Testing Reading Through Summary

Investigating summary completion tasks for assessing reading comprehension ability
For a complete list of titles please visit: http://www.cambridge.org/elt/silt

Also in this series:

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

Barry O'Sullivan

Edited by Cyril J. Weir and Michael Milanovic

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

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

Changing Language Teaching through Language Testing: A washback study
Lijing Cheng

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

Alan Davies

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

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

Edited by Lynda Taylor and Cyril J. Weir

Examining FCE and CAE: Key issues and recurring themes in developing the First Certificate in English and Certificate in Advanced English exams
Roger Hawkey

Edited by Lynda Taylor and Cyril J. Weir

Components of I.2 Reading: Linguistic and processing factors in the reading test performances of Japanese EFL Learners
Toshihiko Shiotani

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
Hamad Kholif 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 Ardzeshir 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
Senior Lecturer, Centre for Research in English Language Learning and Assessment (CRELLA), University of Bedfordshire
and
Consultant, Cambridge English Language Assessment (part of the University of Cambridge)
For my parents – Frank and Kay,
who introduced me to the joy of reading
and taught me to see a world in the words on a page
Contents

Acknowledgements ix
Series Editors’ note xi
Abbreviations xvi

1 Introduction 1
2 The theory of reading comprehension 14
3 An overview of reading comprehension test design 40
4 Using summarising tasks to assess reading comprehension ability 56
5 Designing recall studies to explore readers’ mental representations of two texts 83
6 Using readers’ mental representations to construct summaries of two texts 105
7 Developing summary completion tasks for Texts A and B 173
8 Trialling of summary completion Tasks A and B 187
9 Conclusions and recommendations 206

Appendices
Appendix 1: Text A: Journey by Night (short story) 220
Appendix 2: Text B: The rights and wrongs of treating anorexia (newspaper editorial) 221
Appendix 3: Research protocol for the oral recall study of Texts A and B 222
Appendix 4: Probe questions for the oral recall study of Texts A and B 224
Appendix 5: Sample transcript from oral recall study 225
Appendix 6: Task instructions for written recall study of Texts A and B 227
Appendix 7: Summaries of Texts A and B derived from readers’ oral recalls 229
Appendix 8: Task instructions for trialling summary completion Task A 231
Appendix 9: Task instructions for trialling summary completion Task B 234
Appendix 10: Final answer key for summary completion Task A (Journey) 237
Testing Reading Through Summary

Appendix 11: Final answer key for summary completion Task B
(Anorexia) 239

References 241
Author index 259
Subject index 264
Acknowledgements

The usual advice given to a student on completing their PhD is to publish all or some of it as quickly as possible, either as a monograph or in the form of journal papers. In my own case, professional responsibilities soon took priority and the likelihood of publishing anything from my doctoral dissertation on the testing of reading receded steadily into the distance.

In recent years, however, the focus and outcomes of my earlier research exploring approaches to assessing reading comprehension ability have assumed a fresh relevance and I believe they now represent a timely and useful contribution to our current theory and practice in applied linguistics and language testing. The reason for this is that the past decade has seen a burgeoning interest in the cognitive processing that underpins language use, particularly what distinguishes an expert from a novice user. Increasing attention is also being focused on how far such cognitive processing is adequately represented in approaches to assessing language proficiency. There is growing understanding of the cognitive processes that are typically activated in test and non-test tasks, and of the importance of an appropriate match between the two. This concern for construct validity is particularly relevant when seeking to assess comprehension ability, given that language testers rely heavily upon indirect measures to make visible a process and product located inside the head of the language user.

The nature of comprehension ability and the suitability of approaches for evaluating it are issues with which applied linguists, cognitive psychologists and language testers have grappled for many years. Language testers generally have no difficulty in designing reading tests that activate the lower-level perceptual processes of decoding, lexical access and syntactic parsing, but there remain significant challenges in creating test tasks that successfully activate higher-level conceptual processes of meaning construction and discourse representation in reading, processes which are critical at higher proficiency levels, for example in academic study contexts.

At the University of Cambridge's Research Centre in English and Applied Linguistics (RCEAL), I was fortunate to be able to spend time researching these issues under the supervision of Alastair Pollitt, Senior Lecturer at RCEAL and also Director of the Research and Evaluation Division at the University of Cambridge Local Examinations Syndicate (UCLES). I also benefitted greatly from the research interest and support of Professor Gillian Brown, an eminent authority in text comprehension and discourse analysis,
Testing Reading Through Summary

and Dr John Williams, an experienced cognitive psychologist with expertise in both first and second language acquisition. My overall aim was to investigate what readers actually do when they comprehend a text and to explore the potential for a test task that would reconcile more closely the practice of assessing reading comprehension ability with a sound understanding of the nature of reading comprehension.

Professor Cyril J Weir persuaded me to revisit my PhD research and to remodel the original doctoral thesis in order to make it suitable for publication, by updating the literature in a number of areas and by contextualising the findings for the present day. I am most grateful to him for his encouragement to publish and for his editorial guidance on the manuscript as a whole, as well as to various CRELLA colleagues who assisted with specific chapters: Dr John Field reviewed and advised on Chapters 1 and 2, with regard to the psychological and cognitive dimensions of reading; and Professor Stephen Bax offered input to Chapter 3 on approaches to text and discourse analysis.

It would not have been possible for me to undertake the original research reported in this volume without the support, co-operation and encouragement of various people.

UCLES was extremely generous in providing funding for the doctoral research. In particular I would like to thank Dr Michael Milanovic, Chief Executive of Cambridge English Language Assessment, for the interest and support he showed throughout the study, and indeed throughout my career in language testing over nearly 25 years. Much of my professional life has involved working in close association with Cambridge so it is a privilege and a pleasure to have the opportunity of publishing my work as a volume in the Studies in Language Testing series in the year when Cambridge celebrates its 100th anniversary.

Grateful thanks are also due to Professor Charles Alderson whose own work in the field of reading assessment influenced my early interest and enquiry in this area. Professor Alderson kindly served as my external PhD examiner and his encouraging and insightful comments helped inform revisions to the manuscript for this volume.

Special thanks must be recorded to the staff and students at Hills Road Sixth Form College, Cambridge, who were involved with the oral recall study, and to the staff and pupils at Sawston Village College who took part in the written recall study and in the trialling of the two summary completion tasks.

Finally, the publishers are grateful to the copyright holders for permission to use the copyright material reproduced in this book: Macmillan Caribbean for the use of the short story ‘Journey by Night’ by Undine Giuseppi, and The Independent for the use of the editorial piece ‘The rights and wrongs of treating anorexia’.
Series Editors’ note

Since its inception in 1995, the Studies in Language Testing (SiLT) series has published many PhDs of quality. One of the core purposes of this innovative and now well-established series is to support and promote work in the field of language assessment by enabling the language testing community to benefit from research which makes a contribution to the field but which might not otherwise reach publication. PhDs are selected for inclusion in the series in accordance with certain criteria which include:

- being a contribution to knowledge
- being previously unpublished
- having a sound theoretical basis
- being well-referenced to the literature
- being research-based
- being executed with care and rigour
- demonstrating analysis and interpretation which is well-founded
- having the style of an academic monograph.

The first PhD we published was by Anthony John Kunnan on test taker characteristics and test performance (SiLT volume 2). Eight other PhD theses have been published to date. Caroline Clapham documented the development of IELTS (International English Language Testing System) and looked in particular at the effect of background knowledge on reading comprehension (SiLT volume 4), while Anthony Green investigated the impact of the IELTS writing subtest on English for Academic Purposes pedagogy (SiLT volume 25). James Purpura investigated learner strategy use and performance (SiLT volume 8). Kieran O’Loughlin compared direct and semi-direct tests of speaking (SiLT volume 13) and Angela Hasselgreen looked at testing the spoken English of young Norwegians (SiLT volume 20). Dianne Wall and Liying Cheng both investigated aspects of test washback and impact, with Wall studying its effects on the classroom in Sri Lanka (SiLT volume 22) and Cheng carrying out a study on the classroom in Hong Kong (SiLT volume 21). Toshihiko Shiotsu examined the components of L2 reading ability in the context of Japanese learners of English (SiLT volume 32). A number of these theses were also awarded the Jacqueline Ross TOEFL Outstanding Doctoral Dissertation award. SiLT policy is to publish one PhD for every three or four SiLT volumes and in successfully doing this we have enabled high-quality
Testing Reading Through Summary

doctoral research to reach a wider audience than would normally be expected. In this volume we continue this tradition and publish Lynda Taylor’s PhD thesis on *Testing Reading Through Summary: Investigating summary completion tasks for assessing reading comprehension ability*.

The publication of a study on the use of summary is timely given it was a testing device in vogue for the first three quarters of the 20th century, disappeared from view in the “communicative revolution” of the 1970s and then re-emerged on the global stage at the start of the 21st century.

At the beginning of the 20th century, précis featured widely in many English for Specific Purposes (ESP) and educationally-oriented tests. Robeson (1913), a Master at Eton, describes its use in Civil Service, Army and Navy qualifying examinations, in the commercial and teacher awards and examinations of the Royal Society of Arts and the London Chamber of Commerce, and by the Oxford and Cambridge Schools Examination Board in its Examination for School or Leaving Certificates. He quotes (1913:9) from the London Chamber of Commerce Regulations:

> [T]he object of the précis is to enable any one reading it to be put into possession, in the smallest space of time, of the essential points of the subject to which the documents refer. The characteristics of a good précis accordingly are (a) the inclusion of all that is important and the exclusion of all that is unimportant in the correspondence; (b) the expression of this in a consecutive story as clearly as possible, and as briefly as is compatible with distinctness.

Summary tasks, which by necessity involved reading comprehension at the global text level, were included as test formats in Cambridge examinations from the 1930s onwards. In 1931 a précis of a passage or a poem had been introduced into the English Literature paper in Part B. Typically, candidates had to choose between summarising a passage, which included defining the meaning of words and phrases in the text, and explaining a poem in detail including a focus on style and diction. In addition, by 1936 there was a further summary task in the English Essay paper.

Summaries were viewed in the school system as valuable, integrated tasks and an appreciation of the validity of the task took precedence over any concern with difficulties of marking. The tasks were intended to test comprehension of a whole passage (careful reading at the global as well as the local level) as well as writing ability and this stands in stark contrast to the emphasis on decoding in many tests of reading and in the research literature in the first half of the 20th century.

Summary was to last as a task in the Certificate of Proficiency in English (CPE) right through to 1975. Given the continued use of CPE for university entrance in the 21st century, the demise of such an authentic academic reading-into-writing task may, with the advantage of hindsight, be regretted.
(note, however, its return to favour in 2002 albeit in a reduced intertextual form in the CPE Use of English paper). Cambridge had not been alone in abandoning summary and the well regarded Schools Council Research Studies monograph entitled *The Development of Writing Abilities (11–18)* (Britton, Burgess, Martin, McLeod and Rosen 1975) contains only one brief and fairly disparaging reference to summary on page 47. From a present day perspective, we would argue that (albeit in an integrated format) summary effectively tests the important advanced level reading skill of creating a text level representation, a vital element of academic study, in an authentic manner.

As well as its rebirth in Cambridge examinations, summary has been introduced into the Internet-Based Test of English of a Foreign Language (TOEFL iBT) and the General English Proficiency Test (GEPT) Advanced Reading Test in Taiwan in the 21st century, but relatively little serious research has been carried out on summary as a testing task to date. The literature on the assessment and teaching of reading is immense so it is perhaps surprising to find so few studies available relating to the use of summary as a measurement tool. Their absence is even more surprising given a nascent concern with the construct validity of test formats employed for assessing reading ability in the 21st century.

An overt concern with the constructs being measured in the Cambridge English examinations and their relationship to real-life language use was apparent by the end of the 20th century. The commitment to transparency and the explicit specification of the communicative content of its examinations was further enhanced by Cambridge’s adoption of a socio-cognitive approach to language test design and validation in the first decade of the 21st century; such an approach acknowledges that language use constitutes both a socially situated and a cognitively processed phenomenon and that this must be reflected in language assessment theory and practice.

The increased attention paid to cognitive validity at Cambridge came about as a result of a 10-year project (2003–2013) which saw the publication of the ‘construct-focused’ volumes in the SiLT series (SiLT volumes 26, 29, 30 and 35), guided by Michael Milanovic, Nick Saville, Lynda Taylor, Evelina Galaczi and Cyril J Weir on the editorial steering committee. This ambitious project enabled far greater attention to be paid than previously to the cognitive processing typically activated in test and non-test tasks, and to the importance of an appropriate match between the two. There is now a growing recognition within Cambridge English Language Assessment and its partners, and in the wider international testing community, of the importance for any successful assessment system of seeking and assembling validity evidence on each of the three core aspects of validity: cognitive, context and scoring, which together constitute test construct validity.

Lynda Taylor’s PhD thesis was very much ahead of its time when it was
Testing Reading Through Summary

conceived and written in the mid-1990s given that at the time very few language testers and even fewer examination boards paid any serious attention to the cognitive processing underpinning the tasks they employed. Indeed, the cognitive validity of the tasks used in most tests of reading comprehension is a concept that many language testers and examination boards are still struggling to come to terms with. Lynda Taylor tackled the issues involved in addressing this critical component of construct validity for reading tests head on in her thesis at Cambridge University under the supervision of Alastair Pollitt, then Director of the Assessment Division at the University of Cambridge Local Examinations Syndicate (UCLES), and with the support of the other members of her Research Committee, Gillian Brown, an established authority in text comprehension and discourse analysis, and John Williams, an eminent cognitive psychologist. Lynda investigated a test format that reconciles, more closely than any other alternative format, the practice of assessing reading comprehension ability with our current understanding of the nature of reading comprehension. The editors were thus very pleased when she agreed, after much cajoling, to revisit her thesis a number of years later and remodel it as a book for the SiLT series on the use of summary as a language testing task for measuring reading comprehension ability. They felt it would help address an important gap in the research literature on the testing of reading and further ground the need in test design to take account of what learners actually do when they comprehend a text.

Taylor argues that the main experimental aim of the series of studies she carried out was to investigate the key features of readers’ mental representation of text and to identify how best to develop a summary completion task which directly addressed those understandings. Developing such a task involved exploring different readers’ mental representations of a given text to identify what constituted an adequate verbal summary version of the text in question. Secondly, it required the construction of suitable test items from within the resulting summary which could be used to assess readers’ comprehension of the text.

She describes how readers’ mental representations of two different texts, one narrative and one expository, were explored through a series of studies and how a text-removed summary completion task was developed to accompany each text. The two summary completion tasks were then trialled on a population of readers and the results from this exercise were compared with an independent measure of reading ability for the same population to determine the effectiveness of the text-removed summary completion format as a measure of reading comprehension ability. An accompanying aim of the investigation was to establish some practical guiding principles for the construction of summary completion tasks.

In Chapter 2 of this volume Taylor reviews the development of different theories of reading and text comprehension over the past century. Particular
attention is paid to the active and constructive nature of the comprehension process, in which meaning is constructed by the reader’s cognitive processes interacting with their knowledge base and personal goals. Chapter 3 offers a survey of reading test design, briefly chronicling the historical developments which have led to current practice and it deals in more detail with issues relating to construct validity. In Chapter 4 the rationale for using summary writing tasks as a means of assessing reading comprehension ability is explored along with the problems. An alternative approach – summary completion technique – is considered and the research questions for an empirical study are then presented. Chapter 5 reports on a text recall study designed to investigate readers’ mental representations following the reading of two different texts, Text A (Journey) and Text B (Anorexia). Chapter 6 presents the detailed results of the text recall study for each of these texts, based upon an analysis of readers’ mental representations in terms of text-based (micro-) propositions, summarising (macro-) propositions and additional propositions occurring in their oral recalls. Chapter 7 reports on the design of two text-removed summary completion tasks using summaries derived directly from the readers’ shared mental representations of Texts A and B. Chapter 8 reports the results from trialling the two summary completion tasks with a population of readers and the concluding Chapter 9 summarises the main research findings of the study, discussing their implications and making suggestions for future areas of research.

This volume offers examining boards as well the teacher in the classroom both practical and theoretical support for developing summary completion tasks to assess reading comprehension. In so doing, it affords them the possibility of employing a task which has potentially greater claims to the mantle of cognitive validity than many other formats in common use for assessing reading comprehension ability.

Cyril J Weir and Michael Milanovic
March 2013
Abbreviations

AEB Associated Examining Board
BNC British National Corpus
CEFR Common European Framework of Reference
CPE Certificate of Proficiency in English
DIALANG Diagnostic Language (Assessment)
EAP English for Academic Purposes
EFL English as a Foreign Language
ELM English Language Monitoring (Project)
ELTS English Language Testing Service
ESOL English for Speakers of Other Languages
ESP English for Specific Purposes
GCSE General Certificate of Secondary Education
GEPT General English Proficiency Test
HF High Frequency
IELTS International English Language Testing System
ISLC International Study and Language Centre
KR20 Kuder-Richardson 20
KS3 Key Stage 3
L1 First Language
L2 Second Language
MA Master of Arts
MC Multiple Choice
MCQ Multiple Choice Question
MF Medium Frequency
MR Macro Rule
PDP Parallel Distributed Processing
PTE Pearson Test of English
SP Summarising Proposition
TEEP Test of English for Educational Purposes
TOEFL iBT Internet-Based Test of English as a Foreign Language
TP Text-based Proposition
TTR Type Token Ratio
UCLES University of Cambridge Local Examinations Syndicate
VPA Verbal Protocol Analysis