assessed in a valid and practical manner in a time-delimited academic literacy test. This suggests that we might attempt to cater for both groups in one generic study skills examination, but perhaps we might have to:

- demand a higher standard of performance for postgraduates in terms of depth and breadth in their reading and writing
- include additional bolt-on tasks for postgraduates, for example in relation to critical reading.

We have now provided an introductory background survey of what a number of universities and EAP researchers regard as important in terms of academic reading and reading-into-writing skills in tertiary-medium education at both undergraduate and postgraduate levels. In the rest of this volume we examine more closely the specific cognitive and contextual parameters of academic reading (Chapters 2 and 3) and then establish how IELTS has sought to establish reading proficiency in terms of these (Chapters 4 and 5). Chapter 6 looks at consequential validity in terms of the impact and washback of IELTS, in particular its influence on teaching and learning, while Chapter 7 looks at the criterion-related validity of IELTS in terms of its predictive validity in relation to performance on academic courses and its concurrent validity in relation to comparable tests. The third and final section of the volume moves us beyond the present towards consideration of potential task formats and the criterial features that future tests of academic reading should exhibit. Chapter 8 explores how reading-into-writing tasks can help to maximise the authenticity of an academic reading task, while Chapter 9 considers the growing contribution and benefit in a digital age of new technologies for the valid assessment of academic reading ability.

2 Establishing the cognitive and contextual parameters in academic reading: Insights from empirical research

He who loves practice without theory is like the sailor who boards ship without a rudder and compass and never knows where he may cast. Leonardo da Vinci

Introduction

The starting point for our definition of the construct to be measured in a test of academic reading is to be found in the academic reading activities and the problems these present for students (largely undergraduate, but also sometimes postgraduate) in their first year of tertiary-level study. In this chapter we consider important empirical research that has been carried out into the nature of academic reading comprehension over the past 30 years, particularly its *cognitive processes*, skills and strategies. Relevant *contextual factors* based on the nature of the reading texts students are exposed to in their real-life academic reading activities are then discussed, together with other performance conditions obtaining for reading activities in the real-life academic context, e.g. time constraints or response requirements. Chapter 3 will go on to examine various models of reading that have been proposed to take account of these elements.

Empirical bases to support a model of academic reading

The empirical research data we draw on to evidence the salient parameters that might be included in a model of reading for academic purposes were generated by:

- Weir (1983) who investigated the language activities and associated problems of home and overseas students studying at tertiary level throughout the UK in the early 1980s
- Weir, Yang and Jin (2000) who carried out a similar analysis of academic English reading needs of first year undergraduates in China
- Weir, Hawkey, Green, Ünaldi and Devi (2012a) who carried out a needs analysis of first year undergraduates' and postgraduates' reading needs and problems at the University of Bedfordshire in the 21st century.

The findings of these major research studies are complemented by evidence from the extant literature in this area.

Research Study 1: A survey of the languagerelated study problems of UK university students

As briefly introduced in Chapter 1, Weir (1983) surveyed the academic activities and the problems encountered in coping with language-related study skills in tertiary education in the UK. Responses were elicited from 940 overseas students, 530 British students and 560 academic staff, in respect of 43 postgraduate courses and 61 undergraduate courses. Weir found a good deal of variety but also considerable overlap in reading activities (198–226) and writing activities (226–242) across academic levels and across subject boundaries.

Alderson (1988:222–223) attests to the value of such an empirical approach to needs definition:

Partly in reaction to the non-empirical way in which the ELTS was developed, Cyril Weir spent two years devising questionnaires and observation schedules and gathering data, under the guidance of a working party. He sought to identify information on the study demands placed on overseas students in various educational settings (university and college) in the UK. Weir based his instruments on the Munby Communicative Needs Processor, paying particular attention to the enabling skills within that model. The result was a great deal of information about the frequency with which typical students had to do certain activities and about the difficulties such activities presented. This monumental work is available (Weir 1983) for consultation by future test developers, and it represents a major achievement in empirical needs analysis, such that no similar undertaking need be repeated in the foreseeable future for subjects such as Weir's at least, as it provides a substantial database for EAP test development if one is required.

Writing about the ELTS Revision Project 1986–1989, Alderson and Clapham (1992:163) reveal that:

More than half the applied linguists wanted candidates to be given tasks which were as similar as possible to those they would meet during their future courses. Since several analyses have been carried out into the language needs of tertiary level students (in particular, Weir 1983), we used these for the test specifications and tests . . .

Similarly, Weir's 1983 analysis of the reading requirements of academic study in the UK provided a useful base for the development of the Advanced English Reading Test (AERT) in China (Weir et al 2000:28). The authors comment: The design of the questionnaire, especially the section on EAP reading skills and strategies, was to a large extent based on the "Ordered list of reading comprehension enabling skills in an EAP context" in Weir (1983).

The needs analysis proved similarly useful in Eiken's development of the TEAP in Japan (Taylor 2014, Weir 2014) and the General English Proficiency Test (GEPT) Advanced English test developed by the Language Training and Testing Centre (LTTC) in Taiwan in the 21st century (Wu 2012).

Weir (1983:202–206) detailed the types of reading students have to cope with in an academic context based on the returns to his main study questionnaire from 1,470 students and 560 academic staff in the UK in respect of 104 courses of study:

Question B1/1 Reading intensively for comprehension of total text in:

1.1 Duplicated notes relating to the teaching situation – The greatest frequency of occurrence of this type of reading activity was in science and engineering at both undergraduate and postgraduate level. For science 'A' level and social science students it occurred less frequently though still quite heavily in the "sometimes" category. Very few students recorded that they "never" received these.

1.2 Written questions done in class or assigned for homework – For all disciplines, at all levels, this was an activity which occurred "sometimes" or "often" for most students according to the questionnaire returns. If we link this category with the returns to question B1/1.4, the comprehension of examination questions, then we could argue that, at one time or another, this form of intensive reading activity is an important one for all students.

1.3 Reading to extract specific assignment-oriented information – This appears to be the most important reading task for all students at all levels. It is very frequent at postgraduate level in science and engineering and even more so at both undergraduate and postgraduate levels in the social sciences.

1.4 Examination questions – Obviously frequency is a difficult yardstick to apply here as even when they have to do it relatively infrequently, as was the case for postgraduate mathematicians at Exeter, inability to cope with this form of reading can be fatal in terms of academic success. It is a task most students have to perform at some stage in their programme even if it occurs very infrequently in the first year, as appears to be the case in the social sciences.

1.5 Prescribed texts – This is a vital activity for all postgraduates especially in the social sciences and for undergraduates in this area as well. For 'A' level science, and science and engineering undergraduates, it is a less frequent activity, the majority of the replies falling in the "sometimes" category. Overall there is a very low occurrence of replies in the "never" column across disciplines and levels. Both staff and students saw

this as one of the most important reading activities for most students at least in terms of frequency of occurrence. Few students can escape performing this task.

Question B1/2 Reading to extract main information from the text to get a general idea of a topic:

This again is a task which most students have to perform "sometimes", though some science and engineering undergraduates and science 'A' level students seem to manage to avoid it. It is particularly important for postgraduates especially in the social sciences and for undergraduates in that area as well.

Question B1/3 Reading to extract specific assignment-oriented information:

This appears to be the most important extensive reading task for all students at all levels. It is very frequent at postgraduate level in science and engineering and even more so at both undergraduate and postgraduate levels in the social sciences.

It is interesting to note that this is an activity, which students claim to do far more than the staff recorded. A similar discrepancy occurs in a linked form of reading activity, Question B1/6 concerning the frequency with which texts are read to assess their desirability for intensive study. This is perhaps to be understood in the context of the staff recording their replies only in connection with the students in the courses they teach in the programme specified. They are thus talking only about their own particular courses, whereas the students are talking about the overall programme and are exposed to a far wider set of teaching activities. It is also an activity that perhaps the staff are less likely to know about than some of the others.

Question B1/4 Reading to establish and evaluate the writer's position on a particular issue:

This was not a frequent occurrence in science and engineering subjects though it did happen more frequently at postgraduate level according to the questionnaire returns. In general staff considered it hardly ever occurred at 'A' level or undergraduate level except in the social sciences; even at postgraduate level it occurred "sometimes" rather than "often".

Question B1/5 Reading for purpose of monitoring sources of new information and assessing relevance to course of study:

The returns in the "never" column indicate that most science and engineering students do not have to perform this activity below postgraduate level. Even at the postgraduate level, it is only really a frequent activity (in principle at least) for some social sciences students. It would seem to be an activity which overseas students claim to do more often than British students.

Question B1/6 Reading quickly to assess desirability of text for intensive study:

Together with B1/2 and B1/3 this was an extensive reading task, which a number of students across disciplines and levels claimed they performed frequently. Overall, the replies were more varied on this activity perhaps reflecting differences in preferred learning styles amongst the student body as a whole. Engineering and science undergraduates and science 'A' level students again recorded the highest number of entries in the "never" column, and likewise social scientists in the "often" column. The replies of the post-graduate social sciences students again indicate that extensive reading tasks form an essential part of their learning activities. This is borne out in the observations where social science students, particularly at the postgraduate level, were thought to perform this task the most frequently.

Question B2 Making notes from textbooks:

Science and engineering undergraduates seem to perform this task the least. In the case of the former, this is partly explained by the large numbers taking mathematics among our respondents. The engineers received a lot of duplicated notes and took very few notes for themselves even in lectures, apart from what they copied off the blackboard. Only in the social sciences are there large numbers recorded in the "often" category: approximately 60 and 70% for undergraduates and postgraduates respectively. However, at least 20% of all engineering and science postgraduate and 'A' level science students claimed it was also a frequent activity for them. The majority of "nonsocial science" students reported in the "sometimes" category.

Reading difficulties encountered by students

As well as establishing the frequency of various academic reading activities, Weir (1983:207–213) investigated the problems these activities occasioned for the 1,470 students he surveyed. In the final version of the questionnaire it was decided that, given the receptive nature of this skill, it would be better to ask only the students where they experienced difficulty. Thus, in Question B3 students were asked to indicate how much difficulty they had in each of the following (where applicable):

Question B3/l Reading carefully to understand all the information in a text:

Overall the overseas students experienced slightly more difficulty with this than the British. Few overseas students admitted to experiencing "a lot" of difficulty although roughly a third admitted to "some difficulty".

Question B3/2 Reading to get the main information from a text:

Again, the British students experience slightly less difficulty here than the overseas students and a majority of both consider they have "very little"

or "no" difficulty. Social sciences postgraduates, both British and overseas, would seem to have the least reading difficulties of any group with regard to all the questions in this section.

Question B3/3 Search reading to get information specifically required for assignments:

The overseas students seem to experience more problems than their British counterparts in this task and the majority of their replies fall into the "some" or "very little" difficulty categories. Overall, under 20% claimed they had "no" difficulty at all in this. Many students reported a high frequency of occurrence of this activity so that any difficulties encountered here are significant.

Question B3/4 Critical reading to establish and evaluate the author's position on a particular topic:

Both groups of students, overseas and British, experience a noticeably higher level of difficulty with this than they did in the previous reading tasks, though this is a relatively infrequent activity for all except social science students. It is interesting that overseas social sciences undergraduates admit to having a lot more problems with this as against other reading tasks and this can be put down to the difficulties involved as well as the frequency with which they perform the task.

Question B3/5 Reading quickly:

There is quite a difference here between the British students, who largely record "very little" or "no" difficulty, and the much higher numbers of overseas students admitting to "some" or "a lot" of difficulty. There is a connection with the difficulties experienced in Question B3/3 and given the call made upon the extensive reading skills evidenced above there must be cause for concern in this area. Less than 20% of all overseas students claimed they had "no" difficulty here.

Question B3/6 Making notes from textbooks:

Overall the British students seem to have had less of a problem here, but the gap is quite narrow. The biggest difference between the British and overseas students occurs in science 'A' level, social sciences undergraduate and the postgraduate groups, who also had to cope with this task quite frequently. The least difficulties were experienced by the social sciences postgraduates who had to perform this task the most. Overall, many overseas students would seem to experience "some" difficulty in this area.

Question B3/7 Reading texts where the subject matter is very complicated:

It is not surprising that about 60% of all students admitted to having "some" or "a lot" of difficulty here. British undergraduates in engineering and science experienced nearly the same amount of difficulty as their

overseas counterparts in these circumstances. Very few students, either overseas or British, recorded that they had "no" difficulty in this case.

Students were also given the opportunity (B3/8) to specify any other reading difficulties they had encountered. We have attempted to categorise the replies, made by overseas students only, which expand upon or add to the difficulties noted above. The British students made a much smaller number of comments and they did not raise any problems other than those referred to by overseas students below.

Lexis:

"Meaning of specialist terms."

"Difficulty with words not encountered before."

"Speed is affected by having to look up obscure words in appropriate dictionary."

"Understanding of specific biological or chemical terms when they are new to me or are in an unusual context."

"They sometimes can't understand textbook problems when either an unusual word is used, e.g. beam or ridge, or when situations unfamiliar to them are referred to, e.g. cricket, spin driers."

"Misunderstanding the particular use of language in a certain context." *Text related:*

"Reading very large texts."

Complexity:

"Difficulty in reading abstract subjects like sociology text-books."

"Too literary articles."

"Where the passage is very dense i.e. too many important points in two or three lines."

"No pictures."

"Too much reading to do. I think it would help if we were provided with notes so we didn't have to waste so much time on irrelevant information." Operational:

"Books with very small print."

"When the book has no subject index."

"Poorly written books or articles."

"Questions are often badly worded so that it becomes difficult to interpret what the question means."

Syntax:

Structural:

"Difficulty in understanding formal written English."

"Understanding long complex sentences and following an argument through."

"Difficulty in understanding long sentences."

Knowledge and Culture:

"Instructions for many experiments assume that the student has done something similar previously and also that he has practical experience (he often does not)."

"Their troubles generally stem from weak technical background. How to decide whether a valve is open or closed for example."

"Experience in reading through and following an ordered list of instructions seems sometimes lacking."

"When the set of instructions are too long they can carry out a single instruction, but are often 'floored' by a set of them."

Weir (1983:222–226) summarises the reading skills needed by students in the academic context as follows:

Constituent Enabling Skills

On the basis of the enquiry described above and a survey of the literature, we drew up a list of skills which would seem to be involved in academic reading comprehension tasks. As well as testing reading comprehension in an integrated fashion by linking it with listening and writing tasks we may well wish to test discretely a student's competence in these more individual constituent enabling skills. We have concentrated on what Davies and Widdowson (1974) have called the "structuring" and "interpretation" stages since if the candidate was not beyond the "recognition" stage there would be little point in attempting any test of competence at this level. We have set out below those skills, which we feel were important to our target population in the light of the foregoing discussion. . .

A) Reading Skills

1. Reference skills: (a) Reacting appropriately to typographical features, e.g. punctuation, titles, headings, sub-headings. (b) Skills needed when selecting texts or books and deciding whether contents are relevant to needs i.e. establishing background ethnographic information, e.g. by use of table of contents, preface, index, bibliography.

2. Word perception, decoding: deducing the meaning and use of lexical items through understanding word formation and contextual clues. The concern is not with the specialist technical vocabulary of a particular discipline, these having limited and defined meanings, but rather with what we might call sub-technical vocabulary, high frequency context independent words occurring across disciplines, academic vocabulary which has a common focus in research, analysis and evaluation; the activities which characterise academic work.

3. Understanding relations within the sentence: This especially involves an understanding of sentence structure, modification structure, negation, complex embedding.

4. Understanding relations between parts of a text: (a) Through awareness of grammatical cohesion devices especially reference. (b) Through awareness of lexical cohesion devices especially lexical set/collocation.

5. Understanding relations between parts of text by recognising indicators in discourse: Recognition of "indicators", "clues", "linking signals", "signalling devices", especially those used for introducing an idea, transition to another idea, concluding an idea and anticipating an objection or contrary view. 6. Understanding the communicative value of sentences with and without explicit indicators: This includes the "modal, metalingual and contact functions" of certain linguistic devices. These functions are particularly relevant to the understanding of text in activities of a scientific nature, especially the way they are used to develop various methods of planning and organising information in expository language. Munby (1978, pp.185-189) also offered a list of language micro-functions that can occur in this category and lists them under the following broad headings: 1. Scale of certainty. 2. Scale of commitment. 3. Judgement and evaluation. 4. Suasion. 5. Argument. 6. Rational enquiry and exposition.

7. Understanding conceptual meaning: This involves in particular an understanding of quantity and amount, definiteness and indefiniteness, comparison and degree, time, location and direction, means and instrument, cause, result, purpose, reason, contrast, condition. There is a great need among science and engineering students in particular to understand the ways in which these basic notions are expressed in English in their various grammatical and lexical realisations.

8. Understanding explicitly stated ideas and information.

9. Understanding ideas and information in a text not explicitly stated.

(a) Through making inference, e.g. concerning the context in which it was written, causes, reasons, conclusions, opinions, main ideas. (b) Through understanding figurative language.

10. Separating the essential from the non-essential in a text: Distinguishing the main idea from supporting detail by differentiating especially the whole from its parts, statement from example, fact from opinion, a proposition from its argument.

11. Transcoding information presented in a non-linguistic form, e.g. tables, graphs, diagrams: these methods of presenting information are used frequently in many disciplines because they can convey information in a clear and concise way.

12. Skimming a text i.e. not reading every word. (a) Surveying to obtain the gist of a text or a general impression. (b) Searching the text to locate specifically required information on a single point, multiple points or complete topic. Whereas surveying does not require close scrutiny of the text, in search reading we require both rapid reading followed by intensive study depending on "size" of information sought.

13. Note-making: (a) Extracting salient points for summary – this could be a summary of the whole text, a specific idea or topic in the text of the underlying idea or point of the text.

(b) Selective extraction of relevant points from a text – this could involve the co-ordination of relevant information, the ordered rearrangement of contrasting items or the tabulation of information for comparison and contrast.

(c) Reducing a text through rejection of redundant or irrelevant information or items, e.g. determiners, repetition, compression of examples, use of abbreviations.

Cognitive and contextual parameters: Insights from empirical research

14. Critical evaluation: Assessing the worth of a text and the way information in it has been organised and expressed.

One limitation of Weir's 1983 research is that he mainly focuses on the activities involved in reading in the academic context rather than the performance conditions under which these activities are carried out, such as text length, time constraints, what he would later (2005b) address as the *contextual parameters* of the test task (see the section on contextual parameters below for full discussion of these). Nevertheless, based on Munby's (1978) categories of description and Roger Hawkey's doctoral research (Hawkey 1982), Weir provides some useful information on a number of attendant performance conditions for reading activities in the target academic context (1983:325–333).

Size

Engineer (1977), commenting on the reading comprehension components of tests used internally by British universities to assess the English proficiency of first year foreign students, felt that a tacit assumption had been made in many tests, that a few short reading passages represented an adequate sample of the kinds of reading students are required to do at the tertiary level. The ability to understand short passages of about 150 words, single sentences, or even individual grammatical items, was considered to be equivalent to the ability needed to comprehend larger units of continuous prose. Engineer argued for and illustrated the advantages of using longer passages of over 1000 words. She reached the conclusion that a long passage was not only more representative of academic reading in terms of length, but it actually provided more reliable data regarding candidates' reading ability. Employing cloze procedure and multiple-choice items, Engineer showed that the longer the text, the easier it was to discriminate between different levels of reading ability.

Complexity

During the observations of academic events, it became clear that the range of complexity of text that students were exposed to even in the same course and the difficulties involved in adjudging complexity, meant that no easy answer was available for the question "how complex should the texts selected for the test battery be?"

Our tests were to be aimed at a range of levels from G.C.E. [General Certificate of Education] Advanced to postgraduate students. The first term of most science and engineering undergraduate courses is spent in ensuring that people with disparate science and engineering Advanced level backgrounds are brought up to a common standard. We decided that if we selected texts at a degree of complexity that students would have to face in the first undergraduate term or at G.C.E. Advanced level, then we could be reasonably sure that they would not be too difficult for postgraduate students in terms of complexity of language and subject content. Accordingly, we selected texts from appropriate sources at this level and informally checked their suitability with

groups of overseas and British students, subject specialists and testing experts.

Functional range/Referential range

Despite empirical investigation these dimensions, like that of complexity, defied adequate, precise description. For only if we were writing a test for an extremely limited set of students could we precisely specify the nature of suitable texts. It appears, from our earlier survey (v. pp.132-136) that, across disciplines, most texts students faced were in the middling to high functional and referential range categories. There were, however, sufficient entries in the low category to prevent us from making anything other than a general and none too helpful statement, that most students in the three broad discipline areas have to deal with a wide divergence of texts exhibiting a variety of levels in terms of functional and referential ranges.

We decided, therefore, to pitch texts in the pre-test at a level, in terms of functional and referential range, that one would expect an undergraduate in his/her first term at university to be capable of handling. Taking into account the rather inconclusive evidence available on task dimensions we selected a number of texts and the Project Working Party and groups of language teachers commented on which they thought were the most suitable. In general it was felt that the reading material could not be subject free and it should not be too dense or contain difficult vocabulary unless a glossary could be provided. Preliminary trials were then carried out on a group of 30 G.C.E. second year 'A' level native and non-native speakers and 60 first year undergraduate native speakers to ensure that the various texts were appropriate, as regards our task dimensions, the metalanguage of the rubrics was clear and that sufficient time was available for the completion of the tasks set.

A further problem in text selection

A problem was raised to whether the content of proficiency tests should be subject-specific and if so, how to make it subject-specific.

Davies (1965, p.V.2) commented:

"The EPTB made tentative attempts to include alternative reading and listening tests for scientists and non-scientists. Much detailed work needs to be done in specialist areas, medical, legal, commercial and so on."

The situation has improved only slightly since the inception of E.P.T.B. Analyses of the discourse used in the vast variety of courses under review are still not available. Given this current lack of subject-specific analyses in E.A.P./E.S.T., we were forced to compromise.

One approach would be to use tests of general English structures and vocabulary. Research in this area was carried out by Chaplen (1970) who concluded that non-native speakers following courses in British universities need a firm foundation of everyday English if they are to be able to master the variety of English used in their field of study. Accordingly, the two tests of vocabulary and grammar which he constructed were based on "common-core" English. The importance of Chaplen's study was that he provided some evidence for using tests of common core English to

test students' English language skills, in situations where the testees come from a wide range of disciplines.

We showed above that there was a good deal of common ground between students in different disciplines and at different academic levels, in terms of the types of activity faced in the various study modes, the attendant performance constraints and the levels of difficulty encountered. This does not remove the possibility, however, that the subject matter of the text they are presented with in a test may affect performance (v. Alderson et al. 1982).

We were concerned to investigate whether Science and Engineering students perform better on science texts than they do on non-scientific texts and whether the reverse is true for Arts, Social, Administrative and Business Studies (A.S.A.B.S.) students. To this end we decided to construct a version of the test with texts deemed suitable for scientists/engineers (Session IIB) and one with texts more appropriate to students in the humanities and social sciences disciplines (Session IIA). In addition, we agreed to design a general academic version which was aimed at all students irrespective of discipline (Session I), in which texts are selected from what might be termed "science for everyman" sources. We would trial the different sessions on students from the two broad groups so that all three Sessions: Sessions I, IIA and IIB would be attempted by groups of Science and Engineering and A.S.A.B.S. students.

In Session I we aimed to construct a version suitable for students in the whole of the target group. We selected texts from a general science for everyman area. Having looked at texts in a variety of topic areas we decided to focus on the area of health education for the purposes of the pre-test. This seemed a topic area of relevance to both scientists and non-scientists. Thus we selected a reading passage on smoking and health.

For Session IIA, which is aimed at A.S.A.B.S. overseas students, written texts were chosen which related to:

(a) demographic trends amongst the overseas student population,

(b) problems experienced by overseas students, and

(c) remedial teaching services overseas students might expect to take advantage of.

Session IIB was to be aimed at students studying in science and engineering (Sci./Eng.). For this version written texts were chosen which related to:

(a) practical laboratory instructions,

(b) description of an engineering process, and

(c) a general scientific theory relating to the origins of life on Earth.

At the end of his study Weir (1983) reflects on the potential differences faced by undergraduate and postgraduate students across disciplines. He concludes (1983:549):

In our investigations of the language events and activities students have to deal with in British academic environments and the difficulties they encounter therein, we discovered much that was common between students of different disciplines and at different levels. This did not remove the possibility though that the subject content of texts employed in our test tasks might unduly affect performance. Whilst we attempted to take account of this in our sampling, we were unable to produce any conclusive evidence that students were disadvantaged by taking tests in which they had to deal with texts other than those from their own subject area. The case for a variety of E.S.P. tests therefore remains unproven.

Though clearly some differences occurred in the demands made on, and the problems encountered by, students at undergraduate and postgraduate levels and across disciplines, as mentioned in Chapter 1, these were not considered significant enough following up on this research to warrant a separate test for each academic level or discipline by the Associated Examining Board, an issue to which we return below. In summary, Weir's (1983) study:

- reveals the importance of assessing both careful and expeditious global reading skills in a test of academic reading
- makes clear the need for students to process information at both the whole text and intertextual levels, rather than just comprehending at the sentence or inter-sentential levels, i.e. implies the need to test higher-order reading skills at the discourse level not just lower-order skills
- establishes that contextual parameters of the reading passages used in the test are an important consideration in testing academic reading
- provides clear evidence that home students have many of the academic reading problems experienced by overseas students and both groups need to be screened on entry to university
- provides support for having one academic reading test for the different academic levels and discipline areas.

Research Study 2: Development of the Advanced English Reading Test (AERT) for undergraduates in China

In another major academic English reading test development project, Weir et al (2000) conducted research to establish a specification of operations and performance conditions for the AERT for undergraduates in China. They carried out:

- a needs analysis of reading requirements as perceived by academic staff in Chinese universities
- an analysis of EAP reading tasks in textbooks

- an analysis of EAP reading tasks in commercial tests
- a review of theoretical models of reading in the literature.

We will look briefly at the four strands of the enquiry and summarise the relevant findings for our definition of academic reading.

Needs analysis - Chinese university staff

The empirical study first involved a survey of Chinese undergraduates' EAP reading needs as viewed by 55 teachers of advanced academic reading in English in various universities across China. The following is an account of the discussion of some of the important data arising out of their responses to the questionnaire survey.

From this needs analysis, Weir et al (2000) obtained a general picture of EAP reading needs as viewed by subject teachers of advanced reading in English in China:

- There should be a variety of text types in the AERT, including journal articles, abstracts, and chapters from books. Abstracts and chapters from books were the two text types for EAP reading most highly valued by subject teachers, followed by journal articles.
- The average length of texts varied from one type of text to another with abstracts being the shortest (less than 1,000 words) and chapters from books the longest (ranging from 1,000 to 3,000 words).
- The two most important skills and strategies viewed by subject teachers were 'surveying to obtain the gist' (SKM) and 'understanding explicitly stated ideas' (EXMI). The former is the type of expeditious reading strategy frequently referred to as 'skimming' and the latter is the type of careful reading for global understanding which is termed as 'careful reading for explicitly stated main ideas' in our overview of academic reading skills and strategies.

Analysis of texts and tasks in published EAP teaching materials and tests

For a view of what and how EAP reading was currently taught and tested, Weir et al (2000) carried out a survey of all the major EAP reading textbooks and EAP tests available and used in China at the time in terms of operations (skills/ strategies) and performance conditions (length of texts, reading speed etc.). The analysis of EAP reading teaching and testing tasks had two main aims:

- to identify what skills were covered in EAP reading tests and textbooks
- to identify the performance conditions under which these skills were performed in teaching and testing.

Fourteen EAP reading textbooks, the majority published from 1970 to 1994, were identified (see Weir et al 2000 for details).

Ten major previous EAP tests with a separate reading component dating from the 1960s were identified. These comprised: the University Entrance Test in English for Speakers of Other Languages (UETESOL, two versions), the International English Language Testing System (IELTS), the Test of English as a Foreign Language (TOEFL), the Associated Examining Board's Test of English for Educational Purposes (TEEP), the English Language Testing Service – General Academic Module (ELTS – GA/1), Social Science Module (ELTS – SS), and Technology Module (ELTS – T), the EPTB and the English Language Battery (ELBA).

Both teaching and testing tasks were analysed first of all in terms of the skills employed in EAP reading. Through a literature review, potentially important EAP reading skills for the Chinese higher education context had been identified by Weir et al (2000). These provided the descriptive categories for the analysis of reading at tertiary level in China. The skills fall into four broad categories:

- expeditious reading at the global level
- expeditious reading at the local level
- careful reading at the global level
- careful reading at the local level.

Expeditious reading strategies at the global level include both *skimming* for gist and *search reading* for information on predetermined topics. The expeditious reading strategy at the local level is *scanning* for a specific piece of information through pure matching of the target word or looking for a name, date or number etc. Careful reading skills at the global level can be employed for understanding explicitly stated main ideas, for inferring propositional meanings, and for inferring pragmatic meanings; and finally, careful reading skills at the local level are employed for understanding syntax.

The distinctions of these categories of reading elaborated in Urquhart and Weir (1998) were found to be helpful and a short summary of each is provided below to exemplify the categories Weir et al (2000) used in their analysis of tasks in textbooks and tests. These types of reading are discussed more fully later in this chapter in the section 'Separate testing of the reading skills: A posteriori empirical studies on the AERT'.

· Reading carefully for explicitly stated main ideas

Careful and thorough reading of text for explicitly stated main ideas and important information is an important purpose for reading. We often need to decode the whole of a text to understand it all or to establish its macrostructure. In this mode the reader has to read a text at a careful rate from beginning to end in a linear and sequential fashion with regressions as necessary.

· Reading carefully for implicitly stated main ideas

In some texts the ideas may not be explicitly stated and students need to make propositional inferences using explicit statements in the text to form an inference without recourse to knowledge from outside the text.

· Inferring pragmatic meaning related to a text

Pragmatic inferencing takes place when readers rely heavily on their own schemata and/or opinions to interpret a text.

• Skimming

This involves processing a text selectively to get the main idea(s) and the discourse topic as efficiently as possible. The text is processed quickly to locate important information which then may be read more carefully.

Search reading

This differs from skimming in that the purpose is to locate information on predetermined topic(s), for example in selective reading for writing purposes. It is often an essential strategy for completing written assignments.

Scanning

This involves looking quickly through a text to locate a specific symbol or group of symbols, e.g., a particular word, phrase, name, figure or date. The focus here is on local comprehension and most of the text will be ignored. The rate of reading is rapid and sequencing is not usually observed.

Analysis of skills in teaching tasks

The while-reading teaching activities in the textbooks surveyed by Weir et al (2000) showed an equal focus on the training of the identified major categories of EAP reading skills (see list from Urquhart and Weir 1998), with a slightly lower frequency for inferring pragmatic meanings. Out of the 14 textbooks, the occurrences of these skills and strategies were: nine for skimming, 10 for search reading, 11 for scanning, 10 for understanding explicitly stated main ideas, 10 for inferring propositional meanings, seven for inferring pragmatic meanings.

Post-reading activities required readers to summarise and evaluate the text. Tasks at this stage provided readers with a chance to reflect critically on the text and thus promote interaction between the reader and the text. From the analysis, six out of the 14 textbooks included post-reading tasks in the form of writing summaries (WS).

Analysis of skills in test tasks

Firstly, by arranging the 10 tests in chronological order, it was quite clear that there had been a move away from testing items at the lexico-grammatical levels during the 1990s. The data also indicated clearly that until the late 1970s there was a clear emphasis on careful reading.

Secondly, with the exception of the EPTB and the ELBA, all tests had items which tested scanning and reading carefully for explicitly and implicitly stated main ideas, though the percentage of the items was not always the same even within one. Nearly all papers tested search reading. However, very few items appeared to test candidates' pragmatic knowledge since such items would not be equally fair to all candidates.

Thirdly, there were only a limited number of items testing the skimming strategy, which appeared in a maximum of three out of the 10 tests analysed and a maximum of one item per test. Given that skimming items often test comprehension of the text's discourse topic, and as each passage usually had only one discourse topic, this was perhaps not so surprising. On the other hand, given that the number of texts each test had varied from two to six, with a total of 40 texts for the 10 tests, perhaps more of such items might have been expected. This might have been due to insufficient attention to the strategy and the difficulty of writing such items on the length of passages found in most of these tests.

Analysis of performance conditions in teaching tasks

In analysing the performance conditions of the teaching tasks, six prominent textbooks were selected from the textbooks analysed for the skills and strategies. These comprised: McGovern, Matthews and Mackay's *Reading* (1994), Glendenning and Holmström's *Study Reading* (1992), Lynch's *Reading for Academic Success* (1988), Tomlinson and Ellis's *Reading – Advanced* (1988), Salimbene's *Interactive Reading* (1986), and Arnaudet and Barrett's *Approaches to Academic Reading and Writing* (1984).

Based on the performance conditions for reading tests laid down in Weir's *Understanding and Developing Language Tests* (1993), each of the six books identified was analysed in terms of the following conditions: stated purposes for reading; nature of the texts; rhetorical organisation; propositional features, for example, lexical range, topic areas, and background knowledge; illocutionary features; channel of presentation; size of input/length of text; speed of processing; amount of help given; method factor/response mode; questions/answer in first language/target language (L1/TL); and receptive/ productive.

In their analysis, Weir et al (2000) described the texts in terms of their rhetorical organisation, breaking this term down into *collection of descriptions, causation, problem/solution,* and *comparison* according to Meyer and Freedle's terms (1984). The aim in doing this was to see whether there was any relationship between the type of rhetorical organisation of the text and the type of reading skills/strategies that might be employed in the associated tasks. It was recognised that such information would, obviously, be of great help to textbook writers and test developers at the stage of text selection.

The main difficulty that surfaced was that most texts comprised elements from more than one rhetorical type. Clearly, as Carrell (1984:444) points out:

Most prose consists of combinations of these rhetorical patterns; for example, a folktale may contain description, causation, and timesequenced events (that is, collection) within an overall problem/solution organisation where the protagonist confronts and resolves a problem.

Due to the fact that most of the texts analysed aimed to teach skills of both reading expeditiously and reading carefully, it was felt important to add to the conditions information regarding the amount of control over both the skills employed and the time actually to be spent on each passage/task.

The analysis of the conditions underlying each of the six EAP reading textbooks revealed similarities and differences between the textbooks. Firstly, the purpose of reading was stated clearly in all the textbooks. In general terms, the purpose was for the development of various reading skills and strategies so that students could have access to the necessary information for their academic study in a most efficient manner.

Secondly, texts were usually taken from academic books, journals, periodicals, textbooks and reference books. Topics varied greatly but were of general interest rather than of interest only to candidates majoring in a particular academic field. No special requirement on background knowledge was therefore needed.

Thirdly, channels of presentation were mainly textual with some graphics (table, charts, diagrams). Indices, abstracts, content pages, bibliographies and encyclopaedic entries appeared only in one of the textbooks.

Fourthly, the lengths and the rhetorical organisations of the passages used in the textbooks varied greatly. There was a good coverage of text lengths and rhetorical organisations within one textbook. This explained the wide range of skills and strategies covered by the textbooks.

Fifthly, there appeared to be little or no control over the skills students used on EAP reading teaching tasks. In most cases, teaching tasks which involved the use of both expeditious reading strategies and careful reading skills at both global and local levels were based on one passage. Students were therefore exposed to the same passage more than once in completing the tasks. As a result, it was virtually impossible to control the individual amount of time candidates should spend on each passage for each task, which is the most important factor in determining the use of skills and strategies.

Analysis of performance conditions in test tasks

Each of the 10 EAP tests identified was analysed in terms of the conditions specified above for the teaching tasks. The analysis of the conditions underlying each of the 10 EAP reading tests revealed many differences between the

tests. Firstly, there appeared to be little or no control over the skills candidates used on the EAP reading tests under scrutiny as questions testing both reading quickly and reading carefully were based on one passage.

Secondly, there appeared to be little or no control over the individual amount of time candidates should spend on each passage. In IELTS, TEEP and EPTB times were suggested per passage/section but there was no enforcement of these times. Others simply stated the time allowed for the reading test and left the division of that time to the candidates' discretion.

Thirdly, the length of time given over to the reading tests varied from 15 minutes (EPTB) to 60 minutes (IELTS). The UETESOL did not offer any breakdown for its written paper. Candidates were given 2.5 hours within which to complete three components on writing, editing and reading.

Fourthly, the number of items varied from 27 to 40, the exception being EPTB which comprised a C-test of 50 items and a cloze elide test of 163 items in Version C and 191 in Version D (Davies 1984). The number of texts also varied from two to six, with lengths of between 50 words (C-test) and 1,330 (ELTS –SS).

Fifthly, Weir et al (2000) investigated the rhetorical organisation of each passage (according to the Meyer and Freedle (1984) classification discussed above), the length of each passage, and the skills/strategies tested by each passage. Scanning items appeared more often to be based on *collection of descriptions* passages (13 out of the 22 *collection of descriptions* passages had scanning items); items testing reading carefully for explicitly stated ideas were evenly spread between *causation* (7), *comparison* (6) and a *collection of descriptions* (6); and items testing the inferring of propositional meaning appeared predominantly in *collection of descriptions* (10) although they also appeared in six *comparison* and six *causation* passages.

Test formats were also classified. A total number of 11 types were identified, although only a maximum of seven appeared in any one test (IELTS) and in some tests only one test format was used, for example, multiple-choice questions (MCQs) in the TOEFL test. The 11 types comprised: gap-filling items, information transfer-type items, matching items, MCQs, sequencing items, short-answer questions (SAQs), table completion items, text completion items, true/false/not given (T/F/NG) items, and items in the forms of C-tests and cloze elide tests.

The MCQ format had been popular from the 1960s with all TOEFL and ELTS items analysed utilising this format. The 1970s interest in C-tests and cloze tests was represented by examples of these formats in the EPTB, whilst the 1980s and 1990s saw the appearance of a greater variety of test formats in the EAP reading tests. Firstly, the 1980s witnessed the use of SAQs, sequencing items, and a text completion exercise in the TEEP. Then in the 1990s, other formats in the form of T/F/NG, matching, gap-filling, table completion, and information transfer appeared in IELTS.

An analysis of the data indicated that the multiple-choice format was utilised for the testing of every skill/strategy identified. It was particularly heavily used for testing reading carefully for implicit stated main ideas (more than 70 items out of a total of 176), but also quite often for testing scanning and reading carefully for explicitly stated main ideas (33 and 37, respectively).

With respect to the other test formats used, T/F/NG seemed to lend itself more to the testing of explicitly and implicitly stated main ideas (11 and 12 respectively out of a total of 26), whilst the table completion format appeared to be preferred for the testing of scanning (19 items) and to a lesser extent for testing search reading (10 items). Gap-filling seemed to be used more for the testing of explicitly stated main ideas (12 out of 15 items).

At the micro-linguistic level, inferring lexical meaning and the testing of syntactic structures were mainly tested by means of text completion, C-test and cloze elide procedures (17, 50 and 180 respectively) with a small number of MCQs also being used (10–13 items).

Implications for the development of the Advanced English Reading Test (AERT)

The needs analysis conducted with Chinese university staff in China and the test and teaching tasks analyses summarised above, supplemented with a background literature review, threw light on the development of the test specifications for the AERT. The points arising out of these four strands of the research provided guidance for the development of the operations covered in the AERT.

Reading activities

The study of various models of reading proved to be not so productive as they were mostly premised on only one of the identified types of reading, that is, careful reading. Nevertheless, such research drew attention to the importance of reader-driven processing at the whole-text level and the importance of goal setting. The analyses of testing tasks for EAP reading pointed to a similar range of skills to those discovered in the teaching tasks analysis and provided important implications for the design of the AERT. The findings from all four strands of enquiry resulted in a wider specification of the types of reading to be assessed in the AERT:

- The types of EAP reading activity the research indicated should be included were as follows:
 - > careful reading for global comprehension of main ideas
 - expeditious reading (reading quickly, selectively and effectively) for global comprehension of main ideas

➤ expeditious reading at the local level (scanning for specific details).

- The most flexible test formats for EAP reading tests included MCQ, T/F/NG, and SAQ. Table completion and information transfer were exploited mostly for expeditious reading, with cloze elide, text completion and C-test for careful reading. Matching and T/F/NG were used more for reading at the global level than at the local level. Text completion, C-test and cloze elide were used more for local-level reading than for global-level reading.
- Questions could be placed before or after the text depending on the nature of the questions. For long passages which were supposed to be read quickly using strategies of scanning, search reading and skimming, questions were preferably read in advance so that reading of the text would be more purposeful and realistic. Summarising questions could be put at the end of the passage.

The background literature review, the staff needs analysis and the test and teaching tasks analysis also provided some guidance for the development of the test conditions. Research suggested the following performance conditions had an important effect on performance on test tasks and must be carefully considered by the test developer.

Reading performance conditions

Sources of texts

The needs analyses carried out in China (as in the UK by Weir 1983) suggested that academic journal articles, chapters from subject matter textbooks in English and abstracts were the three most frequently employed sources of texts for EAP reading courses. In the AERT, academic journals, such as *Nature* and *American Scientist*, and textbooks were therefore used to provide texts for the expeditious reading sections (both global and local level) and the careful reading section (global level). For careful reading at the local level, i.e. the test of contextualised lexical meanings, texts were initially selected from the Shanghai Jiao Tong University (SJTU) corpus, which exhibited a high incidence of academic vocabulary.

Length of texts

Passages of various lengths were used to allow the testing of different skills and strategies. Different passages were used for the testing of expeditious and careful reading to make students aware of the flexibility of using different approaches to different texts and different tasks. Almost all Chinese university students are exposed to extensive as well as intensive reading. Most have to read at least chapters from books, which can be up to 3,000 words in length. In the AERT, the length of texts varied according to the purposes of reading. For expeditious reading tasks, longer texts together with the requirement of time limits encouraged the use of appropriate strategies. For careful reading tasks, shorter texts with sufficient time provided ensured that texts would be indeed read carefully. In selection, both longer texts (1,000 words) and shorter texts (about 500 words) were included. In short, the AERT developers sought to ensure that:

- length of text would be appropriate for the intended type of reading and the time allocated consonant with this
- the time allowed for completion would be empirically determined for each reading type tested
- there would be strict enforcement of such time controls at the passage/ reading type level.

Rhetorical organisation of texts

From the analysis of EAP reading teaching textbooks and test papers, it was clear that texts with different rhetorical organisations, e.g., comparison, collection of descriptions, causation, and problem/solution (Carrell 1984), lent themselves better to testing different reading skills, that is, for testing a particular skill there might be an optimal rhetorical organisation.

Problem/solution, causative or comparison texts from journals or textbooks seemed to lend themselves better to testing reading carefully for main idea(s) comprehension than more descriptive texts with lots of detailed information. In careful reading, the texts might not necessarily have clear main ideas for selection and main ideas might have to be constructed through propositional inferencing whereas in skimming and search reading they should be explicit.

Where candidates were expected to skim or search-read lengthier texts, these should ideally have a clear, overt structure and be clearly sequenced with a clear line of argument running through them. A journal article or chapter from a textbook with clear sections and headings, where paragraphs contain topic sentences in initial position which signal the information to be presented, might prove suitable for testing expeditious reading. Problem and solution, causative and comparison texts might have the clearest, tightly organised structures (Carrell 1984, Meyer 1975, Meyer and Freedle 1984). Texts which are overtly organised into sections were also recommended. Texts characterised as collection of descriptions, for example, may have an implicit structure which makes them less suitable for search reading or skimming. Texts without a clear structure may well be authentic but they do not lend themselves easily for use in testing expeditious reading as, just as in real life, they are difficult to follow quickly, to summarise or to make notes on.

Collection of description texts (Carrell 1984, Meyer and Freedle 1984) were considered the best vehicle for testing scanning for specific detail. They were more frequently used for teaching and testing scanning in the EAP reading courses surveyed.

Familiarity with topic

Research had indicated that in selecting texts for EAP reading examinations it was the degree of students' familiarity with the topic that had a major effect on their performance (Khalifa 1997). A crucial part of the AERT development was to ensure that students were reasonably familiar with the topics of each of the texts selected from three broad discipline areas (see more on this below).

Ideally, the topic should be generally accessible, i.e. not too obscure and not too familiar, to all candidates. On the one hand, bias in the content background knowledge can be avoided if all candidates share the necessary background knowledge for reading the text. On the other hand, a certain degree of unfamiliarity is necessary to attract readers' attention, to arouse their interest and, what's more important, to prevent them from answering questions from background knowledge without recourse to the text.

Subject specificity

Research had suggested (Clapham 1996a, Khalifa 1997) that the degree of subject specificity of each text should be a key consideration when developing an EAP reading test. The subject specificity of texts therefore needed to be established carefully in advance. Texts selected were seen to be from one of the three subject groupings, i.e. *science and technology, life science, biology and medicine* and *humanities, business and social studies*, but at the same time they should be accessible to students across all three discipline groupings.

Evidence suggests that it is only when the texts are highly specific that the influence of background knowledge on test performance can be demonstrated (Clapham 1996a, Read 2015). General academic texts taken from the three broad discipline areas in the AERT study (science and technology, life science, biology and medicine, and humanities, business and social studies) were viewed as unlikely to disadvantage students from one discipline against another. In addition, highly subject-specific texts risked diverting the focus of the test from reading skills to subject knowledge. Therefore, preference was for the texts of 'L' (low) level subject specificity but sometimes 'M' (medium) level subject specificity if the topic was quite familiar and the language was not too difficult.

Language difficulty

It is not easy to determine the difficulty level of language in a text. Readability might be a rough index but this criterion has met serious challenge from some reading researchers. Subject teachers' judgements in the context of the AERT development were seen as the best method for ensuring selection of texts of a medium level of language difficulty.

Making decisions on the most suitable texts

AERT was intended to be a test applicable to all undergraduate students who had successfully completed the foundation stage study of English in Chinese universities. All candidates of the test should have a feeling that they are being catered for. Therefore, the topics of the texts should be of at least medium familiarity and the language should not be overly difficult. Most importantly the subject specificity should be kept to the medium level or below. Texts satisfying these conditions were considered if appropriate in terms of the further conditions of length, explicitness of macrostructure and rhetorical organisation.

Separate testing of the reading skills: A posteriori empirical studies on the AERT

There was evidence in both the *a priori* and *a posteriori* validation of the AERT to support the case for maintaining separate testing of the reading skills in the four parts of the test (expeditious global and local, careful global and local). The review of the theoretical literature on processing for the project supported a componential view of reading (Weir et al 2000). The *a priori* studies viz the needs analysis, the teaching and test task analyses all evidenced the separability of reading skills for teaching and testing purposes.

The *a posteriori* empirical studies lent further support to a componential view of reading. Both the factor analysis and the cross tabulations in the first and second trials indicated the separable nature of these skills. The retrospection data and the introspection data offered further support for this position. The introspection data showed that for most of the medium-level students and above in Weir et al's (2000) study, use of different styles of reading was promoted in accordance with the different sections of the test. The weakest students, however, appeared to have only one style – slow careful reading, whatever the skill/strategy being tested.

Weir et al (2000:219) revealed that one significant outcome from the trial test results was that while 322 of the students were able to pass the careful reading part of the test for global comprehension of main ideas with a cut

off score of 9/15 (indicated as *totcareful* horizontally in Figure 2.1), a much smaller number achieved a passing score on the expeditious part (indicated as *totexpedit* vertically). So even if you were a good careful reader, this did not mean you could automatically cope with expeditious reading. The vertical columns in Figure 2.1 show those who passed and failed the expeditious part of the test, and the horizontal columns link these to those passing or failing the careful global reading parts of the test.

	totcareful						
totexpedit	0.0 → 8.00	9.00 → 15.00					
0.00 8.00	582	322					
9.00 	39	125					

Figure 2.1: Careful global reading versus expeditious reading

These problems with expeditious reading are often cited in the research literature. Lin and Yi (1997) found that many international English as a Second Language (ESL) students often required extra time to read their textbooks. Similarly, Phakiti and Li (2011:232) note: 'in reading, studies by Durkin (2004), Goodman (1976) and Reid et al., (1998) found that students whose first language is not English need to spend twice and even three times longer than native speakers to finish reading'.

Perhaps, a cause for even greater concern was found in the number of students (326) who passed the careful reading for global comprehension of main ideas part *(totcareful)* but failed the lexico-grammatical *(totlexis)* part of the battery, as shown in Figure 2.2. If a choice is to be made in what is tested, global items, i.e. comprehension beyond the sentence, must surely take precedence when decisions are taken on what to test. The vertical columns in Figure 2.2 show those who passed and failed the careful local reading part of the test and the horizontal columns link these to those passing or failing the careful global reading parts of the test.

This had important ramifications for the ways scores should be reported. Spolsky (1995) succinctly adumbrated the complex and multidimensional nature of comprehension and stressed the need for full description in reporting results as against a single grade or score. He argued (1995:151): Cognitive and contextual parameters: Insights from empirical research

... we will need to design and use a variety of reading assessment procedures ... to allow us to report on a variety of aspects of the student's ability to understand, and to establish some systematic way of reporting the results on all of them. The differences the student shows across this range of results will inform us at least as much as will the result of adding them together. However good our tests are, a single score will always mislead.

	totcareful							
totlexis	0.0> 8.00	9.00 → 15.00						
0.00 ↓ 8.00	573	326						
9.00 	48	121						

Figure 2.2: Careful local reading versus careful global reading

Given the likelihood that different skills could be taught and then tested through the AERT, it was decided that some form of profiling of these abilities was essential rather than collapsing scores into a single score or grade for reporting purposes.

Although Weir et al's (2000) study was conducted in the Chinese context for testing academic English at the start of the 21st century, the findings have important generic implications for other EAP assessment contexts today. The research suggests that:

- texts used in academic English tests should mirror the contextual variables found in real-life academic texts as far as is feasible
- texts with different rhetorical organisations, e.g., comparison, collection of descriptions, causation, and problem/solution (Carrell 1984), lend themselves better to testing different reading skills, e.g., collection of descriptions texts are more suitable for scanning and inferencing
- expeditious reading must be tested as well as careful reading in academic English tests as students can exhibit different profiles in each
- clear time constraints are required for expeditious tasks and also separate texts from the careful tasks
- overtness of text organisation (markers of importance, textual signposting) must characterise passages used for expeditious tasks

- separability of reading skills should be mirrored in the way results are reported
- global items i.e. comprehension beyond the sentence, must always take precedence over local items
- the test formats employed will affect the type of skills that can be tested; some formats are more suitable for global comprehension items and some for local items, e.g. gap-filling items usually cater for local comprehension.

Research Study 3: A survey of reading for academic purposes at the University of Bedfordshire

Building on the two major studies described above, Weir et al (2012a) carried out an empirical study of the reading habits and difficulties in reading of a large sample of students at a British university. Following a survey of the appropriate literature and an initial open-ended pilot survey of reading habits to establish categories for a main questionnaire study, Weir et al (2012a) employed a structured survey to which 766 students responded.

The study focused on the *purposes* for reading in an academic context, in particular how students read for their assignments. The rubric for Section 3 of the main questionnaire was: *The following purposes for reading are important on my course* followed by four statements of reading purposes. Table 2.1 summarises student responses to the items for the English as an additional language (EAL) and English as a first language (EL1) sub-groups.

The following purposes for reading are	E	AL	EL1		
important on my course:	D (r/o)	D&M (r/o)	D (r/o)	D&M (r/o)	
Searching texts to find information for assignments and exams	55.8% (1)	87.7% (1)	77.9% (1)	95.6% (2)	
Basic comprehension of main ideas	35.5% (4)	79.6% (3)	57.9% (3)	90.1% (4)	
Understand meaning of text as a whole; how main ideas and details relate to each other and author's	37.4% (3)	80.7% (2)	53.9% (4)	97.5% (1)	
purpose Integrating information from different texts for use in assignments, exams	40.5% (2)	78.6% (4)	70.4% (2)	91.1% (3)	

Table 2.1: Responses on the importance of reading purposes across EAL and EL1 groups*

* Scale: D = definitely agree; D&M = definitely and mostly agree; r/o = rank order

Analysing the Likert scale responses across the 468 EAL students and the 298 EL1 students, the main inference was that all four purposes were important to the students. Around 80% of people in the EAL group and over 90% in the EL1 group regarded all these core global reading activities as important. One interesting difference was that the EL1 group ranked text-level representation a lot higher (97.5%, definitely and mostly agree) than the EAL group (80.7%, definitely and mostly agree) and the same was true, though to a slightly lesser extent, of the other higher-order reading skills. It may have been that the EAL group found these higher-level reading skills much more difficult to apply to the texts they read and performed them less often as a result.

Table 2.2 presents a re-analysis of the same data on reading purpose, this time to investigate whether there was significant variation across the 761 students in Year 1 and Year 2 student groups.

The following purposes for reading are	Ye	ar 1	Year 2		
important on my course:	D (r/o)	D&M (r/o)	D (r/o)	D&M (r/o)	
Searching texts to find information for assignments and exams	64.1% (1)	91.2% (1)	67.8% (1)	90.7% (1)	
Basic comprehension of main ideas	42.5% (3)	83.5% (2)	48.7% (3)	84.5% (3)	
Understand meaning of text as a whole; how main ideas and details relate to each other and author's	42.4% (4)	82.2% (4)	47.1% (4)	82.4% (4)	
purpose Integrating information from different texts for use in assignments, exams	52.1% (2)	83% (3)	53.3% (2)	85.5% (2)	

 Table 2.2: Responses on the importance of reading purposes across Year 1 and Year 2 groups*

*Scale: D = definitely agree; D & M = definitely and mostly agree; r/o = rank order

As shown in Table 2.2, there was an increase in the definitely agree figures for each reading purpose between Years 1 and 2. Search reading texts for required information was once again seen as the most important reading purpose across the year groups, with the other specified purposes also agreed as important for the students on their courses. The value of search reading to find information for assignments was clear and the fact that the direct testing of this skill was absent from most major EAP tests including IELTS and TOEFL (see Weir, Chan and Nakatsuhara 2013) was thus a cause for concern. There was no substantive difference between Weir et al (2012a) postgraduate and undergraduate sub-samples in terms of the main purposes for their academic reading, all of which were again agreed to be important by both groups.

The data suggested that the inclusion of a dedicated expeditious reading task to assess search reading for global information should be a *sine qua non*

of any academic reading test. There was also a clear case for inclusion of items testing the higher-level skills of understanding main ideas, how a text is structured and the ability to integrate information from different texts (see discussion of the worrying lack of such items in these areas in IELTS in Chapter 4).

Reading for assignments

Section 4 of the questionnaire, under the rubric *How I read for assignments*, included 16 items. The results for each of these are tabulated in Table 2.3. The table summarises responses in terms of the strength of agreement with each item across the EAL and EL1 groups, with the final two columns making Year 1:Year 2 comparisons. The table contains a considerable quantity of data on the nature of academic reading activities across a range of courses.

The items covered conscious metacognitive strategies by the reader taken before the reading begins, e.g., I think carefully make sure I know exactly what I'll be looking for before I start reading, and while reading is taking place, e.g., While reading I try to relate content to what I know already and judge its value. Many of the strategies specified related to expeditious reading at the global and local levels, e.g., I look at the titles or headings of a text before deciding to read it quickly, I first get overall meaning of text e.g. by reading first paragraph and conclusion and first sentence of other paragraphs and I think carefully of kev words and quickly look for them or words with similar meanings to check if text is worth reading more carefully. Some of the strategies covered careful reading approaches at both the local level, e.g., If I do not know the meaning of a word in a text, I try to work out its meaning, and at the global level, e.g. I try to understand how the text is organised, how the ideas and details connect with each other, I gradually understand what a text is about by reading the sentences slowly and carefully in the order they occur and I read critically to establish and evaluate the author's position on a particular topic.

Evidence from these analyses suggested that the following reading strategies were the most strongly agreed by the questionnaire respondent sample, with differences between the EAL and the EL1, Year 1 and Year 2 sub-groups as stated.

- The strategy *remembering where relevant information is or marking its location for later use in writing my assignment,* is definitely or mostly agreed by 77.1% of the EAL group and 86% of the EL1 sub-groups, and 85.4% of the Year 2 group and 79.2 of the Year 1 students.
- *I think carefully to make sure I know exactly what I'll be looking for before I start reading*, is definitely or mostly agreed by 77.4% of the EAL group and 80.1% of the EL1 sub-groups, and 80.5% and 76.1% of the Year 1 and Year 2 sub-groups respectively.

How I read for assignments		EAL		EL1		Year 1	Year 2
		D (r/o)	D&M (r/o)	D (r/o)	D&M (r/o)	D&M (r/o)	D&M (r/o)
1.	I think carefully to make sure I know exactly what I'll be looking for before I start reading	40.4% (2)	77.4% (2)	38.0% (3=)	80.1% (2=)	80.5% (2)	76.1% (2)
2.	I look quickly through the whole text for a general understanding before doing	28.4% (6)	72.9% (4)	28.5% (8)	68.8% (10)	72.8% (4)	67.3% (8)
3.	I gradually understand what a text is about by reading the sentences slowly and carefully in the order they occur	21.6% (10)	56.7% (14)	23.1% (11)	52.9% (15)	54.8% (14)	56.4% (12)
4.	I remember where relevant info is or mark its location for later use in writing my assignment	36.3% (3)	77.1% (3)	47.6% (1)	86% (1)	79.2% (3)	85.4% (1)
5.	I think carefully of key words and quickly look for them or words with similar meanings to check if text is worth reading more carefully	29.6% (5)	70.6% (5)	38.0% (3=)	81.0% (2=)	70.6 (6)	68.0% (7)
6.	I look at the titles or headings of a text before deaiding to read it quickly	44.2% (1)	82.1% (1)	33.4% (6)	73.3% (9)	81.1% (1)	73.8% (4)
7.	I first get overall meaning of text e.g. by reading first paragraph and conclusion and first sentence of other paragraphs	19.2% (13)	57.6% (13)	16.8% (15)	53.4% (14)	56.4% (13)	55.3% (13)
8.	If I do not know the meaning of a word in a text, I try to work out its meaning	27.6% (7)	63.8% (10)	35.5% (5)	77.5% (7)	69.9% (8)	66.6% (9)
9.	I read a text slowly all the way through even if some parts do not seem relevant to my assignment	8.5% (16)	33.4% (16)	10.6% (16)	30.7% (16)	33.4% (16)	32.6% (16)
10.	I read slowly only those sections of a text I have marked as relevant when going through it quickly before	21.4% (11)	67.6% (6)	21.9% (12)	63.7% (11)	66.7% (11)	65.7% (11)
11.	While reading I try to relate content to what I know already and judge its value	22.8% (8)	67.4% (7)	27.9% (9)	78.9% (5)	71.2% (5)	75.0% (3)
12.	I look back at previous parts of the text to check meaning	18.2% (14)	64.4% (9)	27.8% (10)	75.6% (8)	69.8% (9)	66.3% (10)
13.	I try to understand how the text is organised, how the ideas and details connect with each other	19.4% (12)	61.9% (12)	18.8% (14)	62% (12)	66.6% (12)	52.0% (15)

 Table 2.3: Responses on ways of reading for assignments across EAL and EL1

 and Year 1, Year 2 sub-groups

How I read for assignments	E	EAL		EL1		Year 2
	D (r/o)	D&M (r/o)	D (r/o)	D&M (r/o)	D&M (r/o)	D&M (r/o)
14. I make notes on relevant points from the text as I go along	31.0% (4)	66.4% (8)	41.2% (2)	78.7% (6)	70.3% (7)	72.3% (5)
15. I integrate information from the text I am reading with information from other texts I have already read	22.6% (9)	62.4% (11)	31.3% (7)	79.1% (4)	69.3% (10)	69.7% (6)
 I read critically to establish and evaluate the author's position on a particular topic 	12.9% (15)	51.4% (15)	20.1% (13)	59% (13)	54.3% (15)	55.3% (13)

Table 2.3 (continued)

*Scale: D = definitely agree; D&M = definitely and mostly agree; r/o = rank order

- *I think carefully of key words and quickly look for them or words with similar meanings to check if text is worth reading more carefully* is definitely or mostly agreed by 70.6% of the EAL group and 81% of the EL1 sub-groups, and 70.6% and 68% of the Year 1 and 2 groups respectively.
- *I look at the titles or headings of a text before deciding to read it quickly* is definitely or mostly agreed by 82.1% of the EAL group and 73.3% of the EL1 sub-groups. While this strategy receives a strong endorsement from both sub-groups, it is more strongly agreed with by the EAL sub-group (p<.01), who actually rate it their *most* strongly agreed strategy; there also appears to be a difference between the Year 1 group's top ranking of this strategy and the Year 2 students' fourth ranking of the strategy (p<.01).

Interestingly a number of the more favoured strategies/skills related to expeditious rather than careful reading, and were global rather than local in the sense that they appeared to involve searching *a whole text* for relevance to an actual assignment task rather than understanding at the sentence level. While this finding echoed the conclusions made in Weir's (1983) and Weir et al's (2000) studies, Weir et al's (2012a) survey showed that these reading skills are not only necessary for the EAL students but also the native students. This indicates a need to assess/teach all students, not just the EAL group, these important academic reading skills.

Weir et al (2012a) also considered those academic reading strategies which were not strongly endorsed by the students; they included those that were *less* expeditiously reading related. The following strategies saw the weakest agreements, in terms of their rank orders:

- *I read a text slowly all the way through even if some parts do not seem relevant to my assignment.* This clear example of careful reading was the lowest rated and ranked by all groups (EAL, EL1, Year 1 and Year 2) as measured by all definitely or mostly agree measures.
- *I read critically to establish and evaluate the author's position on a particular topic.* This strategy was low-ranked by all groups. These results tallied closely with what Weir (1983) established in Research Study 1 in this volume about the infrequency of this activity, especially for undergraduates.
- *I gradually understand what a text is about by reading the sentences slowly and carefully in the order they occur.* This strategy was low-ranked (r/o 10/15, positive rating 56.7% and 52.9% for the EAL and EL1 groups respectively, and 54.8% and 56.4% for Year 1 and Year 2 students respectively).

The cross-tabulation of the undergraduate:postgraduate sub-groups in terms of assignment reading strategies suggested that the two groups did not differ in many of these. The main exception appeared to be *look quickly through the whole text for a general understanding before doing anything else* (a stronger tendency for the postgraduates than the undergraduates; definitely and mostly agree percentages 80.7% and 69.7% respectively).

Student difficulties encountered when reading for assignments

Weir et al (2012a) also looked at the *difficulties* encountered by students when reading for their assignments. Overall, as indicated in Table 2.4, it was the similarities rather than the differences across the EAL and EL1 sub-groups that were striking. A similar picture could also be found for the reading problems across Year 1 and Year 2 students. Analysing the Likert scale responses, whether judged in terms of the proportions of definitely agree percentages or the combined definitely and mostly agree categories across the 468 EAL students and the 298 EL1 students, the main difficulties experienced, in rank order, appeared to be with:

- 1. Reading texts where the subject matter is complicated.
- 2. The time available to do the necessary reading.
- 3. Reading lengthy texts.
- 4. Finding relevant information quickly.

Table 2.4 indicates that *time constraints* and *reading texts where the subject matter is complicated* were the major problems for both EAL and EL1 groups. 57.7% of the EAL group and 59.4% of the EL1 group definitely and mostly agreed they experienced difficulty with *time constraints*, and 58.5% of the EAL group and 70.2% of the EL1 group definitely and mostly agreed

Difficulties when reading for		EAL		EL1		Year 1	Year 2
assi	gnments	D (r/o)	D&M (r/o)	D (r/o)	D&M (r/o)	D&M (r/o)	D&M (r/o)
1.	The time available to do the	24.3%	57.7%	31.3%	59.4%	57.0%	61.9%
	necessary reading.	(1)	(2)	(2)	(2)	(2)	(2)
2.	Reading texts where the	18.8%	58.5%	32.2%	70.2%	61.0%	67.7%
	subject matter is complicated.	(2)	(1)	(1)	(1)	(1)	(1)
3.	Words I do not know.	15.6%	44.0%	13.0%	34.6%	41.5%	37.6%
		(3)	(4)	(5)	(6)	(5)	(7)
4.	Sentence structures.	11.1%	35.6%	5.8%	21.6%	32.5%	25.0%
		(8)	(10)	(13)	(15)	(10)	(16)
5.	Finding relevant information	14.1%	43.3%	16.3%	42.9%	43.9%	41.3%
	quickly.	(4)	(6)	(4)	(4)	(4)	(4)
6.	Lengthy texts.	13.0%	47.3%	17.1%	47.2%	45.9%	50.9%
	2 7	(5)	(3)	(3)	(3)	(3)	(3)
7.	Lack of background	12.1%	41.8%	12.4%	35.5%	39.1%	40.5%
	knowledge to understand the content.	(7)	(7)	(6)	(5)	(7)	(5)
8.	Making notes on	8.8%	26.1%	8.6%	19.9%	21.8%	27.8%
	information I will need.	(14)	(17)	(10)	(16)	(17)	(13)
9.	Reading carefully to	10.9%	31.4%	5.2%	23.5%	28.0%	29.0%
	understand the main ideas.	(9)	(15)	(15)	(14)	(15)	(12)
10.	Summarising ideas from a	12.8%	43.5%	11.1%	33.6%	40.9%	38.1%
	text in my own words.	(6)	(5)	(8)	(8)	(6)	(6)
11.	Understanding a detailed	8.0%	37.3%	5.2%	25.3%	32.7%	32.7%
	logical argument.	(16)	(9)	(15)	(11)	(9)	(10)
12.	Reading critically to establish	8.5%	37.9%	8.0%	33.9%	35.6%	36.5%
	and evaluate the author's position on a particular topic.	(15)	(8)	(11)	(7)	(8)	(8)
13.	Relating the content of a text	7.6%	30.7%	4.5%	18.7%	27.3%	24.0%
	to my existing knowledge.	(17)	(16)	(17)	(17)	(16)	(17)
14.	Deciding what is important	10.2%	31.7%	11.9%	32.6%	32.2%	32.1%
	for me and what is not.	(10)	(13)	(7)	(9)	(11)	(11)
15.	Reading a text quickly to	9.5%	34.2%	11.0%	30.4%	32.0%	33.1%
	decide whether I should	(11)	(12)	(9)	(10)	(12)	(9)
	study it carefully.	< <i>'</i>			· /		
16.	Understanding the text as a	8.9%	34.6%	5.5%	23.5%	28.9%	26.4%
	whole: how main ideas and	(13)	(11)	(14)	(12)	(13)	(14)
	details are connected to each other.	()	()	()	()	()	(-)
17.	Integrating information from	9.1%	31.7%	7.6%	23.5%	28.4%	26.4%
	text I am reading with info from other texts I have read.	(12)	(13)	(12)	(12)	(14)	(14)

Table 2.4: Responses on difficulties encountered with reading for assignments across EAL and EL1 and Year 1 and Year 2 sub-groups

*Scale: D = definitely agree; D&M = definitely and mostly agree; r/o = rank order
they experienced difficulty with *reading texts where the subject matter is complicated*. The two items referring directly to these problems were the highest rated and ranked for difficulty for both EAL and EL1 groups, the EL1 group finding significantly *more* difficulty (p<.01) with complicated subject matter, suggesting that first language status is not necessarily the major issue when the study subject itself is complicated. Closely related to the problem students faced with the amount of time available for reading was the difficulty they experienced in *finding relevant information quickly*, a significant problem for 43.3% and 42.9% of the EAL and EL1 groups respectively, and *reading lengthy texts*, a big problem for 47.3% and 47.2% of the EAL and EL1 groups respectively.

Table 2.4 suggests the following potential reading-related difficulties were *not* so highly rated or ranked:

- *relating the content of a text to my existing knowledge* ranked the least or penultimate of the difficulties of both EAL and EL1 groups, with 30.7% and 18.7% respectively definitely or mostly agreeing to finding difficulty
- making notes on information I will need was low-ranked as a readingrelated skill by both groups, only 26.1% and 19.9% respectively of the EAL and EL1 groups definitely or mostly agreeing they saw it as difficult in reading for assignments
- understanding the text as a whole; how main ideas and details are connected to each other and integrating information from text I am reading with info from other texts I have read 34.6% and 23.5% respectively of the EAL and EL1 groups definitely or mostly agreed they saw it as difficult in reading for assignments.

The final two columns in Table 2.4 permitted inferences from the Year 1:Year 2 variable on the matter of student problems in academic reading. Here, as in Table 2.3, Year 1 and Year 2 student responses were compared in terms of percentages of definitely agree plus mostly agree responses and in terms of rank orders of the 17 reading difficulties concerned. As signalled above, it was the *similarity* of the perceptions of the reading problems of the Year 1 and Year 2 students which was notable. This suggested that these problems do not disappear with exposure to a wide range of reading in the first year of study.

Once again, as with the EAL:EL1 group difficulty comparisons above, the same four main problems were identified, in the same rank order by the Year 1 and the Year 2 sub-groups, namely:

- 1. Reading texts where the subject matter is complicated.
- 2. The time available to do the necessary reading.
- 3. Reading lengthy texts.
- 4. Finding relevant information quickly.

Research and Practice in Assessing Academic Reading

Note that, once more, the evidence is strong that students find that the amount of time at their disposal is inadequate to handle the problems of finding relevant information quickly from long and complicated texts.

The most substantial difference among the academic reading problems between our undergraduate and postgraduate student sub-samples appeared to be with difficulty in finding relevant information quickly, a problem for 45.3% of the former group and 31.7% of the latter, suggesting that this problem has been sorted for more students by the time they get to postgraduate level, though still a considerable number would benefit from some remedial intervention here.

The findings on these difficulties encountered by students when reading for their assignments merit serious consideration in the design of academic English tests which claim to be valid indicators of the suitability of the English language proficiency of students seeking to study at British and other EL1 universities. As in Research Studies 1 and 2, it is clear that coping with longer texts expeditiously under time constraints is a significant difficulty for many students and a test that only measures the ability to slowly and carefully comprehend short passages will not account for this important part of the academic reading construct. Students have to demonstrate the ability to process longer texts under time pressure if we are to have a valid picture of their readiness to cope with the demands that will be made on them in academic reading.

For those intending to develop EAP reading tests, Weir et al's (2012a) study provided clear evidence for the necessity of:

- a dedicated expeditious global reading task to extract main ideas from a longer text under time constraints
- · items testing the higher-level skills of understanding main ideas
- items testing how a text is structured
- · items testing the ability to integrate information across different texts
- ensuring there is no ceiling effect on difficulty, e.g., when texts are on academic topics intended for a general audience sourced from newspapers or magazines. Texts must be at the difficulty level of undergraduate texts and automated text analysis by item writers can help to ensure this.

Conclusion

This chapter has examined in detail the focus and findings of three major research studies conducted to investigate empirically the nature of academic reading in English. Taken together, the results of the studies provide us with valuable evidence to inform the development of an academic test of reading.

The studies reveal the importance of assessing both careful and expeditious global reading skills in a test of academic reading. Expeditious reading must be tested as well as careful reading in academic English tests as students can exhibit different profiles in each. Clear time constraints are required for expeditious tasks and also separate texts from the careful tasks. Overtness of text organisation (markers of importance, textual signposting) must characterise passages used for expeditious tasks. The separability of reading skills should be mirrored in the way results are reported.

There is a clear message about the need to focus on global comprehension of texts in each of the studies. They make clear the need for students to process information at both the whole text and intertextual levels, rather than just comprehending at the sentence or inter-sentential levels. This implies the need to test higher-order reading skills at the discourse level, not just lowerorder skills: items testing the higher-level skills of understanding main ideas; items testing how a text is structured; and items testing the ability to integrate information across different texts.

Global items, i.e. comprehension beyond the sentence, must always take precedence over local items as this is at the heart of what students will need to do in academic life. Studies 1 and 3 provide clear evidence that home students encounter many of the academic reading problems experienced by overseas students and both groups need to be screened on entry to university. They provide support for having one academic reading test for the different academic levels and discipline areas.

The studies establish that contextual parameters of the reading passages used in the test are an important consideration in testing academic reading. Texts used in academic English tests should mirror the contextual variables found in real-life academic texts as far as is feasible. There is a need to ensure that there is no ceiling effect on difficulty, e.g., when texts are on academic topics intended for a general audience sourced from newspapers or magazines. Texts must be at the difficulty level of undergraduate texts and automated text analysis by item writers can help to achieve this. Texts with different rhetorical organisations, e.g., comparison, collection of descriptions, causation, and problem/solution (Carrell 1984), lend themselves better to testing different reading skills, e.g., collection of description texts are more suitable for scanning and inferencing. The test formats employed will affect the type of skills that can be tested. Some formats are more suitable for global comprehension items and some for local items, e.g. gap-filling items usually cater for local comprehension.

In Chapter 3 we will extrapolate from our three major research studies on academic reading, linking their empirical findings to theoretical insights from the wider literature on reading. Our aim will be to show where they fit in a model of cognitive and contextual parameters in reading. Additional recourse to the extensive research literature on models of reading will enable us to develop a comprehensive theoretical framework for investigating the cognitive (see Figure 3.1) and contextual aspects (see Table 3.2) of an academic reading test.

B Modelling the construct of academic reading: Insights from theoretical research

He has only half learnt the art of reading who has not added to it the even more refined accomplishments of skipping and of skimming. Arthur Balfour (1905)

Introduction

In this chapter we begin by considering various models of reading that have been proposed in the literature in order to clarify the relationships between the key facets of the academic reading construct. We seek to link the findings from the empirical research described in Chapter 2 to existing models of reading, and in particular to the development of a sociocognitive model which offers a platform for academic reading test development and a framework of reference for analysing and critiquing tests of academic reading.

The question of componentiality

Perhaps the most fundamental issue in the development of a model of academic reading is the *componentiality* of the reading construct. As Weir et al (2000:14) ask: 'Can reading be broken down into underlying skill or strategy components for the purposes of teaching and testing?' The discussion of the empirical data on students' reading activities in Chapter 2 suggests that it can, but additionally the reading research provides examples of what Weir and Porter (1994) refer to as 'unitary', 'bi-divisible' and 'multi-divisible' models of the reading construct. They cite empirical studies supporting the single factor hypothesis including Carver (1992), Lunzer, Waite and Dolan (1979) and Rosenshine (1980). Schedl, Gordon, Carey and Tang (1996), in their TOEFL research report on the dimensionality of the TOEFL reading comprehension items, also provide evidence for the existence of a general reading ability and the essential unidimensionality of the TOEFL Reading test, although they accept that there may be a second factor relating to text content.

Weir et al (2000) and Khalifa and Weir (2009) suggest that part of the reason for the occurrence of a unicomponential view of the reading construct in some research is that product-based studies of reading test scores typically use factor analysis as their measurement tool of choice. Researchers examining construct validity are looking for unidimensionality, and are therefore

foregrounding statistical homogeneity over psychological interpretation of item responses (Henning 1992). Factor analysis is all about *reduction* (Bachman 2004), and may be somewhat insensitive to subtle differences such as those across related reading skills and processes. Factor analysis may thus tend to show apparently different reading skills behaving in similar statistical ways (Lunzer et al 1979, Rosenshine 1980). This may be taken to imply that there is a single broad ability of reading rather than a range of skills and strategies involved in the activity.

However, more process-oriented studies, as reviewed in Chapter 2, clearly suggest the reading construct has more than one dimension. Note the bi-divisible views of reading cited in Weir et al (2000), Carver (1992) and Guthrie and Kirsch (1987), where the two components appear to be reading competence and vocabulary. The Schedl et al (1996) model of the TOEFL Reading test may also be considered bi-dimensional. Coady (1979) offers a three-component model (conceptual ability, language proficiency, background knowledge) as does Bernhardt (1991) (language, literacy, and knowledge including metacognitive strategies such as goal-setting and comprehension monitoring).

These models with a relatively small number of sub-components appear less in tune with recent attempts to define English for Speakers of Other Languages (ESOL) learner and user language proficiency levels more closely in the interests of transnational education and employment mobility (e.g., the CEFR, Council of Europe 2001). In the 21st century, reading skills are increasingly described in comprehensive, multi-componential target language domain terms. As Urquhart and Weir's (1998) matrix in Table 3.1 suggests (see page 71), with its careful and expeditious reading cells, each operationalised through a range of skills at both local and global levels, reading is now seen as a more complex *componential* construct.

Grabe and Stoller (2002) support this view and classify reading processes into higher and lower-level processes. The lower-level processes include word recognition (lexical access), syntactic parsing, semantic proposition formation, and working memory activation. The higher-level processes comprise the formation of a text model of comprehension, a situation model of reader interpretation, background knowledge use and inferencing, and executive control processes which appear to be similar to meta-cognitive strategies.

A further division in the literature into the contrasting categories of *bottom-up* and *top-down* models of reading is also worth brief consideration in our development of an appropriate model for university student reading and its assessment. Bottom-up models tend to operate in terms of a hierarchical written text, from grapho-phonic, phonemic, syllabic, morphemic, word, to sentence levels. According to Dechant (1991:23), readers are assumed first to process 'the smallest linguistic unit, gradually compiling the smaller units to decipher and comprehend the higher units (e.g., sentence syntax)'. Top-down

processing involves the general and domain-specific knowledge that readers can employ to predict text meaning and to comprehend sentences and words within a text (see Bernhardt 1991).

There are also *hybrids*: reading models combining the reasonable insights of both the bottom-up and top-down models. The interactive reading model (e.g. McCormick 1988), developed further by Kintsch (2004) in his construction-integration model of text comprehension, emphasises the reader-driven, purposeful and conscious aspects of reading noted above (see also Weir et al 2000). Further acknowledgement of the reader role in reading is provided in the *interactive-compensatory model* of Stanovich (2000), which suggests that a specific weakness of a reader in a particular skill may be made up for by strengths in others.

A multi-dimensional and dynamic approach

Based on the evidence of theoretical and empirical research into models of reading, we feel that an appropriate model for EAP testing would be a multidimensional, dynamic model of reading, which takes into account as far as possible the global and local levels of reading as well as the metacognitive strategies and the processes involved in understanding texts from various sources for various purposes, employing top-down and bottom-up processing singly or in combination as appropriate.

Alderson (2000) suggests that a problem in actual testing practice is that numerous reading skills probably exist, but are difficult to identify accurately and test separately. We take the different view that a growing body of the research literature suggests that it is possible, with clear specification of terms and appropriate methodology, for testers to reach closer agreement on what skills are being tested. The body of literature includes Bachman, Kunnan, Vanniarajan and Lynch (1988), Weir and Porter (1994), Buck, Tatsuoka and Kostin (1997), Lumley (1993), Teasdale (1989) and Weakley (1993). As Khalifa and Weir (2009) wryly point out, in the DIALANG project individual items are now also viewed by Alderson and his colleagues as being associated with identifiable skills (see Alderson 2005).

Koda (2005) feels that the successful identification of specific components that contribute to reading ability is an important paradigm in the current reading research literature. A componential approach based squarely on a sound theory of processing can be useful in that it provides insight into potential components in reading ability which require our attention if we are to approximate to a valid construct of reading in our reading tests.

Oakhill and Garnham (1988:48) feel that the problem is that much of the earlier research focused on product rather than process in reading. Khalifa and Weir (2009) similarly point out that what was largely absent in the componential approach in the past (leaving aside the later process-oriented

studies) was any serious attempt to relate components to a model of reading ability. They argue that this may stem from an earlier preference for *a posteriori* statistical analysis of construct in the testing community as against an *a priori* approach concerned with both the theoretical underpinnings of a test's construct before it is administered and its contextual validity.

The main criticism of the product-based, *a posteriori*, statistically driven approach is that it was not usually based on a sound analysis of salient cognitive processes that might be initiated by various tasks in a reading test. Furthermore, by its nature, it told us little about what is actually happening when a reader processes text. Further insight may be possible if we attempt to go deeper and examine as far as is possible the actual processing that goes on during reading activities. If we can identify skills and strategies that appear to make an important contribution to the reading process, it should be possible to test these and use the results for reporting on reading proficiency (see Shiotsu 2003, Urquhart and Weir 1998, Weir et al 2000, for a further discussion of these issues).

Understanding the nature of comprehension

In our search for differentiated reading skills and strategies we turn to the theory of what it means to 'comprehend'. Grabe (1991) offers a useful list of component skills in reading on the basis of reading theories (as against an earlier reliance on 'armchair intuition'). We have added some supporting references to his list:

- 1. Automatic recognition skills (see Perfetti 1997).
- 2. Vocabulary and structural knowledge (see Bachman 1990 on grammatical competence, Perfetti 1997 on syntactic parsing and word representation knowledge).
- 3. Formal discourse knowledge (see Koda 2005).
- 4. General and domain knowledge (see Carrell 1983 on formal schemata, Anderson and Pearson 1988 on content schemata, and Kintsch 1998 on domain knowledge).
- 5. Identifying central ideas of a text (see Baumann (Ed) 1986 and Oakhill and Garnham 1988).
- 6. Inferencing skills (see Chikalanga 1990, 1992).
- 7. Metacognitive knowledge (see Urquhart and Weir 1998, Weir et al 2000).
- 8. Skills monitoring (see Carrell, Devine and Eskey (Eds) 1988).

The work of Enright, Bridgeman and Cline (2002) in the TOEFL 2000 project supports this breakdown. Khalifa and Weir (2009) critically add the need to process and integrate information from several texts in a related field and suggest:

The cognitive construction of intertextuality offers a useful heuristic for looking at reading-into-writing at an advanced level and it extends our view of reading beyond the act of comprehension of a single passage (2009:54).

Contemporary theory thus points to the value of a componential approach. The adoption of a multi-componential approach is supported by the literature and the empirical enquiry into the reading activities of university students in the three major research studies reported in Chapter 2. All these data contribute to our descriptions of the activities that should be mirrored in academic reading tests.

By more closely relating these putative skills/strategy components to a cognitive model of academic reading with an empirically grounded base in cognitive psychology, we may be able to better define what should be tested in EAP reading tests. Khalifa and Weir (2009) developed a sociocognitive model for testing reading, which sets out to do exactly this (see Figure 3.1). A number of recent research studies on the validation of reading tasks have accordingly selected this sociocognitive model initially developed by Weir (2005b) and then improved by Khalifa and Weir (2009) as the theoretical basis for their research into the construct being measured by various reading activities (Brunfaut 2016, Brunfaut and McCray 2015, Owen 2016). Owen (personal communication, August 2017) argues:

... a strength of the model is flexibility and falsifiability. Aspects of the model can be "sampled" if specific research agendas focus on metacognitive approaches (e.g. goal-setting) to reading, or purely psycholinguistic, during reading processing approaches. The model is empirically-based, meaning it can be subjected to further empirical scrutiny as the field develops.

In the left-hand column of Figure 3.1 is the meta-cognitive activity of a *goal setter* because, in deciding what **type of reading** to employ when faced with a text (e.g., careful global, or expeditious local), critical decisions are taken which affect the level(s) of processing to be activated in the central core of our model. The various elements of this processing core in the middle column are thus initiated in accordance with decisions taken in the goal setter. The model attempts to define the cognitive processing levels that may be initiated through the various types of reading (see column 2 in Figure 3.1). The *knowl-edge base* displayed in the column on the right-hand side of the model (see Figure 3.1) is drawn upon by elements in the central processing core in line with the intended type of reading and the performance conditions established by the test task. The remainder of this chapter will describe the different elements of the sociocognitive model in greater detail to show how it represents the activity of reading, and of academic reading in particular.



Figure 3.1: A sociocognitive model for developing and validating reading tests

The sociocognitive model – Part 1: Selecting careful or expeditious types of reading

Before reading a text, a decision is usually made by the reader on how they are going to approach the text. In the model above we suggest that this happens in something called the *goal setter*, in the left-hand column of our model. Decisions are taken here on the appropriate type of reading that will take place: *globalllocal* and *carefullexpeditious*.

Careful reading of a text is characterised as slowly and incrementally identifying lexis, parsing syntax, seeking an accurate comprehension of explicit meaning and making propositional or pragmatic inferences, mental model building, finally ending in a construction of a text-level representation (see Figure 3.1).

Careful reading can take place at a local or a global level, i.e. within or beyond the sentence right up to the level of the complete text or across several texts. Careful reading might involve:

- separating explicitly stated main ideas from supporting detail by recognising topic sentences or by recognising lexical indicators of importance
- understanding the development of an argument and/or logical organisation
- generating a representation of the text(s) as a whole.

It might also involve propositional inferencing. This might mean making:

- propositional informational inferences which are either referential, typically answering questions beginning with what and which, or spatio-temporal, typically answering questions beginning with where and when
- propositional explanatory inferences which are concerned with motivation, cause, consequence and enablement and will often answer questions beginning with why and how.

All the information required to make such propositional inferences is recoverable from the text(s). Readers' activities might include:

- discovering the writer's intention
- understanding the writer's attitude to the topic
- identifying the addressee
- distinguishing fact from fiction.

Lastly it might involve pragmatic inferencing, which takes place when readers rely mainly on their own schemata and/or opinions to interpret text(s) (Chikalanga 1990). This might involve making:

- pragmatic informational inferences which are either referential, typically answering questions beginning with what and which, or spatio-temporal, typically answering questions beginning with where and when
- pragmatic explanatory inferences which are concerned with motivation, cause, consequence and enablement and will often answer questions beginning with why and how
- pragmatic evaluative inferences where the reader makes an evaluation on the basis of the content of text(s)

- applying the main idea(s) in the text(s) into other contexts
- evaluating a point of view
- expressing own opinion on the subject.

With reference to their own background knowledge and experience, the readers would try to interpret, respond to, evaluate and possibly apply the writer's message(s) contained in the text(s).

However, despite the frequency with which careful reading for explicit or implicit information occurs, research (e.g. Cohen and Upton 2006, Hawkey 2006, Rosenfeld, Oltman and Sheppard 2004, Urquhart and Weir 1998, Weir 1983) indicates that careful reading alone may be an inadequate reading construct for students in the academic context.

Pugh (1978:20) traces our preoccupation with careful reading back to about 1910 in the United Kingdom, when there was an increased pedagogical interest in exercises on texts requiring close, careful, often iterative textual study rather than any interest in teaching silent extensive reading as a skill *per se*. Texts were seen as vehicles for teaching language (lexis and syntax) rather than the means for improving and utilising a variety of reading skills.

Pugh (1978) contrasts the emphasis on the *speed* of silent reading in 20th century American research with the distinct absence of any such concern for speeded reading in British research in the same period. The pioneering early American work on the speed of reading would seem to have had limited impact in the United Kingdom and there was no sign of it in United Kingdom teaching practice 1900–1970 according to Shayer (1972). Pugh notes of United Kingdom reading researchers up to the 1970s (1978:74): 'They tended not to examine very closely the phenomenon of reading in various ways to achieve various purposes.' Farr, Carey and Tone (1986:62) observed in a similar vein: '... the passage was seen simply as a stretch of prose providing language for comprehension... the concept that readers read in different ways according to their purpose and the type of text was as yet unrecognised.'

Set against this careful reading orthodoxy, Weir (1983), in his research for his TEAP (see Research Study 1 in Chapter 2), generated large-scale survey data suggesting second language (L2) readers in the academic context have particular problems with another type of reading viz expeditious reading: '... how readers process texts quickly, effectively and selectively, i.e. expeditiously, to extract important information in line with intended purposes' (Urquhart and Weir 1998:101). This was later backed up by empirical test data provided by Weir et al's (2000) research for the AERT in China reported in Research Study 2 in Chapter 2. They found that although many students in China were performing adequately in careful reading, this did not necessarily translate into proficiency in expeditious reading. They argued that testing careful reading ability is on its own insufficient as a relevant and comprehensive measure of adequate English reading skills for academic study. Weir et al (2012a) found similar problems with expeditious reading in Research Study 3 (in Chapter 2) where students expressed the most difficulty with the time available to do the necessary reading, reading lengthy texts and finding relevant information quickly.

Urquhart and Weir (1998) assert that given the expectation that students in their university lives need to understand a large part of the domain of knowledge covered by their degree programmes, this entails processing large amounts of text (paper- and web-based) expeditiously (that is quickly, selectively and efficiently) as a precursor to the careful reading which takes place once relevant information has been located (Urquhart and Weir 1998).

Khalifa and Weir (2009) suggest, in their review of the literature on examining reading, that the significant drawback of many process-based models of reading, as well as many of the earlier componential models of reading (Bernhardt 1991, Coady 1979), is that they are nearly all premised on a careful reading model and do not take sufficient account of the different purposes of reading. Hoover and Tunmer (1993:8) observe that the notion of the simple view 'assumes careful comprehension: comprehension that is intended to extract complete meanings from presented material as opposed to comprehension aimed at only extracting main ideas, skimming, or searching for particular details.' Rayner and Pollatsek (1989:439) similarly admit that for most of their account of the reading process they are concerned with the skilled, adult reader carefully reading textbook material and they agree that careful reading models have little to tell us about how skilled readers can cope with other reading behaviours such as skimming for gist (Rayner and Pollatsek 1989:477-478). Most of these earlier reading models therefore fail to account for the processing experience of skilled readers in real-life academic reading activities. As we saw earlier (Weir et al 2012a), the actual academic reading demands faced by students in tertiary-level education are likely to involve expeditious as well as careful reading (see also Weir 1983). Expeditious reading would appear likely to include, for university students, skimming, search reading, and scanning.

Urquhart and Weir (1998) argue that in expeditious reading, the linearity of the text is not necessarily followed. The reader is sampling the text, which can be words, topic sentences or important paragraphs, to extract information on a predetermined topic in search reading or to develop a macrostructure of the whole text as in skimming. The process can be top-down when the reader is deciding how to sample the text and which part(s) of the text are to be sampled; it can also be bottom-up when the reader's attention is on the sampled part(s) of the text that have been identified as worthy of closer scrutiny.

Skimming is generally defined (e.g. Levine, Ferenz and Reves 2000, Munby 1978, Urquhart and Weir 1998, Weir 2005b) as reading to obtain the gist, general impression and/or superordinate main idea of a text. The reader asks: 'What is this text as a whole about?' while avoiding anything which looks like detail. For Urquhart and Weir (1998) the defining characteristics of *skimming*

are: (a) the reading is selective, with sections of the text either omitted or given very little attention; (b) an attempt is made to build up a macrostructure (the gist) on the basis of as few details from the text as possible. The reader is trying to reach the top-level structure of a text, that is, the discourse topic or the superordinate macro-proposition.

This processing of a text selectively to get the main idea(s) and the discourse topic as efficiently as possible might involve both expeditious and careful reading and both bottom-up and top-down processing. The focus may be global or local and the rate of reading is likely to be rapid but with some careful reading. The text is processed quickly to locate important information that then may be read more carefully. Purposes for using this strategy might include:

- to establish a general sense of the text
- to quickly establish a macro-propositional structure as an outline summary
- to decide the relevance of texts to established needs.

Where appropriate to text type, it might involve one or more of the following operationalisations:

- reading titles and sub-titles quickly
- reading the abstract carefully where there is one, e.g., in an academic article
- reading the introductory and concluding paragraph carefully
- reading the first and last sentence of each paragraph carefully
- glancing at words and phrases in particular for discourse cues
- looking for lexical repetition.

For Urquhart and Weir (1998), *search reading* involves quickly locating information on predetermined topics. The reader wants information to answer set questions or to provide data, for example in completing written assignments. It differs from skimming in that the search for information is guided by predetermined topics so the reader does not necessarily have to establish a macropropositional structure for the whole of the text.

The search reading process, like skimming, is rapid and selective and is likely to involve careful reading once the relevant information has been located. Like skimming, bottom-up and top-down processing is therefore involved. Unlike skimming, sequencing is not always observed in the processing of the text although it is likely to be more linear than scanning. The periods of closer attention to the text tend to be more frequent and longer than in scanning. It normally goes well beyond the mere matching of words to be found in scanning activities. It might, for example, include the following operationalisations where appropriate:

- keeping alert for words in the same or related semantic field (the reader is not certain as in scanning of the precise form of these words)
- using formal knowledge of text structure for locating information
- using titles and subtitles
- · reading abstracts where appropriate
- understanding textual warrants.

Scanning involves reading selectively, to achieve very specific reading goals, e.g. finding a number in a directory, finding a particular author's name or the page in a book index where relevant information may be found. The main feature of scanning is that any part of the text which does not contain the pre-selected symbol(s) is dismissed. It may involve looking for specific words/phrases, figures/percentages, names, dates of particular events or specific items in an index at the local word level.

Rosenshine (1980) defines scanning as involving recognition and matching. It is surface-level rather than deep processing of text and is mainly bottom-up processing. There is a rapid inspection of text with occasional closer inspection. Pugh (1978:53) describes it as: 'finding a match between what is sought and what is given in a text, [with] very little information processed for long term retention or even for immediate understanding.' The operationalisations involved might include looking for/matching:

- specific words/phrases
- figures/percentages
- · dates of particular events
- specific items in an index/directory.

Weir, Vidaković and Galaczi (2013) point out that though a great deal of attention has been paid to the assessment of slow, careful, incremental reading, expeditious reading (i.e. fast, efficient, selective reading) has not been explicitly tested in very many high-stakes examinations to this day, despite its inclusion in the CEFR reading descriptors for academic purposes at the C1 level (Council of Europe 2001). The 1991 Cambridge English Certificate in Advanced English (CAE) Reading paper was a notable exception to this, as were the later multiple-matching tasks in the Preliminary English Test (PET) and First Certificate in English (FCE) Reading papers, though these were sadly reduced or removed in the 21st century on validation grounds, partly for reasons of practicality when the exams were streamlined to save time. The Higher Intermediate General English Proficiency Test (GEPT) in Taiwan has had a separate expeditious reading paper since its inception in 2000 (Weir 2005b, Wu 2014, Wu and Lin 2008) but apart from GEPT and the Test for English Majors (TEM) in China (see Zou, Green and Weir 1997) these appear to be the only high-stakes English language examinations which have papers dedicated to testing expeditious reading.

Urquhart and Weir (1998) conclude that the traditional approach to reading adopted by psychologists, language testers and teachers in the UK is for the most part based on a slow, careful, incremental view of reading for comprehension. The lack of research interest in the speed of reading in the United Kingdom that we noted above may partially explain why there have been few attempts to develop models for expeditious reading or to include such reading types in research, tests or teaching as compared to the focus on careful reading. The evidence we cited in Chapter 2 on the criticality of expeditious reading for tertiary-level studies means a change in direction is necessary.

Selecting global or local levels of reading

The literature testifies to shifting perspectives on how to test reading over the course of the last century. Weir et al (2013) explain how for a considerable part of the 20th century the focus in both the teaching and testing of reading was on decoding at the clause and sentence levels (see also Venezky 1984:14). Attention to careful local reading rather than careful global reading (beyond the sentence right up to the level of the complete text or texts) was the norm with no concern either for expeditious (quick, selective and efficient) forms of reading that might be useful for study or leisure purposes.

Venezky (1984:13) argues that, in the US, research on comprehension processes was sparse until the 1950s and even the phrase 'reading comprehension' was seldom found in the psychology literature although references did occur occasionally in relation to methods of teaching and testing. In contrast, great emphasis was put on vocabulary, which was seen as the determining factor in the difficulty of understanding reading materials and was the major concern in intelligence testing.

Urquhart and Weir (1998) argue that by the 1970s this focus was changing. In an editorial in 1980 (*Reading Research Quarterly* 15/2) under the heading 'Why Comprehension?' the editors noted that the earlier emphasis on decoding was attracting much less attention by the 1970s, being replaced by a new emphasis on comprehension. One can thus chart a changing focus in English language reading tests from an initial focus on lower-level decoding processes to a more comprehensive approach in the 1970s that embraced higher-level global comprehension processes as well. The expectation is that 21st century IELTS, along with comparable tests of academic reading, would similarly subscribe to testing global meaning. The evidence from the literature suggests that this might not in fact be the case (see Chapter 4).

In the identification of both a global and a local level at which the reading skills and processes may operate, the critical question is the extent to which the test or the reality requires students to comprehend information within and beyond the sentence (see Alderson 2000). Taking account of recent work in

the EAP testing field, we purposefully include higher-level reading processes at the text or intertextual levels in the careful reading cell in Table 3.1, building on the seminal work of Enright et al (2000) carried out for the TOEFL 2000 revision. Enright et al (2000) argued for including tasks that require processing beyond the level of searching for information and basic comprehension of main ideas in a text and necessitate an understanding of how information in a text as a whole is connected, and how to integrate information from across a variety of texts for use in written assignments or exam essays, i.e. reading must be tested at the whole-text and intertextual levels. Similarly, with regard to reading purpose, Jordan (1997) makes the important connection between global academic reading involving multiple texts and the writing-based tasks or activities for assignments, dissertations, projects or reports, for which the reading is often a preparation.

Maclellan (1997) emphasises that in an academic context, students need to read to learn. They must use an appropriate combination of the skills and strategies that are required for the different purposes of reading in tertiary-level study. Based on the evidence of the three major research studies described in Chapter 2 and the literature on testing advanced-level reading, we advocate adopting a four-cell matrix to EAP test development which systematically distinguishes reading level (within and beyond the sentence) from reading type, a distinction now significant in many of the reading studies and models in the field (see Table 3.1). The matrix is helpful in clarifying the types and level of reading required and those not required in academia according to the research literature. Large-scale investigations of the academic reading needs of students in tertiary education (e.g. Weir 1983, Weir et al 2000, Weir et al 2012a) demonstrate conclusively that in tests of academic reading we need to account for both careful and expeditious, global types of reading. Locallevel reading is less important per se and is necessarily involved in answering global questions anyway.

The sociocognitive model – Part 2: Cognitive processing in academic reading

The cognitive elements of the validity of a reading task are a measure of how closely they elicit the cognitive processing involved in contexts beyond the test itself, i.e. in performing reading task(s) in real life. Khalifa and Weir (2009) drew on the work of authors working within the field of cognitive psychology in order to devise a model of the L1 reading process – supported by empirical evidence – which could be treated as the goal towards which the L2 reader aspires. The various types of reading and the cognitive processes they may give rise to that we have identified from the literature are represented in diagrammatic form in columns 1 and 2 in Figure 3.1.

We dealt with the types of reading in the last section (See Table 3.1). The

	Global level	Local level
Careful reading	 Establishing accurate comprehension of explicitly stated main ideas and supporting details across sentences Making propositional and/or pragmatic inferences Establishing how ideas and details relate to each other in a whole text Establishing how ideas and details relate to each other across texts Critically selecting, connecting and organising ideas for use in writing 	 Establishing accurate comprehension of explicitly stated main idea or supporting details within a sentence Identifying lexis Understanding syntax
Expeditious reading	 Skimming quickly to establish: discourse topic and main ideas, or structure of text, or relevance to needs Search reading to locate quickly and understand information relevant to predetermined needs 	• Scanning to locate specific points of information

Table 3.1: Types of reading

Adapted from Urquhart and Weir (1998:123)

cognitive processes described below form the middle column of our reading model presented in Figure 3.1. These processes are activated in accordance with the decisions taken in the goal setter about the type of reading required. Here we attempt to characterise the reading behaviours available to the competent L1 reader to which the L2 reader might be expected to approximate progressively as their proficiency level in L2 and exposure to reading in L2 improves. Our main interest lies in the **higher-order reading processes** (from inferencing upwards in the central column) as these are what we would expect to see in a test of academic reading just as they dominate real-life reading activities for students in tertiary education (see Research Studies 1–3 in Chapter 2).

The central cognitive processing core

Lower-order processes

Word recognition is concerned with matching the form of a word in a written text with a mental representation of the orthographic forms of the language. Field (2004:234) describes this as 'the perceptual process of identifying the letters and words in a text'.

Oakhill and Garnham (1988:10) explain that the problem of word recognition is to decide which (if any), of all the words you know, the current visual pattern is an instance of. This process is complex for the less experienced reader and takes up considerable processing capacity until they are able to make the automatic connection between written word and mental representation that an experienced reader is able to make. Accuracy and automaticity of word recognition is critical for the skilled reader (Grabe 2004, Perfetti 1997). Automaticity is the result of increasing experience in decoding and of the mind's orientation towards creating processes, which are undemanding upon attention. Word recognition is required in all of the reading types we discussed above.

Lexical access is described by Field (2004:151) as the 'retrieval of a lexical entry from the lexicon, containing stored information about a word's form and its meaning'. The form includes orthographic and phonological mental representations of a lexical item and possibly information on its morphology. The lemma (the meaning-related part of the lexical entry) includes information on word class and the syntactic structures in which the item can appear and on the range of possible senses for the word.

Syntactic parsing covers not only word order, but also word form (morphology) and structural elements (determiners, prepositions, auxiliary verbs etc.). Once the meaning of words is accessed, the reader has to group words into phrases, and into larger units at the clause and sentence level to understand the message of the text. Fluency in syntactic parsing is regarded as important in the comprehension process by a number of authorities (Perfetti 1997). Cromer (1970) illustrates the importance of competence in the syntax of the target language for deriving meaning from text. Researchers like Weir (1983) and Shiotsu and Weir (2007) point to the strong positive correlations between test results in grammar and in reading comprehension.

Establishing propositional (core) meaning at the clause or sentence level is described by Field (2004:225) as: 'an abstract representation of a single unit of meaning: a mental record of the core meaning of the sentence without any of the interpretative and associative factors which the reader might bring to bear upon it.' Propositional meaning is a literal interpretation of what is on the page. The reader has then to add external knowledge to it (see the next paragraph on 'inferencing') to turn it into a message that relates to the context in which it occurred.

Higher-order processes

Inferencing is necessary so the reader can go beyond explicitly stated ideas as the links between ideas in a passage are often left implicit (Oakhill and Garnham 1988:22). Inferencing in this sense is a creative process whereby the brain adds information which is not stated in a text in order to impose coherence. A text cannot include all the information that is necessary in order to make sense of it. Texts usually leave out knowledge that readers can be trusted to add for themselves. If there were no such thing as inferencing, writing a text, which includes every piece of information, would be extremely cumbersome and time consuming.

Building a mental model

Once the reader has processed the incoming sentence and elaborated it where necessary and possible through inferencing, the new information needs to be integrated into a mental representation of the text so far. Field (2004:241) notes: 'Incoming information has to be related to what has gone before, so as to ensure that it contributes to the developing representation of the text in a way that is consistent, meaningful and relevant. This process entails an ability to identify main ideas, to relate them to previous ideas, distinguish between major and minor propositions and to impose a hierarchical structure on the information in the text.' Ongoing meaning representation is provisional and liable to revision as well as updating with new information from the text. Selection may occur whereby stored information is reduced to what is relative or important.

According to Kintsch and van Dijk (1978:374), the propositions representing the meaning of a text are linked together, usually by argument overlap, to form a hierarchical text base. Micro-structures are processed, converted into semantic propositions and stored in the working memory while the cohesion between them is established. As the process moves on, a macro-structure is built up. Background knowledge, stored in long-term memory, is utilised to supply an appropriate schema for the macro-structure, as well as to aid coherence detection in the construction of the micro-structure. Crucial information tends to be at the top levels of this hierarchy, while detailed information is at the lower levels. Field (2004:174) refers to Gernsbacher's (1990) structure building framework:

... the reader maps incoming information on to a current information substructure if it coheres with what is there. If it does not, the reader employs a shifting process which involves creating a new information substructure. Reading is normally supported by paragraphing which in well written texts helps the reader build a meaning structure of the text.

Creating a text-level representation

Field (2004:225) notes how text structure is seen by some as 'a hierarchy of propositions; a set of prominent macro-propositions, beneath which (like subheadings in a table of contents) are grouped micro-propositions of diminishing degrees of importance.' We examined this distinction between global and local meaning previously in the discussion on reading types. At a final stage of processing, a discourse-level structure is created for the text as a whole. The skilled reader is able to recognise the hierarchical structure of the whole text and determines which items of information are central to the meaning of the text. Enright et al (2000:5–6) explain the text model as the representation of rhetorical structure(s) in a text: 'Constructing an organised representation of the text including main points and supporting detail; an

integrated understanding of how supporting ideas and factual details of the text form a coherent whole . . .'

Development of an accurate and reasonably complete text model of comprehension would seem to involve understanding: discourse structure; identifying macro-level relationships between ideas; and which propositions are central to the goals of the text and which are of secondary importance: for example, this might be signalled in a text through foregrounding main information and back grounding of secondary information, important information in first mention position, or marking of thematic information with repetition (see McKoon 1977, Meyer 1975).

Creating an intertextual representation

Lacroix (1999) suggests that the comprehension of complex, multiple texts in a particular domain may require two distinct levels of macro-structural processing to ensure a coherent, condensed structuring of multiple text information. She suggests that the process of macrostructure construction (extracting important information) outlined in Kintsch and van Dijk (1978), which involves identifying and establishing a hierarchy for units of information through the application of transformational macro rules of deletion, generalisation and integration, accounts well for the comprehension of a single text but may not be adequate to represent how mental representations are combined coherently across multiple texts. Lacroix suggests that the additional process of macro-structural organisation (structuring selected information) is necessary for the connection of several text representations through higherlevel semantic links.

Stromso and Braten (2002:211) similarly argue that the 'discourse synthesis' (Spivey 1990) of multiple texts in a specific domain involves 'composing a new text by selecting, organising and connecting content from more than one source text.' The need for an intertextual model, sometimes referred to as a 'documents model' (Perfetti, Rouet and Britt 1999) to account for the production of integrated representations of multiple texts is supported in the work of Britt and Sommer (2004), Goldman (1997, 2003), Hartmann (1995), Perfetti (1997), Perfetti et al (1999), Spivey and King (1989) and Stahl, Hynd, Britton, McNish and Bosquet (1996). As Goldman (2004:344) succinctly puts it: 'the information across texts is part of a larger whole not necessarily specified in any one of the texts'.

Britt and Aglinskas (2002) refer to the heuristics of corroboration, contextualisation and sourcing as being necessary in addition to information integration for document-level reading of multiple historical texts. It may be that future research will reveal further domain-specific requirements for intertextuality. Ünaldi (2010:3), in her investigation of what goes on beyond single text comprehension when readers read multiple texts, points to the higher cognitive demands of the latter: 'Since texts are not normally written to be read in conjunction with other texts, they lack explicit links to facilitate integration of information across texts, and the demands on the reader to form a macrostructure are higher than when reading a single text with its own intratextual coherence.'

Processing in multiple text reading has clear resonances with the cognitive processing that takes place in the knowledge-transforming approach of writing (Scardamalia and Bereiter 1987), where the selecting, connecting and organisation of information from source texts constitute the first cognitive components in the writing process (see Shaw and Weir 2007:Chapter 3). Researchers such as Spivey and King (1989) have shown how competent students interweave texts in writing research papers by utilising source material deemed to have intertextual importance. Stromso and Braten (2002) provide an interesting case study of students' intertextual linking activities in connection with learning from expository textual resources comprising civil law reading materials such as textbooks, a code of laws and legal cases. Hartmann (1995) details a further interesting case study of students constructing a 'mosaic of intersecting texts' on the American Civil War. Both studies illustrate the additional processing required to produce an intertextual representation as compared to that of a single text.

The cognitive construction of intertextuality offers a useful heuristic for looking at reading-into-writing at an advanced level and it extends our view of reading beyond the act of comprehension of a single passage (Hartmann 1995).

We have now looked at each of the levels of processing that may be brought into play as a result of metacognitive decisions taken in the goal setter. A further metacognitive activity may take place after activation of each level of the processing core: test takers are likely to check the effectiveness of their understanding (Sticht and James 1984).

Monitoring

The monitor is the mechanism that provides the reader with feedback about the success of the particular reading process (Urquhart and Weir 1998:105). The nature of the monitoring is contingent on the type of reading, and therefore the monitor is activated in accordance with the original goals of the reader, who might even decide that they have adopted the wrong type of reading and change accordingly.

Self-monitoring is a complex operation, which may occur at different stages of the process (after reading a word, a sentence, a paragraph or a complete text). Pressley and Afflerbach (1995) provide a comprehensive review of monitoring strategies that help regulate comprehension and learning, and evaluating strategies whereby readers reflect or respond in some way to the text. In decoding text, monitoring involves checking word recognition, lexical access, and syntactic parsing. Within meaning building it can involve

determining the success with which we can extract the writer's intentions or the argument structure of the text. While building a mental model there is a need to monitor comprehension to check the viability of the ongoing interpretation. Monitoring chiefly checks the consistency of incoming information against the meaning representation established so far. If the two are in conflict, the reader regresses to check. World knowledge, in the form of schemata in long-term memory, plays an important part in judging the coherence and consistency of what has been understood when it is integrated into the ongoing meaning representation.

Perfetti (1999:197) describes how comprehension monitoring has been found to be ineffective in less skilled readers in a number of studies. Oakhill and Garnham (1988:139–140) argue that the unskilled L1 reader often fails to monitor comprehension or at least makes less use of monitoring strategies, particularly at the comprehension level. Studies have shown that one of the hallmarks of a good reader is the ability to check the meaning representation for consistency. Skilled readers, on failing to understand a part of a text, will take action such as rereading to deal with the problem (see Hyona and Nurminen 2006).

The *knowledge base* displayed in the column on the right-hand side of the model (see Figure 3.1) is drawn upon by elements in the central processing core in line with the intended purpose of reading and the performance conditions established by the test task. The contextual parameters of the text(s), which constitute the conditions under which the process of reading takes place, will determine the demands that are placed on the various elements of the reader's knowledge base. These are examined next.

The sociocognitive model – Part 3: Contextual parameters

If test task performance is to be used to support inferences about performance in the wider domain of real-world tasks, it is essential that both target reading activities and test tasks are comparable in terms both of the cognitive processes required and of the contextual parameters they are performed under. A central assumption in Weir's (2005b) sociocognitive test validation model is that cognitive processing always occurs within and is significantly affected by the *context* it is taking place in. Weir's contextual validity parameters relate the features of the task to the nature of the text that must be processed if the task is to be completed successfully, i.e., to the specific performance conditions under which the reading activities in a test are performed. It requires *situational appropriateness* in both the linguistic and content demands of the text to be processed, and the features of the task setting that impact on task completion.

Similarly, Bachman and Palmer (1996) argue that situational and

interactional authenticities are essential features of valid test tasks. These factors in judging a test's usefulness are analogous to Weir's contextual and cognitive aspects of validity. It is widely accepted that, given the contextual constraints imposed by testing conditions (such as the time available to complete a task, or the length and number of texts that need to be processed), full situational authenticity is generally unrealistic for language assessments. However, contextual features of a test ought to reflect as many of the relevant features of the target reading activity as possible. The literature on the textual parameters that are potential sources of text complexity is daunting and we will only scratch the surface of it here before identifying those parameters that appear to be both useful and applicable for our study.

Bachman, Davidson, Ryan and Choi's (1995) test comparison studies identified as important textual properties such as the nature of text, length, vocabulary, grammar, cohesion, distribution of new information, type of information, topic of discourse, rhetorical organisation and illocutionary acts. Freedle and Kostin (1993, see also Freedle 1997), in a detailed analysis of reading comprehension item difficulty, take into consideration vocabulary, concreteness/ abstractness, subject matter, coherence, length of various segments such as word, sentence, and paragraphs as text-related variables. Fortus, Coriat and Fund (1998) investigated length, number of negations, number of referential markers, vocabulary, grammatical complexity, abstractness, topic, and rhetorical structure as textual variables contributing to the level of difficulty of reading comprehension items. Enright et al (2000) identified three groups of salient textual features to operationalise in test texts: grammatical/discoursal features, pragmatic/ rhetorical features and linguistic variables. Alderson et al (2004) include text source, authenticity, discourse type, domain, topic, nature of content, text length, vocabulary and grammar as relevant features for consideration in selecting texts. Khalifa and Weir (2009) suggest that linguistic demands of task input - reading texts in this case - can be explained in terms of *lexical* and *structural resources*, discourse mode, functional resources, content knowledge, text writer-reader relationships, topic familiarity, cultural knowledge, nature of the text (abstract v concrete) and the subject specificity of the text. Both individually and in combination these contextual parameters are thought likely to impact on the cognitive load imposed upon the reader and affect difficulty of processing in reading.

The text linguistics literature on *complexity* also identifies certain factors as important contributors to the level of difficulty or ease with which a text can be processed and offers certain methodologies for evaluating this. Readability formulae such as Flesch-Kincaid Grade Level involve the calculation of word and sentence length. Although in several studies readability formulae are criticised as being inadequate to reveal textual complexity (see, for example, Gervasi and Ambriola 2002, Masi 2002), they still form the basic aspects in more recent and detailed analyses of textual complexity. Masi (2002) suggests that, together with linguistic and quantitative factors of word and sentence complexity, other semantic and syntactic factors such as *structural embedding*, *content*, *background knowledge of the reader*, and the *type* and *genre* of text, should also be taken into account to reveal a more reliable and predictive measure of text complexity. The latter, however, are hardly measurable by automatic procedures such as computerised calculation.

There appears to be a measure of consensus in the subjective judgements of these different authors on the features to be addressed when considering text complexity. Additionally, there is empirical evidence from studies such as Freedle and Kostin (1993) and Fortus et al (1998) that a subset of the listed characteristics do indeed impact on the difficulty of reading comprehension tests for learners.

From the picture emerging above in the literature, supported by the evidence provided from the three empirical research studies discussed in Chapter 2, it is possible to identify a group of criterial features that suggest themselves as useful for the analysis of texts for testing purposes. In Table 3.2 we list the contextual parameters most likely to have an impact on academic reading test performance.

Task setting	Linguistic demands		
 Response method Order of items Channel of presentation: verbal and/or non-verbal input Taxt length 	 Writer-reader relationship Discourse mode (genre/rhetorical task(s)/patterns of exposition) Lexical complexity – cumulative coverage, academic word level, lexical density, modifiers per noun phrase, mean number of words before main verb Syntactic complexity including sentance length readability. 		
 Time constraints Electronic or paper texts 	 Functional knowledge Nature of information: abstractness/concreteness Content knowledge (subject and culture) 		

 Table 3.2:
 Some context validity parameters for academic reading (test) tasks

Using this framework, we will examine more closely the salient parameters of context validity for a reading test in EAP. Where necessary, recourse to the secondary testing literature will be made to define more closely the dimensions of some of the categories of description in our contextual framework.

Response method

Test response methods have the potential to affect the types of reading and the levels of cognitive processing that can be measured (see Alderson, Clapham and Wall 1995, Kobayashi 1995, Khalifa and Weir 2009). We will examine the range of IELTS formats used for testing reading in EAP at tertiary level in

Chapter 4. It is clear that response method has a significant impact on what can be measured.

Existing frameworks have little by way of advice on this. The CEFR (Council of Europe 2001) offers no guidance on appropriate use of response method, let alone how response method might be linked to level (Alderson et al 2004:10). We will need to look elsewhere to determine which test format is the most suitable for measuring the types of reading and various levels of processing we have identified above.

We will make a distinction between test formats that involve a *selected* response as against those which require a *constructed* response. In selected responses the candidate chooses the answer from a set of options provided at the word, phrase, sentence or paragraph level, and they identify the answer by, for example, putting a line through the letter representing the correct option or encircling the option on an answer sheet which is then usually scanned by an optical mark reader. In constructed responses, candidates have to produce the answer themselves, for example by writing a word, phrase, sentence, or even a short paragraph or an extended text, on a separate answer sheet.

As we will see in Chapter 4, both types have their advantages and drawbacks. Selected responses will inevitably affect the cognitive processes activated in answering an item, though test reliability is enhanced by their use. Constructed responses will involve writing at some level and issues arise in terms of the marking reliability of extended responses. See more discussions of how evidence of high-level reading in writing can be scored in Chapter 4. Given the part reading plays in academic study to generate information for written assignments, this may not trouble us unduly in an academic purposes test.

Order of items

When reading a text carefully we normally construct a representation of each section of text serially and incrementally integrate this with a representation of what we have read up to that point. Given careful reading is a cumulative process, setting the questions according to the order in which the answers are found in the text is thus consistent with this. Hughes (1989:130) argues that 'not to do this introduces too much random variation and so lowers the test's reliability'. Weir (1993:96) agrees and asserts that the sequential 'ordering of the questions helps bring the process of taking the test closer to the way readers would normally process that particular text'.

However, in real life, careful reading often seems to be preceded by expeditious reading especially in an academic context (see Research Study 3 in Chapter 2 for an account of research into reading activities at undergraduate level in the UK that supports this assertion). Academic readers frequently go back to mine carefully various parts of the text, having already grasped the overall message through expeditious reading intended to generate a quick view of the macrostructure (see Shih 1992 for evidence of this). Reading experts indeed advocate this sequence of activities in teaching practice. For example, Nuttall (1996) recommends that activities requiring thorough understanding (e.g. studying the development of an argument, analysing relationships between paragraphs) are best dealt with at the end, while tasks like skimming for overall gist ought to be performed at the beginning of processing a text. This suggests that if both careful and expeditious reading questions are to be set on the same text(s) the expeditious questions should come first, perhaps in their own timed section.

In search reading the academic reader does not necessarily follow the author's sequence in a long text. It may well make the processing associated with this reading type easier if candidates are aware of any sequential ordering of test questions as the search area might thereby be much reduced. Thus, in a reading paper devoted to search reading, questions/answers in a bank should not appear in the order the information occurs in a text.

In scanning, even less of the text needs to be processed and the focus is almost wholly on word recognition, so there is no reason to have the questions in the order the lexical items to be found appear. If questions are placed in order, it diminishes incrementally the amount of lower-level processing necessary to find the particular lexical item.

Channel of presentation

Khalifa and Weir (2009:98) cite research that has shown how text comprehension is influenced by the presence of non-verbal information (e.g., Hegarty and Just 1989, Holliday, Brunner and Donais 1977, Koran and Koran 1980). They argue that when non-verbal information forms an integral part of a text, comprehension can be enhanced. Hegarty, Carpenter and Just (1991:666) argue that when a topic is sufficiently complex a reader is often unable to visualise spatial representation of information without a diagram. Presenting information in a non-verbal form can also help the reader by reducing the need to search for information that has been represented previously (Hegarty et al 1991:660). This body of research demonstrates the importance of the inclusion of non-verbal information as and when appropriate, e.g., when it forms an integral part of the original text. If such assistance is afforded in real-life texts, it should also feature where those texts are employed for testing purposes.

Text length

Overall, when university students were asked by Weir et al (2012a) what they thought helped them to be a successful reader at university, the most frequent responses referred to:

- reading with understanding (39%)
- reading all that you need to read (26%)
- wide reading (18%).

The students in this survey were well aware that they needed to read a lot and diversely. The major problem for them in their studies was coping with the heavy reading load, invariably carried out under time pressure. The length of texts faced in language tests used to measure academic English proficiency is thus a key contextual parameter which will affect the construct validity of those measuring instruments.

Decisions taken on text length in tests must take account of the skills/strategies the test is intended to measure (Alderson 1996, Nuttall 1996). Alderson argues that if candidates are to skim for main ideas, scan for specific information, make judgements about relevance and irrelevance, or distinguish between main points and subsidiary detail, then a long text is needed for these operations to be initiated (see also Nuttall 1996). As discussed in Chapter 2, Engineer (1977) advocated using texts of 1,000 words plus. If expeditious reading is intended and candidates are given a short text, they may employ a careful reading rather than expeditious approach if the time has not been constrained to prevent this.

In the three research studies reported in Chapter 2, a strong case was made for the use of long texts in appropriate contexts on the grounds that these are more representative of required reading in the target situation, at least in terms of length and discourse type (Engineer 1977, Weir 1983, Weir et al 2000, Weir et al 2012a). The beneficial washback effect of the presence of long texts in the test on the teaching that precedes it cannot be ignored. If candidates preparing for a test only have to learn to cope with short texts of, say, a short 10-line paragraph, then they will be ill prepared for the vast amount of text they are expected to cope with once they get to university. At the CEFR C1 and C2 levels, tests are supposed to be able to determine whether candidates can cope with academic study and this means being able to deal with long texts expeditiously as well as being able to develop a mental model and text-level representations. In terms of our reading model, the longer the text, the more demands can be made on both lower- and higher-level processing, and the more types of reading can be catered for.

Text length has an effect on the linguistic resources that will need to be utilised in cognitive processing. In general, the longer the text candidates have to process, the greater the language and content knowledge required (*ceteris paribus* for example the number of idea units). Length of text will also directly affect the processing load involved in building a mental model and/ or a text representation. If short texts are not making the demands on these resources that occur in real-life situations, cognitive validity is compromised (see Skehan 1998).

However, as Alderson et al (2004) have pointed out in relation to the CEFR, distinctions between long and short texts are sometimes inexplicit, nor is it clear how long a text or what time constraints would need to be imposed to reflect successful skimming, scanning and relevance judgements in academic reading. In many ways deciding on these parameters is an empirical issue and careful trialling will be necessary to clearly establish suitable dimensions for time and length.

The issue of test fairness also arises here. Longer texts which approximate more closely to the lengths of texts candidates are exposed to in real life often provide more contextual clues and more support for the reader than do short texts. Weir et al (2012a) provide evidence that that the type-token ratios (TTRs) of longer texts used at undergraduate level are noticeably lower than those of texts used in international high-stakes tests (similar data was found by Malvern and Richards 1997). As a result of the lower proportion of different words, they may in fact be easier to process. TTR is discussed in more detail in the section 'Lexical complexity'.

Weir, Hawkey, Green and Devi (2012b) add that what is perhaps more important than simple text length is the density and complexity of idea units within the text (Bachman 1990). Extending the length of a reading passage while keeping the number of items constant increases difficulty in that it provides a greater number of idea units between which a test taker has to choose when seeking a conceptual match for an item. In effect, it increases the size of the haystack within which the needles have to be found. The number of propositions contained in a text has an effect on text comprehension because of the strain multiple propositions can put on working memory (Kintsch and Keenan 1973).

There is a downside for test administrators in increasing the length of texts in our EAP tests: by using longer texts where the focus is on careful reading more time will need to be made available. This may mean that fewer texts can be included and the range of topics is diminished with possible implications for test bias. However, if the intention is to test skimming and search reading types then this would not be the case as reading in this mode is meant to be quick and selective and is usually severely time-constrained.

Time constraints

Alderson (2000:30) observes that 'speed should not be measured without reference to comprehension, but at present comprehension is all too often measured without reference to speed'.

Fry (1963) suggests that a slow L1 reader reads at a rate of 150 words per minute (wpm), a fair reader at 250 wpm, and a good reader at 350 wpm (see also Carver 1992). Nuttall (1996:56) comments:

... university students in countries where English is a second language may read at about 200 wpm but have been found to study at rates as slow as 60 wpm; presumably the texts were difficult and had to be understood thoroughly... an L1 speaker of English of about average education and intelligence reads at about 300 wpm. The range among L1 speakers is wide: rates of up to 800 wpm and down to 140 wpm are not uncommon.

Weir (2005b:65) points out that:

The time constraints for the processing of text and answering the items set on it will affect the nature of what is being tested. The test developer has to sequence the texts and tasks, and ensure there is enough time allowed for all activities; if time allotment is not carefully planned, it may result in unpredictable performance. If too much time is given in a reading test or is not strictly controlled per section, candidates may simply read a passage carefully and questions designed to test ability to process text expeditiously (i.e. selectively and quickly) to elicit specified information may no longer activate such operations (see Weir et al 2000 for an example of a research project where this happened). If time is more than sufficient in an expeditious reading task, then careful cumulative, linear processing rather than quick selective processing will result. Decisions relating to timing clearly impact on the processing and hence on the theory-based validity of our test tasks. Setting appropriate time limits is best done empirically.

Many exam boards have not sought to control the amount of time spent on individual parts of a reading test because of the practical difficulties of achieving this. The advent of computer-based testing may help here. Computer-based testing facilitates the accurate measurement of expeditious reading skills since it can be used to control the amount of time spent on each task by preventing candidates going back to earlier tasks or spending more than the suggested time on any one activity. However, with paper-and-pencil tests, separate papers would have to be produced to be able to control the amount of time given for doing the task. This has obvious implications for test security, uniformity of administration and increased cost of printing.

Decisions taken on timing clearly impact on the cognitive processing that takes place in our test tasks. At C1 and C2 levels the CEFR states that candidates should be able to read quickly enough to cope with academic studies. This implies a number of tasks at this level need to be performed under restricted time constraints in order to sample expeditious reading skills. Flexibility with regard to types of reading is the mark of a skilled reader. Students in an academic context need to be able to read both expeditiously and carefully. The sheer amount of information they all have to process puts a premium on the former.

Reader-writer relationships

Audience is the intended readership definable by the writer that exists apart from the text. The reader or audience is, according to Grabe and Kaplan (1996:207), 'essential to the creation of text and the generation of meaning'. Nystrand (1989) states that meaning is created between the participants of a discourse and resides in the expectations and assumptions of both the reader and the writer of each other. Weir et al (2012a:54) argue that writing, rather than being an isolated individual action, involves the endeavours of both the reader and the writer and is shaped through mutual assumptions involved in the understanding of rhetorical situations (Hyland 2002:35). Any act of writing is charged with assumptions about the participant relationships and how these are carried out in culturally and institutionally legitimate ways (Hyland 2002:69). Hyland states that 'managing social relationships, then, is crucial in writing as a text communicates effectively only when the writer has correctly assessed both the readers' resources for interpreting it and likely response to it' (2002:69).

Academic articles in peer-reviewed journals will be aimed at an audience perhaps more conversant with the subject matter of the discipline area in question than is the case with an undergraduate first year textbook in the area, and consequently the content knowledge assumed of the reader will be greater in the former. Where the writer feels the audience is familiar with many of the concepts, e.g. in the case of a specialised text in an academic journal where an expert is writing for other experts, a good deal of inferencing and use of background knowledge may be necessary to develop a mental model of the text. In addition, domain-specific low-frequency lexis and more complex grammatical patterns may be used if they are the norm for that genre. All of this will make the text more complex for the reader who does not have the shared linguistic and content knowledge of that discourse community. The processing of text is not simply 'in the text' but depends crucially on all the knowledge that the reader brings to the text.

Discourse mode

Discourse mode according to Weigle (2002:62) includes the categories of genre, rhetorical task, and patterns of exposition:

The genre refers to the expected form and communicative function of the written product; for example, a letter, an essay, or a laboratory report. The rhetorical task is broadly defined as one of the traditional discourse models of narration, description, exposition, and argument/persuasion, as specified in the prompt, while the pattern of exposition (Hale et al 1996) refers to subcategories of exposition or specific instructions to test takers to make comparisons, outline causes and effects and so on.

Genre

Genre is generally understood to encompass 'salient features and conventions which are shaped by communicative purposes' (Hyland 2002:62). It is evident from the literature that specific genres will involve specific conventional features (lexico-grammatical, semantic, and discoursal), which are likely to impinge on the text processing of readers (Bhatia 1997, Hyland 2000). It would seem logical to suggest that if texts to appear in a test are sourced from academic contexts, they are likely to share lexical, syntactic and discourse features with texts encountered at a university.

The following genres, identified through the development of the questionnaires sent to students at a British university (Weir et al 2012a), are seen as relevant to the present analysis. Their pilot open-ended questionnaire elicited 77 responses which confirmed that books were a key source of students' academic reading, but with journals also prominent and a fair number of students doing around half of their academic reading online. The pilot study students offered insightful comparisons between book and online sources of information indicating, for example, that:

- books offer a wider range of sources and more to understand
- print sources may provide deeper information
- print materials tend to be first choice
- online sources may be for interest but not suitable for assignments
- online reading complements, follows up print reading
- the web, with its wide range of information, can offer explanations, clarifications, of questions raised from reading of books
- useful and convenient to have some journals online, but often limited access
- prefer to print out online information, less comfortable reading from screen
- 'don't use OL [online] so much because can't scribble, highlight, take notes so conveniently' (comment from a student)
- online sources less reliable, credible than books, journals.

The pilot questionnaire data provided further evidence that assignment reading is a multi-source task: 34% citing as many as 10–19 sources for an assignment, fairly evenly divided between books, journal articles and websites, although books were more often the main source of reading than the other two. Weir et al (2012a) pursued further the question of difficulties experienced by the students in their academic reading and found that the most frequently referred to pressures were *text difficulty, time* and *information load*.

Weir et al's (2012a) pilot open-ended questionnaire was followed by a structured survey to which 434 students responded online, and 332 in hard

copy format. This high total respondent figure of 766 students is considered adequate for the purposes for which the questionnaire was designed, and the data elicitation methods used.

Table 3.3 summarises student responses on the relative importance of *books, journal articles, reports, the internet, newspapers* and *magazines* in their academic reading.

	EAL		E	L1
	D (r/o)	D&M (r/o)	D (r/o)	D&M (r/o)
Books	54.9% (1)	90% (1)	77.2%(1)	96% (1)
Internet	42.9% (2)	78.3% (2)	51.3% (3)	85.5% (2)
Journals	28.4% (3)	65.6% (3)	59.1% (2)	83.3% (3)
Reports	19.5% (4)	56.9% (4)	26.5% (4)	63.4% (4)
Newspapers	10.3% (5)	44.1% (5)	19.5% (5)	58.4% (5)
Magazines	9.0% (6)	35.0% (6)	11.8% (6)	40.5% (6)

Table 3.3: Sources of information across EAL and EL1 questionnaire respondent groups*

*Scale: D = definitely agree; D&M = definitely and mostly agree; r/o = rank order

Broadly speaking, and in terms both of measurement of responses by the *definitely agree* choice only, and by the sum of the two positive agreement categories, the order of importance of sources was: 1) books, 2) internet sites, 3) journals, 4) reports, 5) newspapers and 6) magazines. Informed by the pilot questionnaire (Study 2 in Chapter 2), a related item on the main questionnaire asked respondents how much reading they actually did online compared with paper print materials. Table 3.4 summarises responses for both EAL and EL1 participants.

	Amount of reading done online				
	0–20%	21-40%	41-60%	61-80%	81-100%
EAL students (n=458)	16.2%	27.5%	30.3%	17.7%	8.3%
EL1 students (n=290)	30.7%	28.3%	23.8%	13.1%	4.1%

Table 3.4: EAL and EL1 group online reading source proportions

The message of the table here is that the EAL students appeared to do rather more of their reading online than did their EL1 colleagues. The *mode* value for the former group was the 30.3% who did 41-60% of their reading online, compared with 23.8% of the latter, 30.7% of whom read 0-20% online.

Cross-tabulated reading source data suggested close agreement between first and second year students, except, perhaps, for almost 11% fewer among the Year 2 sub-group definitely agreeing that internet sources were important on their courses. Table 3.5 summarises the perceptions on reading sources across the two sub-groups.

	Year 1	Year 2
Books	64.2%	64.9%
Internet sites	49.8%	39.0%
Journals	40.7%	41.7%
Reports	22.0%	22.4%
Newspapers	12.9%	17.1%
Magazines	9.0%	12.7%

 Table 3.5: Year 1 and Year 2 group online reading source proportions

The range of academic reading sources and the prominent role played by internet sites in the academic reading of contemporary university students had clear implications for academic reading tests. In terms of substantive differences of perceptions between our undergraduate and postgraduate sub-samples, it was of interest (as well as intuitively credible) to note that a high 83.5% of our postgraduate sub-sample definitely or mostly agreed on the importance of journal articles on their course, compared with 70.7% of our undergraduate sub-group. Similar was the substantially higher proportion of the postgraduate group agreeing the importance of reports on their courses (75.6% of the postgraduate sample compared with 56.8% of the undergraduates).

Electronic or paper-based texts: Same or different?

Reading on a screen was once generally felt to be slower than reading print on paper, with less long-term retention of the material. People tended to read slowly and somewhat inaccurately on early screens. However, the technology, particularly e-paper (for example the Amazon Kindle is an e-book reader with an e-paper display), has improved dramatically, to the point where speed and accuracy are no longer problems, but deeper issues of comprehension and memory still remain (see Tanner 2014 for an overview of the research).

Noyes and Garland (2003) report that students remember more of what they read on paper. Garland and Noyes (2004:51) similarly found that knowledge seems to be better assimilated and more readily retrieved when read in paper format. These results were echoed in an experiment that looked specifically at e-books (Morineau, Blanche, Tobin and Guéguen 2005). Psychologist Erik Wästlund also found that students learned better when reading from paper (Wästlund 2007). Kerr and Symons (2006:13–14) found that children comprehend less efficiently when reading from computer. Wästlund, Reinikka, Norlander and Archer (2005) suggest that reading and working with a computer results in a higher cognitive workload compared with paper. Mangen, Walgermo and Brønnick (2013:65) found: 'students who read texts in print scored significantly better on the reading comprehension test than students who read the texts digitally'.

According to Wolf (2007), electronic reading and immediate access to onscreen information can negatively impact the way the brain responds to text, including reading comprehension, focus and the ability to maintain attention to details like plot and sequence of events. While we might advance in new skills to gather vast amounts of information, Wolf is concerned whether 'the range of attentional, inferential, and reflective capacities in the present reading brain will become less developed' (2007:214). Mangen et al (2013:64) also refer to issues for scoring in digital marking:

The potential effect of presentation medium on reading comprehension is also an issue for essay marking and annotation. In several countries there is currently a shift toward assessors marking digitally scanned copies on screen rather than the original paper copies traditionally used (Coniam, 2011; Johnson, Hopkin, & Shiell, 2011; Johnson & Nádas, 2009; Johnson, Nádas, & Bell, 2010). With extended essays in particular, the potential effect of the presentation medium on reading comprehension becomes an issue. The findings of one recent study addressing this issue (Johnson & Nádas, 2009), suggest that examiners had a weaker recall of essay quality on screen and had greater difficulty recollecting the location of details in these texts. Interview data support the suggestion that the examiners' comprehension was more challenged when reading from screen than from paper (Johnson & Nádas, 2009).

Further research is clearly needed. We need to know: how and to what extent might comprehension of texts differ when displayed on a screen as compared to being printed on paper (see more discussion on this in Chapter 9)?

Rhetorical task

Urquhart (1984) and Barnett (1989) suggest that there is evidence that rhetorical factors in the text as well as the more traditional intra-sentential linguistic factors should be considered in estimating text difficulty (see Barnett 1989:56). Studies looking at the effects of text organisation on comprehension, for example those by Meyer and Freedle (1984), Carrell (1984) and Goh (1990), suggest that problem/solution, comparison, and causation structures are better recalled than collection or description structures. In fact, they argue that comparison and problem/solution text types enhance comprehension more than other types of rhetorical organisation. Koda (2005) cites a number of studies reporting the positive effects of improving text structure and the benefits of explicit training in coherence on comprehension and memory. Freedle (1997) finds that texts subjectively judged to be high in coherence are suitable for main idea reading comprehension items. Such texts lend themselves to testing the expeditious reading types of search reading and skimming. Rhetorical features are therefore an important further consideration in the selection of texts for tests of academic reading.

Rhetorical task refers to 'the primary intent of the author' that guides the reader in understanding the text (Enright et al 2000:20). Enright et al (2000) suggest a useful three-way classification of rhetorical tasks (which they term 'pragmatic features'):

- *Exposition* informs the reader. It may involve descriptions, comparisons, contrasts, explanations and elaborations.
- Argumentation/persuasion/evaluation supports a point of view with reasons, evidence and analysis of an opponent's errors in reasoning. Vocabulary might reflect attitude or perspective and it may be personal in tone. It differs from a balanced, unbiased stance.
- *Historical biographical/autobiographical narrative* tells a story with a defined setting and episodes.

Pattern of exposition refers to 'subcategories of exposition' (Weigle 2002:62), or a specific pattern a writer employs to communicate. Although a single text may include a number of rhetorical moves, it is the overall theme or main point that is targeted through this feature (Enright et al 2000:23). The following patterns are suggested as being worthy of investigation in the literature:

- *Definition/description/elaboration* involves providing full definitions of concepts, describing unfamiliar terminology, elaborating on terms specific to the discipline and clarifying specific uses of the terminology.
- *Illustration* involves providing examples or a short anecdote to fully describe an abstract concept.
- *Classification* involves grouping several items together according to similar features or principles, showing how discrete items belong to a larger group.
- *Comparison/contrast* involves designating distinctions among concepts, particularly regarding their similarity and dissimilarity.
- *Cause and effect* involves analysing causes and effects in relation to an overall point.
- *Problem/solution* involves describing a problem or a series of problems then proposing a solution, which will have a plausible, salutary effect on a course of action.

Research and Practice in Assessing Academic Reading

• *Justify* as used here is similar to the category of *analysis* used by Enright et al (2000). Texts in this category provide evidence to justify a point of view.

Cohen and Upton (2006:7) identified the text types to be found in the new internet-based Test of English as a Foreign Language (TOEFL iBT) examination:

... according to the task specifications (ETS, 2003): With regard to text type, previous TOEFL reading passages "consisted primarily of a particular type of expository text in which a number of discrete facts are loosely integrated and developed" (ETS, 2003, p. 1). Along with expanded length, the texts in the Reading section of the new TOEFL (each test has three texts on different general academic topics) include a broader selection of academic text types, classified by author purpose: (a) exposition, (b) argumentation, and (c) historical biographical/autobiographical narrative. Each of these has at least one structure, such as classification, comparison/contrast, from more than one perspective or point of view (ETS, 2003).

Green, Ünaldi and Weir (2010:200) found that, as most texts beyond 500 words have multiple text patterns, the judges found that it was often not possible, especially with the undergraduate texts, to identify a dominating pattern and this variable is not included in their analysis.

Functional resources

Although early researchers (van Ek and Trim 1998, 2001, Wilkins 1973, 1976), produced a substantial body of work defining notional-functional progression across the CEFR Levels A1 to B2 (A1: Breakthrough, A2: Waystage, B1: Threshold and B2: Vantage), and the empirical work of North (2000) calibrated these functions onto a common scale, only limited work has been carried out on the C levels, which are the target levels for our academic reading test.

A functional progression study for the English Profile Programme carried out by CRELLA at the University of Bedfordshire (Green 2012) focused on the C levels in the CEFR. Green (2012) concludes that the CEFR suggests some qualitative change in the functions that characterise learner language between the A and B levels and the C level. The C level appears to represent not so much an increase in functionality and in the range of contexts for use as increasing ease and subtlety of communication. According to Green (2012:29), at C1 there is a deepening awareness of access to a broad range of language, which allows fluent, spontaneous communication, and, at C2, a level of 'precision, appropriateness and ease' (2012:29) in using the language.
It is apparent in the CEFR that passing from B2 to the C levels should enable the learner to access higher education, professional fields of employment and the literary culture associated with a language.

Green (2012) carried out an empirical study to create a database of the functions and Can Do statements to be found in Cambridge English (then known as Cambridge ESOL) textbooks, tests, scales and syllabuses at the B2 to C levels. As many of the data sources did not distinguish between C1 and C2 they were treated as a composite C level for the purposes of Green's study.

A comparative analysis at the level of Wilkins' (1976) functional categories (Figure 3.2) across functions and Can Do statements suggested that at the C level we see an increase in argument, suasion and rational enquiry and exposition. The proportion of functions is similar at the B2 and C level for emotional relations, while judgement and evaluation and, particularly, personal emotions seem to become less salient at the higher level.

Figure 3.2: Proportion of function words in Wilkins' (1976) categories found by Green (2012) at the B2 and C level (functions and Can Do lists) in the materials database



The following of Wilkins' (1976) functions occurred only at the C level in the lists of functions. In this list, Wilkins' superordinate categories of communicative function are given in parentheses: acknowledgement (emotional relations); disagree (argument); inform (argument); justification (rational enquiry and exposition); proposition (rational enquiry and exposition); recommend (suasion); and threaten (suasion).

Among the Can Do statements, the following occur only at the C level: advocate (argument); assess (judgement and evaluation); demonstration (rational enquiry and exposition); and illustration (rational enquiry and exposition).

The pattern of function words emerging at the C level is suggestive of a shift in focus and points to rational enquiry and exposition, argument, and suasion as being of particular relevance. A similar conclusion was reached by Shaw and Weir (2007) in their retrospective analysis of Cambridge English Writing examinations at these advanced levels.

Lexical complexity

A number of researchers and reading experts have identified potential sources of difficulty arising from the linguistic elements in a text (Nuttall 1996, Perera 1984, Urquhart 1984, Weir 1993). They suggest that lexical as well as grammatical difficulty strongly influence the ease with which a text can be read.

Weir (2005a:292–293) notes, however, that the CEFR provides little assistance in identifying the breadth and depth of productive or receptive lexis that might be needed to operate at the various proficiency levels included in its scales. Some general guidance is given on the learner's lexical resources for productive language use but as Huhta et al (2002:131) point out: 'no examples of typical vocabulary or structures are included in the descriptors'.

Frequency research suggests that learning vocabulary up to about the 5,000-word family level provides rewards in the general ability to use English. Beyond the 5,000 level however, vocabulary becomes increasingly tied to specific topics and/or domains, and so the recommendation is for each learner to focus on their topic-specific technical vocabulary from this point onwards (Nation 2001). Specification of appropriate lexical range would be particularly difficult at the higher levels so perhaps we should not be too critical of the CEFR. At advanced levels of proficiency (C1 and C2), discussion could literally be on any topic. It is impossible to specify a particular set of vocabulary which would enable engagement with a wide variety of topics. This is why the CEFR is so vague about the lexis required at these higher levels.

While it is difficult to specify which words are necessary for any particular language use context, vocabulary research has been more successful at specifying what size of vocabulary is necessary to achieve certain language aims. Khalifa and Schmitt (2010:25) suggest that:

Around 2,000–3,000 word families should supply the bulk of the lexical resources required for basic everyday conversation (Adolphs and Schmitt, 2003). About 3,000 word families is the threshold which should allow learners to begin to read authentic texts. Based partly on Laufer's

(1988) research, it was formerly thought that knowledge of around 5,000 word families would provide enough vocabulary to enable learners to read authentic texts without lexical problems; however, this was based on 95% coverage of texts. Now the consensus is moving toward a view that closer to 98% coverage is necessary for ease of reading which would require a larger vocabulary: something in the area of 8.000–9.000 word families (Nation and Gu, 2007). Of course many words will still be unknown, but this level of knowledge should allow learners to infer the meaning of many of the novel words from context, and to understand most of the communicative content of the text. Beyond this, for a wide L2 English vocabulary, a size of 10,000 word families is the figure most often cited (Hazenberg and Hulstijn, 1996). It is important to note that these sizes are approximations, and ability in English also depends on many other factors, including speaking and reading skills, background knowledge, and strategy use. However the sizes do provide "rules of thumb" which may prove useful for test developers to keep in mind.

A lot of research is now possible using computerised tools to ascertain levels of lexical complexity, for example in school textbooks across the ability range and in our three research studies in Chapter 2, in the texts employed in undergraduate university-level education, and in academic English testing.

Crossley, Greenfield and McNamara (2008:482,488) consider measures for establishing lexical frequency in selecting texts:

Coh-Metrix [Cohm] calculates word frequency information through CELEX frequency scores. The CELEX database (Baayen, Piepenbrock, & Gulikers, 1993) consists of frequencies taken from the early 1991 version of the COBUILD corpus, a 17.9 million-word corpus. For this study, the CELEX frequency score for written words was selected as the lexical-level variable. This measure was selected because frequency effects have been shown to facilitate decoding. Frequent words are processed more quickly and understood better than infrequent ones (Haberlandt & Graesser, 1985; Just & Carpenter, 1980). Rapid or automatic decoding is a strong predictor of L2 reading performance (Koda, 2005). Texts which assist such decoding (e.g., by containing a greater proportion of highfrequency words) can thus be regarded as easier to process . . .

The more frequent the word, the more likely it is to be processed with a fair degree of automaticity, thus increasing reading speed (even among lower level learners) and freeing working memory for higher level meaning building.

Low frequency can be used as a predictor of text difficulty. Weir et al (2012a), in our third research study in Chapter 2, report that a well-established frequency effect in reading results in slower decoding times for less frequent words (Garnham 1985). In addition, a high ratio of low-frequency content words increases the likelihood that a passage will contain a number of words

that are unfamiliar to the test taker. However, too much should not be made of the contribution of unfamiliar words to text difficulty. The fact is that many such words can be decoded by using analogy or derivational morphology; others can be ignored as they are not central to the main argument of the text. The true issue determining difficulty is the transparency of the words rather than necessarily their low frequency.

Percentage of words on Academic Word List

The emergence of computer-assisted analysis of extensive language corpora has facilitated the use of word lists to inform language test development and validation. Of particular value to testing academic literacy are academic word lists that identify words used more commonly in academic than in other contexts, particularly the *sub-technical* vocabulary that occurs across disciplines (Campion and Elley 1971, Coxhead 2000). It is important that texts used in tests reflect the occurrence of such words in academic texts sourced from the university.

VocabProfile (Cobb 2003) enables us to identify the number of academic words in texts based on Coxhead (2000). It is described on Victoria University's website as follows:

The Academic Word List (AWL) was developed by Averil Coxhead as her MA thesis at the School of Linguistics and Applied Language Studies at Victoria University of Wellington, New Zealand. The list contains 570 word families which were selected according to principles. The list does not include words that are in the most frequent 2000 words of English. The AWL was primarily made so that it could be used by teachers as part of a programme preparing learners for tertiary level study or used by students working alone to learn the words most needed to study at tertiary institutions. The AWL replaces the University Word List. (www.victoria. ac.nz/lals/resources/academicwordlist/information.aspx)

Other lexical considerations may also prove helpful in deciding on what texts to use in our academic English tests.

Average syllables per word

The notion that a skilled reader identifies a word purely by its shape has long been discredited according to Weir et al (2012a). They argue that current models of lexical recognition (Rastle 2007) assume that a reader achieves lexical recognition by drawing upon a number of different cues in parallel. A word on the page is matched to an item in the reader's lexicon on the strength of: letter features, letters, digraphs, letter sequences, syllables, and the word as a whole. Of these, the units most easily recognised by a computer program are the syllable and the whole word. Readers take longer to process a multisyllabic word than a monosyllabic one, allowing for frequency effects (Rayner and Pollatsek 1989). The demands of decoding a text at lexical level are thus better measured by counting syllables than by counting whole words.

Lexical density: Ratio of content to function words

Words can be defined as either content words or grammatical function words. A content- or information-carrying word is any verb, noun, adverb or adjective which has a stable and significant lexical meaning. Grammatical function words bind a text together by creating the relationships between the concepts in a sentence. Function words include: auxiliary verbs, numerals, determiners, pronouns, prepositions, and conjunctions.

Lexical density is an index in VocabProfile which measures the proportion of content words to total number of words in a text. Lexical density can be calculated as a percentage by the following formula:

Lexical density = (number of content words/total number of words) X 100

Weir et al (2012a) argue that this measure relates to the processing differences between function words and content words. As the reader's eye moves across the page, it fixates the majority of content words, but only about 40% of function words (Rayner and Pollatsek 1989). The reader of English is able to anticipate and skip functors by detecting them in right parafoveal vision. They are readily recognisable because of their high frequency and short form. The higher the density of a text, the harder the text is to read since accessing the meaning of lexical items requires accessing the mental lexicon; function words can be dealt with directly by pattern matching.

One might perhaps assume there would be an increase in the ratio of content words to function words as texts became more syntactically and lexically difficult (there might, for example, be more complex noun phrases (NPs)). This could possibly slow down processing, in that the reader would need to fixate more. But it would be unlikely to add significantly to cognitive load, since at higher levels of proficiency automatic recognition routines should be in place for the more familiar content words.

Type-token ratio (TTR)

TTR is a useful statistic for examination providers as one facet of text difficulty. It is calculated by:

... dividing the types (the total number of different words) occurring in a text or utterance by its tokens (the total number of words). A high TTR indicates a high degree of lexical variation while a low TTR indicates the

opposite. The range falls between a theoretical zero (infinite repetition of a single type) and one (the complete non-repetition found in a concordance) (Thomas 2005:1).

The higher the TTR, the more demanding the passage is likely to be. A low TTR indicates that words are repeated many times in the text, which should generally increase the ease and speed of text processing.

As the length of the reading passage increases, so does the likelihood of more than one occurrence of any given word. But the range of words also increases. This is particularly the case as (at higher levels) texts begin to approximate more closely to an authentic writing style, where there are constraints against repeating a word in adjacent sentences. To avoid this kind of repetition, writers sometimes exercise a preference for a synonym rather than a pro-form – thus increasing the TTR. This trend can be seen in the presence of the criterion *range of vocabulary* applied in many advanced-level English examinations in the writing section.

Syntactic complexity

Alderson (2000:37) draws attention to the 'importance of knowledge of particular syntactic structures, or the ability to process them, to some aspects of second language reading . . . The ability to parse sentences into their correct syntactic structure appears to be an important element in understanding text'. Shiotsu (2003) investigated components most likely to affect performance in reading for Japanese undergraduates and established clearly the importance of syntactic knowledge. Shiotsu and Weir (2007) used structural equation modelling to demonstrate the relative superiority of syntactic knowledge over lexical knowledge in explaining variance in tests of reading administered to a variety of L2 participants. Crossley et al (2008:482) observe:

[In careful reading] a reading text is processed linearly, with the reader decoding it word by word; but, as he or she reads, the reader also has to assemble decoded items into a larger scale syntactic structure (Just & Carpenter, 1980; Rayner & Pollatsek, 1994). Clearly, the cognitive demands imposed by this operation vary considerably according to how complex the structure . . .

Texts with less complex grammar tend on the whole to be easier than texts with more complex grammar. Berman (1984) investigated how opacity and heaviness of sentence structures could result in increased difficulty in processing. Again, this suggests that a valid test of reading should reflect the syntactic features likely to be encountered in target situation texts.

However, as we noted earlier, text complexity is a balance, an interplay of elements such as propositional content, discourse mode, lexis and sentence

structure. Although longer complex sentences will often be harder than shorter simple ones, if the language used is very elliptical and the lexis used is highly colloquial, short simple sentences may actually be harder to understand than longer sentences.

The CEFR (Council of Europe 2001) provides no guidance on the grammatical range candidates might be expected to cope with in reading tasks at various levels of ability. This has consequences for the use of the CEFR in schools. Keddle (2004:43–44) noted that the CEFR did not measure grammar-based progression and this was problematic in relating the descriptors to the students' achievements. She argued that as a course designer she would have been happier if there were more explicit guidance in relation to grammatical appropriateness at the various levels.

Alderson et al (2004:49) provide four categories of grammatical complexity:

- only simple sentences
- mostly simple sentences
- frequent compound sentences
- many complex sentences.

Alderson et al argue (2004:128) that 'these [categories] aim to provide a general and at the same time a standardised way of describing the grammar in input texts and seem sufficient for the purpose of helping identifying levels'.

Work undertaken by Alderson and Clapham (1992) pointed to a very close relationship between a test of grammar and the IELTS Reading component. Indeed, the relationship was so close that a decision was taken to eliminate the grammar test from the IELTS battery. A similar result had been discovered much earlier by Weir (1983), and the grammar component had similarly been dropped from his TEAP battery despite showing itself to be the best single indicator of proficiency in academic English.

Shiotsu (2003) explored components likely to affect reading test performance for Japanese undergraduates and found that syntactic knowledge played a central role. Shiotsu and Weir (2007), using structural equation modelling, demonstrated the relative importance of syntactic over lexical knowledge in accounting for variance in reading test scores with candidates from a variety of language backgrounds.

A considerable number of indices have been suggested in the literature for the estimation of grammatical complexity (see Ortega 2003, Wolfe-Quintero, Inagaki and Kim 1998). This section will now consider three significant metrics employed in previous research:

• Cohm 37: *average number of words/the mean number of words per sentence*. In general, the longer the sentence the more processing time it takes up (*ceteris paribus*).

- Cohm 41: the *number of modifiers per noun phrase*. Modifiers refer to adjectives, adverbs, or determiners that modify the head noun. Sentences with difficult syntactic compositions have a higher ratio of modifiers. The concern is with the occurrence of complex NPs (these being a recognised feature of academic text).
- Cohm 43: the *mean number of words before the main verb in sentences*. Sentences with a larger number of words before the main verb tend to be more difficult. Structurally opaque texts tend to have proportionally more high-order syntactic constituents and greater numbers of words before the main verb.

Average sentence length

Weir et al (2012a) claim that this index would appear to be a rough measure of both the syntactic complexity and the lexical density of a sentence. Clearly, the number of words in a sentence must often correlate loosely with the sentence's complexity in terms of number of clauses. Alternatively, or in addition. a longer sentence might contain longer and more complex phrases, i.e. might be denser in lexical terms. This measure partly relates to processing at the level of structure building (Gernsbacher 1990) in that the more complex the sentence, the more elaborate is the structure that has to be assembled. If one assumes that longer sentences might also result from longer and more densely packed clauses, then the measure is also an indicator of difficulty of parsing. In parsing, a reader has to hold a series of words in the mind until such time as they reach the end of a clause and can trace a syntactic pattern in the string (Rayner and Pollatsek 1989). The longer the clause, the more words the reader has to hold in the mind. Lewis, Vasishth and van Dyke (2006) suggest that processing items towards the end of longer sentences will be harder, since they usually have to be integrated with items that have occurred earlier on in the sentence. Graesser et al (2001) also suggest that longer sentences tend to place more demands on working memory and are therefore more difficult to process.

Readability

Readability statistics (R1 Flesch Reading Ease and R2 Flesch-Kincaid Grade Level), widely used in test development, are available through Microsoft Word; both measures are based on the relative numbers of syllables, words and sentences found in a text. Flesch Reading Ease scores range from 0 to 100 with lower scores reflecting more challenging texts. A score below 50 is said to require college-level reading skills. The Flesch-Kincaid Grade Level is based on the US school system, with 12 representing the final year of high school and 13 to 16 the college level.

Number of modifiers per noun phrase

The mean number of modifiers per noun phrase is an index of the complexity of referencing expressions. Barker (1998) argues that NPs carry much of the information in a text, and computerised systems that attempt to acquire knowledge from text must first decompose complex NPs to get access to that information.

Graesser, McNamara, Louwerse and Cai (2004) suggest that sentences with difficult syntactic composition have a higher ratio of constituents per NP than do sentences with simple syntax. The presence of modifiers in the form of adjectives or prepositional phrases extends the length of a subject NP, and thus delays the point at which the verb is reached. However, the same argument would clearly not apply in the case of an object NP in a Subject-Verb-Object sentence. Weir et al (2012a) feel that a more satisfying explanation relates to the burden upon parsing: the inclusion of modifiers increases the length and complexity of the string of words that a reader has to hold in the mind while imposing a syntactic pattern upon it.

Mean number of words before the main verb

Graesser et al (2001) claim that sentences with many words before the main verb are taxing on working memory; however, Weir et al (2012a) maintain that this justification is not a convincing one since the authors refer to working memory as a very general notion and do not specify at all how it operates in this case. The best explanation would seem to be a syntactic one associated with parsing. Critical to the parsing of a clause is the verb, which not only provides a predicator for the event being described but also signals the likely syntactic structure of the whole sentence through its valency (Trueswell, Tanenhaus and Kello 1993). The presence of modifiers in the form of adjectives or prepositional phrases extends the length of a subject NP, and thus delays the point at which the verb is reached. An alternative explanation relates to parsing. The words that occur before the verb are the first in a sentence to be analysed, and the longer the subject NP is, the greater the burden imposed at this early stage upon working memory.

Cohesion

While Alderson (2000) notes that an absence of cohesive devices does not seriously damage comprehension when the topic is relatively familiar to readers, it has been argued that explicit cohesive devices help in establishing textual coherence (Goldman and Rakestraw 2000) and that their lack inhibits the recall of texts, being indicative of a less successful mental representation (Ehrlich 1991).

We adopt the Graesser et al (2004:193) definition of cohesion as a property of a text that involves: 'explicit features, words, phrases or sentences that guide the reader in interpreting the substantive ideas in the text, in connecting ideas with other ideas and in connecting ideas to higher-level global units (e.g. topics and themes)'. McNamara, Ozuru, Graesser and Louwerse (2006:573) argue that:

Cohesion arises from a variety of sources, including explicit argument overlap and causal relationships, and can operate between sentences, groups of sentences, paragraphs, and chapters (Givón, 1995; Graesser, McNamara, & Louwerse, 2003).

These cohesive devices cue the reader on how to form a coherent representation. The coherence relations are constructed in the mind of the reader and depend on the skills and knowledge that the reader brings to the situation; hence coherence is a psychological construct, whereas cohesion is a textual construct.

Two forms of textual coherence are estimated by Coh-Metrix: referential cohesion (the extent to which words in the text co-refer) and conceptual cohesion (the degree of similarity between concepts in different parts of a text). The effect of the use of cohesive devices on comprehension is less clear-cut. Cohesion did not prove to be that useful an indicator of level in the studies we considered in the literature; for example Graesser, McNamara and Kulikowich (2011:230) noted that:

... different forms of cohesion are not always positively correlated with grade-level bands. Text cohesion has a small variation over grade level, with a slight decrease for referential cohesion within most text genres and a slight increase for causal cohesion ...

Nature of information

Concreteness/abstractness

Here the concern is with whether the information in the text is abstract, e.g., concerning ethics, love, etc. or concrete e.g., describing the physical contents of a room. Both types may of course be present in the same text. Research indicates that abstract words are in general more difficult to understand than concrete words (Corkill, Glover and Bruning 1988). It seems likely that concrete language is easier to process because it can draw upon the cognitive operations of both verbal and nonverbal (imagery) systems. In contrast, abstract language is restricted to the verbal system. Abstract words are more difficult to process because they are not as imageable as concrete words (Weir et al 2012a). There is some evidence (Bleasdale 1987) that there may be separate lexicons for the two types.

Eddie Williams (personal communication) comments that: 'Concrete nouns are certainly easier to teach via ostensive definitions than abstract

ones, but not necessarily easier to comprehend within text if the reader has not met them before'. Alderson et al (2004:127) comment: 'This dimension, mentioned in the CEFR and also considered useful by the Project team, may prove useful to estimate the difficulty of the input text.'

The concern with the contextual parameter of *nature of information* is with the extent to which the information in a text is concrete (i.e. concerning observable, concrete phenomena) or abstract (i.e. concerning unobservable phenomena such as social institutions) or, at a higher level of abstraction, meta-phenomenal (concerning theoretical treatment of abstract phenomena). Different levels of abstraction may, of course, be found within a single text. Alderson et al (2004:127) see this as a useful feature to consider in estimating text difficulty in relation to the CEFR. Information that is more abstract may prove to be more difficult to process and so divert cognitive resources from language processing. At the same time abstract information often implies a linguistic complexity that may further stretch the L2 reader's resources.

Content knowledge

The contextual parameter of *content knowledge* in the sociocognitive framework proposed by Weir (2005b) suggests that the relationship between the candidate's pre-existing knowledge and the propositional content of a text will affect the way it is processed. Nuttall (1996) puts forward the widely held view that, all else being equal, the greater a reader's knowledge of a text, the easier it should be to process. This has been an area of debate for IELTS since its inception as the five academic subject-specific modules inherited from the ELTS test were reduced, first in 1989, to three and finally, in 1995, to one. The decision to abandon subject-specific modules was taken on the grounds that there was only very limited evidence that it had any effect on text difficulty. Tan (1990) and Clapham (1996a, 1996b) both investigated the effect of content familiarity on candidate performance without finding significant effects on test scores (although Clapham does note an effect for the most specific texts in her corpus). However, Khalifa (1997) made the contrary finding that familiarity with the topic of text can be a good predictor of difficulty. Alderson (2000) also acknowledges the facilitating effect of familiarity with the subject matter and Urquhart and Weir (1998) warn against the danger of using insufficiently specialised texts. It has been suggested that, in order to minimise effects of topic familiarity, test tasks should be based on materials sourced from a variety of academic subject areas (Enright et al 2000).

In our earlier discussion of processing we discussed the focus of items in terms of the explicitness or implicitness of locating the requisite information in the text. Propositional inferencing was considered suitable but pragmatic inferencing was not. However, even if background knowledge is not tested directly *per se* the relationship between the resources necessary to comprehend the text and those possessed by the candidates is an important one.

The relationship between the content of the text and the candidate's existing knowledge (see Douglas 2000) will affect the way it is dealt with. This interaction between the resources of the candidate and those demanded by the task emphasises the symbiotic nature of context and cognitive validity.

According to Hughes, in general testing (1989:93), 'the subject areas will have to be as "neutral" as possible, since the students are from a variety of disciplines'. This statement is also mirrored in Weir (1993:67): 'in those situations where we are writing tests for heterogeneous groups of students, we are by necessity forced to select texts with a wider appeal than is the case when we have a more homogeneous group'. Urquhart and Weir (1998:143) advise that:

The content of a text should be sufficiently familiar so that candidates of a requisite level of ability have sufficiently developed schemata to enable them to process it. A text should not be so arcane or so unfamiliar as to make it incapable of being mapped onto a reader's existing schemata.

As Alderson (2000:29) argues: 'every attempt should be made to allow background knowledge to facilitate performance rather than allowing its absence to inhibit performance'. Neither should a text be too familiar as then there is a danger that the candidate will be able to answer questions without recourse to the text, what Buck (2001:126–127) calls the need for 'passage dependency'.

Enright et al (2000) make the point that if we are to include reading-tolearn activities at the text level then candidates need to be faced with texts that contain information that is new to them. Examinations at the C1 and C2 level are used to judge suitability to handle the language demands of university courses where it would be improbable that students were exposed to text that only contained information they already knew.

Cultural knowledge

Steffensen, Joag-Dev and Anderson (1979), Chihara, Sakurai and Oller (1989), Al-Fallay (1994) and Sasaki (2000) have all provided evidence that *cul-tural knowledge* plays an important role in text comprehension.

In these studies, certain 'key' words – proper nouns, words describing institutions, and words that reflected unfamiliar cultural practices – were changed into words that would be more familiar for the participants. For example, in Chihara et al's (1989) and Sasaki's (2000) studies, which used the same texts, *Joe* was changed to *Hiroshi*, *state* to *prefecture*, and a mother *hugged* rather than *kissed* her son because these changes were felt to reflect a Japanese rather than an American cultural context for the narrative. The resulting texts, because they appeared more familiar to the participants, led to higher scores on a cloze test based on the passage. In this study the judges were asked to look for words that might be associated with a specific culture, including references to:

- names for specific people, places and products (Harvey et al; Rice Krispies; the city of Chicago)
- specific historical events or periods (the Norman Conquest; footballrelated violence in the 1970s)
- local institutions (the probation service; the House of Lords)
- locally familiar objects (breakfast cereals; sharp suits)
- locally situated social practices (window shopping; children in the classroom undertaking problem-solving activities in pairs)
- idiomatic language including culturally specific references (milestone research; professional soap boxes).

Conclusion

In this chapter we have examined various approaches to describing the construct of reading that have been proposed in the literature and we have sought to link the findings from the empirical research described in Chapter 2 to existing models of reading. We have presented a sociocognitive model of reading (see Figure 3.1) which draws upon insights from reading research, assessment theory and cognitive psychology, and we have explained how this can offer a platform for academic reading test development and a framework of reference for analysing and critiquing tests of academic reading, particularly in respect of their cognitive and contextual features (See Figure 3.1 and Table 3.2). Chapters 4–7 form Section 2 in this volume and will explore core sociocognitive parameters from this framework, specifically in relation to the current IELTS Academic Reading test.

Section 2 Academic reading tests in practice: The case of the IELTS Academic Reading Module

Using the framework for analysis developed in Chapter 3 (see Figure 3.1 and Table 3.2), we will examine the cognitive validity (in Chapter 4) and context validity (in Chapter 5) of the IELTS Academic Reading Module in relation to real-life academic reading activities. We will also consider research on the effect IELTS has had on society and individuals in Chapter 6, i.e. its consequential validity, and in Chapter 7 we will examine the relationship between IELTS and its external relationships with other measures of the academic reading construct.

The content of Section 2 will provide an empirical base for discussing some key considerations and potential approaches to assessing academic reading ability in the future in Section 3, whether by means of IELTS or other academic reading tests.

Cognitive validity parameters

Things are not always what they seem. Lewis Carroll (1865)

Introduction

The International English Language Testing System (IELTS) is an international, standardised test of English language proficiency designed for nonnative English language speakers. It is jointly managed by three partner organisations: the British Council, IDP: IELTS Australia and Cambridge Assessment English. It was developed 1986–1989 to replace the earlier English Language Testing Service (ELTS) test, which it took over from (see Davies 2008, Weir and O'Sullivan 2017:Chapter 5). It is now accepted as evidence of English language proficiency by over 9,000 institutions worldwide. Organisations in over 140 countries use IELTS. These include all universities and the vast majority of education providers in Australia, New Zealand and the UK, over 3,000 institutions in the US, as well as most universities in Canada. IELTS is seen by many as the industry standard for the testing of EAP. At the time of writing, it is the only Secure English Language Test approved by UK Visas and Immigration (UKVI) for visa customers applying from both outside and inside the UK.

The IELTS Academic Reading Module currently takes 60 minutes and there are 40 questions to answer. Each question is worth one mark. Sometimes candidates need to give one word as the answer, sometimes a short phrase, and sometimes simply a letter, number or symbol. It is mainly a test of receptive skills, but some limited writing is involved in the short-answer question (SAQ) format, though only brief answers are required, usually no more than two to three words. Candidates must be careful when writing answers on the answer sheet because they will lose marks for incorrect spelling and grammar. This means that the IELTS Reading test is in some respects still a test of writing ability too (albeit in a local sense at the mechanical level), although the earlier integrated link between one of the reading passages and one of the writing tasks was removed in 1995.

According to the official website (www.ielts.org/about-the-test/test-format), there are three reading passages with a total of c.2,150–2,750 words. Individual tasks are not timed. Texts are taken from journals, magazines, books and newspapers. All the topics are of general interest and the texts have been written for a non-specialist audience. The readings are intended to be about issues that are appropriate to candidates who will enter postgraduate or undergraduate courses. At least one text will contain detailed logical argument. One of the texts may contain non-verbal materials such as graphs, illustrations or diagrams. If there are technical terms, which candidates may not know in the text, then a glossary is provided. The texts and questions become more difficult as the paper progresses.

The important issue we deal with in this volume is the extent to which the current IELTS Academic Reading Module remains fit for purpose after 30 years of service, i.e. does it still do the job it was intended to do? Green et al (2010:191) argue:

Providers of tests of languages for academic purposes generally claim to provide evidence on the extent to which students are likely to be able to cope with the future demands of reading in specified real-life contexts. Such claims need to be supported by evidence that the texts employed in the test reflect salient features of the texts the test takers will encounter in the target situation as well as demonstrating the comparability of the cognitive processing demands of the accompanying test tasks with target reading activities.

Whilst one might reasonably expect that high-stake tests such as IELTS will have carefully addressed the cognitive and contextual demands of reading in an academic context, as detailed in the previous chapter, a review of the research literature on IELTS suggests that this might not be the case. The results of various surveys have raised doubts over the comparability of the cognitive demands initiated by the IELTS Academic Reading Module with those encountered in real-life academic study (see Chalmers and Walkinshaw 2014, Hawkey 2006:122–126, 132, 163; Moore, Morton and Price 2007, Owen 2016, Weir et al 2012b). A number of additional concerns have emerged in studies that have carefully analysed the contextual dimensions of IELTS Reading texts and compared these with the texts students encounter in real-life reading contexts (Green et al 2010, Taylor and Chan 2015, Weir et al 2012a).

Reading: principles and process

Principles

To analyse accurately and precisely the nature of the current IELTS Reading tests we need a detailed conceptual framework to help us account for the various elements of the reading construct they have sought to operationalise. To fully explain what is being tested, account has to be taken of the *cognitive* as well as the *contextual* dimensions of test task performance.

Cognitive dimensions

In Chapter 3, having considered the various alternatives, we concluded that we should base a reading task analysis on the cognitive processing model suggested initially by Weir (2005b) and elaborated in Khalifa and Weir (2009). These draw on the external evidence from cognitive psychology concerning the nature of the expertise in reading that examining boards should aim to sample through academic reading test tasks. The concern is with the mental processes readers actually use in comprehending texts when engaging in different types of real-life reading for academic purposes. We start by looking more closely at types of reading and the cognitive processes they give rise to.

Careful reading is intended to extract meaning from presented material at a local or a global level, i.e. within or beyond the sentence right up to the level of the complete text or texts. This approach to reading is based on slow, careful, linear, incremental reading for comprehension. As we saw in Chapter 3, there is also a strong case for taking account of the speed of reading as well as comprehension. Studies into students' reading abilities have indicated that for many readers reading quickly, selectively and efficiently (expeditious reading) poses greater problems than reading carefully and efficiently (Beard 1972, Weir 1983, Weir et al 2000).

Expeditious reading of continuous prose to access desired information in a text is difficult because it demands rapid recognition that is contingent upon sufficient practice in reading at length in the target language. This approach to reading includes skimming, search reading, and scanning. Skimming is generally defined as reading to obtain the overall gist, general impression and/or superordinate main idea of a text. Search reading involves locating information on predetermined topics. The reader only wants the information necessary to answer set questions or to extract data, for example in order to complete written assignments in line with a given rubric. Search reading differs from skimming in that the search for information is guided by predetermined focuses so the reader does not necessarily have to establish a macropropositional structure for the whole of the text. Search reading can take place at both the local and global level. Where the desired information can be located within a single sentence it would be classified as local, and where information has to be put together across sentences it would be seen as global. In both cases the search is for words in the same semantic field as the target information, unlike scanning where exact word matches are sought.

In terms of the cognitive validity of the EAP reading tests under consideration, we are interested in answering a number of questions. Do the cognitive processes required to complete test tasks sufficiently resemble the cognitive processes a candidate would normally employ in real-life tasks that we explored in Chapters 2 and 3? Is the range of processes elicited by test items as comprehensive as those in real-life tasks?

Process

In our review of the IELTS Reading test, we examine the types of reading and the cognitive processes they give rise to according to the model elaborated in Figure 3.1 in Chapter 3, together with the context validity parameters as set out in Table 3.2 in Chapter 3.

The process we followed began with a comprehensive literature search for published information, in particular for independent research, if available, relating to all cognitive and contextual aspects of the IELTS test, in order to gather information on each category in our test analysis framework. We also examined samples of actual tests in order to verify the published claims, or to obtain more detailed information. We drew on published descriptions of the test in order to complete the sections on *Skills Focus* and *Task Description*, *Timing*, and so on, and checked these against samples of each test.

The quantitative and qualitative data reported below were generated in particular by: Weir (1983) in his study of the academic reading needs and problems of tertiary-level students in the UK; Weir et al (2000) in their empirical research for the Advanced English Reading Test in China; Moore et al (2007) in their comparison of reading requirements in IELTS test items and in university study; Green et al (2010) in a comparison of IELTS and UK undergraduate texts; Weir et al (2012a) in their study of the academic reading needs of undergraduates in a British university; Weir et al (2012b) in their large-scale questionnaire-based study of the cognitive reading activities of UK undergraduates; Chalmers and Walkinshaw (2014), who looked at reading strategies in IELTS; Taylor and Chan (2015) in a comparison of IELTS with other major EAP reading tests; Bax's (2015) research into the cognitive processes of multinational readers during an IELTS Reading test; and Owen (2016) in a doctoral thesis examining the cognitive processes activated by IELTS Reading test item types.

The sociocognitive model for reading first published in Weir (2005b) and later refined empirically by Khalifa and Weir (2009) for Cambridge English underpinned these studies (see Chapter 3 for detail). We first examine what reading processes appear to be tested: the cognitive parameters of the IELTS tests.

Cognitive parameters in the current IELTS Academic Reading Module

Expeditious and careful reading activities

Weir et al (2012a) examined the IELTS Academic Reading Module tests to determine the extent to which they were representative of real-life academic reading activities. Their instrument for the analysis of the Academic Reading

tests was derived from the reading strategies, skills and processes reported in Chapter 3. They used a *test task: reading strategy/skills matrix* (see Table 4.1) to describe the activities in IELTS Academic Reading tests.

Types of	Expeditious								
skill	Skimming S		Se	Search reading			Scanning		
	 The read and consistent of the reading over all over a machine of the a machine of the a machine of the all over all	ader locates mprehends hation at the l gist level. ng is selective, ections of the ther omitted en very little on. empt is to build up rostructure whole text st) based on l reading of e of the text sible.	•	The reader l information selectively o topics to an questions, e for related v semantic fie The reader i predetermin does not hav macro prop for the whol Once the rea information question ha and selective careful read	ocates quickly and n predetermined swer set g., by looking ocabulary in the ld. s guided by ed topics and so ve to establish a positional structure of the text. quired to answer a s been quickly sly located, ing will take over	• re•	The reader reads quickly and selectively to achieve very specific reading goals, e.g., looking for a specific word or phrase, date, figure. Limited careful reading may follow this matching activity.		
Types of				Care	eful				
reading skill	EWS	IWS		EAS	IBS	ТМ	SM		

 Table 4.1: Cognitive parameter matrix and reference key for the analysis of

 IELTS Academic Reading tests

EWS: Explicit within sentence. Establishing basic propositional meaning at sentence level through explicitly stated ideas in the text. Basic comprehension questions are used to assess lexical, syntactic, and semantic abilities and the ability to understand important information presented in sentence-level propositions.

IWS: Implicit within sentence. Inferencing by creating information which is not explicitly stated in a sentence. Understanding information in a sentence may require addressing conceptual gaps by constructing a message from both what is explicitly stated and our stored knowledge. Such inferences are necessary for a full understanding of the sentence.

EAS: Explicit across sentences. Establishing meaning through explicitly stated ideas across sentences.

IBS: Implicit between sentences. Inferencing meaning which is not explicitly stated between sentences in a text.

TM: A text model. Creating a text model. Constructing an organised representation of the text including main points and supporting details; an integrated understanding of how supporting ideas and factual details of the text form a coherent whole.

SM: A situation model. Answering questions based on a situation model. Addressing conceptual gaps by constructing a message from both what is explicitly stated and from our stored knowledge. Building a situation model involves the reader forming a representation of the content, relating the contextual information of a text to mental models of corresponding real-life situations.

Two project members were asked to record independently the reading skills they employed to respond to each task on 14 IELTS Academic Reading Module tests, selected as described above. All the IELTS tests used in the analysis are authentic and now in the public domain. We note here, from the 14 complete tests selected for their analysis, that:

- IELTS Academic Reading Module tests contain three separate texts on which candidates must respond to common test tasks
- there are a total of 40 items in each test
- each test contains an average of 3,458 words to read, including the tasks and rubrics
- the average number of words of *reading text* to read in each test is 2,562 words
- the average test text is 854 words long (maximum 1,063 words, minimum 589 words).

The Reading test tasks included: the *matching* of suggested and actual test content; the categorisation of suggested content as *Yes/No* (or *True/False*) or *Not given* in the test text; *gap filling; multiple choice; table* or *other iconic completion*, and *short-answer questions*.

The IELTS Reading test analysts were all informed postgraduate participant project members. They responded to all 42 testlets (14 tests x 3 texts) *qua* IELTS-takers. Table 4.2 below shows the cognitive operations they believed they had used in the process.

The *quantitative* analysis is based on the completion of 14 IELTS tests, comprising 42 texts (at three texts per test), each test with a total of 40 items. The two test taker/analysts thus covered 560 test items, in responding to each of which they were asked to identify the *primary* reading skill they felt themselves to have used. As the two total numbers in the right-hand columns of Table 4.2 are both above 560 (at 562 and 585 respectively for test taker/analysts A and B respectively), the indication, perhaps not surprisingly, is that for a number of items they felt that they applied more than one primary skill.

Modules	,			•		D				D
	Reading	expeditio	usly by:	Readi	ng <i>carefully</i>	for meaning whi	ich is:	Careful	l reading	Totals per test
Test taker/ analyst	skimming	search reading	scanning	explicit within sentences	implicit within sentences	explicit across sentences	implicit between sentences	To construct a text model	To construct a situation model	taker/ analyst
V	0	45	50	277	27	115	45	ю	0	562
В	70	9	93	318	12	57	25	4	0	585
Cognitive skill totals	70	51	143	595	39	172	70	7	0	1,147
Sub-totals: expeditious vs careful reading		264		883						1,147

Research and Practice in Assessing Academic Reading

The first caveat concerning this study is that only two language testing postgraduate students were involved in the analysis and the processes they employed may not necessarily have been representative of the wider population of potential test takers. Owen (2016:16), in critiquing Weir et al (2012a), points to a further consideration in interpreting any data on strategy use:

Purpura (1998) argues that a lack of evidence of specific strategy usage does not mean that particular strategies were not used. Researchers must account for the possibility that mental processes may go unreported by research participants. The research design of any study may attempt to account for this shortcoming by providing more detailed or innovative stimuli to maximise the opportunity for research participants to verbalise or otherwise report their thought processes.

Nevertheless, despite these two limitations, the overall findings in relation to local and global processing of text, and careful and expeditious reading, are indicative and are discussed further below. They form part of the body of evidence we managed to collate from a variety of sources. The similarity of findings with that of other researchers using different methodologies in different contexts determines the degree of faith we might choose to place in them.

The first general finding is the apparent **preponderance in this data of careful reading over expeditious reading activities** for both test taker analysts; 77% of the claimed reading skills (883 out of the total of 1,147) apparently belonging to the former category. This apparent imbalance is perhaps a matter of concern given that the students at the University of Bedfordshire, when they were asked about their actual academic reading purposes and problems in the same study (Weir et al 2012a), saw expeditious reading activities as more appropriate to their needs than careful reading skills. Extracting relevant information quickly from multiple texts on a topic and transforming it for assignments was the core academic activity for all tertiary-medium students.

In this first study, the two analysts reported more than three times as many instances of reading carefully as reading expeditiously. 883 of 1,147 reported reading activities (77%) were instances of participants reading carefully rather than expeditiously, of which 595 (51.6%) were reading explicitly within sentences, with only 242 instances of establishing propositional meaning across sentences (21%) (172 explicit across sentences and 70 implicit across sentences).

One might argue that because IELTS includes 13 or 14 questions relating to each short text, there are ample opportunities for candidates to read the text or parts of it carefully several times to find the information necessary to respond. The longest IELTS text in Weir et al (2012a) has 1,034 words (including the title and glossary) and the shortest has 586. If a participant were to spend about one third of the available time reading the questions and writing the responses, he or she would still only need to read at the very slow rate of around 50 to 75 words per minute in order to read through each text once. IELTS thus allows for very intensive careful reading of material that is often of only modest difficulty (see Chapter 5 on contextual parameters) when compared with the introductory undergraduate readings described in Weir et al (2012a).

A number of points emerge in relation to expeditious reading reported by the two postgraduate analysts. **Scanning for direct matches of words** is the prominent expeditious strategy for both analysts in tackling the IELTS texts, occurring more often than skimming or search reading combined. Secondly, **the patterns of global expeditious reading differed markedly** across participants A and B, although this discrepancy may be due in part to the search activities sharing some characteristics which resulted in participants reporting different skills e.g. skimming vs search reading. Learner awareness of strategy use may well vary across individuals and this again may help explain the discrepancies.

Weir et al (2012b), in their second phase of the study, took another look at the cognitive processing that takes place in IELTS Reading, this time using a more broadly based survey design (see Chapter 2 Research Study 3 for details and Chapter 4). Because of the intensive nature of verbal protocol research, which requires participant training and may generate a very large quantity of data for each individual, studies typically involve no more than a handful of participants. In this second study, Weir et al (2012b) set out to triangulate the detailed protocol data obtained in Weir et al (2012a), with less nuanced data elicited through a questionnaire survey delivered to a large group of participants.

The second study provides evidence that, for many respondents across the different task types, expeditious reading, albeit mainly in the form of scanning for specific words, plays an important role in the way they seek to answer the questions. The most popular test strategy was 2-quickly match words that appeared in the question with similar or related words in the text. This emerged as the most frequently reported strategy on 10 of the 15 test sections with 83% of participants reporting using this strategy at least once. It should be emphasised that this had in fact been the most prevalent expeditious strategy for the participants in Research Study 1 (Weir et al 2012a) as well, confirming the prevalence of this particular expeditious strategy use across the two studies.

Chalmers and Walkinshaw (2014:24, 31), in their study of academic reading, discovered similar levels of **local** expeditious reading behaviour to those reported by Weir et al (2012a, 2012b):

The analysis revealed that participants responded to time pressure, unfamiliar vocabulary and demands on working memory by employing a range of expeditious reading strategies, which focused less on textual comprehension than on quickly locating correct answers. Their comprehension of texts often remained at the "local-literal" level rather than the "global-interpretive" level ... Strategy 2 (quickly matched words that appeared in the question with similar or related words in the text) was clearly the predominant strategy, accounting for 19% of all instances of strategy use. 100% of participants used Strategy 2 at least three times in the test.

They had collected data on local search reading as well as scanning and report (Chalmers and Walkinshaw 2014:25–26):

Search reading is similar to scanning but targets vocabulary related to the semantic field of the words in the question, rather than simply searching for a visual match. Once target words are located, careful reading can take place to establish propositional meaning. Because this type of reading is based upon predetermined topics (e.g., IELTS question items), cognition takes place below the level of building a mental model of textual content and no comprehension of the overall text is necessary. In other words, if the information located answers the question, how it relates to the rest of the text is unimportant.

Owen's doctoral thesis analysed the cognitive activities of students taking IELTS and TOEFL iBT reading tests. He used video recordings of the student reading to facilitate the verbal recall protocols employed. His research (2016:144) provides evidence of the frequent application of (local) expeditious strategies in completing the IELTS Reading Module: 'the total number of codes identified for IELTS was split almost evenly between careful and expeditious reading'.

He found that expeditious reading is not tested separately from careful reading in IELTS, but rather the two appear to co-occur as candidates search quickly for information, which will help them to answer a question by word matching with the question (exact word and/or sematic equivalent) to locate sentences in the text in which the answer might be found. After strategy 2, the next most prominent strategy in participant self-reported behaviour in Weir et al (2012b:231) was strategy 10 - read relevant parts of the text again. This appears as the most popular choice on two test sections and was selected at least once by 77% of participants. This confirms that careful rereading of text when located expeditiously was a common occurrence. The picture of reading in response to IELTS test items that emerges is consistent with the general approach to academic reading reported by student readers in the Weir et al (2012a) protocol study: for many sections of the test, quick and selective reading for the most part at the local level was followed by intensive careful reading of relevant text parts.

The restricted use of global expeditious strategies in IELTS must be a cause for some concern to receiving institutions, as this is clearly the most common reading activity for students in tertiary education (see Chapter 2).

The lack of any serious time constraints, the short nature of most of the passages (see Chapter 5 on contextual parameters of IELTS) and the failure of the item writers to focus on global questions (see the next section on local and global items) means that expeditious global reading is seldom addressed in the IELTS Reading tests.

Bax's (2015:17–18) eye-tracking study of candidates taking IELTS items found significant differences between good and poor readers in terms of expeditious reading:

Unsuccessful students were apparently poor at locating the site of a correct answer in the text, unlike successful students ... Interview data showed that the reason for this was the use by successful readers of relatively conscious metacognitive strategies. Unsuccessful students, however, appeared to use no such conscious strategies, apparently searching relatively randomly through the text to find the location of the answer ... The differences in expeditious reading abilities, as revealed in the eye movement data, matched interview data which suggested that what differentiated successful readers from unsuccessful readers was the fact that most of them used pre-determined and conscious strategies systematically, whereas unsuccessful students for the most part appeared to be more aimless in their approaches.

He concludes (Bax 2015:18):

A further implication for future language test design and development concerns expeditious reading. If, as the findings in this study show, the ability to read expeditiously is an important marker of successful as opposed to unsuccessful readers, then future reading test developers might well choose to give expeditious reading an even more central place in their specifications than they do currently. Given that expeditious reading can be assessed most effectively via computers, which can enforce timed reading components straightforwardly, it may well be that test designers will move even more comprehensively towards computerised modes of testing reading.

The data above suggests that the reading activities tapped by the IELTS Academic Reading Module require modification to more closely represent the academic reading constructs that face university students. There is clearly a need to include texts and tasks that focus directly on students' expeditious global reading skills, in particular search reading at the global level, i.e. beyond the sentence, to stimulate mental model building and text-level comprehension. Candidates in the current tests are evidently employing some local expeditious strategies, but this does not result from tasks being purposefully structured to elicit such strategies. It seems to occur almost invariably as a precursor to careful reading of relevant parts of the text rather than as part of any attempt to expeditiously establish a mental model let alone a text-level representation *per se*.

Weir et al (2012b) conclude that there should be a dedicated section in IELTS to test expeditious global reading (as in for example the GEPT Advanced test in Taiwan, see Wu and Lin 2008). Such a test would impose explicit time constraints that encourage the use of the expeditious global reading strategies essential for university study. This section might include one or two longer texts to be processed under restricted time conditions to facilitate the testing of these global expeditious reading activities. Framing tasks so that the end product is a whole-text representation might facilitate the required processing. If the test is computerised then such an expeditious task is clearly both possible and desirable. Such dedicated tasks are likely to have beneficial washback if they equip students with the strategies to read texts selectively, quickly and efficiently to establish global meaning, as they have to do in real-life reading activities in an academic context.

Further research into comparability of performance on items testing careful and expeditious global reading skills by the target population is necessary (see Weir et al 2000). If a clear need is established to distinguish between the two, it may then be necessary for IELTS to be more proactive in trying to test these expeditious skills in terms of how the test is structured.

Local versus global: A focus on lower-level processes

In their study comparing the IELTS Academic Reading Module to real-life academic activities, Moore et al (2007:2) conclude:

The majority of the IELTS tasks were found to have a "local-literal" configuration, requiring mainly a basic comprehension of relatively small textual units [sentences, inter-sentences, paragraphs] . . . In the academic corpus, a sizeable proportion of tasks had a similar local-literal orientation, but others involved distinctly different forms of engagement, including tasks that required a critical evaluation of material (i.e. more interpretative), or which stipulated reference to multiple sources (i.e. more global) . . . A useful principle to strengthen the test's validity, we argue, would be to push test tasks, where possible, in the direction of the more "global-interpretative" reading modes required in academic study.

Taylor (2012:381–382) picks this up in her review of the impact of IELTS Reading in relation to Moore et al's study:

Most of the IELTS Academic Reading test items were observed to reflect features of reading tasks found in the corpus of academic texts gathered for the study, texts which had as their focus the need for students to understand certain discipline-based concepts. At the same time, however, there was evidence of some divergence between the two domains, with a variety of reading tasks in the academic corpus appearing to require a more critical engagement with material or interaction with multiple sources and viewpoints. These task types and demands were noticeably less evident in the IELTS task corpus under scrutiny...

Earlier concerns in Weir et al (2012a) relating to the number of items that seemed to focus on the sentence level were lessened slightly in their second study (Weir et al 2012b). In the second study, participants most often reported finding the information necessary to respond to the tasks by putting information together across sentences (option 2). Weir et al (2012b:236) conclude:

This was selected most frequently on nine of the 15 test sections and was chosen at least once by 89% of participants. However, there is some evidence that there may nonetheless be a high proportion of test items where the answer can be found within one sentence. Option 1 (within a single sentence) was the most popular selection on four test sections and was chosen at least once by 76% of participants.

In his eye-tracking study mentioned above, Bax (2015:18) set out to look at the difference between global and local items in IELTS. The difficulty he experienced in carrying out this study is insightful for our discussion of IELTS in terms of what he was unable to do, rather than what he did:

... this research study was unable to come to conclusions concerning global reading as opposed to local reading, partly because of the difficulty of isolating suitable global reading items to investigate. It proved more straightforward to investigate local reading. An issue which this raises for reading tests in general is whether they perhaps focus too extensively on local reading, both in terms of quantity of items and also in test validation, to the relative neglect of global reading. One implication of this is that test developers could usefully reconsider the extent to which reading tests succeed in testing global reading, and how we might be able to establish the cognitive validity of such attempts.

Green and Hawkey (2012:302–303) provide insights from some of the item writers in their study on IELTS item writing, which are revealing on the local v global reading issue:

[X] also questioned whether IELTS candidates would need to arrive at a full understanding of the text in order to succeed on the questions, suspecting that in IELTS "half the time the candidates don't read the text from beginning to end because they don't have to" because local details in the text were being tested by the items rather than the overall meaning.

Research and Practice in Assessing Academic Reading

These findings in relation to the amount of text processed to reach an answer in IELTS are confirmed by Owen (2016) in his doctoral thesis, which was concerned with reverse engineering the construct being measured in two internationally prestigious EAP reading tests (IELTS and iBT TOEFL). His thesis is the most extensive study to date of the cognitive processing that occurs when students take the IELTS Academic Reading Module. Owen (2016:361–362) explains how:

An analytical framework based on the cognitive processing core of Khalifa and Weir's reading model (2009) was applied to participant verbalisations. This core represents a hierarchical progression of engagement with text. Lower levels represent local engagement. As one moves up the core, the cognitive load increases. These levels were transformed into a coding framework, with the hypothesis that each of the levels would be able to be applied to participant verbalisations in SRI [Stimulated Recall Interview] if participants received sufficient stimuli.

Table 4.3:	Final coding scheme	e for the proc	essing core of	Khalifa and	Weir's
(2009) mo	del of reading				

Level of processing	Code
Creating an intertextual representation:	[P8]
Construct an organised representation across texts	
Creating a text-level representation:	[P7]
Construct an organised representation of a single text	
Building a mental model:	[P6]
Integrating new information; enriching the proposition	
Inferencing:	[P5s/c]
At word/sentence/clause level	
Inferencing:	[P5w]
At word level	
Establishing propositional meaning:	[P4s]
At sentence level	
Establishing propositional meaning:	[P4c]
At clause level	
Syntactic parsing	[P3]
Lexical access	[P2]

Owen (2016:361–362) describes how:

Khalifa and Weir's (2009) model of reading was applied to participant verbalisations. Each of the levels was given a code and verbalisations were analysed for evidence of the highest level of processing that could be inferred (using a coding algorithm) on the basis of participants' explanation of their actions and responses to specific items. A level of processing from Table 5.1 [Table 4.3 in this volume] was applied to each of the verbalisations. The thesis demonstrated that it is possible to infer the level of cognitive processing used by participants based on whether the unit of reference is a single word, a phrase, a complete sentence, an understanding derived from adjacent sentences, an impression formed from engaging with an extended text, or in response to a level of understanding that has been established from multiple points of reference across multiple paragraphs culminating in a mental model. The level of understanding can be judged by the extent to which the participant rephrases parts of the text in their own words.

Owen (2016:366–367) summarises the identified cognitive processes in the participant verbalisations for both IELTS and TOEFL in Table 4.4.

Higher/lower	Cognitive processing	IELTS	%	Higher/ lower	TOEFL iBT	%	Higher/ lower
Lower	P2	93	39.74	88.03	108	34.84	82.58
processes	P3	28	11.97		31	10.00	
	P4c	40	17.09		66	21.29	
	P4s	45	19.23		51	16.45	
Higher	P5w	2	0.85	11.97	7	2.26	17.42
processes	P5s/c	4	1.71		10	3.23	
	P6	20	8.55		33	10.65	
	P7	2	0.85		4	1.29	
	P8	0	0		0	0	
	Total	234	100	100	310	100	100

Table 4.4: Identified cognitive processes for IELTS and TOEFL*

*Percentages may not sum due to rounding.

Owen concludes (2016:366-367):

The majority of identified cognitive processes were lower-level processes for both tests . . . high-level processing in both tests lies primarily in "forming a mental model" (P6) rather than inferencing or creating a textlevel representation . . . Some of the higher-level processes in Khalifa and Weir's model are under-represented in both IELTS and TOEFL, including inferential reasoning and forming a text-level representation.

Inferencing

In Weir et al (2012a) Study 1, both test taker analysts did find cognitive processing involving implicit meaning at the careful reading level, though in substantially fewer cases (109) than those involving explicit meaning (767), as indicated in Table 4.2. Owen (206:368) found very few examples of items testing inferencing in the IELTS reading items he investigated: 'Positive evidence of this level of processing in IELTS is sparse'.

Research and Practice in Assessing Academic Reading

Moore et al (2007:41–42) drew attention to the fact that the types of materials students need to read on their courses, and the ways they need to go about reading these materials, are subject to a good deal of variation with the technical domains requiring less interpretative reading than the humanities, at least in the early stages of undergraduate life:

... the important type of reading in these more technical disciplines was that related to basic comprehension of material ... it was generally felt that what was crucial in the first year of study in their disciplines was for students to come away with a good working knowledge of foundational ideas and concepts – and not to be spending too much time deciding whether such ideas were valid or not. A number pointed out that whilst more "interpretative" forms of reading were clearly important in students' overall academic development, they had less obvious relevance in the early stages of training in the discipline ... at more advanced levels in these disciplines, the other more interpretative types of reading had a much greater role to play.

The view expressed from the more humanities-oriented areas represented a clear contrast ... A common theme here was that in one's teaching, the more "literal"-based skill areas were taken for granted to some extent, and that much of what students were expected to do simply assumed an understanding of basic concepts in the field ... For these academics, the focus was squarely on the more interpretative reading skills. Among those on the list, the idea of being critical of texts (item 5), and of being able to draw on multiple sources to support an argument (item 7) had particular resonance.

Moore et al (2007:65–66) comment that:

... we saw that the essential task for students in many of the IELTS items was to demonstrate a basic comprehension of the propositional content of reading material. By contrast, the focus of many of the academic tasks was not only to arrive at a basic understanding of material, but also to "work" with this material in order to proffer some interpretation of it. This basic requirement of academic study was well summarised by one informant thus: we typically [want students] to pick out ... the key points in the reading. But we also want them to reflect on what they have read and always relate it to their ... work somehow.

In the academic corpus, it was noted that two types of interpretative reading tasks predominated – what we have termed application and evaluation. In application-related tasks, students were typically required to show how a concept or idea in their reading could be utilised in their work in the discipline; in evaluative tasks, the focus was more on making some explicit assessment of these concepts (e.g. with respect to their validity, worth etc.).

Moore et al (2007:63–64) are at pains to point out the 'difficulties inherent in trying to conceive of some generalist construct of academic reading, one that has clear relevance to all disciplinary contexts'.

Text-level representation

It would be wrong to think that there are no text-level questions in IELTS. Weir et al (2012b:240) report:

Fears that IELTS was not addressing understanding at the whole text level also appear to be ill grounded. 3 (by understanding how information in the whole text fits together) was the most frequent selection on one section.

Moore et al (2007:33) pick out the following example:

Finally, the last question in this sample, Item 12, requires consideration of the whole reading passage – a text consisting of 10 paragraphs (Correct response = D).

12. What is the best title for this passage?

- 1. A The rise of the cinema star
- 2. B Cinema and novels compared
- 3. C The domination of Hollywood
- 4. D The power of the big screen

Owen concurs (2016:327):

"Multiple-choice" items are designed to elicit a range of cognitive processes, as item 1 asks participants to identify the main purpose of the text.

But the point is that text-level questions are few and far between, as Owen's data confirms. Moore et al (2007:34) comment:

Significantly, items . . . requiring test-takers to decide between different possible titles for a reading passage . . . were the only tasks found in the corpus that called for engagement at this whole-text level. A total of five instances of this item type, all in a multiple-choice format, were noted in the overall corpus, accounting for 1% of items.

Owen 2016 (2016:323) also suggests:

It is likely that summary or diagram completion items may require test takers to identify "understand details and/or the main ideas of a part of the text" ("building a mental model" or "creating a text-level

Research and Practice in Assessing Academic Reading

representation") which requires higher-level processing in order to integrate information across multiple sentences and paragraphs.

Moore et al conclude (2007:37):

In summary, the analysis found that a majority of tasks in the IELTS corpus were of a distinctly "local-literal" configuration, requiring mainly a basic comprehension of relatively small textual units (sentences, intersentences, paragraphs). It was noted that for a number of the more common task-types, the required engagement was in fact of a highly "local' and "highly" literal kind (i.e. True/False/Not Given; short answer). Other task types were found to be slightly less "local and literal" in their orientation (i.e. section-summary match), but were thought nevertheless to mainly inhabit the "local-literal" region of our analytical matrix. The only items in our corpus that clearly traversed the "local-literal" domain were certain multiple-choice items that required an appraisal of the reading passage as a whole (e.g., items requiring the selection of a title for a reading passage). It was noted that the Not Given option in True/False questions also required a more global engagement (i.e., in order to establish whether information is in fact present in a text). As was discussed however, items of this type arguably constitute a special case.

Intertextuality: The elephant in the room

Perhaps of most cause for concern is that in all the IELTS research studies, there is almost no evidence of any intertextual reading in IELTS, which is arguably, as we saw in Chapter 2, the most important skill for all undergraduates and postgraduates. Intertextuality had of course been a core innovative feature of the earlier ELTS reading module (1980–1989) and its loss in the ELTS revision 1986–1989 conducted by Lancaster University, which resulted in IELTS replacing ELTS, was never explained (Weir and O'Sullivan 2017). Moore et al (2007:67) draw attention to this deficit:

Another difference noted between the two domains [IELTS and reallife academic] was the quantity of reading required to complete some tasks. As we saw, all tasks in the IELTS corpus were focused on engagement with a single text (the relevant reading passage), and in the case of some task-types, a focus on relatively small components of the text. In contrast, a feature of some of the academic tasks, especially in the more humanities areas, was the need for students to engage with a range of texts. Examples of such tasks were: i) summary tasks which required students to give an account of a variety of sources in relation to a particular topic; and ii) essay tasks requiring the exploration of a range of views as a prelude to students presenting their own views on the topic.

. . . Several informants were sure that it was only through the engagement with multiple sources that students could develop a suitably critical

frame in their studies. As one informant explained it, students might feel they have come across "a perfectly reasonable answer" to a question, but that they are in fact only in a position to presume this if they've had the opportunity to "measure this answer against alternatives".

Moore et al (2007:67-68) note that:

Arguably one of the more significant literacy events in academic study is that which involves the integrating of one's reading on a topic into some related writing activity (Horowitz, 1986a). This was evident in many of the academic tasks analysed in the study, with virtually all of the assignmentstyle tasks in the corpus having a substantive reading component attached to them. A number of informants commented on the importance of this reading-writing nexus, seeing it as an area of particular challenge to students. Concern was expressed here about students' abilities to use and document sources appropriately, along with perceptions about the growing incidence of plagiarism on courses. Several informants noted the absence of these reading-writing connections in the sample IELTS materials provided, and wondered whether this dimension of academic reading could be incorporated into the test somehow ... Arguably, the purpose of a test of reading is to assess students' abilities to process written text. In this context, as we have seen, the actual contents of the reading tend to be somewhat incidental. In university study, by contrast, such content - which relates to study in a discipline - is of paramount importance. Thus, in university study, there is not the same interest in the skills of reading per se; instead acts of reading, as we have seen, are tied intimately to the acquisition, application, and ultimately to the advancement of disciplinary knowledge. This contrast in the role of knowledge in the two domains necessarily entails some quite basic differences in the nature of the texts students need to read, and what it is students need to do when they read them.

Owen (2016:164) also comments on the absence of any items focusing on the top level of processing in Weir's model of reading viz intertextual representation (Khalifa and Weir 2009):

No instances of the final code (P8) were recorded for either test. This is not unexpected, as neither test contains a task which asks participants to combine information from more than one text. All items in both tests relate to a single text only. A focus on inter-textual representation would require two (or more) texts on a single topic so that participants can formulate links across them. As the IELTS test contains only three texts, this would mean that the range of texts presented to test takers would be limited in genre, coupled with the difficulty of designing and specifying item types that require participants to form text-level representations of more than one text, and then demonstrate that they can compare information across texts in some way. This would likely require some form of extended writing activity to demonstrate command of both texts ...

Research and Practice in Assessing Academic Reading

the criticism that IELTS has a limited definition of reading for academic purposes also holds true for the TOEFL iBT.

It seems clear from the extant research that current IELTS tends to focus mainly on **local** lower-order reading skills at the sentence level (Owen 2016), at the expense of **global** higher-order skills such as creating a textual or an intertextual representation, which the research in Chapter 2 suggests is central to study at tertiary level. It is perhaps interesting to reflect on the fact that in the precursor to IELTS, the ELTS test (1980–1989), there were always a number of general questions, which required candidates to read all three passages in the test in order to produce the correct answer to the item, i.e. ELTS tested intertextual skills whereas its successor inexplicably does not. This is demonstrated in these examples from the social studies ELTS module:

- 53 Which Section, if any, deals with a national rather than an international issue?
 A Section 1
 B Section 2
 C Section 3
 D none of the Sections
- 54 Which Sections refer directly to the financial or economic aspects of the subject they are dealing with?A Sections 1 and 2 onlyB Sections 1 and 3 only
 - C Sections 1 and 3 only D Sections 1, 2 and 3
- 55 Which Sections deal in some way with the idea of *control* in human affairs?A Sections 1 and 2 onlyB Sections 1 and 3 onlyC Sections 2 and 3 only
 - D Sections 1. 2 and 3

The need for a multiple reading-into-writing component in IELTS will be explored more fully in Chapter 8 as it is the most situationally and interactionally authentic test task for testing academic reading ability i.e. potentially the most construct valid academic reading task.

Chalmers and Walkinshaw sound a warning note (2014:34–35) in their IELTS study as a result of finding most of the recorded strategy use at the local level in the reading modules:

... the implication is that participants may not have engaged in the degree of cognitive processing suggested by their IELTS sub-score. This is an important consideration for universities using IELTS to measure test-takers' ability to cope with a university-level academic reading load ... In
sum, the participants' focus was often limited to accurately responding to test questions rather than overall textual comprehension, echoing Moore et al.'s (2012) finding that IELTS Academic Reading engaged readers only at a "local-literal" level, involving comprehension of small units of text. So the participants' approach to the Reading test may not reflect Khalifa and Weir's (2009) cognitive processing model, instead resembling Guthrie's (1988) more pragmatic "information location" approach where "the emphasis is on extraction rather than recall, and on selective rather than exhaustive inspection" (p. 182) . . . The findings also raise questions about the validity of IELTS as a test of comprehensive academic reading ability. Most Australian universities require an IELTS 6.0 or higher for entry. Yet reading at university is commonly done to facilitate academic writing (Moore et al., 2012), which involves informational processing far beyond the level of propositional comprehension. If test-taking strategies enable participants to obtain higher scores that do not reflect their real level of reading comprehension skills, universities may need to re-evaluate IELTS as a viable entry pathway. Test construction is another potential issue: additional test task types may need to be incorporated to measure reading comprehension at a global-interpretive level.

Why the prevalence of items testing lower-order reading skills?

Almost two decades ago, Alderson (2000:131) examined the IELTS Academic Reading section and concluded that the reading construct is defined in terms of the 'ability to engage in a set of effective and efficient reading behaviours related to a variety of tasks'. He argues (2000:131, emphases added by authors) that IELTS:

Seeks to sample candidates' ability to perform a number of tasks, although it is not implied that these can be tested in isolation or independently of each other. Such abilities amount to the construct that *at least the original version of IELTS attempted to measure*:

- i identifying structure, content, sequence of event and procedures
- ii following instructions
- iii finding main ideas which the writer has attempted to make salient
- iv identifying the underlying theme or concept

v identifying ideas in a text, and relationships between them, e.g. probability, solution, cause, effect

vi identifying, distinguishing, and comparing facts, evidence, opinions, implications definitions and hypotheses

vii evaluating and challenging evidence

viii formulating an hypothesis from underlying theme, concept and evidence

ix reaching a conclusion, by relating supporting evidence to the main idea x drawing logical inference

(IELTS Specifications, December 1989)

The emphasis here, as one might expect, appears to be on higher-order reading skills in terms of our cognitive framework (see Table 4.3). The specifications for the IELTS revision (RIELTS) in 1993 tell a similar story:

The aspects of language competence for reading in the Revised IELTS (RIELTS) contain the following:

- a) Following and responding to instructions;
- b) Identifying content, and sequence of events and procedures;
- c) Recognising relationships between information items in text, e.g., sequence of events, cause/effect;
- d) Following overall organisation of text and its consequences for interpretation of message;
- e) Finding main idea;
- f) Identifying the underlying theme or concept;
- g) Locating and using factual information in texts;
- h) Re-ordering text information in order to accomplish tasks;
- i) Extracting topic/gist of whole or part of text;
- j) Recognising/completing summaries of whole or part of a text;
- k) Supplying inferences to complete textual gaps;
- 1) Applying text information to extra-textual tasks.

Almost 20 years on, as Alderson is perhaps alluding to in his aside in the last quote ('at least the original version of IELTS attempted to measure'), things have changed somewhat. The tasks laid out in the 2006 Item Writer Guidelines have become a lot more prosaic and the frequent reference to locating *specific information* contrasts sharply with the intentions of the Lancaster team under Alderson who developed the first IELTS academic test in 1989.

The 2006 Item Writer Guidelines spell out the stated aims for each item type (emphases added by authors to highlight the local nature of many items):

MCQ

These items test candidates' ability to a) *scan the text in order to find the relevant part*, and b) read this part in detail in order to understand one of a variety of text content features, e.g. opinion, attitude, tone, purpose, main idea, implication, exemplification etc. Choosing the correct option may depend *on understanding key phrases, or the relationship between a number of words and phrases, at intra- or inter-sentence level*, or beyond. Or it may depend on *understanding emotive lexis* and style, or on understanding global ideas in the text.

Matching information

Candidates need to use a range of reading skills including skimming and *scanning*, and *reading for detail*. Candidates may be asked to find a main idea, a theory, a reference, a summary, a comparison, etc.

Matching headings

Candidates need to use global reading skills to identify main ideas and themes in paragraphs/sections.

Matching people and statements

Candidates need to *scan the text for the options (e.g. names of researchers)* and read these areas of the text in detail *to identify key pieces of information*, which may be opinion-based.

Matching sentence endings

Candidates need to *scan the text to find the relevant part for each sentence beginning*. They need to read this part in detail and understand the main idea, which is then matched to a paraphrased option.

Classification

Candidates need to *scan the text for the categories and read around these in detail to identify key information*. They also need to group this information according to the categories provided.

Identifying Information (True/False/Not Given); Identifying Writer's Views/Claims (Yes/No/Not Given)

Type 1 (True/False/Not Given):

This task tests candidates' ability to a) *scan the text in order to find the relevant part*, and b) read this part in detail in order *to understand factual information*.

Type 2 (Yes/No/Not Given):

This task tests candidates' ability to a) *scan the text in order to find the relevant part*, and b) read this part in detail in order to understand opinionbased information/ideas.

Short-Answer Questions

Candidates need to use skimming and *scanning* skills to locate specific information in the text. They should use words provided in the questions to help them locate the relevant part of the text.

Sentence Completion

Candidates need to use skimming and scanning skills to *locate specific* information in the text. They should use words provided in the sentences to help them locate the relevant part of the text.

Summary Completion

Type 1:

Candidates need to use skimming and *scanning* skills to locate the part of the text that the summary is based on. (It is not necessary to state in the rubric that the summary is based on a certain part of the text unless there is a possibility that candidates will become confused.) *Candidates should*

Research and Practice in Assessing Academic Reading

predict the types of missing words and then read the relevant part of the text in order to identify them.

Type 2:

The testing aim is similar to that for the productive summary task (Type 1), but candidates are expected to select the correct answers from a set of options. This task type is particularly useful for testing the understanding of ideas in the text, rather than discrete information.

Note/Table/Flow-chart Completion and Diagram Labelling

Candidates need to use skimming and *scanning skills to locate specific information in the text. They should use words provided in the task to help them locate the correct answers.*

The latest version of the Item Writer Guidelines (2012) exhibits a similar prevalence of local items. It would be fair to say that around two thirds of the suggested focuses for the reading tasks listed on p.6 of the current Item Writer Guidelines are at the local level.

To try to understand this preponderance of items focusing on the local level, we need to examine what wider influences were at work, especially in UCLES (University of Cambridge Local Examinations Syndicate, now Cambridge Assessment English), when they took over responsibility for IELTS from the British Council in 1989. A more professional orientation to language testing took place with the arrival in UCLES of Peter Hargreaves (1987), and the setting up of an Evaluation Unit led by Michael Milanovic with Nick Saville (1989). By the mid-1990s, a new testing cadre had been recruited including Neil Jones and Lynda Taylor. Alerted by the less than positive findings of the Cambridge-TOEFL comparability study in the late 1980s (Bachman et al 1995), there was a growing realisation that Educational Testing Service (ETS) and its TOEFL examination possessed a more psychometrically sophisticated testing system than UCLES. UCLES from 1913 up to that point had focused on the content validity of its tests and their appropriateness for the language classroom, rather than ensuring these tests had a strong psychometric base (see Weir 2013a).

Lynda Taylor (personal communication, April 2017) suggests that under the leadership of Mike Milanovic, improving scoring validity became a priority for Cambridge English examinations in the 1990s and for IELTS as well when development came under Cambridge English control in the late 1980s. Mike Milanovic, in his interview for the British Council's involvement in assessment history project (Weir and O'Sullivan 2017:220), stressed the need for IELTS to match what he considered the psychometrically superior US TOEFL test:

^{...} we were constantly being compared to TOEFL, which had always been a very professionally produced exam so there was an important

comparator on the field that we could not ignore. ETS has essentially always been and remains an organisation that values research, recognises the importance of psychometrics, recognises the importance of measurement, recognises many of the key elements of assessment that really do matter . . .

Lower-order reading skills tests (see Khalifa and Weir 2009:Chapter 3), tested through MCQ, matching, and gap-filling items would provide the desired reliability (in terms of internal consistency estimates) as a result of their discreteness and the comparative ease of constructing multiple numbers of such items. It was far easier to attain the required target of 40 items for the IELTS Reading Module by including *local* items than to produce 40 items targeting higher-order reading skills. The number of items in a test heavily influences the reliability (internal consistency) coefficient that can be attained for that test *mutatis mutandis*, i.e. the more items a test contains the higher the reliability it is likely to achieve. There was thus a clear predisposition in favour of local items that could be more easily realised and result in a more impressive internal consistency coefficient, the 'industry gold standard' for reporting reliability.

Charge and Taylor (1997:377) report:

It has always been important to maintain adequate reliability in both the objectively and the subjectively marked modules of IELTS. A rigorous process of test production has produced Reading and Listening versions with an average Cronbach Alpha of 0.88, calculated from the performance of over 90,000 candidates on 13 Reading and Listening versions.

With global comprehension items, text-to-item ratios are often much lower, i.e. normally far fewer global items can be written in a single passage as compared to local items since the number of macro propositions/main ideas per text is limited as compared to the number of micro propositions. Furthermore, there can only ever be one overall gist question, viz one superordinate macro proposition in a text. Higher-order reading items are intrinsically harder and more time consuming to write, and more difficult and time consuming to improve in the item development process. Thus if we want to test higher-order skills we might have to be satisfied with fewer items overall or longer tests with more texts than is currently the case.

Because such global items (testing inferencing ability, mental model building or developing a text/multi-text-level representation) are more complex in processing terms than those testing lexical access, syntactic parsing and understanding propositions at the sentence level, the reported internal consistency statistics for tests containing such global items are sometimes lower than when restricting a test to measuring lower-order, more internally consistent items. Where there is a mixture of lower- and higher-order items overall,

Research and Practice in Assessing Academic Reading

internal consistency is often lowered and the global items exhibit lower values than the local.

Weir (2005b:32) raises a concern over the undue dependence on internal consistency in reading test development:

There is a further serious question which might be asked of internal consistency estimates which takes us back to the earlier reliability/validity debate. Would we want a high internal consistency estimate in all cases? Though one might be expected for a multiple-choice test of knowledge of structure or discrete lexical items in the 5,000-word list it might be rather naïve to assume that skills/strategies such as listening and reading are as unitary (see Urquhart and Weir 1998, Grabe and Stoller 2002 who argue for a partially divisible view of reading). If they are divisible, then high internal consistencies presumably would not be expected in the papers testing these skills.

There was clearly an intention to include global items originally in IELTS, as can be seen from the 1989 and 1993 specifications detailed above, but Lynda Taylor (personal communication, April 2017) recalls that it proved far more difficult to produce global items that met the technical standards required in the prevailing zeitgeist in UCLES at that time. TOEFL was the point of comparison, and the major competition/threat for IELTS, and as a result high internal consistency was seen as the *sine qua non*. It was felt that if IELTS was to grow and be sustainable, the reliability of the test had to be unequivocally clear for all to see.

We have examined the cognitive focuses of the current IELTS Academic Reading Module and shown how far the shift from global to local reading items has gone. Given the clear need for first year university students to understand the main ideas in a text and how these relate to each other within and across texts in writing their assignments, this preponderance of items focusing on lower-level reading skills must be a cause for some concern. There is a clear need to ensure that future IELTS tests at the global level include items that require creation of (i) a whole-text interpretation and (ii) intertextual representation (see Chapter 8 for some ideas on how the IELTS partners might meet these requirements). Ability to successfully cope with such critical academic reading activities would be valuable data for receiving institutions wishing to cater for any shortfall in key reading attributes in students post entry.

We next examine if there is anything else in the current IELTS that is acting as a brake on the level of the reading skills that are being measured and if there are any steps that can be taken in any future revision of IELTS to address this.

The limitations of certain test formats: What can be done?

At the risk of being labelled cynical, it might be conjectured that local item types are easier to construct and as a result are often favoured by item writers/ test developers in preference to global items. There is some anecdotal evidence in Green and Hawkey's study of IELTS item writers (2012) to support this. If item writers are left to decide for themselves what item types they will work on, there is always a possibility that an imbalance in the cognitive levels measured by items may result. Such possibilities can be effectively eliminated by tightening up on the procedures for item writing e.g. by specifying the test tasks/level of processing required and obliging item writers to provide evidence that their items are indeed testing higher-order reading skills. A differential fee for global and local item writing would further facilitate a move to requiring more global items. Owen (personal communication, August 2018) supplies further information:

Item types which are viewed as more difficult and time-consuming to write are remunerated at a higher level, regardless of the cognitive processes they activate. E.g. multiple choice are regarded as more difficult as they require distractors, each of which must be unique (i.e. wrong for a different reason), whereas T/F/NG are remunerated at a lower level as only the individual statements require justification. Moving to a model of paying participants according to cognitive processing of items may require extensive item-writer training.

Perhaps an even more pressing issue is raised by Weir et al (2012b) with regard to the limitations certain item types impose on what can be tested. They suggest that certain item types appear to provoke the use of certain processes; for example, gap-filling and information matching seem likely to focus on sentence-level comprehension and are inter-sentential at best. Other item types like MCQ and SAQ can lead to both higher- and/or lower-level processing.

Following on from our research into real-life reading activities in Chapter 2, we would argue that the aim in any future IELTS revision should be to focus on testing higher-order reading skills. Using Owen (2016) as our major informing source we will try and establish which of the current IELTS item types would seem to allow the testing of these higher-order skills and which result in mainly testing at the local level.

Before we start, it is worth establishing that there may be some variability in strategy use even with the same item type. Owen (2016:144) rightly cautions that:

No single item type in either IELTS or TOEFL required the use of an individual strategy for successful completion ... firstly, this is strong

evidence that test takers use a combination of strategies to complete each item, and that it is preferable to speak of relative importance in relation to strategic management of individual test items, rather than claiming an item requires one specific strategy for successful completion.

Owen (2016) looked at the various task types in IELTS and we summarise below his findings in relation to the processing levels they appear most likely to initiate. We also make reference to other critical commentary in the literature on the various methods of testing reading comprehension (Urquhart and Weir 1998, Alderson 2000, Enright et al 2000, Rosenfeld, Leung and Oltman 2001, Grabe and Stoller 2002, Cohen and Upton 2006, Rupp, Ferne and Choi 2006, Khalifa and Weir 2009, Weir 2013a).

MCQs

Owen's research (2016:327) suggests:

... multiple-choice items can be designed to access both higher- and lower-level cognitive processes. For lower-level items, test takers should not engage with the entire text if they are to answer these items quickly and efficiently. For multiple-choice items targeting higher-level processing, test takers need to engage with multiple parts of the text to select the key and eliminate distractors.

Weir (2013b:157-158) argues that:

Multiple-choice items are acknowledged to be an appropriate vehicle in large scale assessments for testing detailed understanding of the text. They are thought to allow more sophisticated elements of text content to be tested, e.g. opinion, inference, argument, in a more controlled way than is possible through open-ended formats.

But he adds the strong caveat that:

There is some concern, however, about the appropriateness of multiple choice questions for activating the higher level processing required in constructing an organised representation of the text . . . For example, an empirical study by Rupp, Ferne and Choi (2006:468–469) questions their value "as composite measures of higher-order reading comprehension"; i.e. their usefulness for assessing comprehension of the macro-structure of a situation model. They conclude (2006:469) that the format may involve the reader in "response processes that deviate significantly from those predicted by a model of reading comprehension in a non-testing context" and they hypothesise (2006:454) that: ". . . responding to MC reading comprehension questions on many standardised reading

comprehension tests is much more a problem-solving process relying heavily on verbal reasoning than a fluid process of integrating propositions to arrive at a connected mental representation of a text."

Field (2013:129) feels the use of MCQs 'imposes quite heavy cognitive demands which go well beyond those that would apply in a non-test context'. Weir (2013b:157–158) argues that the mental model, which would normally be created in reading a text, is:

... affected if candidates try to incorporate all the options provided in an item into an ongoing text representation. The processing that takes place in working out which option fits, and which does not, would bear little resemblance to the way we process texts for information in any of the types of reading we identified in our framework in Figure 2.3 above (see also Farr, Pritchard and Smitten 1990, Nevo 1989 and Rupp et al 2006 for informed research studies on the process of taking multiple-choice tests). However, in practice, the way the question is phrased and the way in which the candidate approaches the task will make a difference to the creation of the mental model . . . The Rupp et al (2006:468) study itself showed that "test-takers first tended to apply macro-level strategies in order to have an overall idea of what the given text and the related questions were about." There is some evidence too from classroom practice, which indicates that candidates are trained to read the text first then look at the items. The presence of exercises in published textbooks focusing on Cambridge English examinations, which train students to read the text before answering the questions, would support this. Also, the fact that Cambridge English presents the text first in careful reading tasks encourages the student to read the text before the questions.

The real value of MCQs may lie in testing inference and gist questions where using open-ended item types might result in a plethora of acceptable responses, thereby making the marking impracticable and/or unreliable. Field (personal communication, July 2017) adds:

I know I inveigh against the many iniquities perpetrated by MCQs but (apart from sharing the modality when used for reading) they also have the advantage that they can target a big variety of higher-level processes: discriminating between ideas – identifying main points – inferring points left unexpressed – interpreting speaker attitude – reporting logical connections – identifying a conclusion reached.

Identifying information (Yes/No/Not Given)

A series of statements relating to different parts of the text are presented. Candidates are required to determine whether these statements are i) given in the text ('yes'), ii) are contradicted by the text ('no') or iii) are unrelated to text content ('not given'). Owen (2016:195) describes how his data on these items evidences:

A high proportion of "lexical access" (fourteen instances) is consistent with strategic management of identifying key words in this item type. Evidence suggests that other frequent levels of processing for correct responses to this item type are "establishing propositional meaning at the clause" and "sentence" levels, with seven and eight instances emerging of each respectively...

Identifying writer's views/claims

Owen (2016) argues that this is a similar item type to 'identifying information'. Candidates are given sentences, each of which restates part of the text. They have to decide whether this statement is an accurate reflection of the stated content in the text ('yes'), contradicts the information in the text ('no'), or the information does not appear in the text ('not given'). The statements reflect claims made by the author or by named individuals within the text. The item type is more likely to be used with texts which present an argument or opposing views or claims. Owen (2016:210) notes:

Four examples of this item type were included in the test instrument (text 3, items 37–40). Only four strategies received ratings of "very high" or "high". These strategies suggest a common approach of identifying key nouns in item stems and identifying relevant parts of the text containing those words. The most commonly observed behaviour was close, careful reading of the text. Participants moved between the text and items frequently, like the previous item type. However, for "identifying the writer's views/claims", participants displayed a greater tendency to use higher level processing . . . Overall, establishing propositional meaning at the clause and sentence levels, and establishing a mental model of part of the text were the most frequently used codes.

Matching information

Matching is a variant on multiple-choice tests and it can take a variety of forms, all of which can be scored objectively. Owen (2016:217) details how:

Items 14-18 are "matching information" items, which require test takers to identify which of a series of sentences reflect accurate statements that are made in the text. In this test, participants are asked to match common beliefs to those that are cited in the text and eliminate those that are not. Participants must select five out of eleven options ... The majority of

the cognitive processes observed for this item type were lower-level processes. Lexical access and syntactic parsing were rated "very high", with establishing propositional meaning at the clause and sentence levels rated as "high".

Matching headings

Owen describes how (2016:226):

Matching heading items present test takers with a series of noun phrases (only title 2 is a clause containing a verb), which they are required to link to the main purpose of each paragraph ... The test developers clearly intend that this item type should encourage test takers to engage with each paragraph carefully, in order to identify the main purpose of each. The main purpose may be included in the first sentence of each paragraph, although this is not universally true, requiring test takers to engage more closely to distinguish between information which is presented as a main argument and information which acts to support that argument. There is strong evidence from the engagement of the two participants that this item type tests a range of both higher and lower cognitive processes ... To link these titles to paragraphs requires a general understanding of the topic of each paragraph and an ability to summarise the main ideas ...

Because of the greater number of options the candidate has to choose from, matching is less prone to successful guessing than MCQs. *Matching headings* with careful construction can cover the important processes in reading a text: main ideas, gist and at higher levels text representation in careful reading and search reading for main ideas in expeditious tasks. Matching is thus a flexible and useful format, enabling the coverage of higher-order reading processes described in our model in Figure 3.1. Everett and Colman (1999:31) commented on this item type in their research study on IELTS preparation materials:

Similarly, the wording of the headings in the many heading-matching tasks may significantly raise the level of complexity of the task. The brevity of many headings, often including nominalisations and passive structures, can be particularly demanding . . . Some have been described as difficult, others were identified as potentially confusing . . . Clearly, there are a number of factors involved in determining the complexity of heading-matching tasks, perhaps the most important being the interpretability of the actual headings. Once again, trialing of these complex tasks would help to identify those that learners may find too difficult, or perhaps confusing.

Sentence completion

Owen (2016:234) explains that candidates have to complete sentences with words from the text to form an accurate paraphrase of a sentence in the text. Sentences relate to detailed points within a paragraph, not the broader argument. Candidates have to search the text for relevant bits to complete each item. Owen (2016:234) records:

One observable strategy for this item type was rated "very high"; "careful local reading". Three other strategies were rated as "high"; returning to the question for clarification (rereads question and/or options); searching for key word/phrase (text) and marking or noting a key noun phrase in the text during careful reading. Item completion procedures then followed a particular pattern. Participants identified key words in the item stem, held them in their working memory and attempted to find identical or similar items in the text . . .

Cognitive processing ratios indicated that this item type targeted lower-level processes. The most common process was lexical access, reflecting the importance of using key terms to identify relevant parts of the text. Once identified, engagement was local (at the sentence level) as the items restated specific parts of the text.

Kintsch and Yarbrough (1982) suggest cloze tests are not sensitive to macro processes but related only to micro processes. Markham's (1985) study showed that cloze procedure does not provide an adequate assessment of inter-sentential comprehension (the ability to build an accurate mental model) which led him to conclude that 'cloze procedure may not yield a valid and reliable assessment of global comprehension in the second language context' (Markham 1985:423). Kobayashi (1995) provided evidence that cloze tests are likely to measure local comprehension whereas open-ended questions can more easily target global comprehension (see also Alderson 1978). Alderson (2000:208-209) claims: "... many cloze items, for example, are not constrained by long range discourse, but by the immediately adjacent sentence constituents, or even the preceding two or three words. Such items will not measure sensitivity to discourse beyond the sentence or even the phrase". In other words, whereas other constructed formats such as SAQs can measure the reader's global comprehension of main ideas of the text and text structure, cloze tests or selective deletion gap-filling items do not necessarily reflect the reader's ability to comprehend beyond the sentence.

Single-item gap filling would appear to measure only a limited part of what might constitute reading proficiency in terms of the processing model we presented in Figure 3.1, namely lexical access, syntactic parsing skills and propositional-level meaning. It does not usually require the higher level of processing involved in text-level reading or ongoing text representation (see, however, Bensoussan and Ramraz (1984) who proposed the deletion of

phrases to try to test understanding of the functions of sentences and the structure of the text as a whole).

These micro-linguistic tests may still have an important role to play in on entry language tests used by universities to screen students after they enrol on their courses. John Read (2015) emphasises the need for testing all students entering a university in the full knowledge that some of them may not require remedial assistance through in-sessional programmes. It is clearly impractical to test all students with a full-scale diagnostic proficiency test. However, a two-tier testing system where everybody sits a general, quick, efficient micro-linguistic test first, which can identify those who will not require further testing or support, seems a practical solution. A number of studies (Weir 1983, Alderson and Clapham 1992, Read 2015) show that such short lexico-grammatical tests can correlate highly with much lengthier full-blown EAP tests such as IELTS and TEEP. Read describes how such restricted tests are used successfully by the University of Auckland in New Zealand to identify those who are already sufficiently proficient to cope with their courses of study without further assistance. They are obviously not able to serve a diagnostic function though and to target problems with reading types or particular skills/strategies, a further full-blown diagnostic reading test is required.

Summary completion

Owen (2016:244, 249) details how:

Participants generally use the summary to identify key words and use these to identify the relevant part of the text. They then read this part carefully, returning to the item stems to gain an understanding of how the summary reflects the text. Key words are underlined in both the summary and the text to facilitate this process. This indicates that the majority of processing will occur within sentences ... the item type targets both high- and low-level processes. Participants relied upon key words to identify relevant parts of the text, and lexical features of the summary and text to identify answers. However, in participants' explanations, there were clear instances in which participants needed to gather evidence from more than one sentence in order to answer confidently due to lexical similarity (items 8 and 10) and understanding the progression of the summary (items 9–10) ... summary or diagram completion items may require ... higher–level processing in order to integrate information across multiple sentences and paragraphs.

Alderson (2000:240–242) seems to be in favour of the gapped summary:

Their task is to restore the missing words, which can only be restored if students have both read and understood the main ideas of the original text. It should, of course, not be possible to complete the gaps without having read the actual text ... Alderson et al (1995:61) conclude that such tests "are difficult to write, and need much pretesting, but can eventually work well and are easier to mark".

Everett and Colman (1999:30) commented on this item type in their research study on IELTS preparation materials:

... many of the summary tasks require an overview of the whole text. An interesting summary task is given in Practice Test 3 in Gibson, Rusek and Swan (1996, p. 67), in which the final paragraph of the reading text has been deleted from the passage and adapted as a reading cloze task. The task demands comprehension of the whole text.

Some IELTS item types were not covered by Owen's doctoral study and we turn to these now.

Short-answer questions (SAQs)

SAQs are generically those that require the candidates to write down answers in the spaces provided on the question paper. This serves to limit the length of the response, even if the number of words required is not specified. The questions set in this format can potentially cover the important information in a text (overall gist, main ideas and important details) as well as an understanding of the structures and lexis that convey this. Activities such as inference, recognition of a sequence, comparison and establishing the main idea of a text, requiring the relating of sentences in a text to other items which may be some distance away in the text, could also be accommodated in this format. In addition to careful reading at both global and local levels, this technique lends itself to testing skimming for gist, search reading for main ideas, scanning for specific information – expeditious reading types. Alderson (2000:227) argues:

The justification for this technique is that it is possible to interpret students' responses to see if they have really understood, whereas on multiple-choice items students give no justification for the answer they have selected and may have chosen one by eliminating others.

In SAQs, the answer has to be sought rather than being provided for the student as in multiple-choice; therefore if a student gets the answer right, one can be more certain that this has not occurred for reasons other than comprehension of the text. In the MCQ format we do not have this guarantee as the option itself rather than the text may have suggested the answer or appeared superior to the other available alternatives. In SAQs, answers need to be worked out from the passage and should not be already known through

existing knowledge, or easily arrived at by matching wording from the question with wording in the text.

Current thinking is that the number of acceptable answers to a question should be limited so that it is possible to give fairly precise instructions to the examiners who mark them (Alderson 2000, Khalifa and Weir 2009, Weir 1993). The mark scheme should allow for the range of semantically acceptable answers. Mechanical accuracy criteria (grammar, spelling, punctuation) would not normally feature in the scoring system as this affects the accuracy of the measurement of the reading construct.

The main disadvantage of this technique is that it involves the candidate in writing, and there is some concern, largely anecdotal, that this interferes with the measurement of the intended construct. The guiding principle for this test format is to keep the answers brief and to reduce writing to a minimum to avoid possible contamination from students having to write answers out in full and for reasons of marker/scoring consistency. Care can be taken in the setting of items to limit the range of possible acceptable responses, for example, by asking for verbatim words from the text as answers. Though given the purposes for reading in the academic context we identified above, this may not be seen as a problem as the candidate will have to do both in their course of study.

In those cases where there is more debate over the acceptability of an answer, for example, in questions requiring inferencing skills, there is a possibility that the variability of acceptable answers might lead to marker unreliability. If such concerns are allowed to unduly influence item choice, issues concerning adequate coverage of representative cognitive processing levels may arise. For example, in the present-day IELTS, item writers are advised to focus SAQs on factual information as otherwise the items end up too open to be workable. This of course is likely to limit the extent of coverage of the processing stages outlined in our model in Figure 3.1 and raise questions relating to a test's cognitive validity (similar problems occur in the testing of listening). For a test to be considered cognitively valid, the cognitive processes required to complete the language tasks must be an adequate resemblance of the cognitive processes a candidate would normally employ in non-test conditions, and be sufficiently comprehensive to be generalisable to that real-world behaviour.

Information transfer

Information transfer involves the transfer of information from one type of layout to another, e.g. from connected text to a table, flowchart, diagram, gapped notes or graph. The new format may still have a verbal element (e.g. labels on a diagram, words in a table, flowchart completion) but additional meaning is provided by the visual aspect. The candidate has to construct the response from the text provided. Cohen and Upton (2006:8) describe an interesting computerised version of this task type at the text-representation level:

For the schematic table, test takers must click and move sentences or phrases into a table to complete a schematic representation [of the passage]. A correctly completed table should reveal an integrated mental model of how the two dimensions fit together conceptually based on the information in the text.

Alderson (2000:248) is concerned that these tests 'may be cognitively or culturally biased' e.g. a candidate may not be familiar 'with tabular presentation of statistical data' and feel they are difficult, thereby raising the issue of bias and unintended variance. The very act of transferring data may involve a level of complexity not there in the original text. Not all students will be faced with such data presentation in their course of studies. Alderson points to further issues if the graphic text has information removed from it, when originally there to be understood in conjunction with the text. Owen (2016:367, 368) concludes:

Some of the higher-level processes in Khalifa and Weir's model are under-represented in both IELTS and TOEFL, including inferential reasoning and forming a text-level representation . . . positive evidence of this level of processing in IELTS is sparse.

The prevalence of items testing at the lower levels of reading in current IELTS is indicative that a historical shift in the focuses of the test has occurred, a direction compounded by the item formats currently used. Serious thought should be given to reducing the number of items/tasks that appear to initiate only lower-order processing (see Table 4.5) to ensure IELTS is only testing global meaning beyond the sentence; a *sine qua non* of academic reading. The current formats testing lower-order reading processes need to be replaced by item types that are likely to require higher-order reading processes.

Some possible new item types for testing higherorder reading processes

A) Cohen and Upton (2006:8) offer a useful example of a matching task at the advanced level in the task specifications for iBT TOEFL:

For the prose summary, test takers are asked to "complete a summary of a text, one or two sentences of which are provided" by selecting three additional sentences from a list of six that express the most important ideas in the passage . . . Distractors include ideas that either are not presented in the passage or are deemed as minor ideas.

Item type	Skill tested (IELTS 2017) according to official websites with the descriptions provided by the partnership (based on Owen 2016)	Level of processing (Owen 2016	Include in future)versions of IELTS
MCQ	A wide range of reading skills, including detailed understanding of specific points to an overall understanding of the main points of the text. Test takers are required to choose the best answer from four alternatives (A, B, C or D), or the best two answers from five alternatives (A, B, C, D or E), or the best three answers from seven alternatives (A, B, C, D, E, F or G). Test takers write the letter of the answer they have chosen on the answer sheet.	Higher and lower	YES
Identifying information T/F/NG	Test takers will be given a number of statements and asked: 'Do the following statements agree with the information in the text?' They are then required to write 'true', 'false' or 'not given' in the boxes on their answer sheets. Assesses the test takers' ability to recognise particular points of information conveyed in the text. It can thus be used with more factual texts.	Lower	NO
Identifying writer's views/ alaims T/E/NC	Assesses the test takers' ability to recognise opinions or ideas, and so it is often used with discussive or accumentative texts	Higher and lower	YES
Matching information	Ability to scan a text in order to find specific information rather than main idea (see <i>Matching</i> <i>headings</i>). Candidates may have to find: specific details, an example, reason, description, comparison, summary or explanation.	Mainly lower	NO
Matching features (Classification)	Candidates are required to match pieces of information to categories, e.g., events to dates. Candidates need to scan the text for the categories and read around these in detail to identify key information. They also need to group this information according to the categories provided	Lower	NO
Matching headings	Ability to recognise the main idea or theme in the paragraphs or sections of a text, and to distinguish main ideas from supporting ones	Higher and lower	YES
Sentence completion	Test takers complete sentences in a given number of words taken from the text. They must write their answers on the answer sheet. The instructions will make it clear how many words/numbers test takers should use in their answers, e.g. 'NO MORE THAN THREE WORDS AND/OR A NUMBER from the passage', 'ONE WORD ONLY' or 'NO MORE THAN TWO WORDS'. Assesses ability to locate detail/specific information in a text.	Lower	NO

 Table 4.5: Summary table of IELTS Reading item types + claimed abilities

 being tested + processing level

Item type	Skill tested (IELTS 2017) according to official websites with the descriptions provided by the partnership (based on Owen 2016)	Level of processing (Owen 2016	Include in future)versions of IELTS
Matching sentence endings	Test takers are given the first half of a sentence based on the text and asked to choose the best way to complete it from a list of possible options. Assesses the test takers' ability to understand the main ideas within a sentence.	Lower	NO
Summary/note/ table/ flowchart completion	Test takers are given a summary of a section of the text, and are required to complete it with information drawn from the text. The summary will usually be of only one part of the passage rather than the whole. The given information may be in the form of: several connected sentences of text (referred to as a summary), several notes (referred to as notes), a table with some of its cells empty or partially empty (referred to as a table), a series of boxes or steps linked by arrows to show a sequence of events, with some of the boxes or steps empty or partially empty (referred to as a flowchart). Assesses ability to understand details and/or the main idea of a part of the text. Because this task type often relates to precise factual information, it is often used with descriptive texts.	Higher and lower	Maybe
SAQs	Questions usually relate to factual information about details in the text. Assesses ability to use scanning skills to locate specific information in the text. They should use words provided in the questions to help them locate the relevant part of the text. The text needs to contain plenty of factual information and detail and words that carry significant meaning, especially nouns/noun phrases. (Other exam boards use SAQs to test higher-	Lower	NO YES
	order mental model building and text-level representation but argumentative text types and answers longer than three words are necessary)	-	
Diagram label completion	Test takers are required to complete labels on a diagram, which relates to a description contained in the text. The instructions will make it clear how many words/numbers test takers should use in their answers. This task type is often used with texts describing processes or with descriptive texts. Assesses ability to understand a detailed description, and to relate it to information presented in the form of a diagram.	Lower	NO

B) In the creation of the British Council International Language Assessment (ILA) test, Weir and his colleagues employed a variant on this idea, as shown in the test item instructions below:

Only six of the statements (A–H) below are true according to the passage.

Choose the six true statements and put them in the correct order as they appear in the passage.

Write your answers on the answer paper. The first one is done for you.

C) Matching paragraphs with a gapped text, as employed in the upper levels of Cambridge English Qualifications, could be a further useful variant for testing academic reading.

A gapped-text task consists of one text from which a number of sentences (e.g. at A2 First), or paragraphs (e.g. at C1 Advanced and C2 Proficiency), have been removed and placed in jumbled order after the text together with a further sentence or paragraph, which does not fit in any of the gaps and functions as an additional distractor. Candidates are required to decide from where in the text each sentence or paragraph has been removed. Each sentence or paragraph may only be used once.

There is a strong argument for the use of such gapped texts/tasks as a response method, especially in terms of placing more demands on cognitive processing of a text at higher levels, in order to distinguish reading ability at these levels from that at lower levels, where other forms of matching may be used. In C1 Advanced, for example, the reader needs to understand the whole text in order to be sure of having completed the gapped-text task correctly. The testing focuses of text structure, text cohesion and coherence require the reader to select an option which fits the text both before and after the gap. This means that it should fit not only the immediate co-text but also fit so that the text after the gap follows on smoothly. Readers need to identify not only a wide range of linguistic devices which mark the logical and cohesive development of a text, but also to understand the development of ideas, opinion and events (over the whole text) rather than the recognition of individual words (and phrases). Finding which paragraph fits into which gap in a given text may require the reader to understand how the text develops from start to finish rather than just the section of text which occurs before and after the particular gap. Again, it remains for empirical research to provide evidence to support this contention but the argument is intuitively satisfying.

- D) Alderson (2000:236–239) suggests using a multiple-choice summary of a text where the candidate has to choose the best summary out of the answers provided. This would avoid the possibility of writing interfering with the measurement of reading.
- E) Another possibility can also be seen in the British Council's ILA test, where a selection of mini-texts each related to a single topic covered in the different paragraphs of the main reading text are provided. The candidate has to match each mini-text with the most appropriate paragraph in the reading passage.

Endnote

It is self-evident from the research reported in Chapter 2 that the main purpose for reading in academic life is to extract information from texts for use in written assignments on courses of study. The formats above that focus on global comprehension should certainly reflect this in terms of the cognitive processing involved. Alderson (2000:49) is still concerned, however, that:

... the only purpose we typically give students for their reading is to answer our questions, to demonstrate their understanding or lack of it ... readers do not usually answer somebody else's questions: they generate and answer their own.

This raises the issue of whether we should give the responsibility back to the student to decide on what is important information to extract from a text. This inevitably leads us in the direction of reading-into-writing tasks. Reading-into-writing activities are well supported in the current research literature on writing assessment (Grabe and Stoller 2002:14) and Pollitt and Taylor (2006) make a convincing argument for this type of task as do Hughes (1989) and Wallace (1997). Khalifa and Weir (2009) describe how they have been used in high-stakes writing tests around the world, for example, from 1989 to 1995 in IELTS and before that from 1980 to 1989 in ELTS, more recently in iBT TOEFL, in the GEPT in Taiwan since 2001, and since the 1980s in TEEP (see Weir 1983), and currently in the Cambridge English C1 Advanced and C2 Proficiency in the revised Writing paper.

The three research studies we reported earlier in Chapter 2 provide support for supplying input in writing tests where provision of stimulus texts reflects the real-life situation as in the writing of university assignments. Likewise, in our theoretical model, the highest level of processing suitable for advanced students is where they have to integrate information across texts to develop a combined representation of the texts they have read. Summary or an integrated reading-into-writing activity would seem to be the most appropriate techniques for measuring such intertextual reading ability. Charge and Taylor (1997:376), in making a case for removing the former link in IELTS between the Reading and Writing papers, argue:

... monitoring of candidates' writing performance suggests that the extent to which candidates exploited the reading input varied considerably. Some candidates drew heavily on the written content of the reading texts, apparently treating the writing task as a measure of their reading ability...

Twenty years on, following the research into reading and writing examined in Chapter 2, this seems more of a positive finding than a negative. Wallace (1997:370) took issue with Charge and Taylor and argued:

This paper queries the removal of a link between reading and writing tasks in the International English Language Testing System (IELTS) examinations since April 1995 on two grounds: that it is prejudicial to the success of some students from a non-English speaking background, whose first language cultures may not provide them with appropriate schemata to be able to perform the writing tasks effectively; and that it is unrealistic in terms of the measurement of important study/linguistic skills required in university study - exactly what this examination was established to assess ... The former IELTS test design seemed much more equitable for students worldwide, especially when compared with its TOEFL equivalent, the Test of Written English (TWE). The latter consists only of a written essay, with no link to any other part of the examination, since it is designed specifically as an optional add-on, and administered as a separate entity. The previous IELTS format, on the other hand, at least guaranteed all students some elements of background input that would be useful for performance of the written task.

Rea-Dickins, Kiely and Yu (2007:42–43) conclude in their IELTS research study on student identity, learning and progression:

... a key capacity in postgraduate study is what we label reading-to-writing. This capacity is not captured in either the reading or writing subtests (and associated test preparation processes), a lacuna which may be considered to weaken the IELTS claim that results are a statement about readiness for academic study.

Reading-into-writing tasks will be discussed fully in Chapter 8 as they clearly represent a situationally and interactionally authentic test task for testing academic reading ability, i.e. they are construct valid tasks for this purpose. It may well be time for those providing academic reading tests to revisit the use of integrated reading-into-writing tasks to better represent the higher-order processing in reading that is the hallmark of academic reading. It is perfectly

plausible to score such tasks from a reading-only perspective according to the relevance and adequacy of the content selected, and the connections made between ideas from the text and their organisation, though of course the availability of a construct valid third piece of writing would be an added bonus.

In Chapter 4 we have considered the cognitive dimensions of the reading task itself; in Chapter 5 we will make reference to the contextual parameters under which the tasks are performed (see Chapter 3 for full explication of these).

5 Contextual validity parameters

To take part in a communicative event is to produce and/or comprehend discourse in the context of a situation and under the performance conditions that obtain it. It is the purpose of a proficiency test to assess whether or not candidates are indeed capable of participating in typical communication events from the specified communication situation(s). Kelly (1978)

Introduction

In this chapter we will make reference to the contextual parameters under which the tasks are performed. We will draw on the contextual parameters taxonomy established in Chapter 3 as being most likely to have an impact on text complexity in reading tests. The interest is in those performance conditions that are likely to influence test task performance when reading activities take place. Are the contextual characteristics of the test task an adequate and comprehensive representation of those that would be normally encountered in academic reading?

We will examine how IELTS has attempted to calibrate test input along a number of dimensions. For example, the length of a text, the length of individual sentences in the text, and the frequency of the vocabulary encountered in the text will all affect ease of reading. Both individually and in combination, contextual parameters are likely to impact on the cognitive demands imposed upon the reader; for example, a text with high-frequency lexis and shorter sentences is likely to be easier to process than a text of the same length on the same topic with a large number of low-frequency lexical items and longer sentences. All other things being equal, a shorter text is likely to be easier to process than a significantly longer text.

Leaving text length and subject specificity aside for the moment, it is now generally agreed that the characteristics of texts employed in a test of reading comprehension for academic purposes ought to reflect as many of the relevant characteristics of the target reading activities as is possible given the situational constraints. Alderson (2000:205) makes the point that in considering task type for reading tests we must always have a clear idea of the **appropriate contextual dimensions** of the texts we are going to use. He suggests it is necessary to base the selection of texts and tasks on what the reader would normally do with the text during and after reading. Such full situational and

interactional authenticity (Bachman 1990), however, is clearly unrealistic for an EAP reading test.

There are several reasons why activities used in high-stakes tests will necessarily differ from processing undergraduate textbooks or articles in professional journals. Academic textbooks are very much longer than any IELTS Reading text can be and different time constraints apply to reading them. Alderson (2000:180) is well aware of this and he acknowledges that reading an article on a course might take up to four hours even if read only once, and that books would take considerably longer, perhaps from two to three weeks. In stark contrast, tests will need to capture a range of common core purposes, reading processes and skills within a strictly limited time frame for reasons of practicality.

Furthermore, as we saw in Chapter 1, the content domain of academic texts will vary considerably (Gardner and Nesi 2013, Nesi and Gardner 2006, Nesi et al 2008). It is not currently possible to accommodate all 13 major genres identified by Nesi and Gardner in one reading test, let alone, without substantial resources, repeatedly produce 13 different equivalent forms that cater for each genre. Given there need to be dozens of different live versions of IELTS for security reasons (a logistical challenge in itself), tailoring texts to suit the subject fields of the candidate is problematic until automation of text selection, test development, test delivery and test marking becomes a reality. We look at this issue of specificity in more detail in the section 'Can one size fit all?' below.

Green et al (2010:193) argue that it is important that the cognitive load placed on students by a text in the reading process should, as far as possible, reflect the load imposed by the texts they will encounter when they begin their studies (with obvious caveats in respect of text length and subject specificity). It is thus useful to establish the parameters of texts students are exposed to in their first year of university as a set of benchmarks.

Focus and methodology

We will review major quantitative and qualitative studies which examined the contextual features of IELTS and some other academic reading test texts. Green et al (2010) report on a text analysis study of 42 passages extracted from 14 core undergraduate textbooks at the University of Bedfordshire (UoB) and their comparability with 42 texts from 14 IELTS Reading tests supplied by Cambridge ESOL (as Cambridge English was known at that time) and Cambridge University Press. Green et al note (2010:197):

The test materials were taken from the Cambridge Examination Papers series (UCLES, 2000, 2002, 2005, 2006, 2007) and from a specimen materials pack provided by Cambridge ESOL (UCLES, 2003). Although

these tests have not been used in live administrations, the material is intended to be representative of the content of the live test. It is produced by trained item writers working to the IELTS specifications and is piloted to ensure that it is at the appropriate level of difficulty.

Working closely with the University of Bedfordshire (then Luton) library staff, they established core first year undergraduate texts in each of the areas where large numbers (3,000+) of international students were studying in the UK according to the most recent Higher Education Statistics Agency (HESA) student record data (2004/5) at the time of the study. The courses taught at the UoB in these high-density areas are shown in Table 5.1.

Table 5.1: Courses taught by the UoB

Advertising, Marketing and Public Relations Biology and Biomedical Sciences Business and Finance Computing and Information Systems Criminology Education Studies Healthcare (Nursing and Midwifery) Human Resource Management Language and Communication (English as a Foreign Language, EFL and Teaching English as a Foreign Language, TEFL) Law Leisure, Tourism and Sports Management Media Arts Psychology Social Sciences and Social Work

The selection of the core undergraduate texts in these areas was made on the basis of:

- those books which had had the most reservations made for them in the last three years and in particular the current year
- those books which were taken out the most in the current academic year
- confirmation by course leaders of key books for each area
- the books students considered to be the most important (as established through the pilot questionnaire and direct enquiry, for more information see Green et al 2010).

As a result, 42 samples of academic text were collected to compare with the 42 IELTS texts. These comprised three extracts from each of the 14 different textbooks – passages extracted at random from the opening chapter, the middle and the concluding chapter. These are core texts that undergraduate

students are expected to get to grips with during their studies at the UoB. The length of extracts (targeted to be between 500 and 1,500 words) aimed to correspond broadly to the length of the texts included in the IELTS Academic Reading test. These sections were selected as self-contained passages that could be understood as coherent standalone texts. Their concern (Green et al 2010:192) was with:

... how closely IELTS Academic Reading texts resemble the texts that first year undergraduates most need to read and understand on arrival at our university – the core textbooks used in their first-year courses. These are the key texts that students will need to be able to follow on arrival at university, before they have had much chance to improve their language skills. The books employed in this study are all established standard textbooks, widely used across British universities: all but one is currently in at least a second edition, with one now in its 15th edition.

The IELTS texts (and the 42 extracts from UoB academic texts) were subjected to a variety of quantitative and qualitative analyses. For example, measures of the quantitative features were obtained through the Web VocabProfile available at www.lextutor.ca (Cobb (2003) supplemented by analysis through WordSmith Tools (Scott 2006), Coh-Metrix Version 2 (Graesser et al 2004, McNamara, Louwerse, Cai and Graesser 2005) and text analysis tools packaged with Microsoft Word for Windows.

Measures of vocabulary included word length (number of characters/ word), type-token ratio (TTR), lexical density, academic words and word frequency levels. For the estimation of grammatical complexity a range of the quantitative measures available through Coh-Metrix was chosen. The Coh-Metrix analysis suite included: average number of words/sentence; the number of modifiers per noun phrase, which concerns the occurrence of complex noun phrases (these being a recognised feature of academic text); and the mean number of words before the main verb in sentences (structurally opaque texts tending to have proportionally more high-order syntactic constituents and greater numbers of words before the main verb). Readability statistics (Flesch Reading Ease and Flesch-Kincaid Grade Level) were calculated using Microsoft Word - both measures being based on the relative numbers of syllables, words and sentences found in a text. Flesch Reading Ease scores range from 0 to 100 with lower scores reflecting more challenging texts. A score below 50 is said to require college-level reading skills. The Flesch-Kincaid Grade Level is based on the US school system, with 12 representing the final year of high school and 13 to 16 the college level. The Crossley et al (2008) readability formula, a potential alternative to traditional readability measures for L2 readers, was also considered.

For the qualitative analyses, Weir et al (2012a) report on how two judges rated the IELTS texts and the undergraduate texts on six criteria. The two

expert judges, with doctorates and experience of teaching and test development in the area of academic literacy, employed Likert scales and categorisation tools to evaluate the texts. In investigating *discourse mode* they included genre (or text source), rhetorical task, pattern of exposition and rhetorical organisation. Each judge independently assigned each text to one of the following genres:

- textbook
- magazine and newspaper article
- research/academic journal article
- report.

The judges also identified the subject area with which each text appeared most closely associated, using the HESA classification of courses of study. Each text was classified by the two judges according to the following discoursal features:

Rhetorical task

- Exposition
- Argumentation/persuasion/evaluation
- · Historical biographical/autobiographical narrative

Pattern of exposition

- Definition/description/elaboration
- Illustration
- Classification
- Comparison/contrast
- Cause and effect
- Problem/solution
- Justify

The two judges also used five-point Likert scales to make a subjective evaluation of the texts on the following features:

- *Rhetorical organisation (1 explicit to 5 not explicit).* This is intended to reflect the ease or difficulty with which the overall propositional pattern of the text is likely to be understood by the reader.
- *Grammatical complexity (1 mainly simple sentences to 5 mainly complex sentences).*
- *Cohesion (1 explicit to 5 not explicit).* An evaluation of the extent to which relations between the ideas were explicitly marked through reference, conjunctions and connectors.
- Content knowledge
 - Subject specificity (1 general to 5 specific). This involved an evaluation of the frequency of technical vocabulary and the extent to which terms were glossed in the text for the general reader.

- *Nature of information (1 concrete to 5 abstract).* An evaluation of the extent to which the text was concerned with concrete observable phenomena.
- *Cultural specificity (1 culture neutral to 5 culture specific).* This involved an evaluation of the frequency of culture-specific content as set out in the literature review above and the extent to which culturally specific references or examples were explained to the general reader.

Quantitative studies

The usefulness of the parameters reported for IELTS and other reading tests below were each discussed individually in Chapter 3. They were selected because they had been empirically shown to affect the complexity of a text.

Word count

Nuttall (1996) argues that a long text is required for candidates to skim for main ideas, scan for specific information, make relevant judgements and distinguish between main points and subsidiary details. Khalifa and Weir (2009) compared the lengths of texts used in the various Cambridge ESOL General English examinations over a 10-year period. Noticeably, a number of longer texts have been used in higher-level tests, presumably in order to increase the range of language candidates are exposed to or to encourage expeditious reading skills. There is a fairly clear increase in the amount of text students are exposed to at the C1 and C2 levels of the CEFR (overall number of words 3,000, maximum for any single text 1,100) as against the lower levels of the CEFR (A2–B2).

In IELTS, there are three different academic domain passages to read, each with accompanying questions. Officially, candidates have to read 2,150–2,750 words in total. Green et al (2010) analysed 42 texts making up 14 IELTS Reading tests. The passages in their study contained 854 words on average (maximum 1,063 words, minimum 589 words).

Sentence length

Average sentence length is often considered an approximate indication of text complexity. Khalifa and Weir (2009) describe how sentence length in the Cambridge ESOL reading examinations increases according to the level of the examination although again there seems to be considerable variation in the lengths of sentences featuring in the tests even at the same level. They argue that attention to this index might ensure greater homogeneity between the texts used at a particular level.

Khalifa and Weir (2009) examined average sentence length and range of sentence lengths in a small corpus of Cambridge ESOL texts (143 texts in

total). On average, the Certificate in Advanced English (CAE) (C1) texts have 18.6 words per sentence and the Certificate of Proficiency in English (CPE) (C2) texts 19.6 words. Green et al (2010) report the sentence length of the undergraduate and IELTS texts they analysed and found that of IELTS to be marginally longer, with undergraduates at 21.47 words per sentence and IELTS at 21.89 words per sentence.

Number of modifiers per noun phrase

Green et al (2010) report the number of modifiers per noun phrase of the undergraduate and IELTS texts they analysed and found them to be very similar, with undergraduates at 0.95 modifiers per noun phrase, and IELTS at 0.98 modifiers per noun phrase.

Number of words before the main verb

Green et al (2010) calculated the number of words before the main verb in the undergraduate and IELTS texts they analysed and found the figure for IELTS to be slightly higher, with undergraduate texts at 4.59 words before the main verb and IELTS at 5.48 words before the main verb.

Readability indices

A gradation in grammatical complexity can be seen in the average Flesch Reading Ease score and Flesch-Kincaid Grade Level estimates obtained for Cambridge ESOL General English examination texts (Khalifa and Weir 2009) (see Gervasi and Ambriola 2002, Klare 1984, Masi 2002 for discussion of the use of these formulae). Though often criticised as inadequate indices of text difficulty in themselves, these formulae still form the basic tools in most detailed analyses of textual complexity (Masi 2002). These two estimates of text complexity are largely based on average number of words in a sentence and average number of syllables per word. The low estimates for texts at CAE and CPE (see Table 5.2) might be a cause for some concern as, according to the CEFR, students at these levels should be capable of processing undergraduate-level texts which Weir et al (2012a) found to be at around 13.5 in terms of Flesch-Kincaid estimates.

 Table 5.2: Readability values for Cambridge ESOL exam texts at C1 and C2 levels

Cambridge ESOL exam level	Flesch Reading Ease score	Flesch-Kincaid Grade Level	Flesch-Kincaid range
CAE (C1)	58.4	9.6	5.7–16
CPE (C2)	57.7	9.9	5.6-16.1

Weir et al (2012a:96) in their study of the IELTS Reading test, found that the IELTS texts, in terms of both the Flesch Reading Ease and Flesch-Kincaid measures, were significantly (p<.05) easier to read than first year undergraduate texts. The difference between the means for IELTS and for undergraduate texts was five points on the 100-point Flesch Reading Ease scale or one year in terms of the Flesch-Kincaid Grade Levels (12.5 as against 13.5).

Figure 5.1 is a box-and-whisker plot summarising the distribution of Flesch-Kincaid reading levels for IELTS and undergraduate texts. The line in the middle of the boxes represents the median, and the upper and lower boundaries of the boxes represent the upper and lower quartiles of the distributions. The figure indicates that the IELTS texts were generally of a similar level of readability to the undergraduate texts, falling within the range of undergraduate text readability. However, one text (Test 8, Text 1), appears as an outlier with a reading grade level of 8. This text, which concerns the construction of Hong Kong airport, has the lower turber of words per sentence of any of the texts analysed and is at the lower extreme for the average number of characters per word (4.5). An implication here may be that using readability formulae could assist the test developers in identifying texts that might fall outside the range of readability typically found in university-level texts.

It is also of interest that no IELTS text had an estimated grade level higher than 16, although undergraduate texts ranged as high as 18. This might be taken as a further indication that even the most difficult of the IELTS texts do not reflect the level of the most challenging of the texts that undergraduates might expect to encounter in their first year of study.

Crossley et al (2008) proposed an alternative readability formula for L2 readers based on vocabulary frequency, similarity of syntax across sentences and referential cohesion. Green et al (2010) report on applying the Coh-Metrix readability formula to IELTS and the undergraduate tests and they are closer than the traditional formulae applied above (SD = standard deviation):

IELTS: Mean 12.60, SD 3.03

Undergraduate: Mean 12.71, SD 3.54

Type-token ratio

The type-token ratio (TTR) is the ratio of different words (types) to the total number of words (tokens). This represents a simple, if rather crude index of the number of different words the reader will need to know to understand a passage. It is generally recommended that a standardised length of text be used in calculating the TTR as the length of a passage will affect the figure obtained (Scott 2006). Although standardised measures are not provided through the Web VocabProfile, they can be obtained through another lexical profiling tool: WordSmith Tools. Green et al (2010) used WordSmith Tools to





find standardised TTRs based on 250-word sections of text and the results are displayed in Figure 5.2. It can be seen that the IELTS texts had a significantly higher mean standardised TTR than the undergraduate texts.

The fact that the standardised TTR in the undergraduate texts is noticeably lower than in the IELTS texts could be taken to indicate that candidates in IELTS are exposed to more demanding texts in terms of this ratio. With coursebooks, for example, one is likely to find more repetition of key words so that the reader is able to develop familiarity with these as they progress through the text.

Green and Hawkey (2012:336), in their study of item writing behaviour in IELTS, offer the following explanation:

... trained writers chose much longer pieces then progressively cut out passages that seemed to repeat information or that included elements that would not be tested. The extent of editing and the desire to avoid repetition perhaps explains why the texts analysed in Weir et al (2010) displayed relatively high type-token ratios in comparison with undergraduate textbooks (indicative of a wide range of vocabulary use and rapid progression of ideas).

Lexical density

Green et al (2010) report the lexical density of undergraduate and IELTS texts they analysed and found them to be broadly similar in this respect: the





undergraduate mean = 0.56, IELTS = 0.57. This may be taken to suggest that IELTS Reading texts in this respect do reflect a similar range of vocabulary to that appearing in undergraduate textbooks.

Lexical range

Khalifa and Weir (2009:133) provide an overview of lexical range across the Cambridge ESOL General English examinations over the last decade. It is noticeable that CPE (C2 level) texts include lexis from right across the first 20k levels of the British National Corpus (BNC). The official BNC website (British National Corpus 2007, emphases in original) describes itself as follows:

BNC is a 100-million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century, both spoken and written. The **written part** of the BNC (90%) includes, for example, extracts from regional and national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, school and university essays, among many other kinds of text. The **spoken part** (10%) consists of orthographic transcriptions of unscripted informal conversations (recorded by volunteers selected from different age, region and social classes in a demographically balanced way) and

spoken language collected in different contexts, ranging from formal business or government meetings to radio shows and phone-ins.

In the Khalifa and Weir (2009:133) overview of lexical range below, CAE characteristics are given on the left and CPE on the right:

K1	78.7	79
K2	8.5	8.5
K3	3.3	3.7
K4	2.3	2.3
K5	1.3	1.1
K6	0.9	0.9
K7	0.7	0.5
K8	0.5	0.5
K9	0.3	0.4
K10	0.3	0.3
Off-list	2.4	1.9

Green et al (2010) record the frequency data for lexis in undergraduate texts. There are fewer words at the K1 level than in any of the IELTS studies referred to in Chapter 2.

K1	74.0
K2	11.89
K3	2.6
K1–K3	88.49

Taylor and Chan (2015) report on lexical frequency data for IELTS:

K1	76.4
K2	11.36
K3	3.26
K1–K3	91.1

Academic lexis

The incidence of academic words in a text proved to be a good predictor of level in Weir et al's (2012c) study of FCE/CAE and CPE texts. CPE clearly exhibits a greater incidence of these semi-technical words.

	Mean	SD
FCE	1.61%	1.26%
CAE	1.63%	1.41%
CPE	5.82%	2.84%

Green et al (2010) report the percentage of academic words they found in the undergraduate and IELTS texts they analysed: undergraduate 10.51%,

IELTS 7.9%. Clearly CAE in particular but also IELTS need to select passages with a greater incidence of academic vocabulary if they are to match this feature of undergraduate texts. Green et al (2010:204–205) comment:

IELTS texts included significantly (p<0.05) fewer sub-technical academic words (IELTS mean = 7.9%, undergraduate mean = 10.5%) t = -4.036 (71.64) p = 0.00, and significantly fewer very infrequent words than the undergraduate texts. The proportion of running words on the AWL was also lower than that found in the corpus of academic texts from which the AWL was derived (10.0%). Although the mean proportion of AWL words occurring in these IELTS texts was higher than the 4% found by Coxhead (2000) in newspaper texts, the lowest proportion found in an IELTS text (2.2%) was closer to the proportion that she found in fiction texts (1.4%) and was just over half of the lowest proportion found in any part of an undergraduate text (4.33%).

The relatively low percentage of AWL words in the IELTS texts may reflect the high proportion of these texts that are sourced from newspapers and magazines . . . The proportion of AWL words varied by IELTS test part, with Part 1 texts having the lowest (7.65%) and Part 3 texts the highest proportion (8.24%) of AWL words. Even in Part 3 of the test, however, coverage of the AWL was lower than in the undergraduate texts.

The undergraduate texts include on average almost four times as many very low-frequency words (words that do not appear on the AWL and have a frequency on the BNC of less than one in 15,000) – IELTS mean = 1.09, undergraduate mean = 4.33, t = -5.853, (61.845) p = 0.00.

What is clear from the exercise and the subsequent discussion between the judges is that IELTS texts often appear to be somewhat journalistic and that newspaper/magazine texts are well represented in the test. The questionnaire responses (Green et al., 2008) had indicated that newspapers and magazines may feature as sources in first year academic reading at the University of Bedfordshire, but books, journals, reports and internet sources were all regarded as more important.

Taylor and Weir (Eds) (2012) similarly report that the kinds of text used in IELTS are those that tend to introduce academic topics to a general audience, often in the form of articles sourced from newspapers or magazines presenting research findings to the general public. Green et al (2010) found that even the most difficult of the IELTS texts did not appear to reach the difficulty level of the most challenging undergraduate texts, suggesting perhaps some sort of ceiling effect for the reading material that is selected for inclusion in the IELTS Academic Reading test.

Weir et al's (2012a) findings in relation to the AWL indicate that IELTS texts typically include a lower proportion of sub-technical academic





vocabulary than the undergraduate texts. Again, investigating coverage of the AWL might assist the test developers in identifying texts that lack representative coverage of sub-technical academic vocabulary. This finding appears, like the findings relating to readability, to suggest that IELTS texts may lack some of the features of academic texts that may cause difficulty for students.

The significantly higher proportion of words in IELTS texts at the 1,000and 3,000-word frequency level may be a corollary of the differences noted in relation to the AWL, with a higher proportion of these more 'general' words appearing in IELTS texts in place of the sub-technical vocabulary more frequently found in the undergraduate texts. The undergraduate texts include on average almost four times as many off-list words (words that do not appear on the AWL or on any 15,000-word frequency level list). Items of this nature include proper nouns and acronyms (Marks & Spencer, Charles, Myanmar, the BBC) as well as neologisms and some of the more technical language found in these texts (applet, compurgation, mediastinum, reusability). It is perhaps unsurprising that such words should appear less often in IELTS texts, which are required to avoid both cultural and subject specificity.

The results of a one-way ANOVA comparing the IELTS and undergraduate texts on the range of contextual parameters are presented in Table 5.3. There were significant (p<.05) differences between IELTS texts and undergraduate texts for readability measures (Flesch Reading Ease and Flesch-Kincaid reading level); standardised TTR; proportion of words on the AWL; proportion of words appearing on the first 1,000-word frequency and 3,000-word frequency levels; and the proportion of infrequent (off-list) words.

		Sum of squares	df	Mean square	F	Sig.
Flesch Reading	Between groups	595.73	1	595.734	4.852	0.030
Ease	Within groups	10067.24	82	122.771		
	Total	10662.97	83			
Flesch-Kincaid	Between groups	21.91	1	21.910	5.150	0.026
reading level	Within groups	348.88	82	4.255		
	Total	370.79	83			
Standardised TTR	Between groups	85.124	1	85.124	5.271	0.024
	Within groups	1324.281	82	16.150		
	Total	1409.405	83			
Proportion of words	Between groups	142.53	1	142.533	16.293	0.000
on AWL	Within groups	717.35	82	8.748		
	Total	859.88	83			
Proportion of words within 1,000-word frequency level	Between groups	169.41	1	169.406	4.783	0.032
	Within groups	2904.60	82	35.422		
	Total	3074.01	83			
Proportion of words within 3,000-word	Between groups	8.58	1	8.576	4.519	0.037
	Within groups	155.63	82	1.898		
frequency level	Total	164.20	83			
Proportion of words outside 15,000-word frequency level (off- list)	Between groups	220.29	1	220.288	34.256	0.000
	Within groups	527.32	82	6.431		
	Total	747.61	83			

 Table 5.3: Analysis of variance of IELTS and undergraduate text contextual parameters*

 $*df = degree \ of \ freedom; \ F = ratio \ of \ the \ two \ mean \ square \ values$

Taylor (2012:380), in Chapter 5 of the SiLT volume of IELTS collected papers (Taylor and Weir (Eds) 2012), concludes the following from the research above carried out by Weir et al (2012a):

Overall, the study found evidence that the IELTS Academic Reading texts fall generally within the parameter ranges exhibited by the undergraduate text corpus that they assembled and analysed. For example, the IELTS texts reflected a similar range of vocabulary to that which appeared in undergraduate textbooks. Comparability was also detected
on the measures of rhetorical organisation, grammatical complexity and cohesion. This is encouraging news for the IELTS test developers and it supports claims made about the validity and usefulness of the test.

However, it is clear that there are a number of quantifiable differences between the texts that students are exposed to in their first year of study and those used in IELTS reading passages. IELTS texts do appear in many respects to be 'easier' than the undergraduate texts students are faced with. The main areas of concern in the above data are that the IELTS texts:

- generally include a lower proportion of sub-technical academic vocabulary than the undergraduate texts
- may not fully reflect the level of readability found among the more challenging academic texts that first year undergraduates might expect to encounter.

These discrepancies require attention to ensure a similar cognitive load is placed on students taking the IELTS Reading test as is the case when reading undergraduate texts.

Qualitative data

Following identification of key textual features that were not susceptible to quantitative analysis, Weir et al (2012a) report on how two judges rated the IELTS texts and the undergraduate texts on six criteria: *rhetorical organisation, subject* and *cultural specificity, abstraction, grammatical complexity,* and *cohesion.* Rates of agreement between the two judges are shown in Table 5.4. Rates of agreement were highest for the more readily observed textual features of rhetorical organisation, grammatical complexity and cohesion, but were also considered acceptable for the more subjective features of subject and cultural specificity and level of abstraction. Where the two judges disagreed, the average of the two ratings was used in the subsequent analysis.

Criteria	Exact	+/- 1
Rhetorical organisation	52%	93%
Grammar	52%	94%
Cohesion	49%	92%
Subject specificity	31%	87%
Cultural specificity	33%	89%
Abstraction	29%	79%

	Table 5.4:	Rates of	agreement	between	the two	judges of	on textual	features
--	------------	----------	-----------	---------	---------	-----------	------------	----------

Research and Practice in Assessing Academic Reading

Table 5.5 shows the results of the non-parametric tests of difference between IELTS and undergraduate texts. Figure 5.4 displays the mean ratings for IELTS and undergraduate texts on each of the six criteria. Results were significant (p<.05) for both subject and cultural specificity. The significant difference between the IELTS and undergraduate texts in relation to *subject* and *cultural specificity* in Table 5.5 no doubt reflects the requirement for IELTS to avoid subject specificity and cultural allusion. No significant differences emerged on the measures of rhetorical organisation, grammatical complexity or cohesion. Although the undergraduate texts appeared to involve greater levels of abstraction, the results for this variable were not significant.

Criteria	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)
Rhetorical organisation	755.5	1658.5	-1.203	0.229
Grammar	788.5	1691.5	-0.914	0.361
Cohesion	716	1619	-1.601	0.109
Subject specificity	323	1226	-5.052	0.000
Cultural specificity	473	1376	-3.706	0.000
Abstraction	686	1589	-1.781	0.075

 Table 5.5: Results of non-parametric tests of difference between IELTS and undergraduate texts

Green et al (2010:206) note:

The cultural specificity found in the undergraduate texts could only be expected to add to the difficulty of reading them for international students. This may indicate that the difficulties with reading arising from lack of background knowledge, reported in the Green et al. (2008) survey, may arise from writers' assumptions about readers' level of cultural knowledge as well as their subject knowledge. Readers hitherto exposed only to relatively culturally neutral texts of the kind found in IELTS might well find the greater cultural specificity of the undergraduate texts to be a further source of confusion.

Taylor (2012:380) echoes this:

IELTS can sometimes lack certain features of academic reading texts that cause students significant difficulty in their studies. One such feature is cultural specificity, and Weir et al [2012a] speculate that readers exposed only to relatively culturally neutral texts of the kind typically found in IELTS (and perhaps in most large-scale, international proficiency tests) might find the greater cultural specificity of undergraduate texts more demanding when they encounter these in their academic studies.

The approach currently taken by the IELTS designers is to avoid content that is dependent on knowledge either of specific discipline areas or of particular





cultures. Item writers are advised to reject texts that might be too technical for the general reader or those that assume knowledge specific to certain cultures (Green and Hawkey 2012).

Can one size fit all?

At the start of this chapter, and indeed in the introductory first chapter, we raised questions about the practicality of catering for subject specificity in IELTS. These doubts are shared by Clapham (1993) and Fulcher (1999), who argue that it is not necessary or perhaps even desirable to employ genuine subject-specific academic texts in an EAP reading test. As we noted in Chapter 1, operational problems in using the genre-based approach can be seen in all the earlier EAP tests designed to test the academic English proficiency of overseas students in the UK. The history of all the major EAP testing projects like ELTS (1975–1989) (see Davies 2008:Chapters 2 and 3 and Weir and O'Sullivan 2017:Chapter 4), IELTS (1989–2017) (see Davies

2008:Chapters 4 and 5 and Weir and O'Sullivan 2017:Chapter 5) and TEAP (1980–2017) (see Weir 1983) demonstrates that attempts to provide discipline-specific options within a testing system are beset by problems not least in terms of content specificity and subject boundaries (for further discussion of these problems see Alderson and Clapham 1992, Alderson and Urquhart 1985, Charge and Taylor 1997, Clapham 1996a, Davidson 1998, Davies 2008, Henning 1988, Read 2015, Weir 1983, Weir and O'Sullivan 2017). The ESP approach, though, in principle, potentially more construct valid, has proven in the past to be simply not feasible logistically for industrial-scale global tests.

We also concluded, in the section on content knowledge in Chapter 2, that there was no consensus in the literature on the value of making subject-specific tests available as happened with ELTS 1980–1989 and in IELTS until 1995 (see for example Clapham 1993, 1996a, 1996b, Fulcher 1999 and Tan 1990, for the arguments against, and Khalifa 1997, Alderson 2000 and Urquhart and Weir 1998 for the arguments in favour of an English for Specific Purposes (ESP) approach).

However, in the new digital age there is a feeling in some quarters, usually prompted by concerns for construct validity, that an ESP approach, providing, for example, domain-specific reading modules, might be revisited (see especially the IELTS research study by Moore, Morton, Hall and Wallis 2015). Advances in digital technology, for example the availability of webbased platforms for test delivery, now make the provision of different versions of a test less problematic at least in terms of logistics. An example of changing attitudes in some test providers is provided by Martine Holland (forthcoming) who informs us that:

A compromise approach, previously taken by ELTS and now taken by the Cambridge English Academic Literacy test (ALT) is to group subjects together by faculty: the latter offers papers tailored for Business and Administration, STEM and Humanities, although this distinction is only at the level of given texts and the question to be answered, rather than marking criteria or task type, and for this reason still follows the study skills model.

There have been a number of fairly recent IELTS research studies that shed light on whether current IELTS is fit for serving a variety of purposes across different domains and we will report briefly on them here. Sedgwick, Garner and Vicente-Macia (2016:29) looked at the workplace language needs of nurses to determine the relevance of the IELTS examination for them and argued:

The test does not assess the same reading abilities as those required in nursing. Nurses have to identify words and phrases, and parse simple syntax in extended narrative. The reading material they deal with – medical notes, handover sheets, forms, charts, checklists, medicine labels

and online definitions – is typically in a restricted code, and includes headings, bullet-pointed lists and brief notes, abbreviations and quantities. In contrast, the IELTS texts require an understanding of features of coherence and cohesion in lengthier texts, interpretation of implicit as well as explicit meaning, syntactic parsing of both simple and complex sentences, and the like. The only extensive nursing-related texts, identified by the current study, are protocols, which are predominantly formulaic and predictable, and policy statements, which arguably nurses should read, but which are not encountered in their daily duties...

The majority of these abilities are essential to the comprehension of extended texts, which, as noted above, are not directly relevant to nurses' reading. A few, however, are relevant: nurses do have to skim and scan text to find information quickly and relate information in medical notes to diagrams and charts in the patient records, they have to identify accurately quantities, as well as expiry dates on medicine labels. . . . Overall, however, the IELTS reading test does not assess nurses' ability to comprehend the kind of written language that they encounter in their work.

Sedgwick et al (2016:33) concluded:

... In general, therefore, the academically-oriented reading skills that are tested in the IELTS are only marginally relevant for nurses.

In revising IELTS, an important issue to consider may be whether the current or any revised generic IELTS can be used appropriately in a variety of domains or whether in the future, variants of IELTS may need to be offered, for example in professional as against academic situations. A number of the IELTS funded research papers have sought to investigate this. Moore et al (2015) looked at the suitability of IELTS for use with those involved in professional work across a range of occupational areas as well as those in the academic domain. Whilst finding a number of similarities in literacy demands between the two groups, they found (2015:1) 'the main differences noted related to the highly transactional nature of professional communications', and suggest:

Two broad options for the future directions of the test would appear to be available to developers. One of these is to continue with the current trend evident in the recently produced materials – that is to pursue the idea of making the test suitably "flexible" so that it has relevance to the two types of cohort considered in the research (i.e. those entering tertiary study and those entering professional employment).

The other option – a more radical one – is to work towards developing a separate IELTS test for general professional employment purposes (emphasis added by authors).

Research and Practice in Assessing Academic Reading

Moore et al (2015:15) are not very sanguine about the current suitability of IELTS for riding both professional and academic horses at the same time and they provide a number of cases where there is cause for concern (emphasis added by authors):

Merrifield (2008) investigated stakeholder attitudes towards the use of IELTS across professional contexts. In her study, the views of key staff from a range of professional associations were explored regarding the suitability of IELTS to assess eligibility for membership of those organisations, and also for professional registration . . . One of the concerns expressed was the extent to which the test is appropriate for testing broader language skills associated with professional practice – "those required for a professional operating as a doctor, nurse, engineer, teacher or accountant" (p 9).

... Read and Wette (2009) investigated the attitudes of a group of overseas-trained health professionals towards the IELTS test in their experiences of meeting the English language requirements for professional registration. The study found that ... IELTS ... had limited capacity to assess "their ability to communicate effectively in clinical settings". Read and Wette note in relation to this response that IELTS is still designed primarily as a test for those entering educational and training contexts, and is not specifically intended "to assess the communication skills required in particular professions" (p 4).

Moore et al (2015:15) also looked at the suitability of IELTS for the teaching profession (emphasis added by authors):

Two studies have investigated stakeholders' perceptions of the role of IELTS: in the selection process of international students for teacher education courses in Australia (Sawyer & Singh, 2011); and for entry for overseas trained teachers into schools in Australia and New Zealand (Murray, Cross & Cruickshank, 2014). While the focus of Sawyer and Singh's study was teacher education (an academic domain), the authors were particularly concerned with the challenges for international students of the teaching practicum component (a professional domain). This study found that the **student-teachers required a wide range of English language/communication skills for practicum classes beyond that needed for academic success, including familiarity with colloquial idiom in a school context and the discipline-specific discourse of particular subjects, as well as the ability to respond spontaneously in classroom interactions with students.**

Focusing more specifically on the skill of reading, Moore et al (2015:37) note:

A key finding regarding reading practices in the professions was the intimate link noted between acts of reading and the undertaking of professional activities – whether these were related to internal organisational processes or to interactions with clients and other external agencies. Thus, as was noted, many reading activities and related genres were concerned with forms of doing: what needs to be done (e.g. staff bulletins, meeting minutes, specifications, contracts, letters of engagement); what is required to be done (e.g. codes of conduct; industry standards; laws and statutes); how to do things (policy manuals; operating procedures, etc.). On this point, a number of participants in describing the challenges of workplace reading, emphasised the importance of being able to extract "the key points" from a text, and to be able to understand exactly what was required "to progress things to a next stage".

... Arguably, this action-orientation of tasks is less [of] a feature of the Academic version of the reading test ... texts used in the Academic module appear to be mainly of an expository nature – that is to say "general interest" texts taken from "books, journals, magazines and newspapers". Related to this, the principal task for candidates in the Academic module appears to be one mainly of identifying specific propositional content in texts (i.e. epistemic in nature).

... In line with the action orientation of much professional work, the study found that reading practices in the professions were oriented more to texts of a deontic nature.

They express concern that (2015:38):

... If, as Kane (2013) argues, an essential part of a test's validity is the extent to which the language elicited on test tasks is comparable with the language used in the relevant real-world domain, then the provisional differences we have observed in this study would seem to pose some challenges for future test development.

The research and development agenda for IELTS will need to consider carefully the uses that are to be made of the test(s) and ensure that IELTS is used to predict abilities which the test is designed to measure. As the authors quoted in this section have forcefully argued (Merrifield 2008, Moore et al 2015, Murray, Cross and Cruickshank 2014, Read and Wette 2009, Sawyer and Singh 2011) further research into professional contexts, including validation research that analyses the language skills required in the professions that use the test for professional registration, will need to be carried out to ensure fitness for purpose across all domains.

Shifting paradigms

Gad Lim, in the introduction to Moore et al (2015:3), points out:

As the researchers note, IELTS was not designed as a specific purpose test for the workplace literacies of particular professions. Rather, it is a test of a person's "readiness to enter a domain of practice" (Taylor, 2007), and users should not expect it to do more than it claims to be, as various IELTS-funded research studies into its use in the professions have painstakingly pointed out...

This concept of 'readiness to enter' appears to be used as an argument for current IELTS to be like it is i.e. general and not based squarely on a putative generic academic domain or professional practice (Taylor 2007). There seems to have been a tacit acceptance within the IELTS partnership from 1989 that IELTS could be allowed to metamorphose into a general English test rather than the test of EAP it had started out as (see Chapter 4 with details of this in regard to the cognitive processes assessed in reading).

The genesis of ELTS is examined in Weir and O'Sullivan (2017:170), who quote Ian Seaton, one of the team working on the forerunner of IELTS, back in 1980:

There was broad agreement about what features any new testing system should have. It should reflect the recent developments in communicative language teaching, have high face validity both for those taking it and those receiving its "scores" and should have a positive feedback into language learning situations. It should try, in its content and skills specifications, to replicate the language used in different academic and training subjects. It should not use only a multiple-choice format, but "authentic" performance in both writing and speaking should be measured as directly and reliably as possible in a test situation.

John Read (2015) offers a possible explanation for the change in direction from its academic heritage. He provides insightful detail of the original use of ELTS (1980–1989) as a diagnostic instrument which enabled the British Council first to select students for awards to study in the UK and then, perhaps more importantly, to indicate to the people responsible for incoming students in Spring Gardens (the British Council HQ in London) how much language help was needed by successful recipients before their programmes of study started (Read 2015:126, Weir and O'Sullivan 2017). Read feels that this 'diagnostic' function was lost in the transformation of IELTS into a preadmission, gate-keeping instrument to be used for testing general proficiency in English rather than academic English *per se*; the so called 'IELTS compromise' post-1995 (Read 2015:4, 29–30), made to cater for the exponential growth of candidates taking the test outwith the training cycle of the British Council.

It is interesting to compare the earlier concern of the ELTS test developers with academic authenticity to the position adopted by Taylor (2007), a former IELTS chief examiner, who suggested that we need to recognise the limits to which a test such as IELTS can simulate (and indeed should be expected to simulate) language use in the target situation. Thus, she notes that 'IELTS

is designed principally to test *readiness to enter* the world of university-level study in the English language' (2007:482, emphasis in original), and does not assume that test takers have already mastered the skills they are likely to need. Taylor goes on to explain that students will often 'need to develop many of these skills during their course of study', including those 'skills . . . specific to their academic domain' (2007:482) (see also Taylor 2012).

Taylor is of course right concerning the domain-specific requirements of socially situated discourse in academia or the multiple domains to be catered for in any genre-based approach (see Murray 2016 and earlier discussion of these approaches in Chapter 1). However, the research evidence does suggest that on entry to a university, students will require the generic academic study skills we have looked at in Chapters 1, 2 and 3. These are the skills that ELTS (and originally IELTS) was tapping into. Murray (2016) is mistaken in his belief that IELTS caters for these academic skills as is clearly demonstrated in the research cited in earlier chapters of this volume.

Given the predisposition towards a 'readiness for study' approach in Cambridge English, it is perhaps not too surprising that by the 21st century, as we saw in Chapter 4, the IELTS Reading test appears to be for the most part a test of general language proficiency at the local level, a far cry indeed from the needs analysis-driven, study skills approach of the ELTS academic English test creators in the British Council 1975–1989 (see Weir and O'Sullivan 2017:Chapter 4).

In any future revision, is IELTS to be simply a general test of a candidate's baseline competence at say B2/C1 levels in the four skills predicting readiness to enter professional or academic situations (rather like Aptis in fact), or is it to be a test of whether the candidate can actually cope with the study skills demands initially faced in a profession or in an academic context? If it is the former, little change is needed. If it is the latter, then clearly more than one 'authentic' needs-driven test will be required and the logistical challenges to be faced that much greater.

We have established in Chapter 3 what an Academic IELTS test should cater for. Moore et al (2015:38) discuss what a 'generic professionally-oriented test' might look like:

... arguably the best option on offer is to rely on a kind of generic experience around some of the broad processes of professional work. Some of these processes have been discussed in the findings, and relate to what we have referred to as "secondary" or "ancillary" interactions; that is to say, various formal and semi-formal communications conducted within organisations to facilitate work activities e.g. making requests to colleagues; reporting on activities completed; getting clarification from supervisors about work processes, and conditions etc.

. . . The current research has only made some provisional observations about the degree of correspondence between the literacy demands of the

test and those related to professional work. A more systematic analysis of items and formats would enable more definite conclusions to be made about the overall suitability of using the test in these contexts, as well as a sense of whether one of the modules – Academic or General Training – is demonstrably more suitable for this purpose than the other. Other useful research would be to further investigate stakeholder responses to professional uses of the test, taking into account the views of graduates, employers, professional associations, and relevant education providers. A number of such projects have already been conducted (Merrifield 2008; Murray et al., 2014; Read & Wette, 2009) or been recently commissioned by IELTS (e.g. Blackmore et al., 2010-2012; Knoch et al., in preparation). The outcomes of such work can assist to establish more clearly both whether there is a need for a new dedicated test, as well as the shape such a test could feasibly take.

The research on academic reading suggests that there may well be some benefit in exploring the possibility and value of developing versions in the future targeting two distinct test taker groups: a more avowedly academic version, and one specifically for professional practice along the lines suggested by Moore et al (2015). It also raises the possibility of continuing alongside these new versions with the present pre-professional, pre-academic version for end users who buy into the argument that IELTS is designed principally to test readiness to enter various target domains, and does not assume that test takers have already mastered the skills they are likely to need (Taylor 2007:482).

We now turn from the complex area of subject specificity to examine some final parameters of IELTS texts.

Discourse mode

Taylor and Chan (2015) note that texts in IELTS are appropriate and accessible to test takers entering undergraduate or postgraduate courses, or seeking professional registration. The kinds of texts used in IELTS are those that introduce academic topics to a general audience, often in the form of articles sourced from newspapers or magazines presenting research findings. These include self-contained reports on developments in science and technology and overviews of academic debates. The IELTS texts often present solutions to problems that are likely to be of interest to the general reader. Texts come from books, journals, magazines, newspapers, reports and online resources, written for a non-specialist audience.

The advantage of the IELTS approach to text selection is that the texts appearing in the test do, based on the limited corpus explored by Weir et al (2012a), have many of the features of the kinds of text encountered by undergraduates. Although there are minor differences attributable to source (word frequency) and length (TTR), the IELTS texts include a vocabulary and a level of grammatical complexity that would place them within the range of texts encountered in the first year of study.

However, while IELTS passages are at a level of difficulty appropriate to university study, they are not as challenging as some of the texts encountered in the more linguistically demanding areas such as the law textbook analysed by Green et al (2010).

Table 5.6 shows the level of agreement between the two judges in assigning the texts to categories for the features of *genre, rhetorical task, pattern of exposition* and *subject area* in Weir et al (2012a).

Criterion	Agreement
Genre	80%
Rhetorical task	80%
Pattern of exposition	73%
Subject area	85%

 Table 5.6
 Level of agreement between two judges on features of genre,

 rhetorical task, pattern of exposition and subject area

The categorisation of texts by genre is set out in Table 5.7. The analysis of the undergraduate texts was straightforward as all were textbooks, but there was some disagreement between the two judges in relation to the IELTS texts. Both agreed that 17 of the texts had been sourced from magazines or newspapers, that seven came from textbooks and that one was a research article. However, the second judge was less likely to identify magazine or newspaper articles as the source, seeing nine of those so identified by the first judge as coming from textbooks and a further seven from research articles. Discussion following the categorisation exercise indicated that some texts had been more difficult than others to categorise and that it was not always clear to the judges whether an individual text had been sourced from a research article, magazine article or textbook. Although some texts had very obvious journalistic features, such as opening paragraphs that served as 'attention grabbers', and one text had the conventional headings of the research article, distinguishing characteristics were not always so easy to locate. A number of texts had little to indicate whether they had been sourced from a newspaper section, from a popular science magazine, from an introductory textbook or from a more specialised academic publication. Green and Hawkey (2012) suggest that it would be of interest to explore how genre is affected by the editing process through which texts are prepared for inclusion in IELTS. It is possible that changes made to texts might have affected the judges' ability to assign them to a genre.

What is clear from Weir et al (2012a) is that IELTS texts often appear to be somewhat journalistic and that newspaper/magazine texts are well represented in the test (see Green and Hawkey 2012 for a similar finding). Green and Hawkey (2012:292–293) summarised the characteristics of target IELTS-type texts as interpreted by the four experienced participants:

The experienced writers seemed to share with the non-experienced group the perception of IELTS texts: subjects of popular interest presented in a formal, report-like format, academic in tone but not so technical that non-specialist readers would be handicapped in understanding them.

Genre		Textbook	Magazine/ newspaper article	Research/ academic journal article	Report
IELTS	Textbook	7			
	Magazine/newspaper article	9	17	7	
	Research/academic journal article			1	
	Report			1	
Undergraduate	Textbook	42			

Table 5.7: Categorisation by genre: Results for judge 1 displayed by row, judge2 by column

As indicated in Table 5.8, both judges agreed that most of the texts were expository in nature – 30 of the IELTS texts and 27 of the undergraduate texts. Both judges also agreed that argumentation and historical/biographical texts were also represented among both sets of texts. In terms of rhetorical task there appears to have been a close match between IELTS and these undergraduate texts. Taylor and Chan (2015) confirm that the IELTS texts were largely historical/biographical, expository or argumentative. Texts were not overly general or specific in terms of content knowledge, fairly neutral in terms of cultural specificity, mostly concrete and contained both verbal and non-verbal information. The texts in IELTS also had rhetorical explicitness.

 Table 5.8: Categorisation by rhetorical task: Results for judge 1 displayed by row, judge 2 by column

Rhetorical task		Exposition	Argumentation	Historical/ biographical
IELTS	Exposition	30	2	1
	Argumentation	3	2	
	Historical/biographical	1		3
Undergraduate	Exposition	27		1
	Argumentation	5	2	
	Historical/biographical	3	1	3

With respect to pattern of exposition, as shown in Table 5.9, the two judges agreed on the classification of 35 of the 42 IELTS texts, but just 26 of the 42 undergraduate texts (shown in bold). Subsequent discussion revealed that IELTS texts were more often felt to reflect one clearly dominant pattern of exposition while the undergraduate texts often involved two or more patterns occurring in sequence. This difficulty may have been caused by the way in which the undergraduate texts were collected; they were extracted from longer texts, often cutting across sections in the textbooks, each of which displayed different patterns.

A further challenge for the judges in identifying patterns of exposition was that the categories are not mutually exclusive – definitions and descriptions often include illustration and a problem–solution text may additionally imply cause–effect. Determining which pattern was dominant in each of the texts investigated did not prove to be straightforward.

The analysis suggested that almost half of the IELTS texts displayed problem-solution or cause-effect patterns while the majority of the undergraduate texts involved elaboration. The selection of texts may have contributed to the difference; the opening chapter of an introductory textbook is often concerned with elaborating the scope of the subject. On the other hand, the brevity of IELTS texts and the high occurrence of newspaper/magazine articles may favour problem-solution and cause-effect patterns of exposition. The use of short texts with relatively clear dominant patterns may also bring its own problems; candidates may not be well prepared to encounter lengthier texts and to cope with transitions and relations between sections that follow different organisational principles.

In classifying the texts according to subject area, the two judges were in complete agreement in assigning the undergraduate texts to subject area and agreed on 35 of the 42 IELTS texts. A broad range of subject areas were represented among the IELTS texts investigated by Weir et al (2012a:145) with social studies, engineering & technology, and business & administrative studies emerging as popular topic areas for the test.

Green et al (2010:207) add a final note of caution with regard to the contextual parameters of the IELTS texts as compared to those found in undergraduate texts:

The IELTS texts include a vocabulary and a level of grammatical complexity that would place them within the range of texts encountered in the first year of study. However, the absence of significant differences on many of the measures investigated does not indicate equivalence and it is noteworthy that the undergraduate texts encompass a broader range of values on 24 of the 29 measures included here and so texts at the level of the most challenging undergraduate textbooks are not represented on the test. While IELTS passages are at a level of difficulty appropriate to university study, they are not as challenging as some of the texts encountered in the more

Table 2.2. Card	goussauou by pauci n or cal		r ognaf tot si	u uspiayeu u	y run, juug			
Pattern of expositio	u	Define/ Describe/ Elaborate	Illustrate	Compare	Classify	Cause/effect	Problem/ solution	Justify
IELTS	Define/describe/elaborate Illustrate	17		-				
	Compare							
	Classify							
	Cause/effect					7		1
	Problem/solution	2				1	11	2
	Justify							
Undergraduate	Define/describe/elaborate	25	1			1	4	7
	Illustrate	1	1					
	Compare	1						
	Classify	2						
	Problem/solution	1	1					1
	Justify	1						

Table 5 0- Cateoorisation by nattern of exnosition: Results for indoe 1 disulaved by row, indoe 2 by column

linguistically demanding areas such as the law textbook analysed for this study.

In terms of contextual parameters, the descriptive framework employed in this study has proved useful in identifying individual IELTS texts with idiosyncratic characteristics that do not match those typically identified with academic text. We feel that this review project offers a methodology whereby such disparities might be identified at the text selection stage.

IELTS item writers: Quis custodiet custodes?

Taylor (2012:386) takes a number of positives from the Green and Hawkey (2012) study on IELTS item writers:

This study provides the field with some valuable insights into the processes of text selection, adaptation and item writing for a test of reading comprehension ability, as well as more generally into the nature of expertise. The differences observed between the experienced and non-experienced groups help to highlight the skills that are required for effective item writing. Overall, the researchers report being favourably impressed by the conscientiousness and professionalism of the trained IELTS item writers that they interviewed and observed, and by the quality of the texts and the items that they produced. This should be a source of encouragement for the IELTS test producers who have undertaken extensive investment over the years to develop rigorous policy and procedures for item writer selection, training and monitoring. It also strengthens the view that such expertise is collective in nature, rather than residing in individuals, and it supports the IELTS partners' decision to have IELTS itemwriting teams based in different parts of the English-speaking world.

However, one finding which is less satisfactory is the extent to which decisions on text content were delegated to the item writers. With multiple versions of IELTS, this raised potential issues related to version comparability. As well as a more finely detailed test specification, this has led to more closely circumscribed and monitored item writing procedures. The various contextual and cognitive parameters for each test must be spelled out clearly in the test specifications, and when test texts and tasks are produced these parameters need to be recorded in close detail by item writers and accompany their submissions.

The variability in text dimensions reported by Green and Hawkey (2012:314–315) for texts submitted by the item writers in their study, in terms of academic word percentages, is a serious wake-up call for the IELTS partners that more needs to be done to ensure text comparability. The fact that it would take item writers a matter of minutes at most to check the contextual parameters of each text (e.g. by using Text Inspector or a similar textual

analysis tool) and record these, underlines the folly of not insisting on this data being provided by item writers when they submit items.

Green and Hawkey's study of IELTS item writers (2011:126) raises some concern about the contextual parameters of texts selected for use in the exam:

... [an] earlier study by Green, Ünaldi and Weir (2010) suggested that the claim that IELTS texts are "similar to those which [students] might need to read on a university course" (IELTS, 2007a) should perhaps be qualified. Texts appearing on the test do not fully reflect the essential first year reading required of first year undergraduates (although they may share certain characteristics with such texts). In this context, it is interesting to note that texts that are specifically designed for academic purposes are not seen by the item-writers to be the most suitable for the test. Catering to intending students from a wide range of disciplines limits the scope for employing genuine (in the sense of "found" and un-adapted) academic texts intended for students working in a specific field. The texts that appear on the test are not reproduced there in their original published form. Rather, they are chosen by item writers with a range of considerations in mind and are extensively reshaped to make them suitable for the test...

The item writers acknowledged that the texts they submit might not fully represent the kinds of texts students might encounter at university, but did not consider this to be a shortcoming. There was an obvious contrast between the length of IELTS texts and the length of the texts that students would need to read . . . "900 words versus a book" as Elisabeth put it. Elisabeth acknowledged that, given the practical restrictions on what could be covered in a one-hour test, "there is a huge amount we don't do of course: dealing with contents [pages], dealing with indexes, dealing with chapters and all that sort of macro stuff. We can't do it". However, Elisabeth defended the IELTS approach, stating that "we are not testing what they may be able to do after a few months at university; we are testing whether they will be able to cope, I think". Because the three texts on the reading test present the test taker with a variety of topics and text types (including more fact-based and more discursive texts), Anne felt that IELTS might "reflect in miniature what [students] have to do [in their university reading]: look at a variety of sources, get key ideas, get attitudes, get opinions".

The choice of formats to be employed should not be decided by item writers alone, especially given the reluctance of some item writers to use MCQ items because these are more time consuming to write (Green and Hawkey 2012:287) or leave out diagrams because these were more difficult to locate or create (Green and Hawkey 2012:299). Green and Hawkey (2012:341) provide further food for thought:

Elisabeth made reference to the discontinued practice of asking item writers to identify the skills being tested by each of their items. Elisabeth had found this difficult but useful and consideration might be given to reintroducing such a practice as a training exercise if not as a routine requirement. It might also be advisable to introduce clearer controls on the range of task types and the range of skills to be targeted for each text.

Green and Hawkey (2012:338-339) note:

The group had discussed at some length the nature of the information that could be targeted using Type 1 MCQ items and the extent to which inferences might be tested using Type 8 T/F/NG items. These discussions left open the possibility that different writers might be targeting different reading skills when using the same item type – as observed in Section 8, each set of T/F/NG items bore a somewhat different relationship to its partner text. This has implications for the comparability of different forms of the test as it makes it more challenging to ensure that every form reflects the required range of reading skills.

Green and Hawkey (2012:340–341) establish a number of critical questions that need to be addressed in future item writer guidelines or related training packages:

- What are the reading skills that the test as a whole is intended to address? And in what proportion?
- Why these reading skills? And how do they relate to the available task types?
- Within each task, what kinds of linguistic relationships should T/F/ NG (and other types of items) have to the text and in what proportion? What are the implications of these for the reading skills being targeted?
- What range of skills should be addressed in each section of the test and what item types should be used to target them?

Endnote

That concludes our analysis of the contextual parameters in current IELTS tests and how they relate to texts in real-life undergraduate reading. In our search for evidence of how IELTS reading compares to external points of reference, in Chapter 7 we will examine research that compares IELTS with EAP tests of 'equivalent' standing. Before that, however, we shall explore in Chapter 6 how IELTS affects the teaching and learning that precedes it – issues of consequential validity, including washback and impact.

6 Consequential validity parameters

Testers must realise that much of the strength of tests lies not only in their technical quality but in their use in social and political dimensions. Studies of the use of tests, as part of test validation on an ongoing basis, are essential for the integrity of the profession. Shohamy (2001:162)

Introduction

In this chapter we consider matters of consequential validity as they relate to IELTS, often referred to as issues of impact and washback, i.e. how the testing of reading in IELTS affects the teaching and learning that precedes it, as well as its wider effect on educational systems and society.

Emerging notions of 'impact' and 'washback'

Bachman and Palmer (1996:29) offer perhaps the most convincing overview, seeing impact as operating on two levels:

- (i) 'a socio-cultural level, in terms of educational systems and society in general' (macro level);
- (ii)'a local and personal level, in terms of the people who are directly affected by tests and their results' (micro level).

Wall's (1997) and Hamp-Lyons' (1997) concept of impact appears to reflect Bachman and Palmer's (1996) version, namely that the test has effects on 'educational systems and society in general' as well as at 'the local and personal level'. Wall (1997) views impact as: 'any of the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system or society as a whole' (1997:291). Hamp-Lyons (1997) defines impact as that which pertains to high-stakes tests whose influence extends to the school, educational systems, and society.

Alderson and Wall's (1996) 'washback' appears to be confined to Bachman and Palmer's point (ii), that is, effects at 'a local and personal level' on people 'directly affected by tests and their results'. Alderson and Wall (1993:121) suggest that the term 'washback' should be limited to the influences the test might have on teaching, teachers, and learning (including curriculum and materials) and this seems now to be generally accepted.

Impact can thus be seen as a superordinate which subsumes washback. Impact is concerned with 'wider influences', with the macro contexts in society, as well as with the micro contexts of the classroom and the school, whereas washback, as a sub-component of impact, focuses rather more narrowly on the latter (see Hawkey 2006 and Hamp-Lyons 2000).

Khalifa and Weir (2009:169) summarise how impact and washback have been defined in various, and sometimes overlapping, ways by different authorities on the subject since the 1990s. With specific regard to the Cambridge English examinations, Weir (2013a:10–11) argues that:

Prior to the late 20th century there is little evidence of any attention being paid in Cambridge to the macro issues of social impact and test use, the consequential aspects of test validity. Nor to be fair, is there much evidence of such a concern in the wider testing field prior to Messick's (1989) seminal publication on validity. It was not until the 1990s that it came onto the radar of most language testers (Alderson and Wall 1996, Bachman and Palmer 1996, Wall 1997, Wall and Alderson 1993). Milanovic and Saville (1996) appears to be the earliest attempt at Cambridge to address the wider impact issues of Cambridge English examinations ... There was, however, an interest in the washback on teaching and learning of its English language tests (impact at the micro level) from the very beginning in 1913 (see Green 2007).

The positive washback of its examinations on what was taught in the classroom has always been important for Cambridge English. The examinations have always been characterised by a close relationship with pedagogy, i.e. curriculum, syllabus, classroom practice and the teaching profession (see Weir and Milanovic (Eds) 2003 for a detailed discussion of this and Weir et al 2017).

Spolsky (2004:305) describes how:

... from its beginning UCLES [University of Cambridge Local Examinations Syndicate, now Cambridge English] accepted the key role to be played in test development by the "stakeholders", in particular those schools in various countries of the world that wished to establish examination centres, mainly for their own students. From the earliest years, the Cambridge test writers and their various committees saw themselves as sharing with the schools not so much an examination as the culmination of a teaching process. Before the word "backwash" had been coined, they regularly asked whether modifications being proposed in the form of the examination would be accepted by the schools.

Bachman et al (1995:131) make a similar point:

... the British examinations system is particularly concerned with promoting positive effects of examinations on curricula and instruction, and

Research and Practice in Assessing Academic Reading

thus is sensitive to including features in its examinations that are consistent with those found in instructional programs.

Weir (2003:5) traces the Cambridge English interest in washback back to Roach in the 1940s:

In one of the very first references to the concept of washback validity, Roach questioned how far examinations act as a stimulus and a focal point for both teachers and taught, and thereby promote the expansion of the studies that they are designed to test.

Messick (1989) argues that it is necessary in test validation studies to establish whether the social consequences of test use and interpretation support the intended testing purpose(s) and are consistent with other social values. Messick emphasises that the appropriateness, meaningfulness, and usefulness of scorebased inferences are a function of the external social consequences of the testing. This view on the potential 'consequences' of testing and test use was developed into a notion of 'consequential validity', as part of a sociocognitive approach to test development and validation (Weir 2005b), according to which a number of key parameters can be considered by test designers and researchers.

Cheng, Watanabe and Curtis (Eds) (2004) argue that test washback and impact should be a major area of concern for educational research. Eight projects are described in their volume including Saville and Hawkey's (2004) account of an IELTS Textbook Washback Study conducted for Cambridge ESOL (see the next section for details). The centrality of test washback and impact in language testing can also be seen in the publication of a number of titles in the SILT series published jointly by Cambridge English and Cambridge University Press (CUP). Between 2005 and 2007 four new volumes appeared focusing on major washback and impact studies carried out by Cheng (2005), Wall (2005), Hawkey (2006) and Green (2007).

Despite a positive and longstanding orientation to test washback at Cambridge English, we had some difficulty in tracing research studies concerned with exploring the washback of IELTS Reading on the teaching and learning of academic reading. We found no research exploring the impact of this IELTS module on wider society. To structure the limited and rather disparate nature of the washback research on IELTS Reading, we will divide the studies up into those which are generally positive in their findings, and those which are more negative.

Generally positive washback

There seems to have been no dedicated study of the impact or washback of the IELTS Reading module apart from those parts in Hawkey's (2006) study

on the impact of IELTS that deal with reading. The IELTS Impact Study was a major long-term programme of research by Cambridge English (then known as Cambridge ESOL) into the impact of IELTS, which ran from 1995. Saville (2012:5) describes how:

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.

The study included three phases: identification of areas to be targeted and the development of instrumentation to collect information which allows impact to be measured (Phase 1); validation of the instruments prior to full-scale implementation (Phase 2); and implementation of the instruments as part of a major survey (Phase 3).

Phase 1 was undertaken by Alderson and his research team of MA students at the University of Lancaster (see Alderson and Banerjee 1996, Banerjee 1996, Bonkowski 1996, Horak 1996, Yue 1997). Phase 2 entailed analyses and pretesting of the draft data collection instruments by the Validation Group (Cambridge ESOL) in conjunction with external consultants including Purpura (Teachers College Columbia), Kunnan (University of California, Los Angeles, UCLA) and Hawkey (University of Bedfordshire). Phase 3 streamlined the original 13 data collection instruments down to five:

- a modular student questionnaire for pre- and post-IELTS candidates, covering language learning background, objectives and strategies
- a language teacher questionnaire embracing teacher background and experience, attitudes towards IELTS, experience of and ideas on IELTS preparation programmes
- an instrument for the evaluation of IELTS-related textbooks and other materials
- a classroom observation instrument for the analysis of IELTS preparation lessons
- a pro forma for receiving information from IELTS administrators on their IELTS experiences and attitudes.

For a comprehensive, reflective account of the complete IELTS impact study the reader is referred to Hawkey (2006). We restrict our focus to reporting a number of the findings that came out of the IELTS impact study (IIS) which are directly relevant to the assessment of second language reading. Specific references to the effects of IELTS Reading are limited in the study and these are concerned mostly with time pressure. Based on the perceptions of student and teacher respondents (2006:122–124), reading was considered to be the most difficult test module and time pressure the most prominent problem, with further concerns expressed about unfamiliarity of topic and difficulty of questions. Perceptions of the relative difficulties of the IELTS macro skill modules (Reading, Writing, Speaking, and Listening) held by both teachers and candidates in the study are reported in Table 6.1.

	Most difficult IELTS module? (%	(0)
Skill	Students	Teachers
Reading	49	45
Writing	24	26
Listening	18	20
Speaking	9	9

Table 6.1:	Student and	teacher	perceptions of	f IELTS	module	difficulty
------------	-------------	---------	----------------	---------	--------	------------

Source: Adapted from Hawkey (2006:122)

The Reading module is seen as clearly the most difficult of the four IELTS test modules across the candidate and preparation course teacher participants. However, it is interesting to note that the Reading test did not appear in the top five reasons given by the 28% of IELTS candidates who felt IELTS was unfair.

The inter-relationships between perceived difficulties emerging from the questionnaire data were investigated further by Hawkey 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 as a significant cause of worry for candidates. Table 6.2 emphasises the dominance of the Reading test module as the most difficult according to IIS test takers and of time pressure as the most prominent problem with the Reading test.

 Table 6.2: 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	Other	Total
Reading	13	20	28	51	14	2	128
Writing	10	10	19	26	8	0	73
Listening	4	7	6	16	4	1	38
Speaking	2	4	6	9	3	1	25

Source: Adapted from Hawkey (2006:123)

Hawkey's investigation (2006:110–111) of prominent IELTS preparation course activities is more revealing about the Academic Reading Module. The figures he quotes suggest that teachers in preparing students for IELTS are carrying out a number of higher-order reading activities in class:

- analysing text structure and organisation students 74% and teachers 90%
- interpreting statistics/graphs/diagrams students 74% and teachers 90%
- reading quickly to get the main idea of a text students 77% and teachers 96%
- learning quick and efficient ways of reading texts students 73% and teachers 93%.

This data might be considered as evidence that the test is achieving beneficial washback in these areas. However, closer inspection of Hawkey's questionnaires to students (2006:186) and to teachers (2006:196) reveals that he only asked about whether higher-order skills took place (yes/no, but not how much) in a normal IELTS preparation class, **not lower-order skills**. This perhaps explains why the focus on lower-order skills in the IELTS examination we noticed in Chapter 4 does not feature here. Nevertheless, this data does suggest positive washback on the teaching of higher-order skills in contrast to our findings on the test itself in Chapter 4.

Similarly, the data he generated on perceived skills coverage in IELTS preparation books (N=43) is revealing (2006:111). The numbers of selections include:

- identifying main points (40)
- identifying overall meaning (38)
- predicting information (36)
- retrieving and stating factual information (34)
- distinguishing fact from opinion (31)
- drawing conclusions (30)
- making inferences (29)
- evaluating evidence (27)
- identifying attitudes (23).

Once again, this may be taken as evidence that in preparing for IELTS students are practising higher-order reading skills that should be beneficial to them in their future courses of study. The only caveat is that the reading-enabling skills listed for the analysis of textbooks in Question 4 of the instrument for the analysis of textbook materials (IATM) (Hawkey 2006:201) do not cover any of the lower-order processing activities which were identified as prominent in the IELTS test in Chapter 4. Respondents were thus **only** asked about whether a subset of higher-order skills was covered (yes/no, but not

how much). Nevertheless, the fact that they were perceived as being covered at all offers some reassurance.

Coleman, Starfield and Hagan (2003:186) investigated the extent to which staff and students thought studying for IELTS *prepared them well for university study* and the results were mixed. Based on survey returns from 429 students and 195 staff, 35.5% of staff and 60.8% of students thought IELTS Reading had done so; 24.7% of staff and 16.4% of students disagreed; the rest were not sure.

Craven (2012:27) reports on reactions of students to the IELTS Reading test:

As with the Listening component, there was a general feeling that the reading in the IELTS test was considerably easier than the reading that was required for university study, not least because the passages were considerably shorter.

The literature that you read during the semester, it's quite harder ... it's quite above the level of the readings that you read during the Test. (Student #10)

To be very frank, I found ... the reading easy ... The reading stuff was really simple, not using lots of vocabulary ... It was really pretty straight forward question ... I found it really easy. (Student #32)

In spite of the differences these students noted in university reading and IELTS test reading, and in spite of the fact that some of the test items were ones students would not encounter in their studies, many of the students felt they were able to transfer reading skills they had developed as part of their studies to their reading for the test. These skills included skimming for the main idea and scanning for details, guessing words from context and recognising paraphrases.

Mickan and Motteram (2008:18) provide evidence of positive washback from the IELTS Reading test on an IELTS preparation programme in the following teachers' comments:

A central focus for reading was attention to the overall meanings of texts:

It's important to keep reading for meaning.

In discussing the approach to the Reading tasks, she instructed students to read the whole passage to find out what it is about before going through the questions: Look at the meaning of the article ... I would strongly recommend that you read the whole article. She recommended an initial reading to determine the topic: ... read whole article ... reading it for gist, for general information, gist, just to get an idea about the passage ... I read it once ... two, I read it again ... what is the main topic ... general meaning ...

The teacher reminded students to concentrate on the meaning of a paragraph rather than single words. [To two students reading an essay written by the teacher] Are you discussing the meaning of the paragraph or just each word? You need to discuss the paragraph, what do you think it means?

Generally negative washback

Everett and Colman (1999) provide a detailed analysis of eight sets of IELTS test practice materials. They conclude (1999:1):

The study also considered the broader issue of the role of the publications in preparing students for IELTS and in the development of language skills in general, with particular reference to preparation for further study. The study finds that IELTS preparation materials should include more texts and tasks that would contribute to the social and academic acculturation of students.

As regards expository texts (Everett and Colman 1999:24):

There were also a number of expository or argumentative texts . . . The opinion was voiced by some of the teachers interviewed that there was not enough of this type of text in the IELTS preparation materials. There are a number of texts which do include discussion and exposition but, as can be seen from the tables, some sets of reading texts contain very few discussion-type texts.

Issues of style also arose (Everett and Colman 1999:26):

As a general rule, and according to teachers interviewed, some journal and magazine style articles with language sometimes bordering on the colloquial were seen as less appropriate for reading and test practice as the texts of a more academic nature . . . Another feature of journal and magazine articles is the "trendy" language found in many of these texts. Such language may appeal to, and be easily accessible to, native speakers who naturally have a wider appreciation for the nuances of the language, but may confuse and frustrate non-native speakers.

A general satisfaction was expressed with the range of item types included in the materials (Everett and Colman 1999:29): Clearly, the sets of practice reading tests do provide a representative selection of item-types and should provide the learners with an adequate understanding of the expectations of the IELTS test . . . The main task types are repeated in a variety of formats throughout the sets of reading practice tests. There are a number of straightforward tasks requiring scanning for specific information, tasks in the form of tables, cloze activities, short-answer and multiple-choice questions. Tasks requiring surveying for gist or key points occur in most sets of practice tests and include the tasks of matching headings and other matching and multiple-choice questions.

Everett and Colman conclude (1999:41):

Reading texts for IELTS practice should be academic in both content and structure. Rather than focusing on providing a wide range of question types, tests should focus on tasks that both extend and test the language skills students will need in their programs of further study. The topics of the texts should be suitable for EAP classes and should provide material that promotes discussion and interaction within the classroom. Students preparing to take the academic version of IELTS usually have a serious purpose for undertaking current studies, and so should be given as much assistance as possible in their social and academic acculturation. Texts that open avenues for discussion on recent issues and topics, related to tertiary studies, constitute useful teaching resources.

Allan (2016) investigated the impact of the IELTS Academic exam, specifically focusing on washback upon learners' test preparation strategies and score gain, and the mediating factors influencing washback when learners in an English as a Foreign Language (EFL) context are not enrolled in test preparation courses. Allan's (2016) study found that the IELTS test generated positive washback on learning the productive skills in the Japanese tertiary context. Moreover, this appears to have led to an increase in test takers' language proficiency, particularly speaking proficiency. However, Allan suggests that IELTS had limited effects on preparation for the Reading module because students were more familiar with that skill. Allan concluded (2016:16):

The IELTS test in this study provided a stimulus, which oriented students towards the study of productive skills, which is a positive step forward for the test takers' development of a rounded language proficiency in the four skills. While the increase in spoken language proficiency cannot be attributed solely to IELTS test preparation, it does provide an indication of the potential for positive washback on learning from the IELTS test in the Japanese context.

Hayes and Read (2004:110–111) present a mixed picture of the washback of IELTS on preparation courses in New Zealand:

This study showed clear evidence of washback effects in the IELTS preparation course at School A. However, they did not seem to be the kind of positive effects envisaged at the outset of this study, in the sense that the teacher and students were narrowly focused on practice of the test tasks, rather than the development of academic language proficiency in a broader sense. By contrast, the course at School B appeared to address a wider range of academic study needs and to promote the students' general language development.

Further detail is provided in Hayes' (2003) doctoral thesis; overall, many participants in the study focused on preparing specifically for the test and not preparing students for academic study *per se*. One teacher commented (2003:217):

This comment by Teacher C summarised their main concerns:

It is just so focused on the exam. It is intense. You don't have to do it this way, but I tend to ask them "How many did you get right" sort of thinking of scores I suppose. I think it prepared them to sit the IELTS exam. I don't think it does a great deal to prepare them [for university]. I think a longer preparation course is infinitely more valuable . . .

Hayes offers an important caveat (2003:226):

It may be unrealistic to expect IELTS to have the kind of positive washback effect [desired] . . . Most earlier research has looked for effects on the teaching and learning of the second language (most commonly, English) in the classroom, but what has been proposed here is that the test should have the effect of reorienting the students' whole approach to academic study. This is probably an unreasonable goal for a three-hour proficiency test to achieve . . .

She concludes (2003:226):

The two "traditional" courses can be considered evidence that IELTS was having a negative effect because the test tasks they focused on were limited in scope as compared to real academic study tasks and they were delivered in a non-communicative, teacher-centred way. However, this may be a narrow perspective, since factors such as the cultural and educational expectations of the students, the commercialisation of education in New Zealand and the way that schools marketed their courses competitively to meet student demand were influencing the outcome. Also, the reliance of tertiary institutions on a single English test score, set at

what is considered by many a rather low level, must be taken into consideration, together with the lack of incentive for students to continue with English language study once they have passed the test. Failed attempts by students to reach the pass standard by relying largely on test strategies prompt students and teachers to realise that more is required, such as grammar development, vocabulary learning and subject knowledge related to typical IELTS topics. This has led to courses such as the one at School B practising a broader range of academic study skills which are not necessarily assessed directly by IELTS tasks.

Impact by design

At the conclusion of his review of IELTS impact studies, Hawkey (2006:163) calls for further IELTS impact studies linking test impact and test performance:

... given evidence of the need for further investigation of the validity of IELTS reading and writing tasks, [other key targets for new research could be based on] observation and/or the views of receiving institution subject lecturers into the nature of reading and writing activities in a range of higher education courses ... linked perhaps to an investigation of candidate reading and writing performances on different reading and writing test tasks [based on the initial enquiry].

As we saw in Chapter 2, the IELTS research programme funded by the British Council and IDP Australia took up Hawkey's recommendations. Two research studies, one on the relationship between the academic reading construct as measured by IELTS and the reading experiences of students in the first year of their courses at a British university, and one on the cognitive processes underlying the academic reading construct as measured by IELTS were carried out in 2006–2008 (see Weir et al 2012a, 2012b and Chapter 2 Research Study 3). These provided a clear picture of the types of reading activities that should be included in any new version of IELTS to encourage attention to these real-life reading activities in preparation courses for university study.

In Weir's (2005b) original version of the sociocognitive approach he makes it clear that the impact of the test is not something that is just investigated after the test (though this is clearly important), but must obviously be considered systematically at the test design stage based on a consideration of who the test takers are and what they need to do in their future target situation. Weir (2005b:51) notes: 'Obviously, the tasks themselves will also be constructed with the overall test population and the target use situation clearly in mind as well as with concern for their theory-based [cognitive] validity'.

In preparing the specification of contextual and cognitive parameters for the test based on clear definition of test taker characteristics, the test developer must inevitably consider the impact of these decisions. Saville (2009) builds on this and argues that his 'impact by design' approach builds on Messick's (1996) concept of what might be termed 'washback by design' (seeking validity by design as the basis for positive washback). Saville notes that, unlike much of the research in the literature to date which has been conducted *post hoc* by external researchers (e.g. many of the contributors to Cheng et al (Eds) 2004), the Cambridge English approach now builds in impact considerations from the start, and seeks to anticipate potential effects and consequences with a commitment to monitoring and changing things as required ('anticipatory impact research'). Saville (2009) itemises the conditions for achieving such impact by design:

- a commitment to assessment as a potentially positive component within dynamic educational and societal processes
- an understanding of context within educational systems a multi-level approach with unpredictable interactions between the wider milieu and local contexts/individuals
- · a sociocognitive approach to learning and assessment
- an explicit and evidence-focused approach to construct definition (cf the importance of the construct in washback research e.g. Green 2007)
- a test development and validation model which allows for the planning of activities over time (cyclical and iterative) and for changes/ innovations to be implemented when necessary
- a commitment to ongoing improvements within a Quality Management System (QMS) approach
- a well-developed view of the constituency of stakeholders and how to involve them effectively in the development and validation of the exams
- anticipation of potential impacts (positive/negative) at various levels (micro/macro), i.e. impact hypotheses
- enhanced methods for 'finding out' and communicating with stakeholders
- collection of adequate data and appropriate analyses both routine and as part of specific instrumental projects to find out what is happening
- use of mixed method approaches (quasi-experimental and constructivist models of research)
- development of a 'toolkit' for making data collection more routine and easier to manage, including storage and analysis of qualitative data such as videos
- a 'theory of action' the ability to deal with change and innovation based on the evidence collected and an understanding of what needs to be done to make things better.

Building on the work of Taylor (2000), a clearer understanding of the constituency of stakeholders who need to be taken into consideration in the process of monitoring impact has emerged in recent years. These are clearly identified by Saville (2009) and were later confirmed by O'Sullivan (2015, 2016).



Figure 6.1: Stakeholders in the testing community

Source: Adapted from Taylor (1999)

Endnote

Having reviewed issues of test impact and washback in relation to the IELTS test, and, where available, to the IELTS Academic Reading Module in particular, we would argue that it is possible to design in advance the positive impact/washback of ongoing review and development of academic reading tests, such as the IELTS Academic Reading test. For example, the procedures for future impact by design of language tests are laid out by Saville (2009). In Chapter 7 we turn our attention to explore how IELTS Reading compares to external points of reference by examining research that compares IELTS with EAP tests of 'equivalent' standing.

Criterion-related validity parameters

A test is said to have criterion-related validity if a relationship can be demonstrated between test scores and some external criterion which is believed to be a measure of the same ability. (ALTE 1998)

Introduction

In this chapter our aim is to explore how IELTS Reading compares to external points of reference or criteria, whether in the form of other tests of academic English or of a descriptive framework of levels of language ability or performance. To do this, we shall examine research that compares IELTS with EAP tests of 'equivalent' standing and with reference frameworks such as the CEFR, as well as any research available concerning the predictive validity of IELTS.

The concept of criterion-related validity

The sociocognitive validation framework includes establishing criterionrelated validity as a form of external evidence of fitness for purpose of scores on a test. It is defined as 'a predominantly quantitative and *a posteriori* concept, concerned with the extent to which test scores correlate with a suitable external criterion of performance with established properties' (Weir 2005b:35). The comparison between scores on the test to be validated and the external criterion of performance may be either concurrent or predictive. Concurrent validity is usually examined by 'comparing scores from a given test with some other measure of the same ability of the candidates taken at the same time as the test' (Shaw and Weir 2007:229), whereas predictive validity involves comparing the test scores with an external measure of the same candidate at a later stage, after the candidate has taken the test. The external measures can be varied, e.g., other test scores, rating by teachers (Alderson et al 1995), candidates' self-assessment (Shaw and Weir 2007), or candidates' course results (Alderson et al 1995).

As the importance attached by test users to test comparability information has increased in recent years, test providers have had to pay greater attention to issues of cross-test comparability – in terms of both the relationship between their own tests and the relationship with tests offered by other examination boards. The ability to relate different tests to one another in meaningful ways provides testers with criterion-related evidence to use for comparability purposes.

We will examine the criterion-related validity of IELTS in detail from the four perspectives listed below:

- cross-test comparability (qualitative and quantitative)
- · equivalence with different versions of the same test
- · comparability with external standards
- comparability with a future criterion performance.

Cross-test comparability (qualitative): The General Medical Council (GMC) Project

Taylor (2004:2) argues that test users want to know how one test compares with other available tests which claim to perform a similar function. University admissions officers want to know how to deal with students who present them with TOEFL, IELTS, CAE (now C1 Advanced) or CPE (now C2 Proficiency) scores; employers need to know how to equate different language qualifications presented by job applicants; educational institutions, teachers and students have to choose which test to take from those on offer. Taylor (2004) points out that there have always been informal as well as formal attempts to compare language proficiency measures; traditionally, comparisons have tended to focus on whether the scores or grades from two different tests are comparable with one another. Bachman et al (1995) emphasised that any comparability study needed to take account of more than just score equivalences. They point out that it must also investigate comparability of test content and performance, a point to which we will return below.

In 2014, Centre for Research in English Language Learning and Assessment (CRELLA), because of its position as the leading university specialist testing centre in the UK, was invited by the UK General Medical Council (GMC) to propose a study to review the suitability and comparability of various English language tests as measurement instruments for providing the GMC with evidence of the English proficiency of internationally trained doctors. The point of comparison was the IELTS academic version, which is the sole test used by the GMC for this purpose. Taylor and Chan (2015:43–46) and the CRELLA team, each in their own specialist testing area (reading/listening/writing/speaking), provided subjective judgements, with reference to a range of sociocognitive parameters, of the comparability of IELTS and a number of potential alternative tests for use by the GMC. These included TOEFL iBT, Cambridge English CAE and CPE, the Pearson Test of English (PTE – Academic), the Occupational English Test (OET), the Michigan Examination for the Certificate in Proficiency in English (ECPE),

the Taiwanese General English Proficiency Test (GEPT Advanced) and City & Guilds International ESOL (IESOL). These tests were selected as potentially appropriate alternatives from a list of 45 English language proficiency tests available at the time. The selection was made on the basis of 13 essential criteria specified by the GMC.

Stephen Bax and Cyril Weir were responsible for the reading part of the report. They describe the process they used to arrive at their decisions on test comparability (Taylor and Chan 2015:42):

Process

In our review of potential alternatives to the IELTS Reading test we therefore examined the types of reading and the cognitive processes they give rise to . . . Data on each of these was entered into a pro-forma template, which was then completed in turn for each of the reading tests examined, including IELTS . . . The process we followed began with a comprehensive literature search for published information, in particular for independent research where available, relating to all aspects of the test in question, in order to gather information for each category in the proforma. We also examined samples of actual tests in order to verify the published claims, or to obtain more detailed information. For example, we drew on published descriptions of each test in order to complete the sections on *Skills Focus* and *Task Description, Timing*, and so on, and checked these against samples of each test.

In categories where no published information was available for a particular test (say, with respect to *Cognitive processing, Goal setting*) two analysts independently evaluated the test in question to reach a judgment. In those cases where a numerical analysis of contextual parameters was needed (e.g. *Readability* and *Lexical Level*), published samples of reading passages from each test were analysed using computer tools such as Compleat Lexical Tutor (www.lextutor.ca), Coh-Metrix (www.cohmetrix.com) and Text Inspector (www.textinspector.com) so as to obtain, for example, lexical and syntactic complexity measures for each test. Given that the published samples of reading texts for certain of the tests are relatively few in number, these figures should be considered indicative rather than authoritative.

Outcomes

When the pro-forma had been completed for each of the reading tests, an expert focus group of three reading specialists reviewed the information for each test in each category and came to a judgement concerning the relative difficulty of the text in each of the categories in comparison with the IELTS Reading paper. The focus group then drew on the individual analyses to form a judgement concerning the relative difficulty of each reading test as a whole, in comparison with IELTS.

The result can be seen in Table 7.1, with the rationale summary following on in Table 7.2.

Table 7.1: Comparison of the IELTS Reading test with the other tests under consideration

Reading test	СРЕ	CAE	PTE –	TOEFL	OET
			Academic	IDI	
Evaluation	>	=	<	=	>
Task features:					
Skills focus	=	=	=	=	=
Cognitive processing:	>	=	=	=	=
Goal setting					
Cognitive processing:	>	=	=	=	=
Levels of reading					
Viewed a second (dimension	_	_		_	
word count/time	_	_	_	_	_
Domain	=	=	=	=	=
Discourse mode	=	=	=	=	=
Content knowledge	=	=	<	<	>
Cultural specificity	=	=	=	=	=
Nature of information	>	>	=	=	=
Presentation	=	=	=	=	=
Lexical level:					
Cumulative coverage	=	=	=	>	>
Academic Word level	<	<	=	=	>
Lexical density	=	=	=	=	=
Mean number modifiers per noun	=	=	=	=	=
phrase					
Mean number of words before main	=	<	<	=	<
verb in main clause		_		_	_
Grammatical level	(_	<u>`</u>	_	_
Average sentence length	>	<	=	<	=
Readability	<	<	=	=	>
Topic	=	=	=	=	>
Rhetorical organisation	=	=	=	=	=
Intended writer/reader relationship	=	=	=	=	>
Reliability ¹ : Estimate of internal cons	sistency of t	test items			
KR20	Less reliable	Less reliable	More reliable	Less reliable	Equally reliable

(Key: = comparable with IELTS; > more demanding than IELTS; < less demanding than IELTS)

¹ An estimate of the reliability of a reading or a listening test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a reading test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the reading/listening test as a whole. Commonly used statistics for computing and reporting reading test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.

Reading test	GEPT Advanced	Michigan ECPE	IESOL and ISESOL C1	IESOL and ISESOL C2
Evaluation	>	<	<	<
Task features:				
Skills focus	>	=	=	=
Cognitive processing:	>	<	<	<
Goal setting		_	_	_
Levels of reading	>	=	=	=
Text features:				
Word count/time	>	<	<	<
Domain	=	=	=	=
Discourse mode	=	<	<	<
Content knowledge	=	=	<	<
Cultural specificity	<	=	=	=
Nature of information	=	=	=	=
Presentation	=	=	=	=
Lexical level:				
The cumulative coverage	>	>	>	>
Academic Word level	=	=	<	<
Lexical density	=	=	=	=
Mean number of modifiers per noun phrase	=	<	<	=
Mean number of words before main	=	<	<	<
Grammatical level	=	>	=	=
Average sentence length	=	<	<	<
Readability	=	<	<	<
Торіс	=	=	=	=
Rhetorical organisation	=	=	=	=
Intended writer/reader relationship	=	=	=	=
Reliability: Estimate of internal consiste	ency of test ite	ems		
KR20	Equally reliable	Equally reliable	Less reliable	Less reliable

Table 7.1 (Continued)

Key: = comparable with IELTS; > more demanding than IELTS; < les	5 5
emanding than IELTS)	

In short, the GEPT Advanced Reading Test in Taiwan and the Cambridge English CPE appear to be more demanding than IELTS in terms of cognitive processing (task features). The evidence we presented in Chapter 3 confirms that IELTS Reading tasks do not seem to match some of the greater cognitive demands made on undergraduates in their reading activities. In terms of contextual validity (text parameters), it is only the OET that appears to be more demanding than IELTS and the differences largely arise because of the

CPE >	CAE =	PTE-A <	TOEFL iBT =	0ET >
The expert focus group judged the CPE Reading test to be more demanding overall than	The expert focus group judged the CAE Reading test to be	Although comparable to IELTS on many measures, the PTE – Academic reading test	The expert focus group judged the TOEFL iBT reading test to be	The expert focus group judged the OET reading test to
the IELTS Reading paper. This is unsurprising given that CPE	comparable to IELTS on all of the main	was judged by the expert focus group to be less demanding	comparable to IELTS on all of the main indices	be more demanding overall than the IELTS
is designed to target a <i>single</i> <i>proficiency level</i> – the C2 level of	indices examined. In terms of two lexical	than IELTS in overall terms. This was due to the fact that	examined. In terms of average sentence length of	Reading paper. This is mainly due to the
the CEFR; therefore, the reading test intentionally reflects texts	measures (i.e. <i>Academic</i> <i>Word level</i> and <i>mean</i>	on a number of important indices, including the <i>time</i>	the reading passages, the content knowledge required,	specific knowledge of healthcare-related
and tasks that are appropriate to this specific level. By comparison,	number of words before main verb in main	<i>duration</i> of the test (maximum of 41 minutes as compared	and <i>cultural specificity</i> , it was adjudged to be less	matters which the test requires of candidates.
IELTS is currently designed to target a range of levels from	<i>clause</i>), it was adjudged to be less demanding	to IELTS' 60) and the <i>length</i> of reading material provided	demanding than IELTS, and in one measure (<i>lexical</i>	This need for specialised knowledge
CEFR B1 to C2. In particular,	than IELTS, and on	(a maximum of 1,130 words	coverage of the texts	is also reflected in, for
the CPE was considered more demanding than the IELTS	one measure (the nature of information)	as compared with IELLS 2,750), as well as several	examined) it was adjudged more demanding. However,	example, some of the more demanding lexis
Reading test in terms of: the	it was adjudged	other indicators, the demands	on balance, and taking into	used, as well as in the
cognitive processing required	more demanding. On balance and taking into	placed upon candidates were adjudged to he less overall than	account all of the indices, the TOEFL reading and	overall readability and
<i>information</i> in the texts; and the	account all the indices,	in the IELTS Reading paper.	IELTS Reading tests were	to pro micasance.
grammatical level and the average	the CAE Reading and IET TS Deading tests	It should, however, be noted	adjudged to be comparable in tarms of the damands	
inputs. Two measures (namely	were adjudged to be	reliability (internal consistency)	they place on the candidate.	
proportion of academic words	comparable in terms of	for PTE - Academic produced		
and <i>readability</i>) were adjudged less demanding than IELTS.	the demands they place on the candidate.	the highest estimate of all the reading tests analysed.		
But on balance, and taking into account all the contextual indices				
and the cognitive demands, the				
CPE Reading appeared more demanding than IELTS Reading.				

Table 7.2: Reading – rationale summary
GEPT Advanced >	ECPE <	IESOL (C1) <	IESOL (C2) <
The expert focus group judged the GEPT Advanced Reading Test to be more demanding overall than the IELTS Reading paper. In particular, the GEPT was considered to be more demanding than the IELTS Reading test in terms of: the skills coverage; cognitive processing required of the condidates (as well as careful reading GEPT has a dedicated paper for expeditious reading); the far greater amount of text that has to be processed; and the cumulative lexical coverage. The GEPT Advanced Reading than IELTS Reading. It should be noted that GEPT is a level-specific test targeted at C1. It was equally reliable.	The expert focus group judged the Michigan ECPE reading test to be equivalent to IELTS in 12 of the main indices examined. In terms of cognitive processing: goal setting; word count; discourse mode; two of the lexical measures; the average sentence length of its reading passages and their readability, it was adjudged to be less demanding than IELTS. In terms of the lexical coverage of the texts examined and grammatical level it was adjudged more demanding, but on balance, and taking into account all of the indices, the Michigan ECPE reading test was adjudged to be less demanding the IELTS. It was equally reliable.	The City and Guilds Cl reading test, although equivalent with IELTS in terms of 12 of the parameters, was adjudged by the expert focus group in overall terms to be less demanding than the IELTS equivalent. This was due to the fact that according to a number of important indices, including the word count/time duration of the test; the length of reading material; cognitive processing: goal setting; discourse mode; content knowledge; three of the lexical measures including academic word coverage; average sentence length and readability, the demands placed upon candidates were adjudged to be somewhat less overall than in the IELTS Reading paper. It was also less reliable.	The City and Guilds C2 reading test, although equivalent with IELTS in terms of 13 of the parameters, was adjudged in overall terms to be less demanding than the IELTS equivalent (which is surprising given that the test providers claim it is targeted at the C2 level). This was due to the fact that in a number of important indices, including the word count/ time duration of the test; the length of reading material; cognitive processing: goal setting; discourse mode; content knowledge, two of the lexical measures including estimence length and readability, the demands placed upon candidates were adjudged to be somewhat less overall than in the IELTS Reading paper. It was also less reliable.

Table 7.2 (continued)

subject-specific nature of the texts it employs. Unless a decision is taken in the light of digital advances to go back to the 1980s and provide subject-specific reading modules as in the ELTS test (see Weir and O'Sullivan 2017), the discrepancies between OET and IELTS need not trouble us unduly (see also the arguments for a single test in Chapters 1, 2 and 3). In terms of our earlier concerns about the ease of IELTS and the low incidence of academic vocabulary as compared to undergraduate texts, the alternatives to IELTS fared even less well by comparison in this research.

In order to probe more deeply into these comparisons, we searched the literature for evidence of more finely nuanced concurrent validity investigations. We could only find a more complete set of data in the literature on one test, which we felt to be a valid alternative to IELTS: CAE. The aim was to see if there was any further indication, in more finely grained empirical data, of changes that might need to be made to the cognitive and contextual parameters of future academic reading tests. The expectation was that key tests of academic reading should be making comparable demands on candidates. Any differences might provide food for thought in terms of need for change. For the contextual parameters, our benchmark was the data arising from the research carried out on undergraduate reading texts by Green et al (2010).

Cross-test comparability (qualitative and quantitative): Concurrent comparison of CAE, IELTS and undergraduate texts

Launched in 1991, the Certificate in Advanced English (CAE, from 2019 known as C1 Advanced) is a widely trusted high-quality test of English for academic and professional purposes. This is demonstrated by the global acceptance of the test by more than 3,000 organisations including education institutions, governments and employers. The test was updated in 1999 and 2008 to keep pace with changes in language teaching and testing, and a further revised version will launch in 2020. It is available in both paper-based and computer-based versions in over 1,300 centres in 113 countries.

C1 Advanced focuses on Level C1 of the CEFR – generally considered as the level required for academic and professional success. It covers all four language skills (reading, writing, listening and speaking) plus knowledge of grammar and vocabulary (in just under five hours). The Reading component (75 minutes) tests the ability to deal with different types of reading text as well as the knowledge and control of the English language. The Writing component (90 minutes) tests the ability to write a variety of text types. C1 Advanced is owned and produced by Cambridge Assessment English (Cambridge English), part of the University of Cambridge. C1 Advanced is increasingly recognised as evidence of English language skills for admission to higher education. According to the Cambridge English website, almost every UK university now accepts C1 Advanced for this purpose.

Green et al (2012) situate CAE in relation to the IELTS texts and the small corpus of essential undergraduate reading texts assembled for the IELTS research project (Weir et al 2012a). Texts for analysis (see Table 7.3) comprised 49 texts from CAE exams, the collection of 14 IELTS Academic Reading tests (42 texts) supplied by Cambridge ESOL (as Cambridge English was then known) for the 2005/6 British Council-funded IELTS Joint Research Committee study reported in Weir et al (2012a) and Green et al (2010), and 14 key undergraduate academic texts (42 extracts) identified by the University of Bedfordshire (UoB) for the same study.

 Table 7.3: Number of texts by source

42
42
49

Green et al (2012) identified 17 indices that appeared relevant to classifying texts by level and then began the next phase of the study, which involved comparisons between CAE, IELTS and undergraduate texts. They were able to map the levels of text difficulty operationalised in the Cambridge ESOL tests to the features of undergraduate texts identified in Green et al's (2010) study. It also provided an indication of whether CAE texts might better reflect features of academic text than do IELTS texts.

A Kruskal–Wallis analysis was used for the three-way comparison. The results are displayed in Table 7.4. Following the Kruskal–Wallis analysis, box-plots were generated to aid interpretation and to provide a graphic representation of the relationship between each of the three text sources and the characteristics under consideration. These showed the range of values for each index for each text source (see Figures 7.1–7.3).

For each set of indices, Weir et al (2012a) identified which displayed significant differences across the three indices (p<.05) and for which of these the IELTS or CAE texts more closely resembled the undergraduate texts in the UoB mini-corpus – texts representing essential reading for new undergraduate students.

Lexical indices

On the key lexical indicators, CAE texts did not generally match the features of academic texts as closely as IELTS. The shorter, more frequent words on CAE suggested that these texts would make easier reading. The only lexical index on which CAE texts appeared closer to the undergraduate norm was

		-								D			
		CAE (n=49)			IELTS	(n=42)		l	Jndergradu	late (n=42		Kruskal– Wallis test
Lexical	Min	Max	Mean	Std. Dev	Min	Max	Mean	Std. Dev	Min	Max	Mean	Std. Dev	
38 Average syllables per word 42 Higher level constituents per	$1.28 \\ 0.61$	$1.76 \\ 0.80$	$1.54 \\ 0.72$	$0.11 \\ 0.04$	$1.45 \\ 0.67$	1.83 0.76	$ \begin{array}{c} 1.66 \\ 0.71 \end{array} $	0.09 0.02	$1.46 \\ 0.65$	2.07 0.81	$1.72 \\ 0.72$	$0.14 \\ 0.04$	$0.000 \\ 0.100$
word 44 Type-token ratio for all content	0.59	0.95	0.75	0.08	0.60	0.77	0.68	0.04	0.51	06.0	0.65	0.08	0.000
46 Celex, logarithm, mean for content words (0-6)	1.94	2.48	2.21	0.15	1.94	2.37	2.17	0.11	1.79	2.48	2.14	0.15	0.144
AWL Off-list	0.00	5.25 3.89	$1.64 \\ 1.05$	$1.41 \\ 0.91$	0.00	5.25 3.89	$1.71 \\ 1.02$	$1.48 \\ 0.91$	$0.12 \\ 0.41$	22.22 11.62	9.36 3.98	4.92 3.06	0.000
Syntactic 27 LSA, Sentence to sentence	0.07	0.32	0.15	0.05	0.10	0.45	0.23	0.07	0.11	0.45	0.26	0.08	0.000
aujacent mean 37 Average words per sentence 41 Mean number of modifiers per	$13.97\\0.54$	29.08 1.37	20.42 0.92	$4.01 \\ 0.20$	16.11 0.71	34.71 1.13	21.89 0.98	$3.73 \\ 0.09$	$13.76 \\ 0.64$	$30.30 \\ 1.24$	21.47 0.95	4.26 0.15	$0.165 \\ 0.149$
noun purase 43 Mean number of words before the main verb of main clause in	2.09	7.41	4.40	1.30	3.32	10.08	5.48	1.35	1.86	7.97	4.59	1.39	0.000
sentences 56 Sentence syntax similarity, all, across paragraphs	0.05	0.12	0.08	0.02	0.06	0.10	0.08	0.01	0.05	0.10	0.07	0.01	0.002
Coherence and concreteness 16 Argument overlap, adjacent, unwoichted	0.10	0.88	0.45	0.16	0.24	0.86	0.52	0.15	0.26	06.0	0.56	0.15	0.003
unweighted 18 Anaphor reference, adjacent,	0.11	0.79	0.41	0.16	0.05	0.59	0.25	0.11	0.00	0.62	0.24	0.13	0.000
21 Anaphor reference, all distances 26 Logical operator incidence score 58 Proportion of content words that overlap between adjacent	0.03 16.71 0.01	0.55 66.27 0.14	0.20 38.58 0.07	$0.12 \\ 11.42 \\ 0.03$	0.02 22.64 0.03	0.22 64.56 0.16	$\begin{array}{c} 0.08\\ 39.28\\ 0.08\end{array}$	0.05 9.65 0.03	$\begin{array}{c} 0.00\\ 20.54\\ 0.04\end{array}$	0.39 71.43 0.17	$0.10 \\ 46.14 \\ 0.10$	0.09 11.78 0.03	$\begin{array}{c} 0.000\\ 0.004\\ 0.000\end{array}$
sentences 60 Concreteness, minimum in sentence for content words	158.00	223.00	167.55	16.72	158.00	195.00	171.83	15.57	158.00	208.00	168.83	15.87	0.447

Table 7.4: Results of independent-samples Kruskal–Wallis test: CAE compared with IELTS and undergraduate textbooks





Figure 7.2: Syntactic indices



43 Mean number of words before the main verb of main clause in sentences







58 Proportion of content



Off-list. Here the gap between IELTS and CAE was only marginal (IELTS mean = 1.02, SD = 0.91; CAE mean = 1.05, SD = 0.91) while the undergraduate texts had a much higher percentage coverage of Off-list words than either test, but also greater variation (mean = 3.98, SD = 3.06). This probably reflects the more technical subject specificity of the undergraduate texts, which, unlike the two tests, do not have to be equally accessible to students of a range of disciplines. It is noticeable that, although IELTS texts were very marginally closer, neither IELTS (mean = 1.71, SD = 1.48) nor CAE texts (mean = 1.64, SD = 1.41) approached anywhere near the percentage coverage of AWL words characteristic of academic texts (mean = 9.36, SD = 4.92). This is a key area which, in our view, should be rectified in the development of any future IELTS revision.

Syntactic indices

On the syntactic indices, CAE again generally appeared to resemble academic texts less closely than IELTS. The exception was Coh-Metrix 43 Mean number of words before the main verb of main clause in sentences. On this index, CAE texts (mean = 4.40, SD = 1.30) quite closely reflected the undergraduate norm (mean = 4.59, SD = 1.39), but IELTS texts (mean = 5.48, SD = 1.35) appeared substantially more complex on this measure.

Text representation indices

On the text-level representation indices, significant differences (p<.05) were found across text sources in all but one case (Coh-Metrix 60 Concreteness, minimum in sentence for content words). IELTS texts were again found to more closely resemble the academic texts. The relatively high scores for the academic texts on indices like Coh-Metrix 58 Proportion of content words that overlap between adjacent sentences (mean = 0.10, SD = 0.03) – compared with CAE (mean = 0.07, SD = 0.03) – suggests that the undergraduate texts possessed greater lexical cohesion than the CAE texts.

Taken together, these findings suggest that the CAE texts did not reflect the nature of academic texts (as represented by the UoB mini-corpus of undergraduate textbooks) as closely as the IELTS texts. The study suggests that selected automated text analysis tools, if employed at the text selection stage, could help the test developers to more closely match the level of the texts employed in CAE to the demands of the academic texts that prospective students might encounter on entering university.

Table 7.5 shows the score comparison between IELTS and *Cambridge English: Advanced* (as CAE was known then) that is presented by Cambridge English Language Assessment (as Cambridge English was known then) (2013:3) and they describe the empirical procedures upon which this comparison is based (2013:3–4) as follows:

IELTS (overall band score)	Cambridge English: Advanced (CAE)
8	80
7.5	74
7	67
6.5	58
6	52
5.5	47
5	41

Table 7.5: Comparing scores on IELTS and CAE

In 2009, we undertook to benchmark C1 level as represented by *Cambridge English: Advanced (CAE)* against *IELTS* scores. For this exercise an empirical validation study was undertaken where registered *IELTS* candidates were invited to also take *Cambridge English: Advanced (CAE)*, and registered *Cambridge English: Advanced (CAE)*, and registered *Cambridge English: Advanced (CAE)* candidates were invited to take *IELTS*, and their scores compared. This counterbalanced design accounted for preparation or motivation-related effects on one test or the other.

For receiving institutions, Cambridge English Language Assessment suggests (2013:4):

Candidates who have secured a *Cambridge English: Advanced (CAE)* grade C are at Level C1 of the Common European Framework of Reference and can be expected to be comparable in ability with candidates who have secured 6.5 or 7.0 in *IELTS*. Where institutions have specified a minimum *IELTS* requirement of 7.0, reference may be made to the standardised score, and a minimum requirement of 67 imposed.

Concurrent equivalence with different versions of the same test

When new versions of an examination are produced, as well as ensuring equivalence in terms of the cognitive and contextual features of reading examinations, examination boards need to adopt an item banking approach underpinned by Item Response Theory (IRT) if they are to demonstrate acceptable statistical equivalence with different versions of the same test.

Test equivalence is established if 'a relationship can be demonstrated between test scores obtained from different versions of a test administered to the same candidates under the same conditions on two different occasions' (Weir 2005b:208. See also Mislevy 1992 for a comprehensive discussion of this area). The ALTE *Multilingual Glossary of Language Testing Terms* (1998:144) offers the following definition of equivalence in test forms: Different versions of the same test, which are regarded as equivalent to each other in that they are based on the same specifications and measure the same competence. To meet the strict requirements of equivalence under classical test theory, different forms of a test must have the same mean difficulty, variance, and co-variance, when administered to the same persons.

The American Educational Research Association, American Psychological Association and National Council on Measurement in Education (1999) further refine the test equivalence definition. It distinguishes between: parallel forms, which should demonstrate equivalence in raw score means, standard deviations, and correlations with other measures for a stated population; equivalent forms, where score conversion techniques or 'form-specific norm tables' are used to compensate for differences in raw score statistics between test versions; and comparable forms, which are very close in terms of content but where the extent of statistical similarity remains unproven. For test providers, of course, it is vital to achieve as complete as possible equivalence across alternate forms of the same test which are produced on different session dates to meet the needs of test users.

Taylor (2004:2) noted that: '...Cambridge ESOL produces different versions – also known as "alternate" or "parallel" forms – of the same examination to be taken on different session dates throughout the year; examinations must clearly be equivalent from session to session in terms of their content coverage and measurement characteristics'. The multilingual glossary (ALTE 1998) notes that equivalence is very difficult to achieve in practice and that considerable effort and expertise goes into ensuring examination equivalence through the implementation of a comprehensive set of standard procedures applied at each stage of examination production (see Saville 2003).

Dictionary of Language Testing by Davies et al (1999) offers a similar definition for equivalence to the one given above and goes on to mention the increasingly common use of IRT analysis and item banking to help with the process of creating equivalent forms. Cambridge Assessment English attempts to achieve version equivalence through establishing a common measurement scale, its item banking system and its item writing and pretesting procedures. In Cambridge English Reading papers, question paper production (QPP) is based on the Local Item Banking System (LIBS), which is a computer-based management and analysis tool developed by Cambridge ESOL to handle the entire production cycle.

Item banking is an application of IRT (Bond and Fox 2001, Wright and Stone 1979). It involves assembling a bank of calibrated items – that is, items of known difficulty. Designs employed for collecting response data ensure a link across items at all levels. The Cambridge ESOL Common Scale, a single measurement scale covering all Cambridge English levels, has been constructed with reference to these objective items. The Cambridge English Common Scale enables different testing events to be related within a single frame of reference, greatly facilitating the development and consistent application of standards.

Items are pretested in specially constructed papers, which include anchor items. Because the anchor items are of known difficulty, the analysis that is carried out on the pretest responses allows the new items to be calibrated onto a logit scale. This logit scale is re-scaled to produce a conventional Cambridge English scale, which is used for examination construction purposes. It is this scale which underpins LIBS. Examinations are constructed from the calibrated tasks in the item bank. Each task, therefore, consists of items of measured (Rasch) difficulty, which are selected from within a specified range to determine the mean difficulty of the task.

Figure 7.4 shows how constructing a single measurement scale requires all the item response data to be linked in some way. Two ways of achieving this are *common person* linking, where a group of learners might, for example take examination papers at two different levels, and *common item* linking, where different examinations contain some items in common. The latter is the basic approach used in pretesting, where each pretest is administered together with an anchor test of already calibrated material.



Figure 7.4: Item banking approach to scale construction

Source: Adapted from Jones and Saville (2007)

Research and Practice in Assessing Academic Reading

Figure 7.4 pictorially represents the basis for the construction of parallel forms of the examinations at the different levels of the system. The common scale which underpins the item bank is based on Rasch scaling, which means that the paper construction must hit a target for the average difficulty of the paper using the difficulty estimates obtained from pretesting.

In operational test production, new items are pretested in specially constructed papers which include anchor items. Because the anchor items are of known difficulty, the analysis that is carried out on the pretest responses allows the new items to be calibrated and linked to the common scale. Using this approach to model the predicted facility of new items, the test construction team is able to make a judgement on how examination tasks are likely to perform under live examination conditions. It also allows for a comparison between the 'live' values and those modelled at the time when analysis takes place for grading the examination.

The standard operating procedures for test construction ensure that tasks selected for IELTS Reading papers fall within the specified range of difficulty and achieve the targeted average for the paper as a whole. These procedures help to ensure comparability of difficulty and maintenance of standards across different versions of the paper and between different administration sessions.

Following our discussion of cognitive and context aspects of validity in Chapters 3, 4 and 5, it must be stressed that high indices of alternate-form reliability alone do not necessarily yield a significant meaning unless supported by evidence of comparability in other aspects of validity as well. For example, inconsistent context or cognitive validity across examination forms may influence examination scores, resulting in bias against particular cohorts as a consequence and affecting examination fairness.

Comparability with external standards

There is increasing pressure on test providers and examination boards to link their examinations to a particular external standard, namely the CEFR. Indeed, within a relatively short period of time the CEFR has become highly influential in Europe and beyond. Khalifa and Weir (2009:202) report that:

Beyond Europe, Education Ministries and other national agencies are increasingly preoccupied with international benchmarking. In Taiwan, for example, the Ministry of Education in 2005 set about establishing a common standard of English proficiency through the adoption of the CEFR. The CEFR is now the external benchmark for its English language examinations and it is compulsory for agencies offering such examinations to provide evidence of linking to the CEFR. Public English language examinations in Chile, China, Colombia, Korea and Japan are investigating similar linking procedures. The United Nations have adopted the CEFR levels for training purposes at UN institutions around the world. In Canada the CEFR has been adopted for national standards in foreign languages, alongside the indigenous framework of the Canadian Language Benchmarks (CLB)... The Council of Europe has attempted to facilitate the alignment process by providing a toolkit of resources, including a draft pilot Manual for relating language examinations to the CEFR and a technical reference supplement to this (Council of Europe 2003, 2004).

Khalifa and Weir (2009:Chapter 7) describe the procedures by which Cambridge English has linked all of its examinations closely to the levels laid out in the external internationally accepted framework of the CEFR. IELTS is linked to this external CEFR standard through the comprehensive and rigorous procedures of familiarisation, specification, standardisation and empirical validation as advised in the CEFR linking documentation (Council of Europe 2003). The CEFR level system provides an interpretative frame of reference for all its examinations including IELTS. These European levels have the advantage of according with the 'natural' proficiency levels familiar to teachers and are supported by the work of the Council of Europe over the last 30 years; this important work is based on a consensus view that adequate coverage is afforded by six broad levels for the purposes of organising language learning, teaching and assessment in the European context (Council of Europe 2001:22–23).

The scale of levels, which is used by Cambridge English, provides a set of common standards and is the basis of the criterion-referenced approach to the interpretation of examination results. Referencing to the criterion is undertaken by means of scalar analyses using the Rasch model to relate the results from the whole range of Cambridge English Qualifications to the global scale of common reference levels of the CEFR (2001:24). The criterion scale provides representations of the external reality, which helps to ensure that the test results are as meaningful and as useful as possible to the key stakeholders (the candidates, their sponsors and other users of examination results).

The CEFR refers to six levels (Council of Europe 2001) for L2 learners of English. A1 and A2 levels describe the ability to read basic or straightforward information in a known area very slowly i.e. very simple sentences or very short predictable texts. B1 level describes the ability to comprehend texts that consist of familiar or high-frequency everyday language: 'understand routine information and articles, and the general meaning of non-routine information within a familiar area' (ALTE Can Do statements, Council of Europe 2001:252); and that scanning for specifics introduces a variety in reading purpose and style and speed of reading for the first time. At B2 level, readers start focusing on the content of texts (e.g., the ideas presented, the writer's attitude, etc.). From Level B2 readers are also expected to be able to process text quickly as well as efficiently i.e. expeditious as well as careful reading is

expected. C1 and C2 levels characterise more mature, proficient readers. Such readers are able to process more abstract texts with structurally and semantically complex language. At C2 the expectation is that candidates can understand everything they read 'including the finer points of complex texts ... complex ideas expressed in complex language' (ALTE Can Do statements, Council of Europe 2001:251–254).

Taylor (2004:2) describes how:

... in 1998 and 1999 internal studies examined the relationship between IELTS and the Cambridge Main Suite Examinations, specifically CAE (C1 level) and FCE (B2 level). Under test conditions, candidates took experimental reading examinations containing both IELTS and CAE or FCE tasks. Although the studies were limited in scope, results indicated that a candidate who achieves a Band 6.5 in IELTS would be likely to achieve a passing grade at CAE (C1 level). Further research was conducted in 2000 as part of the ALTE Can Do Project ... in which Can Do responses by IELTS candidates were collected over the year and matched to grades; this enabled Can Do self-ratings of IELTS and Main Suite candidates to be compared. The results, in terms of mean "Can Do self-ratings", further supported placing IELTS Band 6.5 at the C1 level of the CEFR alongside CAE.

The 2004 conceptual framework was subsequently revised to accommodate IELTS more closely within its frame of reference. Figure 7.5 illustrates how the IELTS band scores, Main Suite, BEC and BULATS examinations were believed to align with one another and with the levels of the CEFR. Note that the IELTS band scores are the overall scores, not the individual module scores.

Taylor (2004:3) noted that this alignment between examinations is based not only on internal research at Cambridge English but also the 'long established experience of examination use within education and society, as well as feedback from a range of examination stakeholders regarding the uses of examination results for particular purposes'. It will continue to be refined as further evidence is generated.

A final caveat is in order. In many ways the CEFR specifications are extremely limited in their characterisation of reading ability at the different levels and they need to be more explicit for testing purposes (see Khalifa and Weir 2009 for discussion of their shortcomings). The work on more closely defining cognitive parameters for reading in Chapter 4 and contextual parameters in Chapter 5 of this volume complements the comprehensive coverage of these areas for Cambridge English Qualifications (Khalifa and Weir 2009), and the findings in both will help make good some of the current deficiencies in the CEFR characterisation across the different levels. This will further ensure that the examinations linked to the CEFR through



Figure 7.5: Alignment of Cambridge English Qualifications and tests with the CEFR

statistical calibration are also comparable in terms of cognitive and contextual attributes.

Comparability with a future criterion performance: Predictive validity

Predictive validity entails the comparison of test scores with another measure of the ability of interest for the same candidates taken some time after the test has been given (Alderson et al 1995). This other measure may consist of scores from some other test (not necessarily language, e.g., degree results), or candidates' self-assessments of their language abilities, or ratings of the candidate by teachers, subject specialists, or other informants (see Alderson et al 1995, Cotton and Conrow 1998, Criper and Davies 1988, Davies 1990, Hill, Storch and Lynch 1999 for exemplification of this). Banerjee (2003) provides a full and innovative discussion of the whole area of predictive validity and argues for considering the external criterion in terms of cost to students and other stakeholders if admitted at various levels of test performance, as against degree of overlap with the criterion course result. Predictive validity is, however, in general beset with difficulties because of the problems of truncated samples and the likelihood of variables that may interfere with the comparison over time.

Daller and Phelan (2013:177–178) explain that low correlations are often a result of truncated samples and a low range of scores being correlated. They argue that those below Band 6 on IELTS are not normally accepted for universities and therefore do not appear in these studies. As a result, the range of available scores is seriously restricted. They add:

Ferguson and White (1993) call this the pre-selection problem: "The prior use of the test to exclude some potential university students leaves one with a truncated sample where the range of scores . . . is curtailed. The likely effect is a depressed validity coefficient" (p. 16). The range of scores is quite important from a purely mathematical point of view. If the variance in one variable is low, any correlation with another variable will be low as well. This reduces the predictive validity of any variable with a low variance. If all scores of an entry exam, for example IELTS scores, are the same then the correlation between these scores and any measure of study success will automatically be zero. This might be the reason that there is only a low or no correlation between the scores of the standardized tests and GPA [Grade Point Average]. Where there is a wide range the predictive power of the standardized test can be quite high (for example Woodrow 2006 . . . but this is not always the case (Feast 2002) . . .

Ingram and Bayliss (2007:5) point out that low correlations in predictive validity studies are the result of the presence of confounding variables (see also Bridgeman, Cho and Di Pietro 2016). Ingram and Bayliss argue (2007:5): 'most predictive studies based on language tests . . . can be criticised on the grounds that it is impossible to account for all the variables'. Cotton and Conrow (1998:72) conclude that there is still a good deal of disagreement on the extent of the relationship between IELTS performance and academic outcomes in the literature: 'the mixed findings of these studies suggest that the relationship between proficiency in English and academic outcome is more ambiguous than one might suppose'.

In reviewing previous comparability studies Weir, Chan and Nakatsuhara (2013:2–3) suggest that:

... in order to achieve more meaningful results of predictive validity, the present study used a variety of more relevant external measures of test takers' performance, for example, their performance on individual written assignments, tests, and examinations on their course work.

Weir et al (2013:28–29) then report on the correlation that was obtained between IELTS performances and real-life performances in their study (see Table 7.6). Their earlier suggestion that richer dependent variable data might enhance correlations is borne out:

Data were collected from 171 students who were studying on a full-time, collaborative, undergraduate programme at the Business School at a British university... The participants' average IELTS reading and writing bands correlated with their average performances on the four selected real-life tasks at .602 (p<.01). When the IELTS Reading and Writing bands were considered separately, there was a correlation of .558 (p<.01) between IELTS reading and the real-life performances, and a correlation of .406 (p<.01) between IELTS writing and the real-life performances. In terms of variance explained, 31.14% variance of the real-life performance was explained by IELTS reading, 16.48% by IELTS writing, and 36.24% by IELTS reading and writing together. In the field of language testing, the range of .25 to .35 is regarded as typical, good correlations between test performances and later real-life measurements (Pollitt, 1988).

		Real-life performances (two writing tasks, one test and one exam)	Exam	Test	Report	Essay
IELTS Reading	Pearson correlation	.558**	.432**	.458**	.397**	.259**
	Sig. (2-tailed)	.000	.000	.000	.000	.001
	Ν	117	169	138	120	171
IELTS Writing	Pearson correlation	.406**	.376**	.350**	.247**	.164*
8	Sig. (2-tailed)	.000	.000	.000	.007	.032
	Ν	117	169	138	120	171
IELTS Reading and	Pearson correlation	.602**	.506**	.504**	.403**	.266**
Writing	Sig. (2-tailed)	.000	.000	.000	.000	.000
	Ν	117	168	138	120	170

 Table 7.6: Correlations between IELTS and real-life performances (Weir, Chan et al (2013)

****** Correlation is significant at the .01 level (2-tailed); ***** Correlation is significant at the .05 level (2-tailed)

Cleland, Dowell, McLachlan, Nicholson and Patterson (2012), following caveats concerning the effect of intervening variables in predictive validity studies, provide some generic advice on interpreting correlations in the medical admissions context, as can be seen in Table 7.7 (see Cheung, McElwee and Emery (Eds) 2017 for an excellent discussion of the complex nature of predictive validity studies and also Pollitt 1988).

Validity coefficient	Interpretation
Above 0.35	Very beneficial
0.21 to 0.35	Likely to be useful
0.11 to 0.20	Depends on circumstances
Below 0.11	Unlikely to be useful

 Table 7.7: Guidelines for interpreting correlation coefficents in predictive validity studies

So the combined reading + writing correlation of IELTS with academic performance is certainly reassuring and compares favourably with that reported for ELTS, its predecessor (see below) and more recently for IELTS. The evidence in the other studies we reviewed is mixed, some finding significant and moderate correlations (=>0.21) between IELTS and academic performance:

Bellingham (1993)	0.52
Cotton and Conrow (1998)	0.36
Davies and Criper (1988) (ELTS)	0.3
Dooley and Oliver (2002)	0.34
Elder (1993)	0.346
Feast (2002)	0.39
Ferguson and White (1992)	0.39
Hill et al (1999)	0.54
Huong (2001)	0.36/0.37 (for reading)
Kerstjens and Nery (2000)	0.262 (for reading)
Weir et al (2013)	0.62
Woodrow (2006)	0.4
Yen and Kuzma (2009)	0.46/0.25 (semester 1/2)
and others not:	
Fiocco (1992)	0.063

However, using Cleland et al's estimates, most would be seen as useful correlations.

Weir (2005b:36) adds a final reminder of the need to consider qualitative, i.e. the contextual/cognitive parameters of a test, as well as quantitative data in relation to test validity:

... it soon becomes clear that if we were to make judgements about the validity of a test for the purposes it was intended, notions of validity posited mainly on predictive and concurrent studies would not be enough, especially given the problematic nature of examining these.

Source: US Department of Labor, Employment Training and Administration (1999)

Despite our caveats concerning predictive validity studies, the current IELTS test appears to be working as well as, if not better than, other EAP tests in respect of this particular facet of validity. Weir et al (2013:28–29) found that average IELTS Reading and Writing bands correlated with their average performances on the four selected real-life tasks at 0.62 (p<.01). So if the current IELTS were to continue being used as a test of language proficiency to assess preparedness for university instruction or the professions, it appears to be as useful as any of the alternative measures as far as its predictive validity is concerned. If more construct valid measures for academic purposes are developed in academic reading tests, such as the IELTS Academic Reading test, one might conjecture that even higher correlation coefficients with target performance outcomes are possible.

Endnote

This brings to a close our review of the evidence-based validity of IELTS in Section 2 of this volume. Drawing upon the comprehensive analysis of the validity-related features of the IELTS Academic Reading test presented in this section, the content of Section 3 will discuss some key considerations and potential approaches to assessing academic reading ability in the future. In Chapter 8 we shall explore how reading-into-writing tasks can help to maximise the authenticity of an academic reading task, while Chapter 9 considers the growing contribution and benefit in a digital age of utilising new technologies for the valid assessment of academic reading ability.

Section 3 Enhancing validity in academic reading assessment

Drawing upon the comprehensive analysis of the validity-related features of the IELTS Academic Reading test presented in Chapters 4–7, the content of Section 3 will discuss some key considerations and potential approaches to assessing academic reading ability in the future.

Chapter 8 will explore how reading-into-writing tasks can help to maximise the authenticity of an academic reading task, particularly to address two perceived shortcomings in current approaches to assessing academic reading, i.e. too great a focus on lower-level rather than higher-level reading processes, and too limited an assessment of global expeditious strategies to locate relevant ideas.

Chapter 9 will consider the growing contribution and benefit of utilising new digital technologies for the valid assessment of academic reading ability, and of academic literacy more broadly. It explores implications for the future development and validation of language tests used for assessing academic reading skills, including potential ethical issues raised. The volume concludes with associated research questions that merit investigation, questions which might help to inform the development of a valuable research agenda for language testers and assessment specialists in this field over the coming years.

8 The use of reading-into-writing tasks to assess academic reading

I never desire to converse with a man who has written more than he has read. Samuel Johnson

Introduction

One of the reasons for carrying out the research for this volume was to provide an empirical basis for informing decisions on the cognitive and contextual parameters in any potential revision of existing academic reading tests such as the IELTS Academic Reading test, or for development of new tests of academic literacy in the future.

In our view, two perceived shortcomings of the current approach to assessing academic reading are: firstly, too great a focus on lower-level rather than higher-level reading processes; and secondly, too limited an assessment of global expeditious strategies to locate relevant ideas. We believe there are ways in which these shortcomings might be successfully addressed, drawing upon the findings of research into academic reading in recent years. This third section of the volume is therefore forward looking. It sets out to examine how we might optimise the authenticity of an academic reading test by linking it to the production of a written essay, in a manner which takes account of our detailed understanding of the nature of the academic reading construct and the innovative technological opportunities now available to test designers in a digital age. Chapter 8 summarises the key empirical research data we can draw upon to establish the salient cognitive and contextual parameters that need to be included in the test specification for a reading test for academic purposes. Chapter 9, the concluding chapter of this volume, will consider the ways in which new technologies might contribute to and benefit the valid assessment of academic reading ability and academic literacy more widely.

Reading-into-writing activity as key to academic literacy

The overwhelming evidence from the research on academic literacy in university and college programmes (as reviewed in Chapter 2) is that students'

writing in these contexts primarily involves the integration of relevant information from source materials that have been read for the purposes of an assignment, for example, in order to summarise the main points contained in a set of texts. Summarisation is 'a very common exercise' in pedagogy (Seidlhofer 1995:2) and is particularly popular in academic reading textbooks (Weir et al 2000). Research to date has shown that summarisation skills are essential for content acquisition and academic success (Friend 2002, Holmes and Ramos 1993, Maclellan 1997, Rea-Dickins et al 2007). Therefore, students' abilities to summarise or integrate ideas from different sources would appear to be a critical focus for assessing a student's academic literacy.

An integrated reading-into-writing task format appears to have certain advantages over the more conventional independent reading task in eliciting higher-level, intertextual reading skills. In particular, a reading-into-writing task:

- gives students a reader's perspective (a goal for comprehending with a communicative intent of its own) which in turn affects the attention allocated to different parts of the texts (Goetz, Schallert, Reynolds and Radin 1983, Just and Carpenter 1980)
- allows/encourages students' autonomy to handle source materials rather than responding to pre-constructed representations (Weigle, Yang and Montee 2013)
- elicits deeper and more academically authentic reading processes as students interpret the full significance of the source material (Plakans 2009a, Weigle et al 2013).

In this chapter, we will look closely at the nature of reading-into-writing and how it is different from reading comprehension in general. We will also argue that the reading-into-writing task type could improve the cognitive validity of an academic reading test, specifically because it is capable of eliciting higher-level reading processes.

Academic reading-into-writing as a construct distinct from reading comprehension

Reading comprehension is often regarded as a prerequisite of academic literacy skills which involve subskills, such as developing a topic, finding and evaluating sources, understanding critical perspectives and argument, integrating primary and secondary sources, and citing sources (Howells 2011). Yu regards reading comprehension as 'a sine qua non' for summarisation (2005). Previous research has shown that success at synthesising materials for writing is related to students' sensitivity to text structure and relations between ideas, which is commonly associated with comprehension (Kintsch and van Dijk 1978, van Dijk and Kintsch (Eds) 1983). However, while reading-into-writing (summarisation) is a natural entailment of reading comprehension (Kintsch and van Dijk 1978, van Dijk and Kintsch (Eds) 1983), it requires students to process beyond the level of comprehension with the additional need to read for evaluation, condensation, and transformation of ideas that have been presented (Hidi and Anderson 1986:473–474).

While students' general reading comprehension and writing abilities are considered to contribute to their academic reading-into-writing ability, research shows that reading-into-writing is a distinct construct in its own right (Ascención-Delaney 2008, Chan 2013, 2018, Watanabe 2001, Yu 2013). For example, Watanabe (2001) reported that both reading comprehension and general writing measures accounted for up to 40% of the variance in students' reading-into-writing ability. However, the reading comprehension measure on its own did not significantly predict the students' reading-into-writing performance. In a study to examine the relationship between the general reading ability of 139 students and their academic reading-into-writing ability, Ascención-Delaney (2008) reported that students' proficiency in reading for comprehension, as measured by the Nelson-Denny Reading Test (Brown, Fishco and Hanna 1993), associated weakly with their ability to summarise academic texts (r=.28). These results, based on analyses of test scores, suggest that a reading test which measures reading comprehension mainly at word and sentence levels would be insufficient as a tool to assess students' ability to integrate multiple reading materials for academic purposes. The next step is to establish the parameters which are additional to those required for reading comprehension per se in order to more adequately measure students' academic reading through a reading-into-writing task.

As we have already established in Chapter 2, it is clear from our survey of the specifications of academic literacy requirements provided by universities themselves that academic reading takes us beyond reading for comprehension of a single text. For example, *Academic Literacy: A Statement of Competencies Expected of Students Entering California's Public Colleges and Universities* (2002), developed by the Intersegmental Committee of the Academic Senates of the California Community Colleges (CASCCC), the California State University, and the University of California, specifies 20 reading skills relevant to academic reading (see Chapter 1). In our view, six of the skills (replicated in the list below) could be more adequately tested through requiring students to write purposefully about the passages they have read rather than having them to respond simply to questions about the material in conventional independent reading item types:

- · read texts of complexity without instruction and guidance
- summarise information
- synthesise information from reading and incorporate it into a writing assignment

- make connections to related topics or information
- · determine major and subordinate ideas in passages
- retain information while searching for answers to self-generated questions.

In a similar vein, we argue that three of the 14 academic reading skills required of students at universities in the UK identified in Weir's (1983) large-scale survey (see Chapter 2) could be better addressed by the integrated format.

1. Separating the essential from the non-essential in a text:

Distinguishing the main idea from supporting detail by differentiating especially the whole from its parts, statement from example, fact from opinion, a proposition from its argument (cf. Munby 1978).

2. Note-making:

(a) Extracting salient points for summary – This could be a summary of the whole text, a specific idea or topic in the text of the underlying idea or point of the text (cf. Beard 1972, Barrett's taxonomy in Clymer 1972, Munby 1978).

(b) Selective extraction of relevant points from a text – This could involve the co-ordination of relevant information, the ordered rearrangement of contrasting items or the tabulation of information for comparison and contrast (cf. Munby 1978, Widdowson 1980).

(c) Reducing a text through rejection of redundant or irrelevant information or items, e.g. determiners, repetition, compression of examples, use of abbreviations (cf. Heaton 1975, Munby 1978).

3. **Critical evaluation**: Assessing the worth of a text and the way information in it has been organised and expressed (cf. Clymer 1972, Davies and Widdowson 1974).

The importance of reading beyond the level of comprehending a single text is also evident in the CEFR (Council of Europe 2001). For example, the skills of processing multiple texts are specified at the B1 level and above – see the descriptors in bold in Table 8.1.

While the concept and importance of integrated skills were recognised in the original CEFR (Council of Europe 2001), the construct was not fully specified at that time. In the CEFR (2001), *mediation* was introduced as one of the four modes of communication – reception, production, interaction, and mediation – in a move away from the traditional four skills (reading, writing, listening, speaking). However, mediation was not developed in the 2001 version of the CEFR and there are no specific Can Do descriptors focusing on integrated skills, except those replicated in Table 8.1.

Nevertheless, in response to a perceived need for a systematic construct definition of integrated skills (Hirvela 2004), in 2018 the Council of Europe

TAULO 0.1. TIII	recoond to entry	้ร ราชวา วเป็นแทน ริน				
CEFR descriptors	AI	A2	B1	B2	CI	C2
Overall reading comprehension (2001:69) Processing text	Can copy out	Can copy out	Can collate short	Can read with a large degree of independence, adapting style and speed of reading to different texts and purposes, and using appropriate reference sources selectively. Can summarise a wide	Can summarise	Can summarise
(2001:96)	single words and short texts presented in standard printed format.	short texts in printed or clearly handwritten format. Can pick out and reproduce key words and phrases or short sentences from a short text within the learner's limited competence and experience.	pieces of information from several sources and summarise them for somebody else. Can paraphrase short written pasages in a simple fashion, using the original text wording and ordering.	range of factual and imaginative texts, commenting on and discussing contrasting points of view and the main themes. Can summarise can summarise extracts from news items, interviews or documentaries containing opinions, argument and discussion. Can summarise the plot and sequence of events in a film or play.	long, demanding texts.	information from different sources, reconstructing arguments and accounts in a coherent presentation of the overall result.

Table 8.1: The skills of processing multiple texts specified in the CEFR (2001)

released a Companion Volume to the CEFR (Council of Europe 2018), and a key feature of this latest addition to the CEFR is a set of descriptors focusing on mediation. According to the CEFR (Council of Europe 2018:33), the two main purposes of mediation include transforming a concept to an audience by linking the current language activity to previous knowledge, breaking down complicated information, adapting language (e.g., from one language to another), and transforming information from one text(s) to another, for example by amplifying a dense text and streamlining a text. To specify the construct involved in these two skills, the new mediation scales are provided in three categories: mediating a text, mediating concepts, and mediating communication. As the latter two are less relevant to our discussion of operationalising academic reading (as in IELTS where students complete the task independently rather than collaboratively with other students), we will focus on the descriptors of mediating a text.

Mediating a text (Council of Europe 2018)

- Relaying specific information
- Explaining data (e.g. in graphs, diagrams, charts etc.)
- Processing text
- Translating a written text
- Note-taking
- Expressing a personal response to creative texts
- Analysis and criticism of creative texts

Mediating a text involves passing on to another person the content of a text to which they do not have access, often because of linguistic, cultural, semantic or technical barriers. Such language use situations are increasingly being incorporated into language curricula. However, the notion of mediation has been extended to include mediating a text for oneself (for example, in taking notes from reading or spoken inputs) or in expressing reactions to texts. This new extension has high relevance to the construct of academic reading outlined in this volume. This move is recognised as a positive step towards the direction of task integration within the sociocognitive approach to construct definition by providing the necessary theories and formalisation to enable the integrated construct to be operationalised for high-stakes assessment purposes (Saville 2018:ix–xii).

In the next section, we will present the empirical evidence showing how integrated tasks could be used to improve the cognitive validity of academic reading tests by eliciting higher-level reading processes.

Assessing high-level academic reading processes using integrated tasks

As supported by the evidence presented in Chapter 2, most independent reading items can only test students' reading processes up to the level of *a text model of comprehension* (Grabe 2009, Grabe and Stoller 2002, Kintsch 1998, Perfetti 1997), where students comprehend the propositions within the text itself and their interrelationships. Integrated tasks, on the other hand, provide a performance condition for students to create *a situation model of interpretation* (Grabe 2009, Grabe and Stoller 2002, Kintsch 1998, Perfetti 1997), where they construct an individual interpretation of the text as well as *a text model*.

Applebee (1984) regards integrated tasks as a path to the 'higher-order' processes, such as combining information from different propositions and integrating material from the text with previous knowledge (Chen and Vellutino 1997, Gough, Hoover and Peterson 1996, Grabe 2009, Kintsch 1998, 2004, Perfetti 1985, 1997, Sticht and James 1984). Reading-into-writing tasks are typically used for two purposes, reading to learn (Carver 1997, Enright et al 2000, Kintsch 1998) or reading to integrate information for writing (Grabe and Stoller 2002) in educational settings. In this volume, we focus on the latter. For readers who are interested in how integrated tasks could promote learning, see Cohen and Upton (2006).

The constructivist tradition in reading research has long examined how students transform content from multiple texts. Bartlett (1932) argues that in reading-into-writing activities, readers actively construct meaning by connecting and integrating prior knowledge with newly encountered information in the reading texts. Van Dijk and Kintsch (1977), based on extensive research on students' processes of summarising narrative stories, proposed that summarisation involves four main processes of *deletion*, *generalisation*, *selection* and *construction*. Kintsch (1998) argues that when reading to integrate, students process the text beyond the level of comprehension to construct elaborate models of the text structure and situation, enabling them to select information from the source text, evaluate it, and use it for writing purposes.

Weigle (2004) defines reading-into-writing as 'a test that integrates reading with writing by having examinees read and respond to one or more source texts' (2004:30). According to Cumming (2013:1), integrated test tasks should 'require learners or test takers to incorporate **substantive** content from source materials in print, audio, and/or visual forms' (emphasis added by authors). Bachman (1990) regards summarisation as the major process involved in integrated test tasks – 'a test requiring test takers to ... summarise the propositional content in a reading passage, will involve the full range of organisational characteristics' (1990:139). Cohen (1994:174) believes integrated test tasks have a natural appeal because they 'simulate real-world tasks in which

Research and Practice in Assessing Academic Reading

non-native readers have to read and write a summary of the main ideas of a text'. With the aim of accounting for the processes involved in integrated reading-into-writing activities, Knoch and Sitajalabhorn (2013:306, emphasis added by authors) provide an expanded definition of integrated tasks as follows:

Integrated writing tasks are tasks in which test takers are presented with one or more **language-rich source texts** and are required to produce written compositions that require (1) mining the source texts for ideas, (2) selecting ideas, (3) synthesising ideas from one or more source texts, (4) transforming the language used in the input, (5) organizing ideas and (6) using stylistic conventions such as connecting ideas and acknowledg-ing sources. The rating scale used to grade such compositions needs to take account of these features specific to integrated writing tasks.

Like Cumming (2013), Knoch and Sitajalabhorn (2013) argue for the essential role of interacting with substantive source materials in reading-intowriting. While there is rich research on how integrated tasks elicit high-level writing processes from writers as they transform knowledge from sources (e.g. Chan 2013, 2017, 2018, Chan, Wu and Weir 2014, Plakans 2008, Plakans and Gebril 2012, Wolfersberger 2013), the reading aspect of the construct elicited by integrated tasks is an under-researched area (Weigle et al 2013). Yu (2013) argues that most studies on integrated tasks have not considered reading as a factor in knowledge transformation. Here we aim to tease out evidence from the literature to illuminate **the nature of the high-level reading processes** involved in integrated tasks.

With the aim of investigating test-taker strategies, Cohen (1994) reported that when required to write a summary, students read the source texts differently than when they read for comprehension (see also Cohen 2006, Cohen and Upton 2007). When reading to summarise, the students tended to analyse and interpret sources at a global level by purposefully selecting and linking information to writing. Esmaeili (2002) also investigated the extent to which test takers employed reading strategies in integrated tasks. He examined the strategies employed by 34 engineering students in thematically linked reading–writing test tasks. The findings show that the students extensively made use of reading strategies, such as mining texts for relevant ideas for writing, borrowing words/phrases, recalling content from reading, and accepting/rejecting viewpoints from reading, on the summary task.

By analysing think-aloud verbal protocols, interviews, and the students' scripts, Plakans (2009a) investigated the reading strategies of 12 L2 students on two integrated reading-into-writing tasks. The findings show that the integrated tasks engage the students in *goal-setting strategies* to clarify their purpose in reading, *comprehension strategies* (which was referred to as 'cognitive processing'), *global reading strategies* such as skimming and identifying

main ideas, *metacognitive self-regulating strategies* to evaluate progress of reading, and *mining strategies*. Plakans (2009a) found that students' use of metacognitive strategies often led to high-level processing of the source texts which are activated by mining strategies including search reading text for ideas to use in writing, rereading text for use in writing, and paraphrasing. Based on the evidence, Plakans (2009a) argues that the integrated tasks show 'test-takers moving beyond comprehension toward the use of the texts in their writing' 2009a:7).

Although coming from the perspectives of test-taking strategy research, the findings of these studies clearly indicate that integrated reading-intowriting tasks engage students in high-level reading processes to transform knowledge. Hirvela (2004) argues that, when reading for writing in academic contexts, L1 students typically employ two kinds of reading: *mining* and *writ-erly* reading, which are not typically activated by traditional reading comprehension items. Mining is the process of reading for the purpose of culling information from a text for a specific goal (Greene 1992). This is referred to as search reading in Urquhart and Weir's (1998) model (see Table 3.1). Writerly reading, on the other hand, is for the purpose of improving one's writing by example, such as looking at word use or considering argument structure. Hirvela (2004) advocates that integrated tasks are the most authentic task type to elicit these higher-level reading processes, and urges for more attention on the reading–writing connections in integrated tasks in research and pedagogy.

To illustrate the reading–writing connections, Hirvela (2004) suggests that reading should be regarded as input while writing is output. His view is similar to Grabe and Kaplan's (1996), who argue that students engage in higher-level reading processes when they need to demonstrate how they understand what they read in writing. They see reading-into-writing tasks as 'reciprocal activities in that the outcome of a reading activity can serve as input for writing, and writing can lead students to further reading resources' (Grabe and Kaplan 1996:297).

The influential research on discourse synthesis conducted by Spivey and colleagues (1984, 1990, 1997, 2001) has laid an important foundation for researchers to examine the processes involved in reading-into-writing activities. Spivey and King (1989:11, emphases added by authors) argue that reading-into-writing tasks engage students in three 'very basic' processes as they transform knowledge gained from the reading in writing.

Some hybrid reading-to-write tasks involve discourse synthesis, a process in which readers read multiple texts on a topic and synthesize them. They **select** content from the composite offered by the sources – content that varies in its importance. They **organize** the content, often having to supply a new organizational structure. And they **connect** it by providing

Research and Practice in Assessing Academic Reading

links between related ideas that may have been drawn from multiple sources.

Following Spivey's work, Plakans (2009b) examined the extent to which reading-into-writing test tasks elicited these discourse synthesis processes from six L2 students in detail. Based on think-aloud verbal protocols, interview responses, and students' scripts, she reported that four of the students followed the discourse synthesis approach as they completed the tasks. She suggested that factors such as general English proficiency and academic experience might have an impact on students' ability to engage with the reading passages at a level required by the integrated tasks.

The literature so far shows that reading-into-writing tasks engage students beyond reading comprehension towards higher-level reading processes to integrate reading in their writing. These higher-level processes include, but are not limited to:

- setting own goals to clarify purpose in reading
- analysing and interpreting multiple texts at a global level (e.g. skimming and identifying main ideas)
- · accepting/rejecting viewpoints from reading
- selecting content from multiple texts
- connecting content from multiple texts (e.g. by providing links between related ideas)
- organising content (e.g. by rearranging priority of ideas, supplying a new organisational structure)
- recalling content from reading
- metacognitive self-regulating strategies to evaluate progress of reading
- writerly reading to improve one's writing (e.g. checking word use, considering argument structure).

After reviewing the wider literature concerning reading-into-writing, we now focus more on the empirical evidence in language testing to examine the extent to which these processes can be operationalised by standardised integrated test tasks.

As mentioned previously, integrated tasks are increasingly being used in standardised language tests to assess students' academic writing ability. However, the number of tests which use integrated tasks to assess academic reading remains small. A study by Weigle et al (2013) investigates a reading/ writing task in Georgia State Test of English Proficiency (GSTEP) which seems to be one of the few exceptions where the integrated task is used to assess both academic reading and writing skills. Although the reading/ writing task uses the short-answer question (SAQ) format (which is usually not considered as an integrated task – see discussions of this item type in Chapter 4), Weigle et al argue that in the context of their study it was regarded as an integrated task, because each student's complete set of responses to the questions were scored holistically, rather than by individual items, on 5-point scales for content, as a measure of reading, and language, as a measure of writing (see Weigle 2004).

By means of think-aloud protocols, retrospective interviews and semistructured interviews, Weigle et al (2013) examined the reading processes of five university students engaged in the reading/writing task which required them to respond purposefully to the reading materials (e.g. to state the main idea of a text, draw inferences, and synthesise information across texts). The results show that the most students in their study engaged in complex and higher-level reading processes to make meaning of the passages (including *connecting to background knowledge, evaluating information, taking a stance and interpreting author's tone*) as well as to construct an intertextual representation of the passages through *identifying main argument of passages, paraphrasing ideas, summarising paragraphs, identifying relationship among ideas.* 'The participants found that without a good understanding of the textual information, they really could not answer the questions' (Weigle et al 2013:40).

As a result, they concluded that 'there is a relationship between performance on the test [reading/writing task] and engagement in higher-level reading processes' (Weigle et al 2013:46). In phase two of the study, Weigle et al (2013) examined the reading processes of 20 participants and reached similar conclusions - that the integrated format engages students in higher level reading processes which cannot usually be elicited by other independent reading item types. The most common processes identified in the students' think-aloud protocols include *paraphrasing* (i.e. think-aloud events which repeat the gist of a sentence in the text without adding additional information), elaboration (i.e. think-aloud events which add meaning to, or embellish, the information in the focal sentence or in the text as a whole), evaluation (i.e. think-aloud events which convey a positive or negative judgement about some aspect of the text), and comprehension problems (i.e. think-aloud events which indicate a lack of understanding). Furthermore, Weigle et al (2013) reported that the strength and completeness of the intertextual models that the students constructed were evident in their writing scripts.

Evidence from standardised language tests

We will now turn to evidence of students' reading processes on four standardised tests, namely, LTTC's GEPT Advanced, Pearson Test of English (PTE) Academic, Internet-based Test of English as a Foreign Language (TOEFL iBT) and Trinity College London's Integrated Skills of English (ISE). Although the integrated tasks in these tests are used to assess students' academic writing ability rather than their higher-level reading skills, we believe that students need to engage in higher-level reading processes in order to successfully complete an integrated reading-into-writing task.

1. TOEFL iBT integrated task

Using textual analysis, Plakans and Gebril (2013) investigated features of source use by 480 students on the TOEFL iBT integrated task. The task involves a comparative summary of listening and reading texts that present differing views on a topic. They found that high-scoring texts included important ideas from both sources (i.e., reading and listening texts in the context of their study) whereas low-scoring texts included ideas mainly from the reading texts and consisted of direct copying of words and phrases. Additionally, through multiple regression analysis, they reported that three features of source text use explained over 50% of the variance in scores across nine score levels: (1) the importance of source text ideas that writers included in their summary; (2) the use of ideas from a reading source text and from a listening text; and (3) the borrowing of exact wording from the source texts (verbatim source use). However, their study did not provide actual evidence of students' processes; the findings indicate how students at different score levels might have engaged in some of the higher-level reading processes, such as search reading for important ideas from multiple sources and connecting these ideas.

2. PTE Academic

Chan (2011) investigated the processes of 10 university students on the PTE Academic task *Summarize Written Text*, which requires test takers to write a one-sentence summary of a passage in 10 minutes (Pearson 2019). Based on three sources of evidence including screen capture data, researchers' observations and stimulated recall protocols, the results show that, on average, test takers spent 33.18% of the task time on reading task instructions and the passage, 9.10% on producing their own text, 12.33% on revising and 19.42% on pausing. 25.97% of the task time remained unused by the test takers. The protocols show that most students engaged in processes of reading the task prompt and brief macro-planning, reading the source text, identifying key ideas, rereading a selected part of the text, organising ideas, translating (i.e. writing), and revising. As a result, Chan (2011) advocates for the important role of discourse synthesis (i.e. selecting, connecting and organising ideas across multiple texts) in the reading-into-writing construct.

3. GEPT Advanced Writing Task 1 (reading-into-writing)

The GEPT Advanced Level test, which was developed and administered by LTTC in Taiwan, targets English learners at the CEFR C1 level (i.e. effective operational proficiency). It aims to build a validity argument in support of the claim that the test assesses students' suitability for pursuing further academic studies or employment where communication in English is essential. The GEPT Advanced Writing Task 1, one of the two writing tasks, requires students to produce a written piece of 250 words that synthesises and summarises information from two given passages, as well as expressing their own opinions and making recommendations, within 60 minutes.

By means of a writing process questionnaire, Chan et al (2014) examined the processes of 160 Chinese students, who were studying in the UK at the time of the study, employed to complete the GEPT Advanced Writing Task 1. There were 48 items in the questionnaire in total. Exploratory factor analyses were used to examine the underlying structure of the cognitive processes elicited by the task. As a result, 11 processes were identified. Here we will report the findings relating to high-level reading processes. According to students' responses to the questionnaire, in a Likert scale from 1 (strongly disagree) to 4 (strongly agree), they read the passages carefully at a global level (M=2.86, SD=0.74), selected relevant ideas from the passages (M=3.22, SD=0.82), organised the ideas selected from the passages (M=2.91, SD=0.62) and generated links among these ideas (M=2.89, SD=0.70). The sub-processes under each process group are listed below.

Careful reading at the global level

- I read through the whole of each source text carefully.
- I read the whole of each source text more than once.

Selecting relevant ideas

- I read some relevant part(s) of the texts carefully.
- I searched quickly for part(s) of the texts which might be useful for the task.
- I took notes on or underlined the important ideas in the source texts.

Organising intertextual relationships between ideas

- I worked out how the main ideas across the source texts relate to each other.
- I prioritised important ideas in the source texts in my mind.
- I worked out how the main ideas in each source text relate to each other.
- I used my knowledge of how texts like these are organised to find parts to focus on.

Connecting ideas and generating new representations

• I linked the important ideas in the source texts to what I know already.

- I made further connections across the source texts while I was writing.
- I developed new ideas while I was writing.
- I developed new ideas or a better understanding of existing knowledge while I was reading the source texts.

In Phase 2 of the study, using the same questionnaire, data was collected from 192 Taiwanese students taking the GEPT Advanced Writing Task 1 under live test conditions. The results show that students in both contexts employed these processes on the task in a similar way. To be specific, the Mann-Whitney U tests indicated no significant differences in the mean ranks of the average rating of all 11 processes, including those listed above.

In a more recent study using keystroke logging and retrospective semistructured interviews, Chan (2017) examined in detail the processes two postgraduates went through when completing the GEPT Advanced Writing Task 1 (a summary writing task). Students' behaviours of switching between the reading task instructions, source texts and writing the essay were analysed in conjunction with detailed keystroke logs (i.e. chronological analysis of text production) and interview data. The results showed that the students read each of the source texts carefully to understand the main ideas of the text before starting to write, and reread the source texts multiple times to select relevant information during writing. The study also revealed noticeable individual differences in how the students transformed ideas from the sources. The higher-performing student in the study was able to select, organise and connect ideas with high automaticity (Field 2004), and as a result, his summarisation processes were less observable in the keystroke logs analysis. In contrast, Chan (2017) found that the lower-scoring student's processes of selecting, organising and connecting ideas were less automatic, which resulted in many occurrences of switching between the source texts and the writing sheet. Although the student explained in the interview that she was summarising ideas from the two sources and interpreting relations between these ideas, there was limited evidence of successful implementation of these high-level processes in her text. The data showed that she 'transformed' the ideas by lifting chunks of texts from sources and 'connected' them by using formulaic expressions, which is a typical feature found in low-scoring performances. The findings of this qualitative study, despite its small-scale nature, add to the evidence supporting the use of the integrated reading-into-writing tasks to assess the higher-level reading processes of selecting, organising and connecting ideas from sources.

4. Integrated Skills of English (ISE)

Trinity College London – an international examination board for the performing arts and English language (see www.trinitycollege.co.uk/site/?id=263)
- commissioned CRELLA to redevelop their ISE exam suite. ISE is a fourskills exam assessing youngsters' and adults' language abilities through an integrated approach. For details of the exam redevelopment project, the reader can refer to Chan, Inoue and Taylor (2015).

The reading and writing paper of the new ISE has two reading tasks (*Long reading* and *Multi-text reading*), one reading-into-writing task and one writing task. In the Multi-text reading task, students are asked to read four texts (including short passages and non-verbal input, e.g. diagrams, tables and graphs) to answer 15 questions in order to demonstrate different reading skills. In the reading-into-writing task, students are asked to write an article by using the information form from the four texts presented in the reading section. Typically, they are asked to explain and/or describe an issue and make recommendations. More details about the exam and sample tasks can be found on Trinity's website (www.trinitycollege.com/site/?id=3194).

Chan (2018) investigated the processes of 104 test takers on the ISE reading-into-writing task at B1, B2 and C1 levels. Students in the study were asked to report the extent to which they believed they used various reading-intowriting processes on a four-point scale with an option for uncertainty (4 – strongly agree; 3 – agree; 2 – disagree; 1 – strongly disagree; 0 – not sure). The findings show that 12 of the 41 sub-processes appear to be essential for successful completion of the reading-into-writing tasks at the three levels. It was found that the integrated tasks engage most students in conceptualisation to understand the task instructions; careful reading to comprehend ideas at the global level; expeditious reading to search for parts of the sources which are relevant to the writing task and reading these parts carefully. Furthermore, the tasks allow students to generate new ideas through writing about the sources. Regarding monitoring and revising, most students across levels checked their writing in relation to task fulfilment, appropriateness of ideas transformed from sources, and language in source use. As a result, Chan (2018:20) concluded that:

... the integrated tasks consistently engage most students at the three levels in knowledge transforming (Bereiter and Scardamalia, 1987) through task representation, higher-order reading processes to select and connect ideas, and revisions.

Another important finding of the Chan (2018) study is that the higherproficiency L2 students tend to employ a wider range of higher-level processes to complete the reading-into-writing tasks than the lower-proficiency ones. For example, the results indicate that organising, one key process of discourse synthesis (Spivey 1984), does not seem to come into play at B1. This has important implications for our discussion of possible approaches to scoring students' higher-level reading abilities on integrated tasks. In short, the above studies on test taker processes on high-stakes integrated tasks provide evidence showing how the integrated task type engages students in higher-order reading processes to transform ideas from multiple texts. This means that the integrated reading-into-writing task type could improve the cognitive validity of an academic reading test. The next critical issue is the extent to which students' intertextual reading skills may be evident in the writing products, and this is discussed in detail below.

Impact of features of integrated tasks on reading processes

In this section, we review the literature regarding how some features of an integrated reading-into-writing task might have impacted on students' higher-level reading processes.

Task purpose

First, there is a general consensus that the purpose for reading affects the processes and strategies engaged (Alexander, Graham and Harris 1998, Brantmeier 2002, Koda 2005, Phakiti 2003). Task purpose (e.g. summarise main ideas, compare and contrast) has an impact on which higher-level reading processes students use. Koda (2005:27) argues that in integrated tasks which require information retention for further interrogation and reflection, the task objective, even presented in a vague sense, provides indispensable guidance during both reading (i.e. how much information needs to be extracted from the text) and post-reading contemplation (i.e. how best to use extracted information in achieving the objective).

Task purpose is likely to influence how students engage in higher-level reading processes because before reading a text a decision is usually made by the reader on how they are going to approach the text. The reader's understanding of the task purpose often functions as a goal for comprehending (Just and Carpenter 1980) which in turn determines the attention allocated to different reading processes (Goetz et al 1983) and the information recalled (Pichert and Anderson 1977, Reynolds 1981). This is referred as the 'goal setter' in our model in Chapter 3, Figure 3.1. When students set goals for their reading, they make decisions on the appropriate type of reading: global/ local and careful/expeditious. In other words, reading-into-writing tasks would engage students in reader-driven, purposeful and conscious aspects of reading (Kintsch 2004). The reading-into-writing task purpose (e.g. to summarise main arguments, to arrange events in priority) allows students to demonstrate their ability in determining the appropriate reading processes to achieve the purpose rather than passively responding to reading items where normally a lower-order reading process (e.g. to locate a fact in a sentence) is intended. Generally speaking, more integrative and analytic tasks (such as making a synthesis from multiple texts as compared to writing a summary of a single text) are found to elicit richer engagement of the texts from readers (e.g. Langer 1980, 1986, Newell 1984).

Genre

As reviewed in Chapter 2, the analysis of EAP reading teaching textbooks and test papers shows that texts with different rhetorical organisations, e.g., comparison, collections of descriptions, causation, and problem/solution (Carrell 1984) lend themselves better to testing different reading skills; that is, for testing a particular skill, there might be an optimal rhetorical organisation. This is also true for reading-into-writing tasks. Some genres are believed to be less demanding to process and summarise than the others. For example, narratives are perceived to be easier to summarise than expository texts because the former tends to contain factual information which is organised in a linear structure whereas the latter tends to contain more complex and unfamiliar ideas (Spivey and King 1989). Argumentative texts, on the other hand, would be more challenging to summarise as this requires high-level reading skills to understand the author's point of view (Brewer 1980, Yu 2008). Taylor (2013) analysed 40 6th form students' recalls of a narrative text and an expository text. The results show that students recalled certain propositions more readily than others and there were greater variations in how students created the mental representation of the expository as compared with the narrative text.

Weigle (1999) compared two common types of nonverbal input: table/ chart and graph. The findings indicate that the table/chart prompt (making and defending a choice based on information presented in a table or chart) tends to elicit traditional five-paragraph essays, while the graph prompt (describing trends in a graph and making predictions based on the information presented in the graph) would elicit several rhetorical angles. In other words, students' approaches to connecting and organising ideas are likely to be different when they summarise from these two prompt types.

Text length and structure

It is generally believed that the longer the text, the more difficult it is to summarise it. Hidi and Anderson (1986) argue that it is more difficult to summarise longer texts because they tend to contain more ideas which lead to more decisions to be made as to which ideas to select. Nevertheless, presence/ absence of a clear overt structure could be more predominant in making a text more or less difficult to transform. For example, a lengthy journal article or chapter from a textbook with clear sections and headings, where paragraphs contain topic sentences in initial position which signal the information to be presented, may well be easier to summarise than short texts without a clear structure and/or a clear sequenced line of argument.

Amount of information

As mentioned above, in general, a text with a higher amount of information should be more difficult to transform than a text with fewer ideas. However, O'Loughlin and Wigglesworth's (1999) study on task difficulty of the IELTS Academic Writing Task 1 (i.e. transforming information from graphs/diagrams/tables) suggests that the impact of amount of information on students' processing might not be straightforward. The study examined, firstly, the extent to which the difficulty of the task is affected by the amount of information provided for the students and, secondly, the extent to which the difficulty of the task is affected by the presentation of the information to the student. Four tasks, which differed in terms of the amount of information the students were required to process to complete the task, were developed for the study. The test score analyses indicated that there were no significant differences in difficulty between the tasks, either in terms of the amount of information presented or in terms of the differences in presentation of the tasks. However, it should be noted that responses from all three proficiency groups (high, middle, low) to the task with less information showed greater complexity in their transformation of the ideas overall than the task with more information. This implies that, when required to transform few ideas, students might connect these ideas at a deeper level and as a result generate more complex representations of the ideas. Similar studies on the impact of amount of information on other text types would be important to enhance our understanding of important factors which may influence the process and product of reading-into-writing.

Linguistic complexity (e.g., lexical and syntactic complexity) might also make a text more difficult or easy to integrate. However, there is no evidence indicating that text complexity might influence students' higher-level reading processes differently from their comprehension processes. We have reviewed the impact of linguistic complexity on comprehension in Chapter 5.

Differences between expert and novice academic readers – implications for scoring

Having established some key features of reading-into-writing tasks in relation to academic reading, we attempt to illuminate the differences between skilled academic readers and unskilled academic readers on reading-intowriting tasks. The aim is to provide insight into how test takers' higher-level reading skills on an integrated task can be scored. Those differences which are more related to academic writing skills, such as abilities to transform the language from sources, are beyond our current focus on reading and hence are excluded.

Generally speaking, readers with higher proficiency use more strategies (Anderson 1991, Phakiti 2003) and strategies that are more global (Block 1986, Carrell 1989, Koda 2005). Plakans (2009a) reported that the higher-scoring L2 students in her study employed more global strategies, such as goal setting by checking the task, skimming for the gist, and asking questions. Higher-scoring students also used more mining strategies, particularly search reading for ideas and rereading to use source texts in their writing. These students were able to pay more attention to setting goals and employing metacognitive, self-regulating strategies when reading, all of which suggests that they were more focused on reading for the purpose of the task than word-level comprehension. In contrast, the low-scoring students focused on word-level comprehension and general compensatory strategies.

Weigle et al (2013) found that higher-performing students showed a tendency to elaborate more in their think-aloud protocols whereas lower-performing students tended to verbalise more comprehension problems (i.e. think-aloud events which indicate a lack of understanding). Plakans and Gebril (2016), in their study exploring the relationship of organisation and connection with scores in TOEFL iBT, reported that higher-scoring students were able to integrate information from all sources whereas weaker students relied heavily on one reading source when they summarised. In addition, when demonstrating their understanding of the course materials, higher-scoring students were more able to organise the relevant contents in a more coherent manner than the lower-scoring students. Similarly, Spivey (1988) reported that characteristics of students' syntheses of sources, especially in relation to organisation, were linked to their ability to read multiple texts.

In a study of a discourse synthesis task of three descriptive texts, Spivey (1983) found that the able stronger readers in the study produced texts which had more content, were more unified and connected, and were based on across-text, important information to a greater extent than the texts of weak readers. Durst (1985, 1987) compared the cognitive processes of high- and average-ability writers on integrated tasks. He found stronger students employed varied and more complex reading operations, focusing on intermediate and global issues in the readings, and tended towards more abstract interpretations of the main ideas with more evaluation (instead of description).

The findings of the above studies indicate that some of the students' higher-level reading abilities can be identified in the features of their writing on integrated tasks. Table 8.2 summarises the differences found between skilled and unskilled academic readers on reading-into-writing tasks.

	Skilled academic readers	Unskilled academic readers
Processes	 employ varied and more complex reading processes focus on intermediate and global issues in the readings tend towards more abstract interpretations of the main ideas with more evaluation are more able to supply a new organisational structure demonstrate more sensitivity to importance of ideas in a single text and 	 employ limited and lower-level processes focus on local issues in the readings tend towards description of ideas
Relevance and adequacy of the content Structure	 across texts include important ideas from all sources (across-text) produce texts which had more important content summarise source information make more use of the organisation of the source texts provide more compact, integrated 	 include ideas mainly from a single source make declarations based on personal knowledge tend to follow the structure of one of the sources
Coherence	 transform ideas from sources into unified and connected ideas 	 produce isolated ideas from sources

 Table 8.2: Differences between skilled and unskilled academic readers on reading-into-writing tasks

(Sources: Cumming et al 2005, Durst 1985, 1987, McGee 1982, Plakans and Gebril 2013, Spivey 1983, 1984, 1992, Watanabe 2001)

In other words, it is perfectly plausible to score the reading element of such tasks according to the relevance and adequacy of the content produced, the way connections have been made between ideas and organised, though the availability of a construct valid piece of writing would be an added bonus and bring other criteria into play if it was to be included as part of the writing assessment.

Conclusions

One of the reasons for researching this book and bringing together its content in a single volume was to provide an accessible empirical basis for informing the ongoing review and development of academic reading tests, such as the IELTS Academic Reading test, particularly with regard to features of cognitive and contextual validity. Drawing upon insights from the comprehensive literature review presented in this volume, we believe we have been able to offer some specific suggestions concerning the future content and format of an academic reading test, in order to produce a test that is fully fit for purpose.

The final chapter of this volume will consider the constantly evolving nature of academic reading ability in a teaching and learning environment that is increasingly characterised by the use of electronic and multimedia resources. It will also explore the growing contribution and benefit of new technologies for the valid assessment of academic reading ability in a digital age, as well as some points for further consideration.

9 Assessing academic reading in a digital age

We shall not cease from exploration And the end of all our exploring Will be to arrive where we started And know the place for the first time. T.S. Eliot (1942)

Introduction

One of the central concerns expressed in this volume has been the priority for tests used for the purposes of university admission or progression to reflect as fully as possible the mental processes engaged by language users (cognitive validity) within academic situations (context validity) – what we have referred to as the sociocognitive approach to test development and validation. The four chapters in Section 2 lay out what applied research has revealed to date about how far a test of academic reading ability, such as IELTS, emulates actual language use within academic study situations.

In this concluding chapter, we consider how, in a digital age, the nature of academic reading itself may be steadily evolving to generate a more complex and multifaceted construct of academic literacy. In addition, we begin to examine the growing contribution and benefit of new digital technologies for the valid assessment of academic reading ability, and of academic literacy more broadly. We shall explore some implications for the future development and validation of language tests used for assessing academic reading skills, including potential ethical issues raised. Towards the end of the chapter we offer some questions that we believe merit focused investigation – questions that could help to shape a preliminary research agenda for language testers and assessment specialists in this field over the coming years.

Academic reading in a digital age

The rapid growth of digitisation, standardised online courses, use of social media, co-operative learning networks and 'big data' over the past 10–15 years has significantly reshaped how university students read today. This inevitably means that the construct of academic reading as defined in Section 1 of this volume, which was based upon a comprehensive review of extensive research on academic reading during the 35-year period from about 1980 up

to 2017, will need to evolve as the context (environment) of academic reading continues to change. Academic reading is no longer confined by presentation mode nor by physical location. Technology has enabled quick and easy access to a vast amount of reading and other material for learners today, including not only text but visual images, audio and video (much of it online). This contrasts strongly with the experience of previous generations of students for whom the predominant academic reading resources were limited to journal articles, books and encyclopaedias physically located within a university or departmental library.

It is clear that university students today read much less paper-based material in hard-copy format, and instead conduct their reading activity using electronic devices such as computers, tablets, and mobile phones (Cartelli 2012). A recent research study conducted by The Harris Poll for Pearson entitled 'Beyond Millennials: The Next Generation of Learners Beyond' (Pearson 2018) reveals that Generation Z (GenZ) learners, i.e. young people currently aged 14-23, use more modern social media or apps for learning than earlier millennial learners (i.e. those now in their mid-20s to mid-30s) who tended to prefer printed materials, such as books and magazines. While millennials and GenZ learners could perhaps be considered as being 'in transition' with regard to digital usage and literacy, Generation Alpha (i.e. those children born between 2010 and 2025 to millennials) will be the first generation entirely born within the 21st century. Looking even further ahead (2040 beyond), they will constitute the iGeneration who use smartphones and tablets naturally. Perhaps they will also have social mores and expectations about technology and how to use it in their lives which are radically different from those we are currently familiar with?

Many universities and higher education organisations now offer online academic courses on internet platforms, such as Massive Open Online Courses (MOOCs). The MOOC movement aims to integrate the connectivity of social networks, the facilitation of acknowledged experts in a field of study and a collection of freely accessible online resources (McAuley, Stewart, Siemens and Cormier 2010), and it has experienced tremendous success since its launch in 2008. For example, by the end of 2016, there were more than 58 million students, 700 universities, and 6,850 courses from various MOOC providers including Coursera, edX and Udacity®. In addition to online courses, most universities now use online learning management systems, such as Blackboard®, to distribute reading materials and manage online discussions, assessments and other learning activities. As a result, a growing number of assignments and assessment tasks in most academic contexts now require students to read on digital screens rather than in print. This large-scale shift to onscreen study and learning might suggest that the medium in which academic reading takes place has been steadily changing, and it may even have potential consequences for the very nature of the mental processes that are engaged by the reader during such onscreen reading. This in turn could have significant implications for the future construct of reading within a sociocognitive approach in terms of the likely interactions occurring between the contextual features and the cognitive processing involved. The spread of multilingualism brings another dimension into the frame, as language repertoires (L1 plus non-target L2) and the neurological dimensions are increasingly being focused on as part of the test taker's ability to process language input and construct interpretations and meanings. Research into these areas remains in its early stages (see some examples below), but the changing nature and impact of both context and processing opens up new and exciting avenues for research given the ways in which it is becoming possible to capture individualised performance data through technologies such as eye-tracking and keystroke logging (though see below for some discussion of possible ethical issues raised).

While there has been growing interest in the potential for new technologies to support academic reading, some early empirical research suggested that tools such as laptop computers, tablets and smartphones actually have little or no added benefit for reading performance in adults (e.g., Ackerman and Goldsmith 2011, Ackerman and Lauterman 2012, Singer and Alexander 2016) or in adolescents (e.g., Mangen et al 2013, Rasmusson 2015). Available research suggests that there can be an increased demand in cognitive load when reading onscreen, possibly due to relative unfamiliarity with the reading devices and specific strategies for reading onscreen. However, the question of whether increased cognitive load might impede reading comprehension when reading onscreen remains largely unanswered. Moreover, appropriate research has not yet been conducted with the population of GenZ 'digital natives', i.e. those who may have bypassed the computer in favour of mobile devices, such as smartphones and tablets. For this population the impact of device unfamiliarity may not apply, and perhaps the nature of their reading may be automaticised in a different manner?

A recent meta-analysis study comparing reading performance on screen and on paper showed that reading comprehension was better when reading was done on paper rather than on screen (Kong, Seo and Zhai 2018). Most of the studies Kong et al included in their meta-analysis measured comprehension in terms of mean scores in a multiple-choice comprehension test/ quiz of a text of about 1,000 words. One of the reasons offered for the differential performance was that the contextual cues available in the digital environments 'hinder cognitive processes, while paper tends to facilitate more effective learning' (Ackerman and Lauterman 2012:1,826). However, the researchers noted that the magnitude of the difference in reading comprehension between paper and screen diminished between studies conducted prior to 2013 and those conducted after 2013. This observation perhaps suggests that greater familiarity or frequency with onscreen reading among readers increases reading facility? (See also Tanner 2014 for an overview of research into processing of screen-based versus paper-based reading.)

To date, therefore, it seems that the jury is out as far as any comparison of reading in the two modes is concerned. Nevertheless, as desktop computers are increasingly being replaced by mobile devices with smaller screens, there is a need for more research into the use of such electronic devices for academic reading, and for academic literacy more widely, i.e. the role of written assignments, oral presentation tasks, etc. New research questions will need to be framed within the digital realities of the 4th Industrial Revolution (Schwab 2017, Seldon 2018), taking into account the generational shift towards greater use of mobile devices as an alternative to traditional academic resources, whether print (books, journals) or electronic media (computers).

In addition to providing varied and dynamic input to learners as part of their tertiary education experience, it is worth noting that modern mobile devices can also collect and process information provided by learners during their academic reading and associated tasks - including the recording of biometric information such as facial gestures, eye-tracking and keystroke logging. The collection and processing of such personal data has growing potential to be used within courses of study to provide learning-oriented feedback and to facilitate access to specific information likely to be of use to the student at their stage of learning. This data-gathering potential is already leading to increasing personalisation in student study, away from fixed time and place learning. While such developments may enable more individualised and interactive learning opportunities, they also raise interesting ethical considerations in relation to the gathering, storage, interpretation and utilisation of personal data, as well as concerns about data privacy and ownership. Ethical considerations associated with the increasingly interactive and potentially intrusive nature of mobile devices become even more acute in light of the rapidly developing fields of cyber-physical systems and of AI, in which the augmentation of human capabilities is a significant goal.

Assessing academic reading in a digital age

Developments in the tertiary education sector over recent years, including the use of new educational technologies, the shift towards online or virtual learning environments, and the growth of transnational higher education all mean that a one-size-fits-all approach is tilting increasingly towards the development of more personalised approaches to teaching and learning. This in turn means that test tools and testing systems for assessing academic reading in a digital age will need to reflect such developments in a variety of ways. Although some progress has been made in recent decades in the migration of a reading test from the traditional paper-based format to a computer-based platform, the capacity of digital technology to transform the assessment of academic literacy has yet to be fully explored and harnessed, and much of the research and development associated with this remains in its early stages. One possible reason for relatively little innovative progress on this front may be that our traditional approach to conceptualising and operationalising a construct of academic reading for assessment purposes requires a radical rethink in the light of how academic literacy is being enacted in a digital age.

Integrated task formats

Assessment of English language proficiency has been traditionally conducted in relation to four separate language skills – reading, writing, listening and speaking. However, in everyday life, communication is rarely achieved through discrete, isolated language activities. For example, university students typically read in order to write an essay, do an oral presentation or participate in a seminar discussion. Though this has always been the case to some degree in the context of tertiary education, it is particularly true in an age where there is now much greater interaction, and even 'blending', between visual and text-based input, and between written and spoken modes of delivery (e.g. email correspondence), and where communication can be remote and asynchronous as well as face-to-face and in real time. As demonstrated in Chapter 8, integrated task formats (e.g. reading-into-writing or reading-intospeaking) offer relevant and appropriate ways of reflecting the reality of tertiary education and, in theory at least, modern technology should facilitate improved approaches to assessment of academic skills.

To better reflect what language users do in real life, the CEFR (Council of Europe 2001) specifies language use in terms of four modes of communication - reception, production, interaction and mediation. However, as mentioned in Chapter 8, the construct of mediation was not clearly illustrated (with scales and descriptors) until very recently, with the publication of the Companion Volume in 2018 (Council of Europe 2018). The CEFR sees mediation as one of the main functions of language use in schools, universities and social contexts where the user/learner acts as a social agent to transform and (co-)construct meaning from one language to another, from one modality to another, and from one party to another. Such mediation of language in order to bridge, convey and construct meaning in real life goes far beyond the four independent language skills as they are more traditionally conceived and operationalised in most English language tests. According to the CEFR, there are three major forms of mediation: mediating a text (such as processing text, relaying information, note-taking, analysis and evaluation of texts as discussed in Chapter 8); mediating concepts (e.g. to construct meaning collaboratively); and mediating communication (e.g. to facilitate meaning in transnational contexts). To date, most of these skills cannot be operationalised using the more traditional independent task formats that testers have been used to using. However, Saville (2018:xi) anticipates that the newly developed description concerning mediation 'opens up new avenues for language learning and assessment tasks that can be integrated in innovative ways'.

As argued in Section 2, assessments of academic reading should at least use flexible item formats (e.g. MCQs and multiple-matching as opposed to gap-fill tasks) frequently and constructively to target *higher-level* reading skills at global level, which is a critical element in skilled academic reading. Technological solutions should make this easier to accomplish, e.g. using 'click and drag' items or editing/highlighting tasks. It is also important to include test items which require students to connect a series of short texts to the paragraphs of a longer text (to test the ability to form a discourse-level representation) as well as summarising main ideas from multiple sources. Once again, in theory at least, digital technology should make it easier to operationalise these approaches, avoiding the need for extensive printed material for the student to manipulate. In the digital age, if an assessment of academic reading aims to tap into some of the integrated and interactive reading skills as illustrated by the new mediation descriptors, one possibility would be to use scenario-based task formats which can incorporate features of mediation in academic contexts. For example, registered test takers could be allowed access 24 hours in advance to multiple reading texts that will be available to them during a mediation task (e.g. to summarise a lecture for a peer who missed the class), so that they can familiarise themselves with them. Registered test takers could have a 'word pad' access for note-taking, which would be available to them again for reference during the communication task. Some of these integrated and interactive features of an academic reading task could be enhanced by adopting innovative assessment solutions, as discussed below.

Innovative assessment solutions

As reading on screen and on smaller mobile devices has become the norm in most academic contexts, one might reasonably expect an assessment of academic reading to involve some actual online reading. Computer-based test delivery has several advantages in replicating academic reading in real life. First, it could help to enforce more rigid time restrictions that require the use of expeditious reading strategies during the assessment as university students search for relevant, useful information from a vast range of online reading materials. Secondly, an online, restricted source repository within a set time frame (of perhaps 48 hours) could easily be created for scenario-based tasks such as the example provided above. Thirdly, a computer-based test delivery system could allow higher-level students a degree of autonomy in selecting and handling texts. One possibility would be a password-protected site which could be the hosting platform for students to create, as part of the test response (e.g. an essay, a report or a presentation), a small-scale 'portfolio' of reading and writing (and also listening and speaking) materials on an assigned topic. The portfolio of materials created by students could receive formative assessment and feedback, in addition to a traditional summative set of scores on the response.

A computer-based delivery system would also allow systematic and innovative ways of collecting learner/learning data for ongoing monitoring, evaluation and feedback. However, simply adopting a computer-based delivery mode for a test of academic reading ability risks failing to capitalise on some of the other enhancements that might be possible thanks to modern technology. For example, artificial intelligence (AI) has become ubiquitous in a number of aspects of higher education study, from predicting spelling and words in writing, to personal assistants in learning apps or smartphones, or even to making high-stakes student admission decisions using algorithms. So how might AI be applied to support innovative assessment solutions? Could AI be used more in rating, e.g., from automated scoring of dichotomous and multiple-choice reading items to more sophisticated content grading? What might be some of the adaptive solutions in reading assessment which respond to the individual profile and needs of a particular student (perhaps at a particular stage in their study journey) - placing greater emphasis on certain reading skills, assigning to them genres that they haven't yet mastered, and generally helping them to improve their academic reading at their own pace? Could AI be used to complement current item analysis approaches, providing feedback to task design as it monitors student processing, performance and progress? Perhaps more 'out-of-the-box' thinking on the part of test designers and developers is required for assessments of academic reading, and academic literacy more widely, to harness the new opportunities and potential offered by a digital age?

Data-driven individualised experience

If more data is available concerning an individual learner's reading processes, behaviours and habits, we might also then be in a position to develop targeted assessments which better adapt to a test taker's ability level and characteristics. Technology now offers us some highly sophisticated means of collecting learner data of various types.

Eye-tracking data, for example, can be used to inform students' processing of reading tests (Bax 2013, Bax and Chan 2016, 2019). With the popularity of reading on screen, it is now much easier to collect longitudinal data on reading behaviours such as the genres and levels of books read on electronic devices as well as speed of comprehension of texts in various contexts (Matsuo et al 2018, Roelfsema, Denys and Klink 2018). Advances in neurological science mean that it is now possible to measure the distribution of students' mental effort corresponding to learning content as detected by the classifier using learners' electroencephalography (EEG) data. Drawing upon Cognitive Load Theory, Lin and Kao (2018) examined students' mental efforts when they were learning on MOOCs. They argued that the data can effectively facilitate our understanding of students' mental efforts in online learning contexts to enable automatic feedback in relation to segments of learning materials on which they have expended heavy mental effort.

Using sophisticated data-processing and statistical techniques such as computational natural language processing, machine learning or AI (Subramanian et al 2018), data on learner reading could be used to develop a form of personalised profile. This profile could then be used to construct feedback, and create tasks and activities that are more tailored and appropriate to individuals (Buckley and Doyle 2017, Dafoulas, Maia, Samuels-Clarke, Ali and Augusto 2018).

A more dynamic system

In order to achieve some of the features mentioned above, we anticipate that a new functioning testing system for the assessment of academic reading would have to be far more flexible and dynamic than what is currently offered by most large-scale standardised testing systems. O'Sullivan (2015:3) argues that a dynamic testing system should not be static and invariable but allow changes to be implemented 'in the course of actively engaging with the needs of test users, and in taking into consideration the recommendations generated by an active validation research agenda'.

It is also possible that such a system could draw on the experience and expertise of gamification, which applies the mechanics, dynamics and aesthetics associated with games to non-game contexts (Simoes, Redondo and Vilas 2013). Gamification has excited considerable interest in educational contexts as a way to increase student engagement, motivate and promote learning and facilitate students in the development of sustainable life skills such as financial literacy (Decos 2015) and information literacy (Buckley and Doyle 2017, Markey et al 2008). Might gamification also be appropriate for enhancing the development and assessment of academic literacy?

The creation of a new functioning testing system for the assessment of academic reading in a digital age as envisaged above will clearly require considerable additional work in the fields of computational intelligence, psycholinguistics, and language testing theory and practice. This volume seeks to make a small contribution to this endeavour by mapping out the territory and the journey that have already been covered with regard to the assessment of academic reading ability over the past half century.

The sociocognitive framework has served us well as a convenient and convincing framework of reference for examining the key features of a test that can help to make it fit for purpose, as well as for mapping out an ongoing agenda for research and validation. Over more than a decade Weir has sought to show how various aspects of validity (cognitive, contextual, scoring, criterion-related and consequential) can assist language testers to conceptualise and operationalise the 'unitary validity' approach first proposed by Messick (1989, 1995). Messick himself acknowledged the need to break down unitary validity into more discrete components, in his case six 'distinguishable aspects of construct validity' (Messick 1995:744): theoretical rationales including process models of task performance; content relevance and representativeness; scoring structure; generalisability of score properties and interpretations; external aspects including convergent and discriminant evidence; and consequential aspects.

Evidence pertinent to all these aspects needs to be integrated into an overall validity judgement to sustain score inferences and their action implications . . . which is what is meant by validity as a unified concept (Messick 1995:747).

The approach adopted in this volume is an interactionalist position which sees the academic reading construct as residing in the interactions between the underlying cognitive ability, the context of use and the process of scoring, but set within a wider framework of the appropriateness, meaningfulness and usefulness of score inferences.

Having said this, there remains much work to be done to explore and better understand the way in which the different validity components or aspects interact with one another to generate the desired outcome. With this in mind, we conclude our volume by summarising some of the essential questions that appear to merit investigation. In themselves, the questions below do not constitute a coherent research agenda; nevertheless, they may help to inform and shape the development of such an agenda for language testers and assessment specialists over the coming years. For convenience and ease of reference, the questions are grouped according to the six components of test validity of the sociocognitive framework as laid out in Weir (2005b).

Academic reading and academic literacy: Informing a future research agenda?

Test taker characteristics

- What are the characteristics of test takers in a digital age in relation to their physical, psychological and experiential characteristics (cf. O'Sullivan 2000)?
- Is it possible to identify new and emerging test taker characteristics that are specific to 'digital readers'?

- How can technology help to better cater for individual variation across test takers without compromising test fairness?
- What new forms of digital arrangement can be made available to test takers with special learning needs such as dyslexia, hyperlexia and dysgraphia?

Context validity

- How is the nature of academic reading impacted by the presence of more visual material in the form of images and/or video clips that accompany text-based input (e.g. YouTube videos, TED Talks, online lectures, etc.)?
- Do university students read the same academic materials (e.g. in terms of text length, genres, complexity of ideas, integration of word and image) when reading on screen and on paper? What are the new 'genres' of academic texts in a digital age?
- How do lecturers and students use mobile devices to support reading motivation and extend comprehension?
- To what extent does digital literacy, i.e. the ability to both understand and use digitised information (Gilster 1997:2), influence students' academic study and language use? Should digital literacy be part of the construct of academic reading in a digital age, and if so, how should this be defined and operationalised?
- How do academic literacy tasks change in a digital age? What academic reading skills are necessary to perform these tasks pre-entry to higher education and which ones are to be acquired post-entry?
- Should administration of academic reading assessment be shifted from a more homogeneous and centralised to a more tailored and user-oriented approach?
- How can technology best be applied to task design in order to elicit specific types of reading skills, e.g. expeditious reading and search reading?

Cognitive validity

- How do web-based tools, such as search engines, change the ways students locate relevant information (search reading) for academic purposes?
- How do students process information gathered from different web-based sources (e.g. paraphrasing others' ideas, critical evaluation of sources/ ideas, synthesis of multiple texts) for academic purposes?

- Would some reading skills become more essential to be successful in academic reading in a digital age? For example, would goal setting (i.e. selecting appropriate type of reading) be increasingly important when students are exposed to a vast amount of information? If so, what would be the best ways to assess these skills?
- Does reading on smaller mobile devices make it easier or more difficult to comprehend and retain ideas, and why is that?
- What specific strategies, if any, do readers adopt when reading academic texts on smaller mobile devices?

Scoring validity

- As the use of new formats and task types is anticipated in academic reading tests in a digital age, what would be the best ways to rate or evaluate the performance elicited by these tasks to reflect the intended construct?
- What kinds of rater (human and machine) training are needed?
- How could technology provide us with more sophisticated approaches to scoring, score processing and score interpretation, perhaps using complex algorithms or data mapping?

Consequential validity

- While it is feasible to trace students' reading behaviours within and beyond test conditions through the use of data analytics, what are the ethical considerations of using such data?
- How should we set standards and codes of practice regarding the use of students' reading data? How should consent to use this data be established? Is students' reading data on social media in academic/educational contexts private or public? What is the potential harm to students if their reading behaviour data (i.e. preference of particular genres) and biometric information are misused or abused?
- Using learning analytics and algorithms, would we be helping the higher education sector to make better decisions regarding the level of test takers' reading proficiency?

Criterion-related validity

• As the assessment industry moves towards a more individualised and personalised approach to testing and assessment, what are the

implications for achieving 'equivalence' with different versions of the same test and the current item banking approaches? What would be the best ways to control and 'equalise' the features/levels of reading input?

• What are the implications for cross-test comparability, especially between existing academic reading tests and new generation academic reading tests that take account of the role of technology in academic reading?

Validity as a unitary concept

- How might it be possible to explore the interactions between the different components of the sociocognitive framework? Are certain interactions more critical/dominant than others?
- How might new technologies facilitate the collection and analysis of data to achieve this?

This list of questions is not intended to be exhaustive, but simply to highlight some key avenues for investigation. Answers to questions such as these will have direct implications for the way in which academic reading skills will be assessed most effectively in the years that lie ahead, and could help to shape and guide a relevant and fruitful research agenda for the future. Our aim is that others will build upon our efforts thus far and so lead us on to ever better ways of assessing academic literacy.

References and further reading

- Ackerman, R and Goldsmith, M (2011) Metacognitive regulation of text learning: On screen versus on paper, *Journal of Experimental Psychology: Applied* 17 (1), 18–32.
- Ackerman, R and Lauterman, T (2012) Taking reading comprehension exams on screen or on paper? A metacognitive analysis of learning texts under time pressure, *Computers in Human Behavior* 28 (5), 1,816–1,828.
- Alderson, J C (1978) A Study of the Cloze Procedure with Native and Non-Native Speakers of English, PhD thesis, University of Edinburgh.
- Alderson, J C (1988) New procedures for validating proficiency tests of ESP? Theory and practice, *Language Testing* 5 (2), 220–232.
- Alderson, J C (1996) The testing of reading, in Nuttall, C (Ed) *Teaching Reading in a Foreign Language* (Second edition), London: Heinemann, 212–228.
- Alderson, J C (2000) Assessing Reading, Cambridge: Cambridge University Press.
- Alderson, J C (2005) *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*, London: Continuum.
- Alderson, J C and Banerjee, J V (1996) *How might impact study instruments be validated*², unpublished manuscript commissioned by UCLES.
- Alderson, J C and Clapham, C M (1992) Applied linguistics and language testing: A case study of the ELTS Test, *Applied Linguistics*, 13 (2), 149–167.
- Alderson, J C and Urquhart, A H (1985) The effect of students' academic discipline on their performance on ESP reading tests, *Language Testing* 2 (2), 192–204.
- Alderson, J C and Wall, D (1993) Does washback exist?, *Applied Linguistics* 14 (2), 115–129.
- Alderson, J C and Wall, D (1996) Special Issue: Washback, *Language Testing* 13 (3).
- Alderson, J C, Clapham, C M and Wall, D (1995) Language Test Construction and Evaluation, Cambridge: Cambridge University Press.
- Alderson, J C, Figueras, N, Kuijper, H, Nold, G, Takala, S and Tardieu, C (2004) The development of specifications for item development and classification within the Common European Framework of Reference for Languages: learning, teaching, assessment: Reading and listening, available online: eprints.lancs. ac.uk/44/1/final_report.pdf
- Alexander, P A, Graham, S and Harris, K R (1998) A perspective on strategy research: Progress and prospects, *Educational Psychology Review* 10 (2), 129–154.
- Al-Fallay, I (1994) Limiting bias in the assessment of English as a foreign language: the impact of background knowledge on the proficiency of Saudi Arabian students learning English as a foreign language, unpublished PhD thesis, University of New Mexico.
- Allan, D (2016) Investigating washback to the learner from the IELTS test in the Japanese tertiary context, *Language Testing in Asia* 6 (7), 1–20.

ALTE (1998) *Multilingual Glossary of Language Testing Terms*, Studies in Language Testing volume 6, Cambridge: UCLES/Cambridge University Press.

American Educational Research Association, American Psychological Association and National Council on Measurement in Education (1999) *Standards for Educational and Psychological Testing*, Washington, DC: American Educational Research Association.

Anderson, N J (1991) Individual differences in strategy use in second language reading and testing, *The Modern Language Journal* 75 (4), 460–472.

Anderson, R and Pearson, P (1988) A schema-theoretic view of basic processes in reading comprehension, in Carrell, P, Devine, J and Eskey, D (Eds) *Interactive Approaches to Second Language Reading*, Cambridge: Cambridge University Press, 37–55.

Applebee, A N (1984) Writing and reasoning, *Review of Educational Research* 54 (4), 577–596.

Arnaudet, M L and Barrett, M E (1984) *Approaches to Academic Reading and Writing*, New Jersey: Prentice Hall.

Ascención-Delaney, Y (2008) Investigating the reading-to-write construct, *Journal* of English for Academic Purposes 7, 140–150.

Ata, A W (2015) Knowledge, education, and attitudes of international students to IELTS: A case of Australia, *Journal of International Studies* 5 (4), 488–500.

Bachman, L F (1990) *Fundamental Considerations in Language Testing*, Oxford: Oxford University Press.

Bachman, L F (2004) *Statistical Analyses for Language Assessment*, Cambridge: Cambridge University Press.

Bachman, L F and Palmer, A S (1996) *Language Testing in Practice*, Oxford: Oxford University Press.

Bachman, L F, Kunnan, A, Vanniarajan, S and Lynch, B (1988) Task and ability analysis as a basis for examining content and construct comparability in two EFL proficiency test batteries, *Language Testing* 5, 128–159.

Bachman, L F, Davidson, F, Ryan, K and Choi, I-C (1995) An Investigation into the Comparability of Two Tests of English as a Foreign Language: The Cambridge–TOEFL Comparability Study, Studies in Language Testing volume 1, Cambridge: UCLES/Cambridge University Press.

Banerjee, J V (1996) *The design of the classroom observation instruments,* Cambridge: UCLES internal report.

Banerjee, J V (2003) *Interpreting and using proficiency test scores*, unpublished PhD thesis, Lancaster University.

Barker, K (1998) A trainable bracketer for noun modifiers, in Mercer, R E and Neufeld, E (Eds) Advances in Artificial Intelligence: Proceedings of the 12th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, Vancouver: Springer, 196–210.

Barnett, M (1989) *More Than Meets the Eye*, Englewood Cliffs: Prentice Hall Regents.

Baron, N S (2013) Redefining reading: The impact of digital communication media, *PMLA* 128 (1), 193–200.

Baron, N S (2015) *Words Onscreen: The Fate of Reading in a Digital World*, New York: Oxford University Press.

Baron, N S, Calixte, R M and Havewala, M (2017) The persistence of print among university students: An exploratory study, *Telematics & Informatics* 34, 590–604.

Bartlett, F C (1932) *Remembering: A Study in Experimental and Social Psychology*, Cambridge: Cambridge University Press.

- Baumann, J (Ed) (1986) *Teaching Main Idea Comprehension*, Newark: International Reading Association.
- Bax, S (2013) *Readers' Cognitive Processes During IELTS Reading Tests: Evidence from Eye Tracking*, ELT Research Papers 13–06, London: British Council.
- Bax, S (2015) Using eye-tracking to research the cognitive processes of multinational readers during an IELTS reading test, *IELTS Research Report Series* 2, 1–21.
- Bax, S and Chan, S (2016) Researching the Cognitive Validity of GEPT High-Intermediate and Advanced Reading: An Eye Tracking and Stimulated Recall Study, LTTC-GEPT Research Reports RG-07, available online: www.lttc.ntu. edu.tw/lttc-gept-grants/RReport/RG07.pdf
- Bax, S and Chan, S (2019) Using eye-tracking research to investigate language test validity and design, *System* 83, 64–78.
- Bayliss, A and Ingram, D E (2006) *IELTS as a Predictor of Academic Language Performance*, available online: aiec.idp.com/uploads/pdf/BaylissIngram%20 (Paper)%20Wed%201630%20MR5.pdf
- Beard, R (1972) *Teaching and Learning in Higher Education*, Harmondsworth: Penguin.
- Bellingham, L (1993) The relationship of language proficiency to academic success for international students, *New Zealand Journal of Educational Studies* 30 (2), 229–232.
- Bensoussan, M and Ramraz, R (1984) Testing EFL reading comprehension using a multiple-choice rational cloze, *Modern Language Journal* 68 (3), 230–239.
- Berman, R (1984) Syntactic components of the foreign language reading process, in Alderson, J C and Urquhart, A (Eds) *Reading in a Foreign Language*, London: Longman, 139–159.
- Bernhardt, E (1991) Reading Development in Second Language: Theoretical, Empirical and Classroom Perspectives, New Jersey: Ablex Publishing Corporation.
- Bhatia, V (1997) Applied genre analysis and ESP, in Miller, T (Ed) *Functional Approaches to Written Text: Classroom Applications*, Washington DC: USIA, 134–149.
- Biber, D, Conrad, S M, Reppen, R, Byrd, P, Helt, M, Clark, V, Cortes, V, Csomay, E and Urzua, A (2004) *Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus*, TOEFL[®] Monograph Series MS-25, Princeton: Educational Testing Service.
- Blackmore, J, Farrell, L, Devlin, M, Arber, R, Gribble, C and Rahimi, M (2010/2012) Investigating Stakeholder Responses to Changing Skilled Migration Policies for Australian International Graduates, Melbourne/Canberra: Deakin University/Australian Research Council.
- Bleasdale, F A (1987) Concreteness-dependent associative priming: Separate lexical organization for concrete and abstract words, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13 (4), 582–594.
- Block, E (1986) The comprehension strategies of second language readers, *Tesol Quarterly* 20 (3), 463–494.
- Bond, T G and Fox, C M (2001) *Applying the Rasch Model*, Hillsdale: Lawrence Erlbaum.
- Bonkowski, F (1996) *Instruments for the assessment of teaching materials*, unpublished MA assignment, Lancaster University.
- Brantmeier, C (2002) Second language reading strategy research at the secondary and university levels: Variations, disparities, and generalizability, *The Reading Matrix* 2 (3), 1–14.

- Brewer, W F (1980) Literary theory, rhetoric, and stylistics: Implications for psychology, in Spiro, R J, Bruce, B C and Brewer, W F (Eds) *Theoretical Issues in Reading Comprehension: Perspectives from Cognitive Psychology, Linguistics, Artificial intelligence and Education,* Hillsdale: Laurence Erlbaum, 221–239.
- Bridgeman, B, Cho, Y and Di Pietro, S (2016) Predicting grades from an English language assessment: The importance of peeling the onion, *Language Testing* 33 (3), 307–318.
- British National Corpus (2007) British National Corpus, available online: www. natcorp.ox.ac.uk
- Britt, M A and Aglinskas, C (2002) Improving students' ability to identify and use source information, *Cognition and Instruction* 20 (4), 458–522.
- Britt, M A and Sommer, J (2004) Facilitating textual integration with macrostructure focusing tasks, *Reading Psychology* 25 (4), 313–339.
- Brown, J, Fishco, V and Hanna, G (1993) Nelson-Denny Reading Test: Manual for Scoring and Interpretation, Forms G & H, Chicago: Riverside.
- Brunfaut, T (2016) Looking into Reading II: A Follow-up Study On Test-takers' Cognitive Processes While Completing Aptis B1 Reading Tasks, British Council Validation Series VS/2016/001, London: The British Council.
- Brunfaut, T and McCray, G (2015) Looking Into Test-Takers' Cognitive Processes Whilst Completing Reading Tasks: A Mixed-method Eye-tracking and Stimulated Recall Study, British Council Validation Series AR-G/2015/001, London: The British Council.
- Buck, G (2001) Assessing Listening, Cambridge: Cambridge University Press.
- Buck, G, Tatsuoka, K and Kostin, I (1997) The sub-skills of reading: Rule-space analysis of a multiple-choice test of second language reading comprehension, *Language Learning* 47 (3), 423–466.
- Buckley, P and Doyle, E (2017) Individualising gamification: An investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market, *Computers and Education* 106, 43–55.
- Buell, J G (1992) TOEFL and IELTS as Measures of Academic Reading Ability: An Exploratory Study 2, paper presented at the 26th Annual Meeting of the Teachers of English to Speakers of Other Languages, Vancouver, Canada, 4–7 March 1992.
- Buell, J G (1993) TOEFL and IELTS as measures of academic reading ability, Southern Illinois working papers in linguistics and language teaching, 1–17.
- Bullock, A (1975) A Language for Life: Report of the Committee of Inquiry Appointed by the Secretary of State for Education and Science Under the Chairmanship of Sir Alan Bullock, London: HM Stationery Office.
- Cambridge Assessment English (2016) Principles of Good Practice. Research and Innovation in Language Learning and Assessment, Cambridge: Cambridge Assessment English.
- Cambridge English Language Assessment (2013) Comparing scores on Cambridge English: Advanced (CAE) and IELTS, Cambridge: Cambridge English Language Assessment.
- Campion, M and Elley, W (1971) An Academic Vocabulary List, NZCER: Wellington.
- Carrell, P L (1983) Three components of background knowledge in reading comprehension, *Language Learning* 33 (2), 183–207.
- Carrell, P L (1984) The effects of rhetorical organisation on ESL readers, *TESOL Quarterly* 18 (3), 441–469.

- Carrell, P L (1989) Metacognitive awareness and second language reading, *The Modern Language Journal* 73 (2), 121–134.
- Carrell, P L, Devine, J and Eskey, D (Eds) (1988) Interactive Approaches to Second Language Reading, Cambridge: Cambridge University Press.
- Cartelli, A (2012) Current Trends and Future Practices for Digital Literacy and Competence, Hershey: IGI Global.
- Carver, R P (1992) Reading rate: Theory, research and practical implications, Journal of Reading 36, 84–95.
- Carver, R P (1997) Reading for one second, one minute, or one year from the perspective of rauding theory, *Scientific Studies of Reading* 1 (1), 3–43.
- Chalmers, J and Walkinshaw, I (2014) Reading strategies in IELTS tests: Prevalence and impact on outcomes, *English Australia Journal* 30 (1), 24–39.
- Chan, S (2011) Demonstrating cognitive validity and face validity of PTE Academic Writing Items Summarize Written Text and Write Essay, *Pearson Research Notes*, 1–20.
- Chan, S (2013) *Establishing the construct validity of EAP reading-into-writing tests*, unpublished PhD thesis, University of Bedfordshire.
- Chan, S (2017) Using keystroke logging to understand writers' processes on a reading-into-writing test, *Language Testing in Asia* 7 (1), 1–27.
- Chan, S (2018) *Defining Integrated Reading-into-Writing Constructs: Evidence at the B2–C1 Interface*, English Profile Studies volume 8, Cambridge: UCLES/ Cambridge University Press.
- Chan, S, Wu, R, and Weir, C J (2014) Examining the context and cognitive validity of the GEPT Advanced Writing Task 1: A comparison with real-life academic writing tasks, *LTTC-GEPT Research Reports* 3, 1–102.
- Chan, S, Inoue, C and Taylor, L (2015) Developing rubrics to assess the readinginto-writing skills: A case study, *Assessing Writing*, available online: uobrep. openrepository.com/uobrep/handle/10547/621934
- Charge, N and Taylor, L (1997) Recent developments in IELTS, *ELT Journal* 51 (4), 374–380.
- Chen, R S and Vellutino, F (1997) Prediction of reading ability: A cross-validation of the simple view of reading, *Journal of Literacy Research* 29, 1–24.
- Cheng, L (2005) Changing Language Teaching through Language Testing: A Washback Study, Studies in Language Testing volume 21, Cambridge: UCLES/Cambridge University Press.
- Cheng, L, Watanabe, Y and Curtis, A (Eds) (2004) *Washback in Language Testing: Research Context and Methods*, Hillsdale: Lawrence Erlbaum.
- Cheung, K Y F, McElwee, S and Emery J (Eds) (2017) *Applying the Sociocognitive Framework to the BioMedical Admissions Test: Insights from Language Assessment*, Studies in Language Testing volume 49, Cambridge: UCLES/Cambridge University Press.
- Chihara, T, Sakurai, T and Oller, J (1989) Background and culture as factors in EFL reading comprehension, *Language Testing* 6 (2), 143–151.
- Chikalanga, I (1990) *Inferencing in the Reading Process*, unpublished PhD thesis, University of Reading.
- Chikalanga, I (1992) A suggested taxonomy of inferences for the reading teacher, *Reading in a Foreign Language* 8 (2), 697–709.
- Clapham, C M (1993) Is ESP testing justified?, in Douglas, D and Chapelle, C (Eds) *A New Decade of Language Testing Research*, Washington, DC: TESOL, 257–271.

- Clapham, C M (1996a) *The Development of IELTS: A Study of the Effect of Background Knowledge on Reading Comprehension*, Studies in Language Testing volume 4, Cambridge: UCLES/Cambridge University Press.
- Clapham, C M (1996b) The academic modules: reading, in Clapham, C and Alderson J C (Eds) *IELTS Research Report 3: Constructing and Trialling the IELTS Test*, London: British Council, UCLES/IDP Education Australia, 49–68.
- Clapham, C M (2000) Assessment for academic purposes: where next?, *System* 28 (4), 511–521.
- Cleland, J, Dowell, J, McLachlan, J, Nicholson, S and Patterson, F (2012) *Identifying best practice in the selection of medical students,* available online: www.gmc-uk.org/about/what-we-do-and-why/data-and-research/research-andinsight-archive/identifying-best-practice-in-the-selection-of-medical-students
- Clymer, T (1972) What is reading? Some current concepts, in Robinson, H M (Ed) Innovation and Change in Reading Instruction, Sixty-seventh Yearbook of the National Society for the Study of Education Part II, Chicago: University of Chicago Press, 7–29.
- Coady, J (1979) A psycholinguistic model of the ESL reader, in Mackay, R, Barkman, B and Jordan, R R (Eds) *Reading in a Second Language*, Massachusetts: Newbury House, 5–12.
- Cobb, T (2003) VocabProfile, available online: www.lextutor.ca
- Cohen, A D (1994) English for academic purposes in Brazil: The use of summary tasks, in Hill, C and Parry, K (Eds) *From Testing to Assessment: English as an International Language*, New York: Longman, 174–204.
- Cohen, A D (2006) The coming of age of research on test-taking strategies, Language Assessment Quarterly 3 (4), 307–331.
- Cohen, A D and Upton, T (2006) *Strategies in Responding to the New TOEFL Reading Tasks*, TOEFL[®] Monograph Series 33, Princeton: Educational Testing Service.
- Cohen, A D and Upton, T (2007) 'I want to go back to the text': Response strategies on the reading subtest of the new TOEFL[®], *Language Testing* 24 (2), 209–250.
- Coleman, D, Starfield, S and Hagan, A (2003) The attitudes of IELTS stakeholders: Student and staff perceptions of IELTS in Australian, UK and Chinese tertiary institutions, *IELTS Research Reports* 5, 160–235.
- Corkill, A J, Glover, J A and Bruning, R H (1988) Advance organizers: Concrete versus abstract, *Journal of Educational Research* 82 (2), 76–81.
- Cotton, F and Conrow, F (1998) An investigation of the predictive validity of IELTS amongst a group of international students studying at the University of Tasmania, *IELTS Research Reports* 1, 72–115.
- 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 (CEFR): A Manual. Preliminary Pilot Version, Strasbourg: Language Policy Division.
- Council of Europe (2004) Reference supplement to the Preliminary Pilot Version of the Manual for Relating language examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment, DGIV/EDU/LANG (2004) 13, Strasbourg: Language Policy Division.

- Council of Europe (2018) Common European Framework of Reference for Languages: Learning, Teaching, Assessment, Companion Volume with New Descriptors, Strasbourg: Education Policy Division.
- Coxhead, A (2000) A new academic word list, *TESOL Quarterly* 34 (2), 213–238.
- Craven, E (2012) The quest for IELTS band 7.0: Investigating English language proficiency development of international students at an Australian university, *IELTS Research Reports* 13, 1–61.
- Criper, C and Davies, A (1988) *ELTS Validation Project Report*, London: The British Council.
- Cromer, W (1970) The difference model: A new explanation for some reading difficulties, *Journal of Educational Psychology* 61 (6), 471–483.
- Crossley, J, Greenfield, J and McNamara, D S (2008) Assessing text readability using cognitively based indices, *TESOL Quarterly* 42 (3), 475–493.
- Cumming, A (1989) Writing expertise and second-language proficiency, *Language Learning* 39, 81–141.
- Cumming, A (2013) Assessing integrated writing tasks for academic purposes: Promises and perils, *Language Assessment Quarterly* 10 (1), 1–8.
- Cumming, A, Kantor, R, Baba, K, Erdosy, U, Eouanzoui, K and James, M (2005) Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL, Assessing Writing 10 (1), 5–43.
- Dafoulas, G, Maia, C C, Samuels-Clarke, J, Ali, A and Augusto, J C (2018) Investigating the role of biometrics in education – the use of sensor data in collaborative learning, paper presented at the 12th International Conference on e-Learning, Madrid, 17–19 July 2018.
- Dahan Golan, D, Barzillai, M and Katzir, T (2018) The effect of presentation mode on children's reading preferences, performance, and self-evaluations, *Computers and Education* 126, 346–358.
- Daller, M H and Phelan, D (2013) Predicting international student study success, *Applied Linguistics Review* 4 (1), 173–193.
- Davidson, F (1998) Book review: The development of IELTS: A study of the effect of background knowledge on reading comprehension, *Language Testing* 15 (2), 288–292.
- Davies, A (1965) *Proficiency in English as a second language*, unpublished PhD thesis, University of Birmingham.
- Davies, A (Ed) (1968) Language Testing Symposium: A Psycholinguistic Approach, Oxford: Oxford University Press.
- Davies, A (1984) Validating three tests of English language proficiency, *Language Testing* 1, 50–69.
- Davies, A (1990) Principles of Language Testing, Oxford: Blackwell.
- Davies, A (2008) Assessing Academic English: Testing English Proficiency 1950–1989 – The IELTS Solution, Studies in Language Testing volume 23, Cambridge: UCLES/Cambridge University Press.
- Davies, A and Criper, C (1988) *ELTS Research Report 1(i)*. *ELTS Validation Project Report*, London: The British Council/UCLES.
- Davies, A and Widdowson, H G (1974) Reading and writing, in Allen, J P B and Corder, S P (Eds) *Techniques in Applied Linguistics*, The Edinburgh Course in Applied Linguistics 3, 155–201.
- Davies, A, Brown, A, Elder, C, Hill, K, Lumley, T and McNamara, T (1999) Dictionary of Language Testing, Studies in Language Testing volume 7, Cambridge: UCLES/Cambridge University Press.

- Dechant, E (1991) Understanding and Teaching Reading: An Interactive Model, Hillsdale: Lawrence Erlbaum.
- Decos, L (2015) Opportunity for banks to utilize gamification as a tool to promote financial education to children, available online: www.theseus.fi/bitstream/ handle/10024/90868/Liliia_deCos.pdf?sequence=1&isAllowed=y
- Dillon, A (1992) Reading from paper versus screens: A critical review of the empirical literature, *Ergonomics* 35 (10), 1,297–1,326.
- Dooley, P (1999) An investigation into the predictive validity of the IELTS Test as an indicator of future academic success, in Martin, K Stanley, N and Davison N (Eds) *Teaching in the Disciplines/ Learning in Context*, Perth: The University of Western Australia, 114–118.
- Dooley, P and Oliver, R (2002) An investigation into the predictive validity of the IELTS test as an indicator of future academic success, *Prospect* 17 (1), 36–54.
- Douglas, D (2000) Assessing Language for Specific Purposes: Theory and Practice, Cambridge: Cambridge University Press.
- Drum, P A, Calfee, R C and Cook, L K (1981) The effects of surface structure variables on performance in reading comprehension tests, *Reading Research Quarterly* 16 (4), 486–514.
- Dunlea, J (2015) Validating a set of Japanese EFL proficiency tests: Demonstrating locally designed tests meeting international standards, PhD thesis, University of Bedfordshire.
- Durkin, K (2004) Challenges Chinese Students Face in Adapting to Academic Expectations and Teaching/learning Styles of UK Masters Courses, and How Cross-cultural Understanding and Adequate Support Might aid them to Adapt, London: British Council.
- Durst, R (1985) Cognitive and linguistic dimensions of secondary students' analytic writing, unpublished PhD thesis, Stanford University.
- Durst, R (1987) Cognitive and linguistic demands of analytic writing, *Research in the Teaching of English* 21, 347–376.
- Editorial (1980) Why comprehension? *Reading Research Quarterly* 15 (2), 181–182.
- Ehrlich, M F (1991) The processing of cohesion devices in text comprehension, *Psychological Research* 53 (2), 169–174.
- Elander, J, Harrington, K, Norton, L, Robinson, H and Reddy, P (2006) Complex skills and academic writing: A review of evidence about the types of learning required to meet core assessment criteria, *Assessment & Evaluation in Higher Education* 31 (1), 71–90.
- Elder, C (1993) Language proficiency as a predictor of performance in teacher education, *Melbourne Papers in Language Testing* 2 (1), 68–85.
- Engineer, W D (1977) An Investigation of a Reading Model for English as a Second Language, unpublished PhD thesis, University of Edinburgh.
- Enright, M, Grabe, W, Koda, K, Mosenthal, P, Mulcany-Ernt, P and Schedl, M (2000) TOEFL 2000 Reading Framework: A Working Paper, TOEFL[®] Monograph Series 17, Princeton: Educational Testing Service.
- Enright, M, Bridgeman, B and Cline, F (2002) *Prototyping a test design for a new TOEFL*, paper presented at Research in support of the development of new TOEFL, symposium conducted at the meeting of the National Council on Measurement in Education, New Orleans, April 2002.
- Esmaeili, H (2002) Reading-to-write reading and writing tasks and ESL students' reading and writing performance in an English language test, *The Canadian Modern Language Review* 58, 599–622.

- ETS (2010) ETS Linking TOEFL iBTTM Scores to IELTS® Scores A Research Report, available online: www.ets.org/s/toefl/pdf/linking_toefl_ibt_scores_to_ ielts_scores.pdf
- Everett, R and Colman, J (1999) A critical analysis of selected IELTS preparation materials, *IELTS Australia Research* 5, 1–84.
- Farr, R, Carey, R and Tone, B (1986) Recent theory and research into the reading process: Implications for reading assessment, in Orasanu J (Ed) *Reading Comprehension: From Research to Practice*, Hillsdale: Lawrence Erlbaum, 135–149.
- Farr, R, Pritchard, R and Smitten, B (1990) A description of what happens when an examinee takes a multiple-choice reading comprehension test, *Journal of Educational Measurement* 27 (3), 209–226.
- Feast, V (2002) The impact of IELTS scores on performance at university, International Education Journal 3 (4), 70–85.
- Feng, Q and Chen, L (2016) Study on teaching methods of reading comprehension strategies by comparison between TEM-4 reading comprehension and IELTS Academic Reading comprehension, *Journal of Language Teaching and Research* 7 (6), 1,174–1,180.
- Ferguson, G and White, E (1992) A Predictive Validity Study of IELTS a report commissioned by UCLES, Institute of Applied Linguistics, University of Edinburgh.
- Field, J (2004) Psycholinguistics: The Key Concepts, London: Routledge.
- Field, J (2013) Cognitive validity, in Geranpayeh, A and Taylor, L (Eds) Examining Listening: Research and Practice in Assessing Second Language Listening, Studies in Language Testing volume 35, Cambridge: UCLES/ Cambridge University Press, 77–151.
- Fiocco, M (1992) English proficiency levels of students from a non-English speaking background: A study of IELTS as an indicator of tertiary success, unpublished research report, Curtin University of Technology.
- Fortus, R, Coriat, R and Fund, S (1998) Prediction of item difficulty in the English subtest of Israel's inter-university psychometric entrance test, in Kunnan, A (Ed) Validation in Language Assessment: Selected Papers from the 17th Language Research Colloquium, Long Beach, Hillsdale: Lawrence Erlbaum, 61–87.
- Freedle, R (1997) The relevance of multiple-choice reading test data in studying expository passage comprehension: the saga of a 15-year effort towards an experimental/correlational merger, *Discourse Processes* 23, 399–440.
- Freedle, R and Kostin, I (1993) *The Prediction of TOEFL Reading Comprehension Item Difficulty for Expository Prose Passages for Three Item Types: Main Idea, Inference, and Supporting Idea Items,* TOEFL[®] Research Reports RR-93-13, Princeton: Educational Testing Service.
- Friend, R (2002) Summing it up teaching summary writing to enhance science learning, *The Science Teacher* 69 (4), 40–43.
- Fry, E (1963) *Teaching Faster Reading: A Manual*, Cambridge: Cambridge University Press.
- Fulcher, G (1999) Assessment in English for Academic Purposes: Putting content validity in its place, *Applied Linguistics* 20 (2), 221–236.
- Ganobcsik-Williams, L (2001) *Teaching writing at university: A survey of staff perspectives*, paper presented at Teaching Writing in Higher Education: An International Symposium, Warwick Writing Programme, the University of Warwick, March 2001.

- Ganobcsik-Williams, L (2004) A Report on the Teaching of Academic Writing in UK Higher Education, London: Royal Literary Fund.
- Gardner, S and Nesi, H (2013) A classification of genre families in university student writing, *Applied Linguistics* 34 (1), 25–52.
- Garland, K J and Noyes, J M (2004) CRT monitors: Do they interfere with learning? *Behaviour and Information Technology* 23 (1), 43–52.

Garnham, A (1985) Psycholinguistics: Central Topics, New York: Methuen.

Gebril, A and Plakans, L (2013) Toward a transparent construct of readingto-write tasks: The interface between discourse features and proficiency, *Language Assessment Quarterly* 10, 9–27.

Gebril, A and Plakans, L (2016) Source-based tasks in academic writing assessment: Lexical diversity, textual borrowing and proficiency, *Journal of English for Academic Purposes* 24, 78–88.

Geisler, C, Kaufer, D and Steinberg, E (1985) The unattended anaphoric "this": When should writers use it? *Written Communication* 2, 129–155.

Geranpayeh, A and Taylor, L (Eds) (2013) *Examining Listening: Research and Practice in Assessing Second Language Listening*, Studies in Language Testing volume 35, Cambridge: UCLES/Cambridge University Press.

- Gernsbacher, M A (1990) Language Comprehension as Structure Building, Hillsdale: Lawrence Erlbaum.
- Gervasi, V and Ambriola, V (2002) Quantitative assessment of textual complexity, in Merlini Barbesi, L (Ed) *Complexity in Language and Text*, Pisa: PLUS-University of Pisa, 197–228.
- Gibson, C and Rusek, W (1992) *The validity of an Overall Band Score of 6.0* on the IELTS Test as a Predictor of Adequate English Level Appropriate for Successful Academic Study, unpublished MA thesis, NCELTR.
- Gilster, P (1997) Digital Literacy, New York: Wiley Computer Publishing.
- Glendenning, E H and Holmström, B (1992) *Study Reading: A Course in Reading Skills for Academic Purposes*, Cambridge: Cambridge University Press.
- Goetz, E, Schallert, D, Reynolds, R and Radin, D (1983) Reading in perspective: What real cops and pretend burglars look for in a story, *Journal of Educational Psychology* 75, 500–510.
- Goh, S (1990) The effects of rhetorical organisation in expository prose on ESL readers in Singapore, *RELC Journal* 21 (2), 1–13.

Goldman, S R (1997) Learning from text: Reflections on the past and suggestions for the future, *Discourse Processes* 23, 357–398.

Goldman, S R (2003) Learning in complex domains: When and why do multiple representations help? *Learning and Instruction* 13 (2), 239–244.

Goldman, S R (2004) Cognitive aspects of constructing meaning through and across multiple texts, in Shuart-Ferris, N and Bloome, D M (Eds) Use of Intertextuality in Classroom and Educational Research, Greenwich: Information Age Press, 317–351.

- Goldman, S R and Rakestraw, J (2000) Structural aspects of constructing meaning from text, in Kamil, M, Rosenthal, P, Pearson, P and Barr, R (Eds) *Handbook of Reading Research III*, Hillsdale: Lawrence Erlbaum, 311–335.
- Goodman, K S (1976) Reading: A psycholinguistic guessing game, in Singer, H and Rudell, R B (Eds) *Theoretical Models and Processes of Reading*, Newark: International Reading Association, 497–508.

Gough, P B, Hoover, W A and Peterson, C L (1996) Some observations on a simple view of reading, in Cornoldi, C and Oakhill, J (Eds) *Reading Comprehension Difficulties: Processes and Intervention*, Hillsdale: Lawrence Erlbaum, 1–13.

- Grabe, W (1991) Current developments in second language reading research, *TESOL Quarterly* 25 (3), 375–406.
- Grabe, W (2004) Research on teaching reading, *Annual Review of Applied Linguistics* 24, 44–69.
- Grabe, W (2009) *Reading in a Second Language: Moving From Theory to Practice*, Cambridge: Cambridge University Press.
- Grabe, W and Kaplan, R B (1996) *Theory and Practice of Writing: An Applied Linguistics Perspective*, New York: Longman.
- Grabe, W and Stoller, F L (2002) *Teaching and Researching Reading*, Harlow: Pearson Education.
- Graesser, A C, Karnavat, A B, Daniel, F K, Cooper, E, Whitten, S N and Louwerse, M M (2001) A computer tool to improve questionnaire design, in *Statistical Policy Working Paper 33, Federal Committee on Statistical Methodology*, Washington, DC: Bureau of Labor Statistics, 36–48.
- Graesser, A C, McNamara, D S, Louwerse, M M and Cai, Z (2004) Coh-Metrix: Analysis of text on cohesion and language, *Behavioral Research Methods*, *Instruments, and Computers* 36, 193–202.
- Graesser, A C, McNamara, D S and Kulikowich, J M (2011) Coh-Metrix: Providing multilevel analyses of text characteristics, *Educational Researcher* 40 (5), 223–234.
- Graham, J G (1987) English language proficiency and the prediction of academic success, *TESOL Quarterly* 23 (1), 505–521.
- Green, A (2007) *IELTS Washback in Context: Preparation for Academic Writing in Higher Education*, Studies in Language Testing volume 25, Cambridge: UCLES/Cambridge University Press.
- Green, A (2012) Language Functions Revisited: Theoretical and Empirical Bases for Language Construct Definition Across the Ability Range, English Profile Studies volume 2, Cambridge: UCLES/Cambridge University Press.

Green, A and Hawkey, R (2011) Re-fitting for a different purpose: A case study of item writer practices in adapting source texts for a test of academic reading, *Language Testing* 29 (1), 109–129.

- Green, A and Hawkey, R (2012) An empirical investigation of the process of writing Academic Reading test items for the international English Language Testing System, in Taylor, L and Weir, C J (Eds) *IELTS Collected Papers 2: Research in Reading and Listening Assessment*, Studies in Language Testing volume 34, Cambridge: UCLES/Cambridge University Press 270–378.
- Green, A, Ünaldi, A and Weir, C J (2008) The cognitive processes of second language academic readers, LLAS Pedagogic Research Fund Project report.
- Green, A, Ünaldi, A and Weir, C J (2010) Empiricism versus connoisseurship: Establishing the appropriacy of texts for testing reading for academic purposes, *Language Testing* 27 (3), 191–211.
- Green, A, Weir, C J, Chan, S, Taylor, L, Field, J, Nakatsuhara, F and Bax S (2012) *Textual features of CAE reading texts: CAE texts compared with reading texts from FCE, CPE and IELTS and with essential undergraduate textbooks,* Research report for Cambridge English Language Assessment.
- Greene, S (1992) Mining texts in reading to write, *Journal of Advanced Composition* 12 (1), 151–170.
- Gu, X (2013) Verbal Protocol Analysis of Test-takers' Processes of Taking IELTS Academic Reading Test, paper presented at British Council New Directions English: Language assessment in China conference, Chongqing University, November 2013, available online: www.britishcouncil.cn/sites/default/files/ topic_2_-gu_xiangdong_0.pdf

- Guthrie, J (1988) Locating information in documents: Examination of a cognitive model, *Reading Research Quarterly* 23 (2), 178–199.
- Guthrie, J and Kirsch, I (1987) Distinctions between reading comprehension and locating information in text, *Journal of Educational Psychology* 79, 220–228.
- Hale, G, Taylor, C, Bridgeman, B, Carson, J, Kroll, B and Kantor, R (1996) *A study of writing tasks assigned in academic degree programs*, TOEFL[®] Research Reports RR-95-44, available online: onlinelibrary.wiley.com/doi/ epdf/10.1002/j.2333-8504.1995.tb01678.x
- Hamp-Lyons, L (1997) Washback, impact and validity, *Language Testing* 14 (3), 295–303.
- Hamp-Lyons, L (2000) Social, professional and individual responsibility in language testing, System 28, 579–591.
- Hartmann, D K (1995) Eight readers reading: The intertextual links of proficient readers reading multiple passages, *Reading Research Quarterly* 30 (3), 520– 561.
- Hawkey, R (1982) An investigation of inter-relationships between cognitive/affective and social factors and language learning: A longitudinal study of 27 overseas students using English in connection with their training in the United Kingdom, PhD thesis, University of London.
- Hawkey, R (2006) Impact Theory and Practice: Studies of the IELTS Test and the Progetto Lingue 2000, Studies in Language Testing volume 24, Cambridge: UCLES/Cambridge University Press.
- Hawkey, R and Weir, C J (2008) Language, study skill and related issues facing international students in the first year of their MPhil/PhD studies: Relevance to institutional language, academic and other support services, research report for Pedagogical Research Fund for Languages, Linguistics and Area Studies in Higher Education (2007–2008).
- Hayes, B (2003) *IELTS preparation in New Zealand: An investigation into the nature of the courses and evidence of washback*, unpublished PhD thesis, Victoria University of Wellington.
- Hayes, B and Read, J (2004) IELTS test preparation in New Zealand: Preparing students of the IELTS academic module, in Cheng, L, Watanabe, Y and Curtis, A (Eds) Washback in Language Testing: Research Contexts and Methods, Hillsdale: Lawrence Erlbaum, 97–112.
- Hazenberg, S and Hulstijn, J H (1996) Defining a minimal receptive second language vocabulary for non-native university students: An empirical investigation, *Applied Linguistics* 17 (2), 145–163.
- Heaton, J B (1975) Writing English Language Tests, Harlow: Longman.
- Hegarty, M and Just, M A (1989) Understanding machines from text and diagrams, in Mandl, H and Levine, J (Eds) *Knowledge Acquisition from Text and Picture*, Amsterdam: North-Holland, 171–194.
- Hegarty, M, Carpenter, P A and Just, M A (1991) Diagrams in the comprehension of scientific texts, in Barr, R, Kamil, M L, Mosenthal, P and Pearson, P D (Eds) *Handbook of Reading Research II*, New York: Longman, 641–668.
- Henning, G (1988) The influence of test and sample dimensionality on latent trait person ability and item difficulty calibrations, *Language Testing* 5, 83–99.
- Henning, G (1992) Dimensionality and construct validity of language tests, Language Testing 9 (1), 1–11.
- Hewings, A and Hewings, M (2001) Approaches to the study of disciplinary variation in academic writing: Implications for syllabus design, in Hall, D R and Hewings, A (Eds) *Innovation in English Language Teaching*, 71–83.

- Hidi, S and Anderson, V (1986) Producing written summaries: Task demands, cognitive operations, and implications for instruction, *Review of Educational Research* 56, 473–493.
- Hill, K, Storch, N and Lynch, B (1999) A comparison of IELTS and TOEFL as predictors of academic success, *English Language Testing System Research Reports 2, 52–63.*
- Hirvela, A (2004) *Connecting Reading and Writing*, Ann Arbor: The University of Michigan Press.
- Holland, M (forthcoming) *Testing Academic Literacy in Reading and Writing for Undergraduate University Admissions*, unpublished MA in Applied Linguistics, University of Bedfordshire.
- Holliday, W G, Brunner, L L and Donais, E L (1977) Differential cognitive and affective responses to flow diagrams in science, *Journal of Research in Science Teaching* 14 (2), 129–138.
- Holmes, J L and Ramos, R G (1993) Study summaries as an evaluation instrument: Questions of validity, *English for Specific Purposes* 12 (1), 83–94.
- Hoover, W and Tunmer, W (1993) The components of reading, in Thompson, G, Tunmer, W and Nicholson, T (Eds) *Reading Acquisition Processes*, Clevedon: Multilingual Matters, 1–19.
- Horak, T (1996) *IELTS impact study project*, unpublished MA assignment, Lancaster.
- Horowitz, D M (1986a) What professors actually require: Academic tasks for the ESL classroom, *TESOL Quarterly* 20, 445–460.
- Horowitz, D M (1986b) Essay examination prompts and the teaching of academic writing, *English for Specific Purposes* 5, 107–120.
- Howells, E (2011) *Literature: Reading to Write*, Boston: Pearson Education, Inc. Hughes, A (1989) *Testing for Language Teachers*, Cambridge: Cambridge
- University Press.
- Hughes, A, Porter, D and Weir, C J (1988) *ELTS Research Report 1(ii)*. *ELTS Validation Project: Proceedings of a conference held to consider the ELTS Validation Project Report*, London: The British Council/UCLES.
- Huhta, A, Luoma, S, Oscarson, M, Sajavaara, K, Takala, S and Teasdale, A (2002) DIALANG – A Diagnostic Language Assessment System for Learners, in Alderson, J C (Ed) Case Studies of the Use of the Common European Framework, Strasbourg: Council of Europe, 131–145.
- Humphreys, P, Haugh, M, Fenton-Smith, B, Lobo, A, Michael, R and Walkinshaw, I (2012) Tracking international students' English proficiency over the first semester of undergraduate study, *IELTS Research Reports* 12 (1), 1–41.
- Huong, T T T (2001) The predictive validity of the international English language testing system (IELTS) test, *Post-Script* 2 (1), 66–96.
- Hyland, K (2000) *Disciplinary Discourses: Social Interactions in Academic Writing*, Harlow: Longman.
- Hyland, K (2002) *Teaching and Researching Writing*, Applied Linguistics in Action Series, London: Longman.
- Hyland, K (2005) Metadiscourse, London: Continuum.
- Hyona, J and Nurminen, A M (2006) Do adult readers know how they read? Evidence from eye movement patterns and verbal reports, *British Journal of Psychology* 97 (1), 31–50.
- IELTS (2005) *IELTS Handbook*, London/Cambridge: British Council/IDP: IELTS Australia/UCLES.
- Ingram, D E and Bayliss, A (2007) IELTS as a predictor of academic language

performance, available online: aiec.idp.com/uploads/pdf/BaylissIngram%20 (Paper)%20Wed%201630%20MR 5.pdf

Intersegmental Committee of the Academic Senates of the California Community Colleges, the California State University, and the University of California (2002) Academic Literacy: A Statement of Competencies Expected of Students Entering California's Public Colleges and Universities, available online: senate. universityofcalifornia.edu/_files/reports/acadlit.pdf

Johns, A M (1993) Reading and writing tasks in English for academic purposes classes: Products, processes and resources, in Carson, J G and Leki, I (Eds) *Reading in the Composition Classroom*, New York: Newbury House, 274–289.

Johnson, M and Nádas, R (2009) Marginalised behaviour: Digital annotations, spatial encoding and the implications for reading comprehension, *Learning, Media and Technology* 34 (4), 323–336.

Jones, C, Turner, J and Street, B (Eds) (1999) *Students Writing in the University: Cultural and Epistemological Issues*, Amsterdam: John Benjamins.

Jones, N and Saville, N (2007) Scales and frameworks, in Spolsky, B and Hult, F M (Eds) *The Handbook of Educational Linguistics*, Malden: Blackwell, 495–509.

Jordan, R (1997) English for Academic Purposes: A Guide and Resource Book for Teachers, Cambridge: Cambridge University Press.

Just, M A and Carpenter, P A (1980) A theory of reading: From eye fixations to comprehension, *Psychological Review* 87 (4), 329–354.

Kaplan, A M and Haenlein, M (2016) Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster, *Business Horizons* 59 (4), 441–450.

Keddle, J S (2004) The CEF and the secondary school syllabus, in Morrow, K (Ed) Insights from the Common European Framework, Oxford: Oxford University Press, 43–54.

Kelly, R (1978) On the construct validation of comprehension tests: An exercise in applied linguistics, unpublished PhD thesis, University of Queensland.

Kerr, M A and Symons, S E (2006) Computerized presentation of text: Effects on children's reading of informational material, *Reading and Writing* 19 (1), 1–19.

Kerstjens, M and Nery, C (2000) Predictive validity in the IELTS test: A study of the relationship between IELTS scores and students' subsequent academic performance, *IELTS Research Reports* 3, 85–108.

Khalifa, H (1997) A study in the construct validation of the reading module of an EAP proficiency test battery: Validation from a variety of perspectives, unpublished PhD thesis, University of Reading.

Khalifa, H and Schmitt, N (2010) A mixed-method approach towards investigating lexical progression in Main Suite Reading test papers, *Research Notes* 41, 19–25.

Khalifa, H and Weir, C J (2009) *Examining Reading: Research and Practice in Assessing Second Language Reading*, Studies in Language Testing volume 29, Cambridge: UCLES/Cambridge University Press.

Kintsch, W (1998) Comprehension: A Paradigm for Cognition, New York: Cambridge University Press.

Kintsch, W (2004) The construction-integration model of text comprehension and its implications for instruction, in Ruddell, R and Unrau, N (Eds) *Theoretical Models and Processes of Reading*, Newark: International Reading Association, 1,270–1,328.

Kintsch, W and Keenan, J M (1973) Reading rate and retention as a function

of the number of propositions in the base structure of sentences, *Cognitive Psychology* 5, 257–279.

- Kintsch, W and van Dijk, T A (1978) Toward a model of text comprehension and production, *Psychological Review* 85 (5), 363–394.
- Kintsch, W and Yarbrough, J C (1982) Role of rhetorical structure in text comprehension, *Journal of Educational Psychology* 74 (6), 828–834.
- Klare, G R (1984) Readability, Handbook of Reading Research 1, 681-744.
- Knoch, U and Sitajalabhorn, W (2013) A closer look at integrated writing tasks: Towards a more focused definition for assessment purposes, *Assessing Writing* 18 (4), 300–308.
- Knoch, U, May, L, Macqueen, S, Pill, J and Storch, N (2016) Transitioning from university to the workplace: Stakeholder perceptions of academic and professional writing demands, IELTS Research Reports Online Series, available online: www.ielts.org/-/media/research-reports/ielts_online_rr_2016-1.ashx
- Kobayashi, M (1995) *Effects of Text Organisation and Test Format on Reading Comprehension Test Performance*, unpublished PhD thesis, Thames Valley University.
- Koda, K (2005) Insights into Second Language Reading: A Cross-Linguistic Approach, New York: Cambridge University Press.
- Kong, Y, Seo, Y S and Zhai, L (2018) Comparison of reading performance on screen and on paper: A meta-analysis, *Computers and Education* 123, 138–149.
- Koran, M L and Koran, J (1980) Interaction of learner characteristics with pictorial adjuncts in learning from science text, *Journal of Research in Science Teaching* 17, 477–483.
- Krishnan, K S (2011) Careful vs. expeditious reading: The case of the IELTS reading test, *Academic Research International* 1 (3), 25–35.
- Lacroix, N (1999) Macrostructure construction and organisation in the processing of multiple text passages, *Instructional Science* 27, 221–233.
- Langer, J A (1980) Relation between levels of prior knowledge and the organization of recall, in Kamil, M and Moe, A J (Eds) *Perspectives in Reading Research and Instruction*, Washington, DC: National Reading Conference, 28–33.
- Langer, J A (1986) Learning through writing: Study skills in the content areas, Journal of Reading 29, 400–406.
- Laufer, B (1988) What percentage of text-lexis is essential for comprehension?, in Lauren, C and Nordmann, M (Eds) *Special language: From Humans to Machines*, Clevedon: Multilingual Matters, 316–323.
- Lauterman, T (2003) Lexical coverage of spoken discourse, *Applied Linguistics* 24 (4), 425–438.
- Lea, M (1994) I thought I could write until I came here: Student writing in higher education, in Gibbs, G (Ed) *Improving Student Learning: Theory and Practice*, Oxford: Oxford Centre for Staff Development, 216–226.
- Lea, M and Stierer, B (Eds) (1999) *Student Writing in Higher Education: New Contexts*, Buckingham: Open University Press/Society for Research in Higher Education.
- Lea, M and Street, B V (1998) Student writing in higher education: An academic literacies approach, *Studies in Higher Education* 23 (2), 157–172.
- Lea, M and Street, B V (2000) Student writing and staff feedback in higher education: An academic literacies approach, in Lea, M and Stierer, B (Eds) *Student Writing in Higher Education: New Contexts*, Buckingham: Open University Press/Society for Research in Higher Education, 32–46.

- Levine, A, Ferenz, O and Reves, T (2000) EFL academic reading and modern technology: How can we turn our students into independent critical readers? *TESL EJ* 4 (4), 1–9.
- Lewis, L R, Vasishth, S and van Dyke, J (2006) Computational principles of working memory in sentence comprehension, *Trends in Cognitive Science* 10 (10), 447–454.
- Lillis, T (2003) Student writing as 'academic literacies': Drawing on Bakhtin to move from 'critique' to 'design', *Language and Education* 17 (3), 192–207.
- Lin, F-R and Kao, C M (2018) Mental effort detection using EEG data in E-learning contexts, *Computers & Education* 122, 63–79.
- Lin, J-C G and Yi, J K (1997) Asian international students' adjustment: Issues and program suggestions, *College Student Journal* 31 (4), 473–451.
- Livingston, J (1997) Metacognition: An Overview, London: Routledge/Kegan Paul.
- Lumley, T (1993) The notion of subskills in reading comprehension tests: An EAP example, *Language Testing* 10 (3), 211–234.
- Lunzer, E, Waite, M and Dolan, T (1979) Comprehension tests, in Lunzer, E and Gardner, K (Eds) *The Effective Use of Reading*, London: Heinemann, 37–71.
- Lynch, E (1988) *Reading for Academic Success: Selections from Across the Curriculum*, London: Macmillan.
- Maclellan, E (1997) Reading to learn, Studies in Higher Education 22, 277-288.
- Malvern, D and Richards, B (1997) A new measure of lexical diversity, in Ryan, A and Wray, A (Eds) Evolving Models of Language: Papers from the Annual Meeting of the British Association for Applied Linguistics held at the University of Wales, Swansea, September 1996, British Studies in Applied Linguistics volume 12, Clevedon: Multilingual Matters, 58–71.
- Mangen, A, Walgermo, B R and Brønnick, K (2013) Reading linear texts on paper versus computer screen: Effects on reading comprehension, *International Journal of Educational Research* 58, 61–68.
- Markey, K, Swanson, F, Jenkins, A, Jennings, B J, St Jean, B, Rosenberg, V, Yao, X and Frost, R L (2008) The effectiveness of a web-based board game for teaching undergraduate students information literacy concepts and skills, *D-Lib Magazine* 14 (9/10), available online: mirror.dlib.org/dlib/september08/ markey/09markey.html
- Markham, P (1985) The rationale deletion cloze and global comprehension in German, *Language Learning* 35 (3), 423–430.
- Masi, S (2002) The literature on complexity, in Merlini Barbesi, L (Ed) Complexity in Language and Text, Pisa: PLUS-University of Pisa, 117–145.
- Matsuo, K, Yamada, M, Bylykbashi, K, Cuka, M, Liu, Y and Barolli, L (2018) Implementation of an IoT-Based E-Learning Testbed: Performance Evaluation Using Mean-Shift Clustering Approach Considering Four Types of Brain Waves, paper presented at 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA), Kraków, 17–18 May 2018, available online: ieeexplore.ieee.org/stamp/stamp. jsp?tp=&arnumber=8418073
- McAuley, A, Stewart, B, Siemens, G and Cormier, D (2010) *The MOOC model for digital practice*, available online: davecormier.com/edblog/wp-content/ uploads/MOOC_Final.pdf
- McCormick, T (1988) *Theories of Reading in Dialogue: An Interdisciplinary Study*, New York: University Press of America.
- McGee, L M (1982) Awareness of text structure: Effects on children's recall of expository text, *Reading Research Quarterly* 17, 582–590.

- McGovern, D, Matthews, M and Mackay, S E (1994) *Reading*, New York: Prentice Hall International.
- McKoon, G (1977) Organisation of information in text memory, *Journal of Verbal Learning and Verbal Behavior* 16, 247–260.
- McNamara, D S, Louwerse, M M, Cai, Z and Graesser, A C (2005) *Coh-Metrix version 1.4,* available online: cohmetrix.memphis.edu/
- McNamara, D S, Ozuru, Y, Graesser, A C and Louwerse, M M (2006) Validating Coh-Metrix, in Sun, R and Miyake, N (Eds) *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, Austin: Cognitive Science Society, 573–578.
- McNamara, D S, Graesser, A C and Louwerse, M M (2013) Sources of text difficulty: Across the ages and genres, in Sabatini, J P and Albro, E (Eds) Assessing Reading in the 21st Century: Aligning and Applying Advances in the Reading and Measurement Sciences, Lanham: R&L Education, 89–116.
- Merrifield, G (2008) An impact study into the use of IELTS as an entry criterion for professional associations: Australia, NZ and USA, *IELTS Research Reports* 8, 1–43.
- Messick, S A (1989) Validity, in Linn, R L (Ed) *Educational Measurement* (3rd edition), New York: Macmillan, 13–103.
- Messick, S A (1995) Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning, *American Psychologist* 50 (9), 741–749.
- Messick, S (1996) Validity and washback in language testing, *Language Testing* 13, 241–256.
- Meyer, B J F (1975) *The Organisation of Prose and its Effect on Memory*, Amsterdam: North-Holland Publishing Company.
- Meyer, B J F and Freedle, R (1984) Effects of discourse type on recall, *American Educational Research Journal* 21 (1), 121–143.
- Mickan, P and Motteram, J (2008) An ethnographic study of classroom instruction in an IELTS preparation program, *IELTS Research Reports* 8, 1–26.
- Milanovic, M and Saville, N (1996) *Considering the impact of Cambridge EFL examinations*, Cambridge: UCLES internal report.
- Mislevy, R J (1992) *Linking Educational Assessments: Concepts, Issues, Methods, and Prospects*, Princeton: Educational Testing Service.
- Moore, T and Morton, J (1999) Authenticity in the IELTS Academic Module Writing test: A comparative study of Task 2 items and university assignments, *IELTS Research Report* 2, 74–116.
- Moore, T and Morton, J (2005) Dimensions of difference: A comparison of university writing and IELTS writing, *Journal of English for Academic Purposes* 4, 43–66.
- Moore, T, Morton, J and Price, S (2007) Construct validity in the IELTS academic reading test: A comparison of reading requirements in IELTS test items and in university study, *IELTS Research Reports* 11, 1–89.
- Moore, T, Morton, J, Hall D, and Wallis, C (2015) Literacy practices in the professional workplace: Implications for the IELTS reading and writing tests, *IELTS Research Reports Online Series* 2015 (1), 1–46.
- Morineau, T, Blanche, C, Tobin, L and Guéguen, N (2005) The emergence of the contextual role of the e-book in cognitive processes through an ecological and functional analysis, *International Journal of Human-Computer Studies* 62 (3), 329–348.
- Munby, J (1978) *Communicative Syllabus Design*, Cambridge: Cambridge University Press.
- Murray, N (2016) Standards of English in Higher Education: Issues, Challenges and Strategies, Cambridge: Cambridge University Press.
- Murray, J C, Cross, J L and Cruickshank, K (2014) Stakeholder perceptions of IELTS as a gateway to the professional workplace: The case of employers of overseas trained teachers, *IELTS Research Reports Online Series* 1, 1–78.
- Nakatsuhara, F (2014) A Research Report on the Development of the Test of English for Academic Purposes (TEAP) Speaking Test for Japanese University Entrants – Study 1 & Study 2, available online: www.eiken.or.jp/teap/group/ pdf/teap_speaking_report1.pdf
- Nation, P (2001) *Learning Vocabulary in Another Language*, Cambridge: Cambridge University Press.
- Nation, P and Gu, P Y (2007) Focus on Vocabulary, Sydney: NCELTR.
- Nesi, H and Gardner, S (2006) Variation in disciplinary culture: University tutors' views on assessed writing tasks, in Kiely, R, Clibbon, G, Rea-Dickins, P and Woodfield, H (Eds) Language, Culture and Identity in Applied Linguistics, British Studies in Applied Linguistics volume 21, London: Equinox Publishing, 99–117.
- Nesi, H, Gardner, S, Thompson, P, Wickens, P, Forsyth, R, Heuboeck, A, Holmes, J, Hindle, D, Ebeling, S, Leedham, M and Alsop, S (2008) An Investigation of Genres of Assessed Writing in British Higher Education: Full Research Report, ESRC End of Award Report, RES-000-23-0800, Swindon: Economic and Social Research Council.
- Nevo, N (1989) Test-taking strategies on a MC test of reading comprehension, Language Testing 6 (2), 199–217.
- Newell, G (1984) Learning from writing in two content areas: A case study/ protocol analysis, *Research in the Teaching of English* 18, 265–287.
- North, B (2000) *The Development of a Common Framework Scale of Language Proficiency*, New York: Peter Lang Publishing.
- Noyes, J M and Garland, K J (2003) VDT versus paper-based text: reply to Mayes, Sims and Koonce, *International Journal of Industrial Ergonomics* 31 (6), 411–423.
- Nuttall, C (1996) Teaching Reading Skills in a Foreign Language, London: Heinemann.
- Nystrand, M (1989) A social interactive model of writing, *Written Communication* 6, 66–85.
- O'Loughlin, K and Wigglesworth, G (1999) *Investigating task design in Academic Writing prompts*, Research Report to IELTS Steering Group, Round 5 IELTS Joint-Funded Research.
- O'Sullivan, B (2000) *Towards a model of performance in oral language testing*, unpublished PhD thesis, University of Reading.
- O'Sullivan, B (2002) Investigating variability in a test of second language writing ability, *Research Notes* 7, 14–17.
- O'Sullivan, B (2015) *Aptis Test Development Approach*, Aptis Technical Report TR/2015/001, London: British Council.
- O'Sullivan, B (2016) Adapting tests to the local context, in Saida, C, Hoshino, Y and Dunlea, J (Eds) *British Council New Directions in Language Assessment: JASELE Journal Special Edition*, Tokyo: British Council.
- Oakhill, J and Garnham, A (1988) *Becoming a Skilled Reader*, Oxford: Basil Blackwell.
- Ortega, L (2003) Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing, *Applied Linguistics* 24 (4), 492–518.

- Owen, N (2016) An evidence-centred approach to reverse engineering: Comparative analysis of IELTS and TOEFL iBT reading sections, unpublished PhD thesis, University of Leicester.
- Parry, S (1998) Disciplinary discourse in doctoral theses, *Higher Education* 36 (3), 273–299.
- Pearson (2018) Beyond Millennials: The Next Generation of Learners Beyond, *Global Research and Insights* (August), 1–21.
- Pearson (2019) *PTE Academic Score Guide Version 10*, available online: pearsonpte.com/wp-content/uploads/2017/08/Score-Guide.pdf
- Perera, K (1984) Children's Writing and Recording, Oxford: Basil Blackwell.
- Perfetti, C A (1985) Reading Ability, Oxford: Oxford University Press.
- Perfetti, C A (1997) Sentences, individual differences, and multiple texts: Three issues in text comprehension, *Discourse Processes* 23, 337–355.
- Perfetti, C A (1999) Comprehending written language: a blueprint of the reader, in Brown, C M and Hagoort, P (Eds) *The Neurocognition of Language*, New York: Oxford University Press, 167–208.
- Perfetti, C A, Rouet, J F and Britt, M A (1999) Toward a theory of document representation, in van Oostendorp, H and Goldman, S R (Eds) *The Construction of Mental Representations During Reading*, London: Lawrence Erlbaum Associates, 99–122.
- Phakiti, A (2003) A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance, *Language Testing* 20 (1), 26–56.
- Phakiti, A and Li, L (2011) General academic difficulties and reading and writing difficulties among Asian ESL postgraduate students in TESOL at an Australian university, *RELC Journal* 42 (3), 227–264.
- Pichert, J and Anderson, R (1977) Taking different perspectives on a story, Journal of Educational Psychology 69, 309–315.
- Plakans, L (2008) Comparing composing processes in writing-only and readingto-write test tasks, Assessing Writing 13, 111–129.
- Plakans, L (2009a) The role of reading strategies in integrated L2 writing tasks, Journal of English for Academic Purposes 8 (4), 1–15.
- Plakans, L (2009b) Discourse synthesis in integrated second language writing assessment, *Language Testing* 26, 561–585.
- Plakans, L and Gebril, A (2012) A close investigation into source use in integrated second language writing tasks, Assessing Writing 17 (1), 18–34.
- Plakans, L and Gebril, A (2013) Using multiple texts in an integrated writing assessment: Source text use as a predictor of score, *Journal of Second Language Writing* 22 (3), 217–230.
- Plakans, L and Gebril, A (2016) Exploring the relationship of organization and connection with scores in integrated writing assessment, *Assessing Writing* 31, 98–112.
- Pollitt, A (1988) Predictive validity, in Hughes, A, Porter, D and Weir, C J (Eds) (1988) ELTS Research Report 1(ii). ELTS Validation Project: Proceedings of a Conference Held to Consider the ELTS Validation Project Report, Cambridge: The British Council/UCLES.
- Pollitt, A and Taylor, L (2006) Cognitive psychology and reading assessment, in Sainsbury, M, Harrison, C and Watts, A (Eds) Assessing Reading: From Theories to Classrooms, Cambridge: National Foundation for Educational Assessment, 38–49.
- Pressley, M and Afflerbach, P (1995) Verbal Protocols of Reading: The Nature of Constructively Responsive Reading, Hillsdale: Lawrence Erlbaum.

- Pugh, A K (1978) Silent Reading: An Introduction to its Study and Teaching, London: Heinemann Educational Books.
- Purpura, J (1998) Investigating the effects of strategy use and second language test performance with high and low-ability test takers: a structural equation modeling approach, *Language Testing* 15, 339–379.
- Rasmusson, M (2015) Reading paper-reading screen: A comparison of reading literacy in two different modes, *Nordic Studies in Education* 34 (1), 3–19.
- Rastle, K (2007) Visual word recognition, in Gaskell, M G (Ed) The Oxford Handbook of Psycholinguistics, Oxford: Oxford University Press, 71–87.
- Rayner, K and Pollatsek, A (1989) *The Psychology of Reading*, Englewood Cliffs: Prentice Hall.
- Read, J (2000) Assessing Vocabulary, Cambridge: Cambridge University Press.
- Read, J (2015) Assessing English Proficiency for University Study, Basingstoke: Palgrave Macmillan.
- Read, J (2016) Post-admission Language Assessment of University Students, New York: Springer.
- Read, J and Wette, R (2009) Achieving English proficiency for professional registration: The experience of overseas qualified health professionals in the New Zealand context, *IELTS Research Reports* 10, 181–222.
- Rea-Dickins, P, Kiely, R and Yu, G (2007) Student identity, learning and progression: The affective and academic impact of IELTS on 'successful' candidates, *IELTS Research Reports* 7 (2), 1–78.
- Reid, I, Kirkpatrick, A and Mulligan, D (1998) Framing Reading, Perth: National Center for English Language Teaching and Research with the Center for Literacy, Culture and Language Pedagogy at Curtin University of Technology.
- Reynolds, M J (1981) Communicative Syllabus Design: The Topic and Task Approach, Cambridge: Cambridge University Press.
- Roach, J O (1944) The Cambridge Examinations in English, A Survey of their History, Cambridge Assessment Archives, PP JOR 1/1a xxi.
- Roelfsema, P R, Denys, D and Klink, P C (2018) Mind reading and writing: The future of neurotechnology, *Trends in Cognitive Sciences* 22 (7), 598–610.
- Rosenfeld, M, Leung, S and Oltman, P K (2001) *The Reading, Writing, Speaking,* and Listening Tasks Important for Academic Success at the Undergraduate and Graduate Levels, TOEFL[®] Monograph Series 21, Princeton: Educational Testing Service.
- Rosenfeld, M, Oltman, P K and Sheppard, K (2004) *Investigating the Validity* of *TOEFL: A Feasibility Study Using Content and Criterion-related Strategies*, TOEFL Research Report, RR- 71, Princeton: Educational Testing Service.
- Rosenshine, B (1980) Skills hierarchies in reading comprehension, in Spiro, R, Bruce, B and Brewer, W (Eds) *Theoretical Issues in Reading Comprehension*, New Jersey: Lawrence Erlbaum Associates, 535–554.
- Rupp, A A, Ferne, T and Choi, H (2006) How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective, *Language Testing* 23 (4), 441–474.
- Salimbene, S (1986) Interactive Reading, London: Newbury House Publishing.
- Sasaki, M (2000) Effects of cultural schemata on students' test-taking processes for cloze tests: A multiple data source approach, *Language Testing* 17, 85–114.
- Saville, N (2003) The process of test development and revision within UCLES EFL, in Weir, C J and Milanovic, M (Eds) *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002*, Studies in

Language Testing volume 15, Cambridge: UCLES/Cambridge University Press, 57–120.

- Saville, N (2009) *Developing a model for investigating the impact of language assessment within educational contexts by a public examination provider*, unpublished PhD thesis, University of Bedfordshire.
- Saville, N (2012) Applying a model for investigating the impact of language assessment within educational contexts: The Cambridge ESOL approach, *Research Notes* 50, 4–8.
- Saville, N (2018) Series Editors' note, in Chan, S, Defining Integrated Reading-into-Writing Constructs: Evidence at the B2–C1 Interface, English Profile Studies volume 8, Cambridge: UCLES/Cambridge University Press, ix–xii.
- Saville, N and Hawkey, R (2004) The IELTS impact study: Investigating washback on teaching materials, in Cheng, L, Watanabe, Y and Curtis, A (Eds) *Washback in Language Testing: Research Contexts and Methods*, Hillsdale: Lawrence Erlbaum, 73–96.
- Sawyer, W and Singh, M (2011) Learning to play the 'classroom tennis' well: IELTS and international students in teacher education, *IELTS Research Reports* 11, 1–54.
- Scardamalia, M and Bereiter, C (1987) Knowledge telling and knowledge transforming in written composition, in Rosenberg, S (Ed) Advances in Applied Psycholinguistics, Volume 2: Reading, Writing and Language Learning, Cambridge: Cambridge University Press, 142–175.
- Schedl, M, Gordon, A, Carey, P and Tang, K (1996) An Analysis of the Dimensionality of TOEFL Reading Comprehension Items, TOEFL Research Reports No. RR-95-27, Princeton: Educational Testing Service.
- Schwab, K (2017) *The Fourth Industrial Revolution*, New York: The Crown Publishing Group.
- Scott, M (2006) Oxford WordSmith Tools 4.0, available online: www.lexically.net/ downloads/version4/html/index.html
- Sedgwick, C, Garner, M and Vicente-Macia, I (2016) Investigating the language needs of international nurses: insiders' perspectives, *IELTS Research Report Series* 2, 1–38.
- Seidlhofer, B (1995) Approaches to Summarization: Discourse Analysis and Language Education, Tübingen: Gunter.
- Seldon, A (2018) The Fourth Education Revolution: Will Artificial Intelligence Liberate or Infantilise Humanity?, Buckingham: The University of Buckingham Press.
- Shaw, S D and Weir, C J (2007) Examining Writing: Research and Practice in Assessing Second Language Writing, Studies in Language Testing volume 26, Cambridge: UCLES/Cambridge University Press.
- Shayer, D (1972) *The Teaching of English in Schools, 1900–1970*, London: Routledge/Kegan Paul.
- Shih, M (1992) Beyond comprehension exercises in the ESL academic reading class, TESOL Quarterly 26 (2), 289–318.
- Shiotsu, T (2003) Linguistic knowledge and processing efficiency as predictors of L2 reading ability: A component skills analysis, unpublished PhD dissertation, University of Reading.
- Shiotsu, T and Weir, C J (2007) The relative significance of syntactic knowledge and vocabulary breadth in the prediction of second language reading comprehension test performance, *Language Testing* 23 (4), 99–128.

- Shohamy, E (2001) *The Power of Tests: A Critical Perspective on the Uses of Language Tests*, Harlow: Pearson Education.
- Simoes, J, Redondo, R P D and Vilas, A F (2013) A social gamification framework for a K-6 learning platform, *Computers in Human Behavior* 29 (2), 345–353.
- Singer, L M and Alexander, P A (2016) Reading across mediums: Effects of reading digital and print texts on comprehension and calibration, *Journal of Experimental Education* 85 (1), 1–18.
- Skehan, P (1998) Task-based instruction, *Annual Review of Applied Linguistics* 18, 268–286.
- Sommers, N (2000) Something more, something deeper: The experience of undergraduate writing, paper presented at the 51st Annual Conference on College Composition and Communication, Minneapolis, Minnesota, April 2000.
- Spivey, N N (1983) *Discourse synthesis: Constructing texts in reading and writing*, unpublished doctoral dissertation, University of Texas at Austin.
- Spivey, N N (1984) *Discourse Synthesis: Constructing Texts in Reading and Writing*, Outstanding Dissertation Monograph, Newark: International Reading Association.
- Spivey, N N (1987) Constructing constructivism: Reading research in the United States, *Poetics* 16, 169–192.
- Spivey, N N (1988) Comprehending and composing: The synthesis of comparative texts, unpublished manuscript, Carnegie Mellon University.
- Spivey, N N (1990) Construing constructivism: Reading research in the United States, *Written Communication* 7 (2), 256–287.
- Spivey, N N (1992) Discourse synthesis: creating texts from texts, in Hayes, J R, Young, R E, Matchett, M, McCaffrey, M, Cochran, C A and Hajduk, T (Eds) *Reading Empirical Research Studies: The Rhetoric of Research*, Hillsdale: Lawrence Erlbaum, 469–512.
- Spivey, N N (1997) *The Constructivist Metaphor: Reading, Writing and the Making of Meaning*, San Diego: Academic Press.
- Spivey, N N (2001) Discourse synthesis: Process and product, in McInnis, R (Ed) Discourse Synthesis: Studies in Historical and Contemporary Social Epistemology, Westport: Praeger, 379–396.
- Spivey, N N and King, J R (1989) Readers as writers composing from sources, *Reading Research Quarterly* 24 (1), 7–26.
- Spolsky, B (1995) Measured Words, Oxford: Oxford University Press.
- Spolsky, B (2004) Review: Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002, *ELTJ* 58 (3), 305–309.
- Stahl, S A, Hynd, C R, Britton, B K, McNish, M M and Bosquet, D (1996) What happens when students read multiple source documents in history?, *Reading Research Quarterly* 31 (4), 430–456.
- Stanovich, K (2000) Progress in Understanding Reading: Scientific Foundations and New Frontiers, New York: The Guildford Press.
- Steffensen, M, Joag-Dev, C and Anderson, R (1979) A cross-cultural perspective on reading comprehension, *Reading Research Quarterly* 15, 10–29.
- Sticht, T G and James, J H (1984) Listening and reading, in Pearson, P D, Barr, R, Kamil, M and Mosenthal, P (Eds) *Handbook of Reading Research*, New York: Longman, 293–317.
- Stoynoff, S (2009) Recent developments in language assessment and the case of four large-scale tests of ESOL ability, *Language Teaching* 42, 1–40.
- Stromso, H I and Braten, I (2002) Norwegian law students' use of multiple

sources while reading expository texts, *Reading Research Quarterly* 37 (2), 208–227.

- Subramanian, R, Wache, J, Khomami Abadi, M, Vieriu, R L, Winkler, S and Sebe, N (2018) ASCERTAIN: Emotion and personality recognition using commercial sensors, *IEEE Transactions on Affective Computing* 9 (2), 147–160.
- Tan, S (1990) The role of prior knowledge and language proficiency as predictors of reading comprehension among undergraduates, in de Jong, J and Stevenson, D (Eds) *Individualising the Assessment of Language Abilities*, Clevedon: Multilingual Matters, 214–224.
- Tanner, M J (2014) Digital vs. print: Reading comprehension and the future of the book, SLIS Student Research Journal 4 (2), available online: scholarworks. sjsu.edu/slissrj/vol4/iss2/6
- Taylor, L (1999) Constituency matters: responsibilities and relationships in our testing community, paper presented at the Language Testing Forum, University of Edinburgh, 19–21 November 1999.
- Taylor, L (2000) Stakeholders in language testing, Research Notes 2, 2-4.
- Taylor, L (2004) IELTS, Cambridge ESOL examinations and the Common European Framework, *Research Notes* 18, 2–3.
- Taylor, L (2007) The impact of the joint-funded research studies on the IELTS writing test, in Taylor, L and Falvey, P (Eds) *IELTS Collected Papers: Research in Speaking and Writing Assessment*, Studies in Language Testing volume 19, Cambridge: UCLES/Cambridge University Press, 479–492.
- Taylor, L (2012) The impact of the funded research studies on the IELTS Academic Reading test, in Taylor, L and Weir, C J (2012) *IELTS Collected Papers 2: Research in Reading and Listening Assessment*, Studies in Language Testing volume 34, Cambridge: UCLES/Cambridge University Press, 379–387.
- Taylor, L (2013) Testing Reading Through Summary: Investigating Summary Completion Tasks for Assessing Reading Comprehension Ability, Studies in Language Testing volume 39, Cambridge: UCLES/Cambridge University Press.
- Taylor, L (2014) A report on the review of test specifications for the reading and listening papers of the Test of English for Academic Purposes (TEAP) for Japanese university entrants, available online: www.eiken.or.jp/teap/group/pdf/ teap_rlspecreview_report.pdf
- Taylor, L and Chan, S (2015) *Reviewing the suitability of English language tests* for providing the GMC with evidence of doctors' English proficiency, available online: www.gmc-uk.org/about/research/28134.asp
- Taylor, L and Weir, C J (Eds) (2012) IELTS Collected Papers 2: Research in Reading and Listening Assessment, Studies in Language Testing volume 34, Cambridge: UCLES/Cambridge University Press.
- Taylor, S E (1965) Eye movements in reading: Facts and fallacies, *American Educational Research Journal* 2, 187–202.
- Teasdale, A (1989) Introspection and judgmental approaches to content validation: A study using the Test in English for Educational Purposes, unpublished MAAL dissertation, Department of Linguistics, University of Reading.
- Tercanlioglu, L (2004) Postgraduate students' use of strategies in L1 and ESL contexts: links to success, *International Education Journal* 5 (4), 562–570.
- Thomas, D (2005) *Type-token ratios in One Teacher's Classroom Talk: An Investigation of Lexical Complexity*, available online: www.birmingham.ac.uk/ Documents/college-artslaw/cels/essays/languageteaching/DaxThomas2005a. pdf

- Thorndike, E (1917a) Reading as reasoning: a study of mistakes in paragraph reading, *Journal of Educational Psychology* 8, 323–332.
- Thorndike, E (1917b) The understanding of sentences: A study of errors in reading, *The Elementary School Journal* 18 (2), 98–114.
- Tomlinson, B and Ellis, R (1988) *Reading Advanced*, Oxford: Oxford University Press.
- Trueswell, J C, Tanenhaus, M K and Kello, C (1993) Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths, *Journal of Experimental Psychology: Learning, Memory and Cognition* 19 (3), 528–553.
- UCLES (2011) Methodology used for benchmarking Cambridge English: Advanced to IELTS, available online: www.cambridgeenglish.org/images/28892-caecomparison-methodology.pdf
- UCLES (2013) Comparing Scores on Cambridge English: Advanced and IELTS, Cambridge: UCLES.
- Ünaldi, A (2010) Investigating reading for academic purposes: Sentence, text and multiple texts, PhD thesis, University of Bedfordshire.
- Urquhart, A (1984) The effect of rhetorical ordering on readability, in Alderson, J C and Urquhart, A (Eds) *Reading in a Foreign Language*, London: Longman, 160–181.
- Urquhart, A and Weir, C J (1998) *Reading in a Second Language: Process, Product and Practice*, Essex: Pearson Education Ltd.
- US Department of Labor, Employment Training and Administration (1999) *Testing and Assessment: An Employer's Guide to Good Practices*, available online: wdr.doleta.gov/opr/fulltext/document.cfm?docn=6032
- Van Dijk, T A and Kintsch, W (1977) Cognitive psychology and discourse: Recalling and summarizing stories, in Dressler, W U (Ed) Current Trends in Text Linguistics, New York: de Gruyter, 61–80.
- Van Dijk, T A and Kintsch, W (Eds) (1983) *Strategies of Discourse Comprehension*, New York: Academic Press.
- Van Ek, J and Trim, J L (1998) *Waystage 1990*, Cambridge: Cambridge University Press.
- Van Ek, J and Trim, J L (2001) Vantage, Cambridge: Cambridge University Press.
- Venezky, R L (1984) The history of reading research, in Pearson, P D (Ed) Handbook of Reading Research, New York: Longman, 3–38.
- Wall, D (1997) Impact and washback in language testing, in Clapham, C and Corson, D (Eds) *Encyclopædia of Language and Education Volume 7: Language Testing and Assessment*, Dordrecht: Kluwer Academic, 291–302.
- Wall, D (2005) The Impact of a High-stakes Examination on Classroom Teaching: A Case Study Using Insights from Testing and Innovation Theory, Studies in Language Testing volume 22, Cambridge: UCLES/Cambridge University Press.
- Wall, D and Alderson, J C (1993) Examining washback: the Sri Lankan impact study, *Language Testing* 10 (1), 41–69.
- Wallace, C (1997) IELTS: Global implications of curriculum and materials design, ELT Journal 51 (4), 370–373.
- Wallace, M (1980) English for Academic Study Series: Study Skills for Academic Writing, London: Prentice Hall.
- Wästlund, E (2007) *Experimental studies of human-computer interaction: Working memory and mental workload in complex cognition*, unpublished PhD, University of Gothenburg.
- Wästlund, E, Reinikka, H, Norlander, T and Archer, T (2005) Effects of VDT and paper presentation on consumption and production of information:

Psychological and physiological factors, *Computers in Human Behavior* 21, 377–394.

- Watanabe, Y (2001) *Read-to-write tasks for the assessment of second language academic writing skills: Investigating text features and rater reactions*, unpublished doctoral dissertation, University of Hawaii.
- Weakley, S (1993) *Procedures in the content validation of an EAP proficiency test of reading comprehension*, unpublished MATEFL Dissertation, CALS, University of Reading.
- Weigle, S C (1999) Investigating rater/prompt interactions in writing assessment: Quantitative and qualitative approaches, *Assessing Writing* 6 (2), 145–178.
- Weigle, S C (2002) Assessing Writing, Cambridge: Cambridge University Press.
- Weigle, S C (2004) Integrating reading and writing in a competency test for nonnative speakers of English, *Assessing Writing* 9, 27–55.
- Weigle, S C, Yang, W and Montee, M (2013) Exploring reading processes in an academic reading test using short-answer questions, *Language Assessment Quarterly* 10 (1), 28–48.
- Weir, C J (1983) Identifying the language needs of overseas students in tertiary education in the United Kingdom, unpublished PhD thesis, Institute of Education, University of London.
- Weir, C J (1990) Communicative Language Testing, London: Prentice Hall.
- Weir, C J (1993) Understanding and Developing Language Tests, New Jersey: Prentice Hall.
- Weir, C J (2003) A survey of the history of the Certificate of Proficiency in English (CPE) in the twentieth century, in Weir, C J and Milanovic, M (Eds) Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002, Studies in Language Testing volume 15, Cambridge: UCLES/Cambridge University Press, 1–56.
- Weir, C J (2005a) Limitations of the Common European Framework for developing comparable examinations and tests, *Language Testing* 22 (3), 281–300.
- Weir, C J (2005b) *Language Testing and Validation: An Evidence-based Approach*, Basingstoke: Palgrave Macmillan.
- Weir, C J (2013a) An overview of the influences on English language testing, in Weir, C J, Vidaković, I and Galaczi, E D, Measured Constructs: A History of Cambridge English Language Examinations 1913–2012, Studies in Language Testing volume 37, Cambridge: UCLES/Cambridge University Press, 1–102.
- Weir, C J (2013b) The measurement of reading ability, in Weir, C J, Vidaković, I and Galaczi, E D (2013) *Measured Constructs: A History of Cambridge English Language Examinations 1913–2012*, Studies in Language Testing 37, Cambridge: UCLES/Cambridge University Press, 103–179.
- Weir, C J (2013c) *The construct of academic literacy*, preliminary discussion paper commissioned by Cambridge English Language Assessment.
- Weir, C J (2014) A research report on the development of the Test of English for Academic Purposes (TEAP) writing test for Japanese university entrants, available online: www.eiken.or.ip/teap/group/pdf/teap writing report.pdf
- Weir, C J and Milanovic, M (Eds) (2003) Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002, Studies in Language Testing volume 15, Cambridge: UCLES/Cambridge University Press.
- Weir, C J and O'Sullivan, B (2017) Assessing English on the Global Stage. The British Council and English Language Testing, 1941–2016, Sheffield: Equinox Publishing Ltd.
- Weir, C J and Porter, D (1994) The multi-divisibility or unitary nature of reading:

The language tester between Scylla and Charybdis, *Reading in a Foreign Language* 10 (2), 1–19.

- Weir, C J, Yang, H and Jin, Y (2000) An Empirical Investigation of the Componentiality of L2 Reading in English for Academic Purposes, Studies in Language Testing volume 12, Cambridge: UCLES/Cambridge University Press.
- Weir, C J, Hawkey, R, Green, A, Ünaldi, A and Devi, S (2012a) The relationship between the academic reading construct as measured by IELTS and the reading experiences of students in their first year of study at a British university, in Taylor, L and Weir, C J (Eds) *IELTS Collected Papers 2: Research in Reading and Listening Assessment*, Studies in Language Testing volume 34, Cambridge: UCLES/Cambridge University Press, 37–119.
- Weir, C J, Hawkey, R, Green, A and Devi, S (2012b) The cognitive processes underlying the academic construct as measured by IELTS, in Taylor, L and Weir, C J (Eds) *IELTS Collected Papers 2: Research in Reading and Listening Assessment*, Studies in Language Testing volume 34, Cambridge: UCLES/ Cambridge University Press, 212–269.
- Weir, C J, Hawkey, R, Bax, S, Chan, S, Field, J, Green, A and Taylor, L (2012c) Report to Cambridge ESOL UK on the contextual parameters of CAE examinations, unpublished project report.
- Weir, C J, Chan, S and Nakatsuhara, F (2013) Examining the criterion-related validity of the GEPT advanced reading and writing tests: Comparing GEPT with IELTS and real-life academic performance, *LTTC-GEPT Research Report* 1, 1–43.
- Weir, C J, Vidaković, I and Galaczi, E D (2013) Measured Constructs: A History of Cambridge English Language Examinations 1913–2012, Studies in Language Testing volume 37, Cambridge: UCLES/Cambridge University Press.
- Widdowson, H G (1980) Applied linguistics: The pursuit of relevance, in Kaplan, R B (Ed) On the Scope of Applied Linguistics, Rowley: Newbury House, 74–87.
- Wilkins, D A (1973) *Grammatical, Situational and Notional Syllabuses*, London: British Council.
- Wilkins, D A (1976) Notional Syllabus, Oxford: Oxford University Press.
- Winetroube, S (1997) *The Design of the Teachers' Attitude Questionnaire*, Cambridge: UCLES.
- Wisker, G (2007) The Postgraduate Research Handbook: Succeed With Your MA, MPhil, EdD and PhD, Basingstoke: Palgrave Macmillan.
- Wolf, M (2007) Proust and the Squid, New York: Harper Collins.
- Wolfe-Quintero, K, Inagaki, S and Kim, H-Y (1998) Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity, Honolulu: University of Hawai'i Press.
- Wolfersberger, M (2013) Refining the construct of classroom-based writing-fromreadings assessment: The role of task representation, *Language Assessment Quarterly* 10 (1), 49–72.
- Woodrow, L (2006) Academic success of international postgraduate education students and the role of English proficiency, *University of Sydney Papers in TESOL* 1 (1), 51–70.
- Wright, B D and Stone, M H (1979) Best Test Design, Chicago: MESA Press.
- Wu, R Y F (2012) Establishing the validity of the General English Proficiency Test Reading component through a critical evaluation of alignment with the Common European Framework of Reference, unpublished PhD thesis, University of Bedfordshire.
- Wu, R Y F (2014) Validating Second Language Reading Examinations:

Establishing the validity of the GEPT through alignment with the Common European Framework of Reference, Studies in Language Testing volume 41, Cambridge: UCLES/Cambridge University Press.

- Wu, J R W and Lin, A C W (2008) Assessing English Proficiency at Advanced Level: Testers' Feedback to Teaching, paper delivered at 25th International Conference of English Teaching and Learning, The Language Teaching and Training Centre, Taipei, 2008.
- Yen, D and Kuzma, J (2009) Higher IELTS score, higher academic performance? The validity of IELTS in predicting the academic performance of Chinese students, *Worcester Journal of Learning and Teaching* 3, 1–7.
- Yu, G (2005) Towards a model of using summarization tasks as a measure of reading comprehension, unpublished doctoral thesis, University of Bristol.
- Yu, G (2008) Reading to summarize in English and Chinese: A tale of two languages?, *Language Testing* 25 (4), 521–551.
- Yu, G (2013) The use of summarization tasks: Some lexical and conceptual analyses, *Language Assessment Quarterly* 10 (1), 96–109.
- Yue, W (1997) An investigation of textbook materials designed to prepare students for the IELTS test: a study of washback, unpublished MA dissertation, Lancaster University.
- Zou, S, Green, R and Weir, C J (1997) *The Test for English (TEM) Validation Study*, Shanghai: Shanghai Foreign Language Education Press.

Author index

A

Ackerman, R 244 Afflerbach, P 75 Aglinskas, C74 Alderson, J C 11, 12, 23, 33, 60, 69, 77, 78, 70, 81, 82, 96, 97, 99, 101, 102, 127–128, 134, 138, 139, 140, 141, 142, 146, 149, 150, 166, 180, 183, 193, 213 Alexander, P A 236, 244 Al-Fallay, I 102 Ali, A 249 Allan, D 188 Alsop, S 7, 8, 150 ALTE 207-208, 211, 212 American Educational Research Association 208 American Psychological Association 208 Ambriola, V 77, 155 Anderson, NJ 239 Anderson, R 61, 102, 236 Anderson, V 223, 237 Applebee, A N 227 Arber, R 172 Archer, T 88 Arnaudet, M L 38 Ascención-Delaney, Y 223 Augusto, J C 249

B

Baba, K 240 Bachman, L F 59, 60, 61, 76, 77, 82, 130, 150, 180–182, 194, 227 Banerjee, J V 183, 213–214, Barker, K 99 Barnett, M 88 Barolli, L 248 Barrett, M E 38 Bartlett, F C 227 Baumann, J 61 Bax, S 110, 117, 119, 159, 195, 201, 248 Bayliss, A 214 Beard, R 109, 224 Bellingham, L 216 Bensoussan, M 138 Bereiter, C 75 Berman, R 96 Bernhardt, E 59, 60, 66 Bhatia, V 85 Blanche, C 87 Bleasdale, FA 100 Block, E 239 Bond, T G 208 Bonkowski, F 183 Bosquet, D 74 Brantmeier, C 236 Braten, I 74, 75 Brewer, WF 237 Bridgeman, B 61, 84, 214 British National Corpus 158 Britt, MA74 Britton, BK 74 Brønnick, K 88, 244 Brown, J 223 Brunfaut, T 62 Bruning, R H 100 Brunner, L L 80 Buck, G 60, 102 Buckley, P 249 Bylykbashi, K 248

C

Cai, Z 99, 152 California State University 14–16, 223 Cambridge Assessment English 11 Cambridge English Language Assessment 207 Campion, M 94 Carey, P 58, 59 Carey, R 65 Carpenter, P A 83, 93, 96, 222, 236 Carrell, P L 39, 43, 44, 47, 57, 61, 88, 237, 239 Carson, J 84

Cartelli, A 243 Carver, R P 58, 59, 82, 227 Chalmers, J 108, 110, 115, 116, 126 Chan, S 49, 108, 110, 159, 172, 174, 194, 195, 201, 214, 215, 223, 228, 232, 233, 234, 235, 248 Charge, N 12, 131, 146-147, 166 Chen, R S 227 Cheng, L 182, 191 Cheung, KYF 215 Chihara, T102 Chikalanga, I 61, 64 Cho, Y 214 Choi, H134, 135 Choi, I-C 77, 130, 181, 194 Clapham, C M xi, xii, 11, 12, 23, 44, 78, 97, 101, 139, 140, 165, 166, 193, 213 Cleland, J 215, 216 Cline, F 61 Clymer, T 224 Coady, J 59, 66 Cobb. T 94, 152 Cohen, A D 65, 90, 134, 142, 227, 228 Coleman, D 186 Colman, J 137, 140, 187, 188 Conrow, F 213, 214, 216 Cooper, E 98, 99 Coriat, R 77, 78 Corkill, A J 100 Cormier, D 243 Cotton, F 213, 214, 216 Council of Europe xii, 5, 59, 68, 79, 97, 211, 212, 224, 225, 226, 246 Coxhead, A 94, 160 Craven, E186 Criper, C 213, 216 Cromer, W72 Cross, J L 168, 169, 172 Crossley, J 93, 96, 152, 156 Cruickshank, K 168, 169, 172 Cuka, M 248 Cumming, A 227, 228, 240 Curtis, A 182, 191

D

Dafoulas, G 249 Daller, M H 214 Daniel, F K 98, 99 Davidson, F 12, 77, 130, 166, 181, 194 Davies, A x, 4, 11, 12, 29, 32, 40, 107, 165, 166, 208, 213, 216, 224 Dechant, E 59 Decos, L 249 Denys, D 248 Devi, S 22, 48, 49, 52, 53, 56, 65-66, 70, 80, 81, 82, 84, 85, 93, 94, 95, 98, 99, 100, 108, 110, 114, 115, 116, 119, 121, 152, 155, 156, 160, 162–164, 172–175, 190, 201 Devine, J 61 Di Pietro, S 214 Dolan, T 58, 59 Donais, EL 80 Dooley, P 216 Douglas, D 102 Dowell, J 215, 216 Doyle, E 249 Dunlea, J ix Durkin, K 46 Durst, R 239, 240 E Ebeling, S 7, 8, 150 Ehrlich, M F 99 Elder, C 208, 216 Elley, W94 Ellis, R 38 Emery, J 215 Engineer, WD 81 Enright, M K 61, 70, 73, 77, 89, 90, 101, 102, 134, 227 Erdosy, U 240

Eskey, D 61 Esmaeili, H 228

F

Eouanzoui, K 61

Farr, R 65, 135

Ferenz, O 66 Ferguson, G 214, 216

234

Feast, V 214, 216

Ferne, T 134, 135

Fiocco, M 216

Fishco, V 223 Forsyth, R 7, 8, 150

Fox, C M 208

Friend, R 222

89

Fortus, R 77, 78

Everett, R 137, 140, 187, 188

Field, J 71, 72, 73, 135, 159, 201,

Freedle, R 38, 40, 43, 44, 77, 78, 88,

Figueras, N 77, 79, 82, 97, 101

Frost, R L 249 Fry, E 82 Fulcher, G 165, 166 Fund, S 77, 78

G

Galaczi, E D 68 Ganobcsik-Williams, L9 Gardner, S 3, 7-9, 12, 13, 150 Garland, K J 87 Garner, M 166, 167 Garnham, A 60, 61, 71, 72, 76, 93 Gebril, A 228, 232, 239, 240 Geranpayeh, A vii Gernsbacher, MA 73, 98 Gervasi, V 77, 155 Gilster, P 251 Glendenning, EH 38 Glover, JA 100 Goetz, E 222, 236 Goh, S 88 Goldman, S R 74, 99 Goodman, KS46 Goldsmith, M 244 Gordon, A 58, 59 Gough, P B 227 Grabe, W 59, 61, 70, 72, 73, 77, 84, 89, 90, 101, 102, 132, 134, 146, 227, 229 Graesser, A C 93, 98, 99, 100, 152 Graham, S 236 Green, A 22, 48, 49, 52, 53, 56, 65, 66, 70, 80, 81, 82, 84, 85, 90, 91, 93, 94, 95, 98, 99, 100, 108, 110, 114, 115, 116, 118, 119, 121, 123, 133, 150, 151, 152, 154, 155, 156, 157, 159, 160, 162, 163, 164, 165, 172, 173, 174, 175, 177, 178, 179, 181, 182, 190, 191, 200.201 Green, R 68 Greene, S 229 Greenfield, J 93, 96, 152, 156 Gu, X 223 Guéguen, N 87 Guthrie, J 59, 127

Η

Hagan, A 186 Hall, D 166–172 Hamp-Lyons, L 180, 181 Hanna, G 223 Harris, K R 236 Hartmann, D K 74, 75 Hawkey, R 20, 22, 31, 48, 49, 52, 53, 56 65, 66, 70, 80, 81, 82, 84, 85, 93, 94, 95, 98, 99, 100, 108, 110, 114, 115, 116, 118, 119, 121, 123, 133, 152, 155, 156, 157, 159, 160, 162, 163, 164, 165, 172, 173, 174, 175, 177, 178, 179, 181, 182, 183, 184, 185, 190, 201 Hayes, B189 Heaton, J B 224 Hegarty, M 80 Henning, G 12, 59, 166 Heuboeck, A 7, 8, 150 Hewings, A 9 Hewings, M 9 Hidi, S 223, 237 Hill, K 208, 213, 216 Hindle, D 7, 8, 150 Hirvela, A 224, 229 Holland, M 166 Holliday, WG 80 Holmes, J 7, 8, 150 Holmes, J L 222 Holmström, B 38 Hoover, WA 66, 227 Horak, T183 Horowitz, D M 5, 125 Howells, E 222 Hughes, A 79, 102, 146 Huhta, A 92 Huong, TTT 216 Hyland, K 84, 85 Hynd, CR 74 Hyona, J 76

I

Inagaki, S 97 Ingram, D E 214 Inoue, C 235 Intersegmental Committee of the Academic Senates of the California Community Colleges (CASCCC) 14–16, 223

J

James, J H 75, 227 James, M 240 Jenkins, A 249 Jin, Y 22, 23, 34, 35, 36, 37, 38, 40, 45, 47, 52, 58, 59, 60, 61, 65, 70, 81, 83, 109, 110, 118, 222 Joag-Dev, C 102 Jones, N 209 Jordan, R 70 Just, M A 80, 93, 96, 222, 236

K

Kantor, R 240 Kao, C M 249 Kaplan, R B 84, 229 Karnavat, A B 98, 99 Keddle, JS97 Keenan, J M 82 Kello, C 99 Kelly, R 149 Kerr, MA 88 Kerstjens, M 216 Khalifa, H vii, viii-ix, 13, 44, 58, 60, 61, 62, 66, 70, 77, 78, 80, 92, 101, 109, 110, 120, 121, 125, 127, 131, 134, 141, 142, 146, 154, 155, 158, 159, 166, 181, 210, 211, 212 Khomami Abadi, M 249 Kielv, R 147, 222 Kim, H-Y 97 King, J R 74, 75, 229, 237 Kintsch, W 60, 61, 73, 74, 82, 138, 222, 223, 227, 236 Kirsch, I 59 Klare, G R 155 Klink, PC248 Knoch, U 228 Kobayashi, M 78, 138 Koda, K 60, 61, 70, 73, 77, 89, 90, 93, 101, 102, 134, 227, 236, 239 Kong, Y 244 Koran, J 80 Koran, ML 80 Kostin, I 60, 77, 78 Kuijper, H 77, 79, 82, 97, 101 Kulikowich, J M 100 Kunnan, A 60 Kuzma, J 216

L

Lacroix, N 74 Langer, J A 237 Lauterman, T 244 Lea, M 3, 7, 8, 10–13 Leedham, M 7, 8, 150 Leung, S 134 Levine, A 66 Lewis, L R 98 Li, L 46

2

Lillis, T 7, 10 Lin, A C W 68, 118 Lin, F-R 249 Lin, J-C G 46 Liu, Y 248 Louwerse, M M 98, 99, 100, 152 Lumley, T 60, 208 Lunzer, E 58, 59 Luoma, S 92 Lynch, B 60, 213, 219 Lynch, E 38

M

Mackay, SE 38 Maclellan, E 70, 222 Maia, CC 249 Malvern, D 82 Mangen, A 88, 244 Markey, K 249 Markham, P138 Masi, S 77, 155 Matthews, M 38 Matsuo, K 248 McAuley, A 243 McCormick, T 60 McCray, G 62 McElwee, S 215 McGee, L M 240 McGovern, D 38 McKoon, G74 McLachlan, J 215, 216 McNamara, DS 208 McNamara, T 93, 96, 99, 100, 152, 156 McNish, M M 74 Merrifield, G 168, 169, 172 Messick, S A vii–xi, 181, 182, 191, 250 Meyer, B J F 38, 40, 43, 44, 74, 88 Mickan, P186 Milanovic, M 181 Mislevy, R J 207 Montee, M 222, 228, 230, 231, 239 Moore, T 108, 110, 118, 122–127, 166-172 Morineau, T 87 Morton, J 108, 110, 118, 122-127, 166 - 172Mosenthal, P 70, 73, 77, 89, 90, 101, 102, 134, 227 Motteram, J 186

Mulcany-Ernt, P 70, 73, 77, 89, 90, 101, 102, 134, 227 Munby, J 30, 31, 66, 224 Murray, J C Murray, N 3, 12, 168, 169, 171, 172

Ν

Nakatsuhara, F 49, 201, 214, 215 Nation, P 92 National Council on Measurement in Education 208 Nery, C 216 Nesi, H 3, 7–9, 12, 13, 150 Newell, G 237 Nicholson, S 215, 216 Nold, G 77, 79, 82, 97, 101 Norlander, T 88 North, B 90 Noyes, J M 87 Nurminen, A M 76 Nuttall, C 80–83, 92, 101, 154 Nystrand, M 84

0

O'Loughlin, K 238 O'Sullivan, B ix, 11, 12, 107,124, 130, 165, 166, 170, 171, 192, 200, 249, 250 Oakhill, J 60, 61, 71, 72, 76 Oliver, R 216 Oller, J 102 Oltman, P K 65, 134 Ortega, L 97 Oscarson, M 92 Owen, N 62, 108, 110, 114, 116, 120–126, 133–140, 142–144 Ozuru, Y 100

Р

Palmer, A S xii, 76–77, 180, 181 Patterson, F 215, 216 Pearson 232, 243 Pearson, P 61 Perera, K 92 Perfetti, C A 61, 72, 74, 76, 227 Peterson, C L 227 Phakiti, A 46, 236, 239 Phelan, D 214 Pichert, J 236 Plakans, L 222, 228, 229, 230, 232, 239, 240 Pollatsek, A 66, 95, 96, 98 Pollitt, A 146, 215 Porter, D 58, 60 Pressley, M 272 Price, S 108, 110, 118, 119, 122–127 Pugh, A K 65, 68

R

Radin, D 222, 236 Rakestraw, J 99 Ramos, R G 222 Ramraz, R 138–139 Rasmusson, M 244 Rastle, K 94 Rayner, K 66, 95, 96, 98 Rea-Dickins, P 147, 222 Read, J 12, 44, 139, 166, 168, 169, 170, 172 Redondo, R P D 249 Reinikka, H 88 Reves, T 66 Reynolds, M J 236 Reynolds, R 222, 236 Richards, B82 Roelfsema, P R 248 Rosenberg, V 259 Rosenfeld, M 65, 134 Rosenshine, B 58, 59, 68 Rouet, J F 74 Rupp, A A 134, 135 Ryan, K 77, 130, 181–182, 194

S

Sajavaara, K 92 Sakurai, T102 Salimbene, S 38 Samuels-Clarke, J 249 Sasaki, M 102 Saville, N 181, 182, 183, 191, 192, 208, 209, 226, 247 Sawyer, W 168, 169 Scardamalia, M 75, 235 Schallert, D 222, 236 Schedl, M 58, 59, 70, 73, 77, 89, 90, 101, 102, 134, 227 Schmitt, N 92 Schwab, K 245 Scott, M 152, 156 Sebe, N 249 Sedgwick, C 166, 167 Seidlhofer. B 222 Seldon, A 245 Seo, Y S 244 Shaw, S D vii, 75, 92, 193

Shayer, D 65 Sheppard, K 65 Shih, M 80 Shiotsu, T 61, 72, 96, 97 Shohamy, E180 Siemens, G 243 Simoes, J 249 Singer, L M 244 Singh, M 168, 169 Sitajalabhorn, W 228 Skehan, P81 Sommer, J 74 Sommers, N 9 Spivey, N N 74, 75, 229, 230, 235, 237, 239, 240 Spolsky, B 46, 181 Stahl, SA74 Stanovich, K 60 Starfield, S186 Steffensen, M 102 Stewart, B 243 St Jean, B 249 Sticht, T G 75, 227 Stoller, F L 59, 132, 134, 146, 227 Stone, MH 208 Storch, N 213, 216 Street, B V 3, 7, 8, 10–13 Stromso, H I 74, 75 Subramanian, R 249 Swanson, F 249 Symons, S E 88

T

Takala, S 77, 79, 82, 92, 97, 101 Tan, S 101, 166 Tanner, M J 87, 245 Tanenhaus, M K 99 Tang, K 58, 59 Tardieu, C 77, 79, 82, 97, 101 Tatsuoka, K 60 Taylor, L vii, 12, 24, 108, 110, 118, 130, 131, 132, 146, 147, 159, 160, 162, 164, 166, 169, 170, 171, 172, 174, 177, 192, 194, 195, 201, 208, 212, 235, 237 Teasdale, A 60, 92 Thomas, D96 Thompson, P 7, 8, 150 Tobin, L 87 Tomlinson, B 38 Tone, B 65 Trim, JL 90

Trueswell, J C 99 Tunmer, W 66

U

UCLES 150–151 Ünaldi, A 22, 48, 49, 52, 53, 56, 65, 66, 70, 74, 80, 81, 82, 84, 85, 90, 93, 94, 95, 98, 99, 100, 108, 110, 114, 115, 116, 119, 121, 150, 151–160, 164, 172–175, 178, 190, 200, 201 University of California 14–16, 223 Upton, T 65, 90, 134, 142, 227, 228 Urquhart, A H 12, 36, 37, 59, 61, 65–71, 75, 88, 92, 101, 102, 132, 134, 166, 229 US Department of Labor, Employment Training and

V

van Dijk, T A 73, 74, 222, 223, 227 van Dyke, J 98 van Ek, J 90 Vanniarajan, S 60 Vasishth, S 98 Vellutino, F 227 Venezky, R L 69 Vicente-Macia, I 166, 167 Vidaković, I 68 Vieriu, R L 249 Vilas, A F 249

Administration 216

W

Wache, J 249 Waite, M 58, 59 Walgermo, B R 88, 244 Walkinshaw, I 108, 110, 115, 116, 126 Wall, D xi, xii, 78, 140, 180, 181, 182, 193, 213 Wallace, C 146, 147 Wallis, C 166-172 Wästlund, E 87, 88 Watanabe, Y 182, 191, 223, 240 Weakley, S 60 Weigle, S C 84, 89, 222, 227, 228, 230, 231, 237, 239 Weir, C J see References and further reading section for Weir, C J titles as this author has too many citations to list here

Wette, R 168, 169, 172 White, E 214, 216 Whitten, S N 98, 99 Wickens, P 7, 8, 150 Widdowson, H G 29, 224 Wigglesworth, G 238 Wilkins, D A 90–92 Winkler, S 249 Wisker, G19 Wolf, M 88 Wolfe-Quintero, K 97 Wolfersberger, M 228 Woodrow, L 214, 216 Wright, B D 208 Wu, J R W 68, 118 Wu, R Y F 24, 68, 228, 233

Y

Yamada, M 248 Yang, H 22, 23, 34, 35, 36, 37, 38, 40, 45, 47, 52, 58, 59, 60, 61, 65, 70, 81, 83, 109, 110, 118, 222 Yang, W 222, 228, 230, 231, 239 Yao, X 249 Yarbrough, J C 138 Yen, D 216 Yi, J K 46 Yu, G 147, 222, 223, 228, 237 Yue, W 183

Z

Zhai, L 244 Zou, S 68

Subject index

Approaches to academic reading

Generic study skills 1, 3–7, 9, 10, 11, 12, 13, 20, 171

- Genre based 1, 3, 7–9, 10, 11, 12, 13, 63, 78, 84–88, 100, 125, 150, 153, 165, 169, 171, 173, 174, 237, 248, 251, 252
- Socially situated 1, 3, 7, 10–11, 12, 13, 84, 103, 171, 180–182

English language tests

- Advanced English Reading Test (AERT) (1995) 23, 34–48, 65, 110
- Aptis (2012-) 171
- Certificate in Advanced English (CAE) (1991–) 68, 155, 159, 160, 194–207, 212
- Certificate in Proficiency in English (CPE) (1913–) 155, 158, 159, 194, 196–199
- English Language Testing Service (ELTS) (1980–1989) 11, 23, 36, 40, 101, 107, 124, 126, 146, 165, 166, 170, 200, 216
- English Proficiency Test Battery (EPTB) (the Davies Test) (1965– 1982) 11, 32, 36, 38, 40
- First Certificate in English (FCE) (1975–) 68, 159, 212
- General English Proficiency Test (GEPT) (2000–) 24, 68, 118, 146, 195–199, 231, 233–234

International Language Assessment (ILA) (2009–) 145, 146

- Internet-based Test of English as a Foreign Language (TOEFL iBT) (2005–) 90, 116, 121, 126, 194–199, 231, 232, 239
- Test in English for Academic Purposes (TEAP) (1979–1982) 4, 11, 20, 24, 65, 97, 166
- Test in English for Educational Purposes (TEEP) (1983–) 36, 40, 139, 146

- Test of English as a Foreign Language (TOEFL) (1964–) 36, 40, 49, 58, 59, 61, 70, 90, 121, 130–134, 147, 194
- Trinity College London (1977–) Integrated Skills in English (ISE) 231, 234–236

Key concepts in academic reading assessment

- Authenticity: interactional and situational xii–xiii, 43, 76–77, 126, 147, 148, 149–150, 170–171, 211, 221–241
- Background/content knowledge (subject and culture) 38, 39, 44, 54, 59, 65, 73, 78, 84, 93, 101–103, 164, 231
- Careful reading 4, 17, 26, 34–43, 45–48, 50–54, 56, 57, 59, 62–71, 79–83, 96, 109–118, 121, 135–140, 199, 211–212, 233–236
- Channel of presentation 38, 39, 7, 80
- Cognitive processing viii–xiii, 5, 22, 61–63, 70–83 108–110, 115, 120–127, 133–141, 145, 146, 170, 190, 195–199, 228, 233, 239, 244
- Cognitive parameters in reading 103, 110–132, 177, 190, 194, 212, 216
- Cohesion 29, 77, 99–100, 145, 153, 156, 163–167, 206
- Common European Framework of Reference for Languages (CEFR) 5, 59, 68, 79, 81, 82, 83, 90–92, 97, 101, 154, 155, 193, 198, 200–213, 224–226, 233, 246
- Consequences xi, 182, 191
- Contextual parameters in reading/ performance conditions viii, x, xiii, 1, 5, 21, 22–57, 76–103, 149–179, 195, 200, 212, 221
- Discourse mode 77, 78, 85, 96, 153, 172–177, 196, 197, 199

- English for Academic Purposes (EAP) vii, x, 3, 4, 11, 20, 23, 24, 34–49, 56, 60, 62, 70, 78, 82, 107, 109, 110, 120, 139, 150, 165, 166, 170, 179, 188, 192, 193, 217, 237
- English for Specific Purposes (ESP) 11, 12, 166
- Establishing propositional meaning 36, 37, 38, 40, 43, 63, 64, 67, 71, 72, 96, 101, 109, 111, 114, 116, 120, 122, 127, 136, 137, 138, 153, 169, 227
- Ethical considerations 219, 244–245, 252
- Expeditious reading 34, 35, 36, 39–57, 49, 62–71, 79–83, 89, 109–118, 127, 140, 154, 199, 211, 219, 221, 235, 236, 247, 251
- Fairness 38, 82, 184, 210, 251
- Functional knowledge and resources 32, 77, 78, 90–92
- Genre 7–9, 10, 11, 12, 13, 63, 78, 84–87, 100, 125, 150, 153, 165, 169, 171, 173, 174, 237, 248, 251, 252
- Goal setter 62, 63, 71, 75, 236
- Global reading 34–36, 39, 41, 42, 45–52, 56, 57, 59, 60, 62–64, 67, 69–73, 100, 109, 114–121, 124, 126, 127–133 138, 140, 142, 146, 166, 200, 211, 219, 221,228, 230, 233, 235, 236, 239, 240, 247
- Higher-order reading skills 34, 49, 57, 71–76, 126, 128, 131, 133, 134, 137, 142, 147, 185, 227, 235, 236
- Inferencing 37, 43, 47, 57, 59, 61, 63, 64, 71, 72, 73, 101, 111, 120–123, 131, 141
- Integrated skills 29, 146, 147, 222, 224–240, 246, 247
- Intertextual representation 34, 57, 62, 63, 70, 74–75, 120, 124–127, 132, 146, 222, 231, 233, 236
- Length of texts 31, 35, 38, 39, 40, 42–43, 45, 53, 54, 55, 66, 77, 78, 80–82, 90, 96, 98, 99, 109, 149, 150, 152, 154–156, 167, 172, 175, 178, 196–199, 238, 251
- Lexical access 59, 63, 72, 75, 120, 131, 136, 137, 138
- Lexical complexity 78, 82, 92-94
- Local reading 36, 37, 39, 42, 45–52, 57, 59, 60, 62–73, 109, 114–121, 124, 126, 127–133, 138, 140, 171, 236, 240

- Lower-order reading skills 34, 71–72, 126, 127–132, 142, 185, 236
- Mediation 5, 6, 63, 224, 226, 246, 247
- Mental model building 63, 64, 73, 76, 81, 84, 112, 116, 117, 118, 120, 121, 123, 131, 135, 136, 138, 142, 144
- Models of reading 5, 10, 11, 22–41, 58–77, 81, 120–121, 125–127, 229
- Monitoring 25-26, 51, 69, 75-76
- Nature of information, abstract/ concrete 77, 78, 89, 100–101, 154, 163–165, 212, 239, 240
- Patterns of exposition 30, 78, 84, 89–92, 153, 173–176, 187
- Reading difficulties 23, 24, 26–33, 43, 45, 48, 49, 53–60, 66, 69, 77, 78, 82, 83, 85, 88, 92–101, 109, 153–156, 161, 164, 184, 208–210
- Reading-into-writing tasks 6, 7, 14, 21, 62, 75, 126, 146, 147, 217–241
- Real-life approach xi, xii, xiii, 22, 47, 57, 66, 71, 80, 81, 105, 108–112, 118, 133, 146, 179, 190, 215, 217
- Response method xi, 78-79, 145
- Rhetorical organisation 38, 39, 40, 43, 45, 47, 57, 77, 89, 153, 163–165, 196, 197, 237
- Scanning 5, 36, 37, 38, 40, 41, 42, 44, 47, 57, 66, 67, 68, 71, 80, 82, 109, 111, 113, 115, 116, 128–130, 140, 144, 186, 188, 211
- Search reading 4, 27, 30, 36–43, 49, 66, 67, 71, 80, 82, 89, 109, 111, 115–117, 137, 140, 229, 232, 239, 251
- Skilled/unskilled academic readers 4, 66, 72, 73, 76, 83, 94, 238–240, 247
- Skimming 4, 5, 17, 30, 35–38, 42, 43, 58, 66, 67, 71, 80, 82, 89, 109, 111, 113, 115, 128–130, 140, 186, 228, 230, 239
- Sociocognitive model of reading 58, 62–103, 110
- Syntactic complexity xi, 78, 96–98, 195, 238
- Subject specificity 44, 77, 149, 150, 153, 161, 163–165, 172, 206
- Syntactic parsing 59, 61, 63, 72, 75, 120, 131, 137, 138, 167
- Test taker characteristics 190–191, 239, 248, 250–251

- Text-level representation/Discourse representation 49, 63, 73–74, 81, 118, 120, 121, 123–125, 131, 142, 144, 205, 206
- Text selection 32, 38, 150, 172, 177, 206
- Time constraints 22, 31, 33, 47, 53, 56,
- 57, 77, 78, 82–83, 117, 118, 150 Tonio familiarity 77, 101
- Topic familiarity 77, 101
- Word recognition 59, 63, 71–72, 75, 80 Writer-reader relationship 77, 78

Key events

- ELTS (1980–1989) 11, 36, 107, 124, 126, 146, 166, 170–171, 200
- ELTS Revision Project (1986–1989) 23, 107, 124

Key organisations

- British Council 107, 130, 145, 146, 170, 171, 190, 201
 Cambridge Assessment English (Cambridge English, formerly UCLES and Cambridge ESOL) 68, 91, 92, 107, 110, 130, 132, 135, 145, 146, 150, 154, 155, 158, 166, 171, 181, 182, 183, 191, 192, 194, 197, 200, 201, 206, 207–213
- Council of Europe 5, 59, 68, 79, 97, 211–212, 224–226, 246–247
- CRELLA, University of Bedfordshire 90–91, 194–195, 234–235
- Educational Testing Service (ETS) 130–131
- **IDP: IELTS Australia 107**
- Lancaster University 124, 128, 183

Test systems

Administration x, xi, 83, 150–151, 163, 210, 251 Calibrated items 208–210

- Comparability with external standards xii, 194, 210–213
- Comparability with a future criterion performance 194, 213–217
- Concurrent equivalence with different versions of the same test 207–210
- Criterion-related validity viii, xi, xiii, 21, 193–217, 250, 252
- Cross-test comparability/equivalence 193–206, 253
- Delivery 150, 166, 246, 247, 248
- Impact 180-192
- Internal consistency xiii, 131–132, 196–198
- Item banking 207-210, 253
- Item Response Theory (IRT) 207, 208
- Item writing 56, 57, 117, 119–120, 128–130, 133, 141, 151, 157, 165,
- 128–130, 135, 141, 151, 157, 16, 177–179, 208
- Marker reliability 79, 135, 141
- Online test taking 245–250
- Pretesting 140, 183, 208, 209, 210
- Rasch analysis 209–211
- Rater training 192, 252
- Scoring validity (including reliability) viii, ix, xi, 130, 252
- Security 12, 83, 150
- Stakeholders in the testing community x, 168, 172, 181, 191, 192, 211, 212, 214
- Test delivery 150, 166, 247, 248
- Test specification xii, 5, 14, 23, 34, 41, 90, 117, 127, 128, 132, 142, 151, 170, 177, 190, 208, 211, 212, 221
- Validation vii, viii–x, xii, xiv, 1, 45, 62, 68, 76, 94, 119, 169, 180, 182, 183, 191, 193, 207, 211, 219, 242, 249, 250
- Washback xi, xii, 21, 81, 118 179, 180–192