

li85

Linguistic Insights

Studies in Language and Communication

Erik Castello

# Text Complexity and Reading Comprehension Tests

Peter Lang

Based on the analysis of a specially compiled corpus of internationally recognized English as a foreign language (EFL) reading tests at different levels of proficiency, this volume explores the relation between the complexity of written texts and the difficulty of reading comprehension tests. It brings together linguistic investigations into the text-inherent complexity of the tests and a study of the data derived from their administration to groups of Italian university students. The study of text complexity draws on corpus linguistics, text linguistics and systemic functional linguistics. Both quantitative and qualitative analyses are carried out on the language used in the reading texts and in the related tasks that make up the corpus of tests. The assessment of test difficulty, on the other hand, is informed by research on language testing, and, in particular, by findings and methodologies of Classical Test Theory and Item Response Theory. Relevant aspects of these theories are used to analyze and interpret both the data obtained from the administration of the tests and the data collected by means of feedback questionnaires completed by test takers. The application of such diverse methodologies and the subsequent comparison of the results of the analyses has brought out interesting correlations between text-inherent complexity, perceived test difficulty and actual test difficulty.

**Erik Castello** is a tenured researcher in English Language and Translation and teaches English at the University of Padua, Italy. He holds a doctorate in English linguistics and his research interests include corpus linguistics, systemic functional linguistics, language testing, and discourse analysis.

**Text Complexity and Reading Comprehension Tests**



# Linguistic Insights

Studies in Language and Communication

Edited by Maurizio Gotti,  
University of Bergamo

Volume 85

## ADVISORY BOARD

Vijay Bhatia (Hong Kong)  
Christopher Candlin (Sydney)  
David Crystal (Bangor)  
Konrad Ehlich (Berlin / München)  
Jan Engberg (Aarhus)  
Norman Fairclough (Lancaster)  
John Flowerdew (Leeds)  
Ken Hyland (London)  
Roger Lass (Cape Town)  
Matti Rissanen (Helsinki)  
Françoise Salager-Meyer (Mérida)  
Srikant Sarangi (Cardiff)  
Susan Šarčević (Rijeka)  
Lawrence Solan (New York)  
Peter M. Tiersma (Los Angeles)



PETER LANG

Bern • Berlin • Bruxelles • Frankfurt am Main • New York • Oxford • Wien

Erik Castello

# Text Complexity and Reading Comprehension Tests



**PETER LANG**

Bern • Berlin • Bruxelles • Frankfurt am Main • New York • Oxford • Wien

**Bibliographic information published by Die Deutsche Bibliothek**

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available on the Internet at <http://dnb.ddb.de>.

**British Library and Library of Congress Cataloguing-in-Publication Data:**

A catalogue record for this book is available from *The British Library*, Great Britain.

**Library of Congress Cataloging-in-Publication Data**

Castello, Erik.

Text complexity and reading comprehension tests / Erik Castello.

p. cm. – (Linguistic insights ; v. 85)

Includes bibliographical references and index.

ISBN 978-3-03911-717-8 (alk. paper)

1. English language–Study and teaching–Italian speakers. 2. English philology–Study and teaching–Italy. 3. Reading comprehension–Ability testing. 4. English language–Ability testing. I. Title.

PE1129.I7C37 2008

428.4028'7–dc22

2008035625

Published with a grant from Università degli Studi di Padova (Italy), Dipartimento di Lingue e Letterature Anglo-Germaniche e Slave, as one of the products of the ministerially co-funded project /Englishmultiweb/, Padua unit of the national project /eColingua/.

ISSN 1424-8689 (Print edition)

ISBN 978-3-03911-717-8

E-ISBN 978-3-0351-0276-5

© Peter Lang AG, International Academic Publishers, Bern 2008

Hochfeldstrasse 32, Postfach 746, CH-3000 Bern 9, Switzerland

[info@peterlang.com](mailto:info@peterlang.com), [www.peterlang.com](http://www.peterlang.com), [www.peterlang.net](http://www.peterlang.net)

All rights reserved.

All parts of this publication are protected by copyright.

Any utilisation outside the strict limits of the copyright law, without the permission of the publisher, is forbidden and liable to prosecution.

This applies in particular to reproductions, translations, microfilming, and storage and processing in electronic retrieval systems.

Printed in Germany

# Contents

Thanks and Acknowledgements .....	9
1. Introduction .....	11
1.1 Background and Purpose of this Study .....	11
1.2 The Theoretical and Methodological Framework.....	13
1.3 Outline of Content .....	21
2. A Corpus of Reading Comprehension Tests.....	23
2.1 The DIALANG Sub-corpus.....	24
2.2 The TOEFL Sub-corpus.....	26
2.3 The IELTS Sub-corpus .....	30
2.4 The PET Sub-corpus.....	32
2.5 The FCE Sub-corpus.....	34
2.6 The CAE Sub-corpus .....	36
3. Quantitative Investigation into the Complexity of the Texts of the Corpus.....	41
3.1 The Type/Token Ratio .....	42
3.2 Measuring Lexical Density and Lexical Variation .....	48
3.2.1 Measures of Lexical Density.....	49
3.2.1.1 Lexical Items vs. Grammatical Words .....	53
3.2.1.2 Lexical Density (Ure's method) in the Corpus.....	58
3.2.2 A Measure of Lexical Variation.....	64
3.3 Measuring Vocabulary Difficulty .....	71
3.3.1 Range 32 and the Underlying Research .....	72
3.3.2 Using Range 32 and Base Word Lists for the Analysis of the Corpus.....	75
3.4 Measuring Grammatical Intricacy and Halliday's method of Measuring Lexical Density .....	87

3.4.1 Halliday's Method of Calculating Lexical Density (LDH), Grammatical Intricacy, the Grammar of the Clause Complex, and Embedding .....	88
3.4.1.1 The Clause Complex .....	89
3.4.1.2 Embedding of Clauses .....	94
3.4.2 Further Investigation of the Lexical Density of the Texts and an Investigation of their Grammatical Intricacy .....	97
3.5 Measuring Readability .....	103
3.5.1 Readability Formulae .....	103
3.5.2 Investigating the Readability of the Corpus .....	106
3.6 On the Relationship between the Variables .....	113
 4. Text Complexity, Participant Identification and Cohesion in the Corpus .....	123
4.1 Participant Identification in English: from Discourse Semantics to Lexicogrammar .....	124
4.1.1 Participant Identification in English: a Review of some Relevant Contributions .....	124
4.1.2 Phoricity and the Tracking System Network .....	129
4.1.3 The Participant Identification System Network .....	140
4.2 Semantics, Lexicogrammar and the Textual Metafunction: a Systemic Functional Perspective.....	147
4.2.1 The Semantic and the Lexicogrammatical Strata in Systemic Functional Linguistics .....	148
4.2.2 The Textual Metafunction: Cohesion and Thematic Structure in English .....	150
4.2.2.1 Cohesion in English.....	151
4.2.2.2 Thematic Structure .....	153
4.3 Participant Identification and the English Nominal Group.....	157
4.4 The Role of Participant Identification and of Textual Cohesion in the Corpus .....	164
4.4.1 Test 12.....	165
4.4.2 Tests 2, 25 and 11.....	170
4.4.2.1 Test 2 .....	171



4.4.2.2 Test 25 .....	174
4.4.2.3 Test 11 .....	181
4.4.3 Test 14 .....	184
4.4.4 Test 22 .....	189
4.4.5 Tests 15, 3 and 13 .....	193
4.4.5.1 Test 15 .....	194
4.4.5.2 Test 3 .....	202
4.4.5.3 Test 13 .....	203
4.4.6 Tests 18, 19, 23, and 21 .....	206
4.4.6.1 Tests 18 and 19 .....	206
4.4.6.2 Tests 23 and 24 .....	207
 5. Assessing Test Difficulty: Data Analysis of Tests and Feedback Questionnaires .....	 211
5.1 Setting the Scene for this Part of the Study .....	211
5.2 Reliability and Validity in Language Testing .....	213
5.3 Test and Questionnaire Administration and Data Analysis ....	216
5.4 Test Battery 1 (from Test 01 to Test 10) .....	220
5.5 Test Battery 2 (from Test 18 to Test 21) .....	240
5.6 Test Battery 3 (Tests 11, 12, 22, and 23) .....	251
5.7 Test Battery 4 (from Test 12 to Test 16) .....	260
5.7.1 From Test 13 to Test 16 .....	262
5.7.2 Test 12 and the Perception of the Difficulty of the Tests in Battery 4 .....	 271
5.8 Test Battery 5 (Tests 17, 24 and 25) .....	278
5.9 Relationship between Test Takers' Performance, their Perception of Text/Task Difficulty and Results of the Quantitative Investigation .....	  289
 6. Conclusions .....	 295
 References .....	 309
Appendix .....	321
Index .....	351



## Thanks and Acknowledgements

The author and publisher are grateful to the following for permission to reproduce copyright material as specified below. It has not always been possible to identify the sources of all the reading texts used in the tests analysed or to contact the copyright holders, and in such cases the author and the publisher would welcome information from the copyright owners.

### Tests 1 to 10

© DIALANG <<http://www.dialang.org/english/index.htm>>.

### Tests 11 and 12

Excerpts pp. 345-346 and 351-352. From *Longman Complete Course for the TOEFL Test, Preparation for the Computer and Paper Tests* by Deborah Phillips. Copyright © 2001 by Addison-Wesley Longman, Inc. Reprinted by permission of Pearson Education, Inc.

### Test 13

Test taken from *Heinemann TOEFL Preparation Course*. Copyright © M.K. Mahnke, Carolyn Duffy & Jacqueline Flamm, 1996. Reproduced with permission of Macmillan Publishers Ltd.

### Tests 14, 15, 16, 17, 19, 22, and 23

University of Cambridge Local Examinations Syndicate, *Cambridge IELTS 2 Student's Book with Answers*, (2001), *Preliminary English Test 2 Student's Book*, (2001), *Cambridge First Certificate in English 4 Student's Book*, (2001), © UCLES, reprinted with permission of the publisher (Cambridge University Press) and UCLES.

### Test 17

*The Independent* for the extract from “Green wave washes mainstream shopping” by David Nicholson-Lord.

Tests 18, 20, 21, 24, and 25

University of Cambridge ESOL Examinations. Excerpts adapted from “Sample paper R&W PET.pdf”, downloaded from <[http://www.candidates.cambridgeesol.org/cs/digitalAssets/105746\\_pet\\_sample\\_papers.zip](http://www.candidates.cambridgeesol.org/cs/digitalAssets/105746_pet_sample_papers.zip)> and “Sample Paper R CAE.pdf” downloaded from <[http://www.candidates.cambridgeesol.org/cs/digitalAssets/106850\\_cae\\_sample\\_papers.zip](http://www.candidates.cambridgeesol.org/cs/digitalAssets/106850_cae_sample_papers.zip)>, © University of Cambridge ESOL Examinations.

# 1. Introduction

This study originates from the practical need that is common to all those who have to face the problem of creating reading-comprehension language tests: the choice of the texts that are suited to a given level of linguistic competence and the construction of equally suitable tasks to assess the candidates' comprehension of the texts. The background and purpose of this study will be illustrated in section 1.1, the theoretical and methodological framework adopted will be sketched out briefly in section 1.2 and an outline of the contents will be given in section 1.3.

## 1.1 Background and Purpose of this Study

This study brings together two areas of research in an investigation of text complexity and English-language test difficulty. One aspect of this study is more strictly linguistic and is related to attempts in linguistics – most importantly text-linguistics, functional linguistics and cognitive linguistics – to define and characterise the notion of ‘text complexity’. The other aspect is related to research into the notions of ‘test difficulty’ and ‘task difficulty’ carried out in the field normally called applied linguistics and particularly in research on language testing. In order to investigate these phenomena, a corpus made up of twenty-five reading comprehension tests was designed and analysed. The tests in this 10,613-word long corpus offer the material for analyses of text complexity and of test difficulty. Specifically, they consist of an input text and of one or more accompanying tasks and have been elaborated by their authors to assess the reading skills of students of English as a second or foreign language at different levels of linguistic proficiency. Statistical methods were made use of to

quantify, wherever possible, both the inherent complexity and the (receiver-oriented) difficulty of the corpus.

The distinction made in this study between ‘complexity’ and ‘difficulty’ derives from Merlini Barbaresi (2003: 23), who distinguishes between ‘text inherent complexity’ on the one hand and ‘the (receiver-oriented) notion of text difficulty’ on the other. She defines text complexity as:

a multifaceted quality of the text starting with the speaker’s locutionary and illocutionary planning, but adjustable and negotiable during the text development.

Difficulty, on the other hand, is defined as “a predictable and explainable perlocutionary effect, subject to situational variables”. She then goes on to say that:

Text complexity is a dynamic configuration resulting from the contributions of complex phenomena, as they occur at the various text levels. Text is here viewed as a system, whose components are interacting to achieve a global effect.

Merlini Barbaresi (2003: 29) also hypothesises that ‘markedness’:

can be measured by means of a grid of criteria, that complexity is the result of the amount and values of marked phenomena in the specific text and that difficulty derives from the value of complexity, but it is relativised by contextual variables.

For Merlini Barbaresi (2003: 33), text complexity is an ‘emergent cumulative quality’, due to the interplay of the various linguistic sub-systems a text is made up of. She underlines how, in order to effectively, economically, and ‘optimally’ convey the intended meanings, compromise solutions between the various textual sub-systems are necessary, and marked choices among the options available within the sub-systems might have to be made. The theoretical framework that Merlini Barbaresi proposes to account for the interplay of marked choices at different discourse and textual levels is the theory of ‘naturalness/markedness’, which she elaborated beginning in Merlini Barbaresi (1988), on the basis of the works by, for example, Peirce (1965), Dressler (1985), and Dressler, Mayerthaler, Panagl, and

Wurzel (1987)<sup>1</sup>. In Merlini Barbaresi's model, text complexity is, in particular,

a global assessment obtained by evaluating the number and mutual impact of the various marked phenomena. Its degree is proportionally dependent on the percentage of marked vs. unmarked phenomena, and on their markedness values. A large proportion of marked phenomena in a text will account for a high degree of global complexity, but it will not allow a conception of global markedness. From the receiver's perspective, complexity is evaluated in terms of processing difficulty, but this value is situationally biased, i.e. depending on many variables, whereas markedness values are not. Such factors as text type, micro- and macro-goals, addressee and his/her approach to the text are powerful variables, which come in to regulate complexity and relative processing difficulty. (Merlini Barbaresi 2003: 34)

However, Merlini Barbaresi (2003: 29) admits that their research project on markedness as complexity in texts has not yet achieved "a definition of graded values of markedness, complexity and difficulty and of their mutual significance," although "this remains a prominent goal in the research project". Indeed, although the great potential that a markedness framework might offer for this research was clearly recognized, it was decided that an attempt to use it as the basis of this study would be premature. The approach, methods and procedures described in 1.2 below were, therefore, chosen and applied.

## 1.2 The Theoretical and Methodological Framework

The theoretical and methodological framework for this study was, as said above, selected and designed with in mind the investigation of both the text-inherent complexity and the processing (and test) difficulty of the twenty-five tests of the corpus. The investigation into their inherent complexity was carried out by applying both quantita-

---

1 The notion of markedness has been used by linguists belonging to different schools and traditions. For instance, some structural linguists of the Prague School applied it to the fields of phonology and grammar (Givón: 1990: 946), while Givón (1990: 945-986) applied it to syntax.

tive and qualitative text- and corpus-analyses. Specifically, on the one hand methods and findings of systemic functional linguistics, corpus linguistics, text linguistics, and psycho-linguistics were drawn upon for the qualitative analysis, and on the other some calculations already available in the literature were made use of to obtain quantitative data: the ‘type/token ratio’, ‘lexical density’, ‘lexical variation’, ‘vocabulary difficulty’, ‘readability formulae’, as well as ‘grammatical intricacy’ and the calculation of ‘lexical density’ following Halliday’s (1989) method. These last two calculations, in particular, require hand tagging, but they are also amply described by various researchers, including the present author (Castello 2002 and 2004).

The phenomena investigated were captured by using thirteen measures in all:

- The type/token ratio;
- The standardised type/token ratio;
- Lexical density calculated following Ure’s method;
- Lexical density calculated following Halliday’s method;
- Lexical variation;
- Lexical variation calculated on the basis of the number of word families in a text;
- Grammatical intricacy;
- The ratio of secondary clauses in hypotactic structure to independent, paratactically related clauses, and primary clauses in hypotactic structures;
- The Flesch Reading Ease index;
- The Flesch-Kinkaid Grade Level index;
- The Gunning-Fog index;
- The percentage of academic words;
- The percentage of words that do not belong to any frequency list.

The type/token ratio is the ratio of the number of different words (types) to the number of running words (tokens) in a given text or corpus. This index indicates how often, on average, a new ‘word-form’ appears in the text (McEnery and Wilson, 1996: 158), and can be used as a way of measuring vocabulary and lexical diversity in a text or corpus. For the comparison of texts of different length the standardized type/token ratio is more informative (Scott 2006).



Lexical density (Halliday 1987: 60) is “the proportion of lexical items (content words) to the total discourse.” It can be measured either as the ratio of lexical items to the total running words or as the ratio of the lexical items to the clauses in a text. Grammatical intricacy is the ratio of the total number of ranking clauses to the total number of clause complexes. Halliday (2004: 654) considers lexical density and grammatical intricacy two different kinds of complexity, the former associated typically with written language and the latter with spoken language.

Lexical variation is a measure of the lexical repetition or lexical diversity in a text (Gibson 1993: 157). This will be measured in two ways. The first is the ratio of the total number of different lexical words in a text to the total number of lexical words. The second, on the other hand, is the ratio of the total number of word families in the text to the total number of lexical words. The type/token ratio, lexical density, grammatical intricacy, and lexical variation can all be considered measures of text complexity.

The so-called readability formulae represent a method of assigning a numerical estimate of ‘readability’ to a text. Bruce and Rubin (1988: 5) define readability as ‘ease of reading’, ‘interest’ or ‘ease of understanding’, which are receiver-oriented criteria for difficulty or lack of difficulty. However, the readability formulae used in this study – i.e. the Flesch Reading Ease formula, the Flesch-Kinkaid Grade Level formula, and the Gunning Fog Index – although experimented on readers, only take into consideration text-inherent factors (e.g. word length, sentence length), and, therefore, they can be included in this overall attempt to quantify the complexity of the present corpus.

It was also decided to concentrate on the area of text cohesion and, in particular, the problem of ‘participant identification’ and lexical cohesion. *Participant identification* is the term used by Martin (1992: 93-157) to indicate the textual resources that allow readers and listeners to track the identities of participants, that is, to introduce people and things into a text and keep track of them once there. Participant identification is, in particular, concerned with the ‘phoric’ relationships between the items which allow readers or listeners to recover the correct referent. In the present study, the analysis of participant identification was carried out to identify the tracking resources used in the tests of the corpus, and not just in the texts. The

focus was, therefore, on the resources used to identify the relevant participants which occur in both the input texts and in the accompanying tasks. The analysis of cohesion in chapter 4 is qualitative rather than quantitative, and takes into consideration various phenomena which might be considered elements of complexity. Although the findings on text cohesion have not been quantified, indications have emerged of marked and complex types of cohesion, where it was foreseeable that there could be difficulty in properly identifying a participant retrieved through cohesive devices. These indications were confronted with the actual test results and questionnaire answers.

The investigation into the difficulty of the language tests in the corpus was carried out by applying language testing methodologies. Specifically, some language testing procedures were made use of, firstly, to properly administer the tests to a sample population of 379 readers/test takers, and analyse their test results, and, secondly, to devise an evaluative feedback questionnaire about the difficulty of the texts and tasks under investigation and analyse and interpret the data collected.

As stated above, in language testing one of the main areas of research concerns test and task difficulty. Test and task difficulty, in particular, is not only related to the characteristics of the task itself (e.g. answering a multiple-choice question), but also to the characteristics of the readers and to those of the texts to be read. Text, task (and test) and readers are, therefore, conceived of as interrelated and interdependent variables. In this respect, Alderson (2000: 32-33) maintains that the 'reader' and the 'text' are the two main 'constellations' of variables involved in the process of reading, and the two of them are seen as interacting with each other. As he points out, the readers and their characteristics have important effects on the reading 'process' – i.e. the interaction between the reader and the text – as well as on the 'product' of reading – i.e. the 'understandings' which readers end up with – because:

different readers will develop somewhat different understandings of what a text 'means'. This is at least in part because a text does not 'contain' meaning which is waiting to be discovered by an able reader. Rather, meaning is created in the interaction between a reader and a text: the text has what Halliday (1979) and Widdowson (1979) call meaning potential, and the potential is realised – in the product of understanding – only by readers reading. Since

[...] readers' knowledge and experiences influence the realisation of this meaning potential, and since readers may differ in their knowledge and experiences, then the products of reading will also necessarily differ. (Alderson 2000: 6)

Similarly, according to Urquhart and Weir (1998: 112-113):

Readings may differ in the case of readers from different cultures, either ethnic or professional, or in the case of the same reader at different times, with different knowledge or different preoccupations. Such differing readings, which are generally not under the control of the readers, Urquhart terms 'interpretations'. The reading product may also vary according to a dimension controlled by the reader's purposes. As Candlin points out, the reader 'may decide to glimpse at the text, extracting 'gist', or work conscientiously through it, satisfying himself that he has made sense of all of it'. In Urquhart (1987) such variations are labelled 'comprehensions'.

From this perspective then the different understandings or comprehensions of a text – i.e. the readers' realisations of the text's meaning potential – can be thought of as the perlocutionary effects of a given text on different readers, and, following Merlini Barbaresi's definition of text difficulty given above, as related to the difficulty of a text. Furthermore, since text, readers, and task are interdependent variables, text difficulty and task difficulty also depend on each other and on the characteristics of the readers.

Alderson (2000: 33-60) reviews some of the main contributions and findings concerning the readers' characteristics which are relevant to language testing. According to his classification, the main variables that have been investigated in this area are: the reader's knowledge, e.g. background knowledge, knowledge of language, schemata<sup>2</sup>; the

---

2 Alderson (2000: 34) writes that a distinction is often made between 'formal schemata' and 'content schemata'. Quoting Carrell (1983), he writes that formal schemata include "knowledge of language and linguistic conventions, including knowledge of how texts are organised, and what the main features of particular genres are." Content schemata, by contrast, include "knowledge of the world, including the subject matter of the text." He also writes that "content schemata can be divided into background knowledge – i.e. knowledge which may or may not be relevant to the content of a particular text – and subject-matter knowledge, which is directly relevant to text content and topic. Moreover, some researchers have focused upon certain aspects of

reader's motivation; the strategies that readers use when processing texts; the reader's skills and abilities; the reader's purpose and motivation in reading; the reader's affect, e.g. the emotional state induced by a testing situation; and other stable characteristics, such as sex, social class, occupation, and whether the readers are beginning or fluent. As far as text variables related to text complexity are concerned, Alderson (2000: 60-79) lists: 'text topic and content', 'text type and genre', 'literary and non-literary texts', 'text organisation', 'text readability', 'typographical features', 'verbal and non-verbal information', the 'medium of text presentation', and 'traditional linguistic variables'. These include, for example, 'syntactic complexity', "the opacity and heaviness of the constituent structure of sentences which make it difficult for readers to parse sentences", and 'vocabulary difficulty.' Some of these variables are investigated in chapter 3, where an attempt is made to quantify and capture some aspects of the complexity of the texts which form the corpus for this study (see chapter 2). However, Alderson warns us that inherent text complexity alone is not a sufficient concern for those investigating test difficulty and says that:

The language of texts would seem, *prima facie*, highly relevant to the testing and assessment of reading. The interesting thing about much of the research is that a common-sense assumption proves too simplistic, and that identifying text variables which consistently cause difficulty is a complex task. Clearly at some level the syntax and lexis of texts will contribute to text and thus to task difficulty, but the interaction among syntactic, lexical, discourse and topic variables is such that no one variable can be shown to be paramount. Moreover, even the ability to guess words from the context has to be seen in context: the context of the reader, and other variables in the text. (Alderson 2000: 70-71)

Alderson's concerns are shared, and in fact the choice of asking different groups of Italian university students to read some texts suited to their respective levels of language proficiency and answer related test questions (see chapter 5) was made partly to take into account their characteristics as readers.

According to Alderson (2000: 85), the difficulty of a task, or item, or test question is "very simply measured by the proportion of

---

background knowledge, in particular that knowledge which is common to a particular culture, or culture knowledge". (Alderson 2000: 34)

candidates getting the answer correct compared with those getting it wrong". The causes of task difficulty are manifold and depend mainly on the characteristics of the tasks. As the testing 'techniques' or 'test methods' used in the tests of the corpus are mainly 'multiple-choice', 'multiple-matching', 'true/false', and 'true/false/not given' (see chapter 2), the two main types of task characteristics that will be taken into account and investigated in this study are the so called "item variables that may influence the reading item difficulty measure" and the "text-by-item overlap variables that may influence the reading item difficulty measure" (Freedle 1997: 402). The former includes such item characteristics as correct choice length, incorrect choice plausibility, 'stem'<sup>3</sup> length', and 'stem-unique content words'<sup>4</sup> (Alderson 2000: 88), which are specific to multiple-choice questions. The latter includes the so called "lexical overlap between the necessary information and the correct option" (Buck, Tatsuoka and Kostin 1997: 450), where necessary information means "the information in the text which the reader must understand to be certain of the correct answer" (Buck, Tatsuoka and Kostin 1997: 437). An investigation into these task characteristics, which, however, are also to a certain extent text characteristics, is carried out in chapter 4. As said above, this chapter analyses the ways participants are identified in the input texts of the corpus and in the accompanying tasks, that is, in the overall tests. This analysis, complemented by an analysis of the cohesive resources used in these tests, was carried out to gain insights into the complexity of a given test dealt with as a text.

Another important characteristic of tasks which might influence text difficulty has to do with the type of reading that is required. There are, in fact, different 'purposes' of reading that readers/test takers consciously 'adopt' to carry out the specific tasks in a test. It can be claimed, in particular, that different purposes of reading need different 'types of reading' of a given text, that different 'readings' in turn require one to process the text in quite different ways, and that these differences can cause variations in processing or text difficulty. In this

---

3 'Stem' is the first part of multiple choice question, appearing before the optional choices.

4 Unique content words are those "nouns, verbs or modifiers that appear in the question but not in the passage" (Alderson 2000: 88).

respect, Urquhart and Weir (1998: 115) propose to break down reading into reading types according to the skills and strategies<sup>5</sup> that need to be deployed to reach specific goals. To this end, they distinguish between ‘careful’ and ‘expeditious’ reading and between ‘global’ and ‘local’ reading. By combining these reading typologies, they further distinguish between the four reading types represented in Figure 1.1.

	<i>Global</i>	<i>Local</i>
<i>Expeditious</i>	A. Skimming quickly to establish discourse topic and main ideas. Search reading to locate quickly and understand information relevant to predetermined needs.	B. Scanning to locate specific information; symbol or group of symbols; names, dates, figures or words.
<i>Careful</i>	C. Reading carefully to establish accurate comprehension of the explicitly stated main ideas the author wishes to convey; propositional inferencing.	D. Understanding syntactic structures of sentence and clause. Understanding lexical and/or grammatical cohesion. Understanding lexis/deducing meaning of lexical items from morphology and context.

Figure 1.1. Matrix of reading types (adapted from Urquhart and Weir 1998: 123).

As can be seen, by reading a text expeditiously and globally (type A) one seeks to either establish the discourse topic and the main ideas in a text (i.e. skimming) or to locate and understand information relevant to predetermined needs (i.e. search reading). Expeditious and local reading (type B), by contrast, aims to locate specific information such as symbols or names, whilst the purpose of careful and global reading (type C) is to accurately comprehend explicitly stated main ideas and to infer ‘propositional’ meaning. Finally, by reading carefully and locally (type D), one processes the immediate co-text and occasionally

---

5 According to Urquhart and Weir (1998: 96-98), there is confusion in the literature as to what distinguishes a skill from a strategy. Some of the differences they list are the following ones: strategies are reader-oriented, whilst skills are text-oriented; strategies represent conscious decisions taken by the reader, skills are deployed unconsciously; strategies, unlike skills, represent a response to a problem.

the wider context in order to deduce the meaning of unknown lexical items, the reference encoded by grammatically cohesive items, or the syntactic structures of sentences or clauses. In chapter 2 information about the specific skills and/or strategies assessed by each item or group of items in the tests of the corpus is given. It must be said, however, that “there is still no consensus on divisibility of skills”, and that in the literature three positions are common: “the first is that reading is a unitary skill; the second is that reading is multidivisible, even though there is no agreement on how many skills might be empirically distinguishable; the third is that there is a two-way split” (Alderson 2000: 95).

Whenever the data deriving from this study could be handled quantitatively, statistical techniques were used. This is true for the test results and the questionnaire data, which are analysed in chapter 5, as well as for the analysis of text complexity presented in chapter 3. When possible and relevant, in chapters 3, 4, and 5 attempts were also made to ‘triangulate’<sup>6</sup> the various types of data.

### 1.3 Outline of Content

In chapter 2 information about the twenty-five reading comprehension tests which form the corpus for this study is given. Information is also given about the testing systems which they belong to, viz. the on-line diagnostic system called DIALANG, the testing systems TOEFL and IELTS, and University of Cambridge ESOL Examinations’ Preliminary English Test (PET), First Certificate in English (FCE) and Certificate in Advanced English (CAE).

In chapter 3 a series of quantitative analyses carried out into the texts of the corpus is reported on. They all aim to capture some aspects of the complexity of these texts and to make some inferences

---

6 In social research the term ‘triangulation’ is used to indicate the use of data from different sources and obtained by means of different methodologies to increase the quality control and representativeness of a study (see, for example, Bailey 1999: 38, and 5.1 below).

about the difficulty of the tests. The linguistic variables analysed are: the ‘type/token ratio’, ‘lexical density’, ‘lexical variation’, ‘vocabulary difficulty’, ‘grammatical intricacy’, and ‘text readability’. In order to carry out these analyses, some indices and formulae are made use of, as well as word frequency lists based on large corpora and software designed to process them. Hand tagging of the corpus is performed in order to calculate grammatical intricacy and lexical density following Halliday’s method. For each of these procedures and variables the relevant literature is reviewed.

Chapter 4 focuses on some textual aspects of the texts under enquiry which it was thought might be revealing in terms of complexity. In particular, since most of the tasks are ‘closed-ended’ questions (e.g. multiple-choice, true/false questions), some relations and interactions between the texts and the accompanying tasks are also analysed. The main textual phenomena investigated are ‘participant identification’ – i.e. the lexico-grammatical realisations of the ways participants are introduced and presumed in the texts and in the tasks – and lexical cohesion. When relevant the contribution of another two aspects of cohesion, thematic structure and lexical cohesion, are taken into account.

Chapter 5 begins with a discussion of the procedures followed and the choices made in the administration of the tests. It then reports on how further information about text and task difficulty was obtained from the administration of a feedback questionnaire about the test takers’ perception of the difficulty of each text and of the accompanying tasks. The rest of chapter 5 provides an analysis and discussion of the test results and of the data derived from the feedback questionnaire. In the discussion each piece of data is related to each other and to the data obtained from the analyses reported on in the previous chapters.

The final chapter presents the main conclusions arrived at in this study on the basis of the results obtained by applying the various methods of analysis adopted. The theoretical and practical implications of the results obtained in this study as for text complexity and test difficulty are also discussed. Finally, areas for future research are identified.



## 2. A Corpus of Reading Comprehension Tests

The corpus the present study is based on is composed of twenty-five reading comprehension input texts and of the accompanying questions/tasks to be answered/performed by test-takers. They were selected from various testing systems that aim to assess the reading skills of students of English as a second or foreign language at different levels of competence and for different purposes. The majority of them were chosen from free sample materials available on-line, i.e. University of Cambridge ESOL Examinations' Preliminary English Test (PET), First Certificate in English (FCE) and Certificate in Advanced English (CAE)<sup>1</sup>, and from commercially available preparation courses for the same and other English language examinations, i.e. Test of English as a Foreign Language (TOEFL)<sup>2</sup>, the International English Language Testing System (IELTS)<sup>3</sup>, and University of Cambridge ESOL Examinations' PET and FCE tests. Some other texts and tasks, on the other hand, were taken from the

- 
- 1 University of Cambridge ESOL Examinations (Cambridge ESOL) is part of Cambridge Assessment <<http://www.cambridgeassessment.org.uk>>. Cambridge ESOL offers a wide range of exams for learners and teachers of English: General English exams, Professional English exams, Young Learners English, and Teaching Awards. The General English Exams offered are: KET (*Key English Test*), PET (*Preliminary English Test*), FCE (*First Certificate in English*), CAE (*Certificate in Advanced English*), CPE (*Certificate of Proficiency in English*), and *Certificates in ESOL Skills for Life* <<http://www.cambridgeesol.org/exams/index.html>>. Websites last visited on 25<sup>th</sup> April 2008.
  - 2 The *Test of English as a Foreign Language* (TOEFL) measures the ability of non-native speakers of English to use and understand English in college and university settings. It is developed and delivered by ETS (Educational Testing Service) <<http://www.ets.org/toefl/>>, last visited on 25<sup>th</sup> April 2008.
  - 3 The *International English Language Testing System* (IELTS) is recognised as "secure, valid and reliable indicator of true to life ability to communicate in English in education, immigration and professional accreditation". IELTS is "jointly managed by British Council, IDP: IELTS Australia and the University of Cambridge ESOL Examinations (Cambridge ESOL)" <[http://www.ielts.org/general\\_pages/about\\_us.aspx](http://www.ielts.org/general_pages/about_us.aspx)>, last visited on 25<sup>th</sup> April 2008.

reading component of DIALANG, a web-based diagnostic testing system funded by the European Commission<sup>4</sup>.

These tests were chosen and focused on for various reasons. Firstly, they are all valid, reliable<sup>5</sup> and constantly monitored tests: evidence of their validity and reliability is provided by studies that are regularly made publicly available (e.g. on the websites devoted to these testing systems). Secondly, although different from each other, their constructs<sup>6</sup> for reading seemed to be relevant for the kind of readers under investigation in this study, i.e. Italian university students of English (including both language and non-language majors) and for the sample population that was actually asked to read the texts, perform the tasks and give feedback on them for this research project<sup>7</sup>. Thirdly, with the exception of TOEFL, the tests range across different levels of language proficiency in English, which helped decide which tests were suited to the current level of English language knowledge of each group of students. In the following paragraphs of this chapter a description of the most relevant characteristics of the testing systems under enquiry and of the texts and tasks selected will be provided. The complete tests are presented in the Appendix.

## 2.1 The DIALANG Sub-corpus

DIALANG is a web-based language diagnostic and self-assessment system based on the *Common European Framework of Reference for Languages: Learning, teaching, assessment* (CEF) (Council of

---

4 See Alderson (2000: 125-128) and the DIALANG website <<http://www.dialang.org/english/index.htm>>, last visited on 25<sup>th</sup> April 2008.

5 For a definition of the concepts of ‘validity’ and ‘reliability’ see 5.2 below.

6 “A construct is a psychological concept, which derives from a theory of the ability to be tested” (Alderson 2000: 118). Further information about the concept of construct and construct validity can be found in 5.2 below.

7 The nationality of some of the students that took the tests is actually not Italian. All test-takers, however, were enrolled on undergraduate or post-graduate degree courses at the University of Padua.

Europe, 2001). It aims to assess reading, listening, writing and speaking skills as well as the knowledge of grammatical structures and vocabulary in 14 different European languages by means of separate testing modules. It also has a self-assessment component to be taken before each module, which helps the system ‘tune’ the test to the test takers’ ability<sup>8</sup>. The results test takers are given comprise the level they attained on the Council of Europe scale, an indication of what that level means, i.e. what people can usually do at that level, and a review of the responses to the items listed by sub-skill. DIALANG is therefore principally meant to help language learners diagnose and control their learning progress and increase their language learning awareness. In the reading component “questions and tasks vary from easy to demanding, such that questions asking for specific facts are usually easier than questions requiring synthesis, analysis or inference” (Alderson 2000: 128). For this research project 10 input texts and the corresponding task(s) were chosen from those presented to the present author, who sat the DIALANG reading component himself on 4<sup>th</sup> May 2005. The 30 items taken were ‘screen captured’ and so was the feedback received on each one of them and the item review by sub-skill obtained at the end of the module. The selected items were then named and numbered as reported in Figure 2.1. Information about the order in which each one of them was administered, the sub-skill each item is claimed to test and the test method used to elicit test-takers’ responses is also provided. As can be seen, four questions are asked about DIALANG text n. 1.

Alderson (2000: 128) specifies that<sup>9</sup>:

the test is divided into ‘overall’ and ‘analytic’ sections. The overall section includes items which ‘tap overall reading’ – presumably meaning main ideas, gist or the like – but can also include items which tap one or more of the three ‘skills’ focused on in the analytic section. These are: ‘identifying main

---

8 For more information see <<http://www.dialang.org/english/index.htm>>, last visited on 25<sup>th</sup> April 2008.

9 Alderson (2000: 128) actually quotes from the DIALANG Assessment Specifications for Reading Comprehension, Version 6, 18<sup>th</sup> February 1998. He writes that “the Assessment Specifications themselves merely refer to the Common European Framework, the use of which is detailed in a separate document, the DIALANG Assessment Framework (DAF)” (Alderson 2000: 125).

idea(s)/information/purpose; reading for specific detail/information'; and 'inferencing/going beyond the literal meaning (including lexical inferencing).

<i>DIALANG tests</i>	<i>Order</i>	<i>Sub-skills tested</i>	<i>Test methods</i>
1. <i>Eye problems</i>	11-14	identifying main idea	multiple matching
2. <i>Our hands met</i>	5	inferencing	multiple choice
3. <i>It should be fun</i>	10	inferencing	multiple choice
4. <i>Ecotourism</i>	20	identifying main idea	multiple choice
5. <i>Digital cameras</i>	6	inferencing	multiple choice
6. <i>Travel agency</i>	16	inferencing	multiple choice
7. <i>Bangkok</i>	17	inferencing	multiple choice
8. <i>Station</i>	2	inferencing	multiple choice
9. <i>Suzanne Somers</i>	3	inferencing	multiple choice
10. <i>Injuries</i>	25	identifying main idea	multiple choice

Figure 2.1. Information about the DIALANG items chosen for the corpus: their name, order of administration, the sub-skill tested and the test methods used.

## 2.2 The TOEFL Sub-corpus

The Test of English as a Foreign Language (TOEFL) is “intended to measure English proficiency broadly interpreted, without the engagement of any special background knowledge or specific reference to use [...], and would thus be considered a more general purpose language test” (Douglas 2000: 14). It can currently be taken in two different formats: the paper-based TOEFL test and the Internet-based TOEFL test (TOEFL iBT)<sup>10</sup>. The computer-based TOEFL test format is no longer available. The paper-based TOEFL test consists of three separately timed sections, plus a writing test. These sections are called ‘Listening Comprehension’, ‘Structure and Written Expression’, Reading Comprehension’, and ‘Writing (*Test to written*

---

10 The information about the TOEFL tests given in this section was obtained from <<http://www.ets.org/toefl/>>, last visited on 25<sup>th</sup> April 2008, unless otherwise specified.

*English*)'. The computer-based TOEFL test used to measure English language proficiency in 'Listening', 'Structure', 'Reading' and 'Writing' (essay); the Listening and Structure sections are computer-adaptive and new types of questions were added to the Listening and Reading sections (ETS 2005: 3). The TOEFL Internet-based test (TOEFL iBT) assesses all four language skills: speaking, listening, reading, and writing. It is delivered through the Internet and requires one to combine two or more of the above skills to respond to a question.

As can be seen, in all three versions of the test the reading component is present, although the time limit within which the module is to be completed and the number and types of questions vary. In particular, the time limit for the 'Reading Comprehension' module of the paper-based TOEFL test is 55 minutes, the number of texts to be read is 5 and the questions about them are 50 (Mahnke and Duffy 1996: 310, Phillips 2001: 343). By contrast, the maximum time allowed for the completion of the 'Reading' component of the computer-based TOEFL test is 70 minutes, the number of texts to read is 4 and the questions about them are 44 (Phillips 2001: 350). A characteristic that the TOEFL reading tests share with many other reading tests is that they consist of "a number of (usually short) passages, each accompanied by comprehension questions. Since the questions all relate to the same text, they are often regarded as a subtest and are referred to increasingly as 'testlets'" (Alderson 2000: 109).

Enright *et al.* (2000: 18) recommend that the tasks in the TOEFL 2000<sup>11</sup> reading test "be based on a variety of text material. Some could be based on a single text, ranging in length from a short paragraph to a lengthy selection; others could draw on multiple texts, also ranging in length". They also specify that:

---

11 "The TOEFL 2000 project is a broad effort under which language testing and Educational Testing Service (ETS) will evolve into the 21<sup>st</sup> century" (Enright *et al.* 2000: i). Enright *et al.* (2000: 49) claim that the most obvious improvement they can make over the reading test is to "articulate the construct [they] want to test and to link the proposed test design to that construct". Furthermore, they say that "one of the goals of the TOEFL 2000 project is to identify variables contributing to task difficulty in order to improve the interpretability of test scores" (Enright *et al.* 2000: 49).

any subject area that is typical of academic study could provide appropriate material for the reading test. The current TOEFL test covers a range of very general academic topics broadly classified as topics related to Arts, Humanities, Social Sciences, Physical Sciences, or Life Sciences. It seems appropriate to continue to include as much topic variety as possible in the new test. As with the current test, however, care should be taken to ensure that specialized knowledge of a particular field is not necessary to understand the information presented in the passages. (Enright *et al.* 2000: 16)

Enright *et al.* (2000: 20) further recommend that reading passages “be classified according to their dominant pragmatic and rhetorical features”, where by dominant pragmatic features they mean the primary intent of the author and by rhetorical features the higher level organization of the text. According to the parameter ‘pragmatic features’, Enright *et al.* (2000: 20-23) classify the types of text materials as ‘exposition’, ‘argumentation/persuasion/evaluation’ and ‘historical biographical/autobiographical narrative’. By contrast, different rhetorical patterns help them categorize texts as ‘definition’, ‘illustration’, ‘classification’, ‘comparison/contrast’, ‘cause/effect’, ‘problem/ solution’, and ‘analysis’ (Enright *et al.* 2000: 23-28). Finally, the types of tasks to be used in the TOEFL 2000 test should test what they call the ‘four purposes for reading’, i.e. ‘reading to find information’, ‘reading for basic comprehension’, ‘reading to learn’ and ‘reading to integrate information across multiple texts’ (Enright *et al.* 2000: 31-37).

As TOEFL 2000 is a computer-based test, Enright *et al.* (2000: 37) also list an array of response formats which should be used by item writers for “the TOEFL 2000 reading test and integrated tasks”:

- multiple choice;
- open response formats;
- click on a word, phrase, or sentence in the text or graphic;
- click on and drag a word, phrase, or sentence in the text or graphic;
- complete a chart, graph, or table;
- create a chart, graph, or table;
- extended written and/or spoken response.

For this study two texts and the corresponding testlets were selected from those proposed by two preparation courses for the paper-based

TOEFL version and another one from a preparation course for the computer-based TOEFL version. Specifically, text 11 was taken from Phillips (2001: 345) and named 'The next artist', and text 13 was selected from those proposed by Mahnke and Duffy (1996: 311) and named 'One of the most mysterious things'. They are both meant to be used to practise for the paper-based TOEFL version, whereas text 12, taken from Phillips (2001: 351-352) and called 'The final battle', belongs to the reading component of the computer-based TOEFL test. The preparation courses from which the texts and their testlets were taken also provide a classification of the 'sub-skills' tested by each item (Mahnke and Duffy 1996: 585, Phillips 2001: 349). In Figure 2.2 information about the texts, the sub-skills each item is claimed to test and the test method used is given.

<i>TOEFL</i>	<i>Sub-skills tested</i>	<i>Test methods</i>
<i>11. The next artist (paper-based)</i>		
11.01	answer transition question correctly	multiple choice
11.02	recognize the organisation of ideas	multiple choice
11.03	answer stated detailed questions correctly	multiple choice
11.04	determine meanings from word parts	multiple choice
11.05	answer implied detail questions correctly	multiple choice
11.06	use context to determine meanings of difficult words	multiple choice
11.07	find definitions from structural clues	multiple choice
11.08	determine meanings from word parts	multiple choice
11.09	find 'unstated' details	multiple choice
<i>12. The final battle (computer-based)</i>		
12.01	answer main idea questions correctly	multiple choice
12.02	find pronoun referents	click on a word in the text
12.03	use context to determine meanings of difficult words	multiple choice
12.04	answer stated detailed questions correctly	multiple choice
12.05	answer stated detailed questions correctly	multiple choice
12.06	determine where specific information is found	click on a sentence in the text
12.07	find 'unstated' details	multiple choice
12.08	use context to determine meanings of difficult words	click on a word or phrase in the text

12.09	answer implied detail questions correctly	multiple choice
12.10	determine where to insert a piece of information	click on a 'square' in the text
12.11	recognize the organisation of ideas	click on a paragraph in the text
<i>13. One of the most mysterious things (paper-based)</i>		
13.01	topic and main ideas	multiple choice
13.02	facts and details	multiple choice
13.03	purpose	multiple choice
13.04	referents	multiple choice
13.05	facts and details	multiple choice
13.06	vocabulary in context	multiple choice
13.07	vocabulary in context	multiple choice
13.08	inferences	multiple choice
13.09	organization	multiple choice
13.10	attitude	multiple choice

Figure 2.2. The TOEFL texts chosen for the corpus, the sub-skills tested by each item and the response formats (test methods) used.

## 2.3. The IELTS Sub-corpus

Texts 14, 15 and 16 and their respective tasks were chosen from the General Training Reading module of IELTS (International English Language Testing System), and text 17 from the Academic Reading module of the same testing system. In particular, the texts and the corresponding tasks were selected from the examination papers found in UCLES (2001): text 14 from UCLES (2001: 98-99), text 15 from UCLES (2001: 102-103), text 16 from UCLES (2001: 116-117), and text 17 from UCLES (2001: 83-86). IELTS is mainly “intended to fulfil English language requirements for entry to English-medium universities, for non-native speakers of English” (Alderson 2000: 130). It consists of four modules: all candidates have to take the same Listening and Speaking modules, but can choose to take either the Academic or the General Training Reading and Writing modules.



The General Training version of the test is for “candidates taking the test for entry to vocational or training programmes not at degree level, for admission to secondary schools and for immigration purposes”, while the academic version is “for candidates taking the test for entry to undergraduate or postgraduate studies or for professional reasons” (UCLES 2001: 1). Both the Academic and the General Training Reading modules consist of three sections based on three reading passages and 40 questions. The time limit for both of them is 60 minutes and their sections are claimed to be of increasing difficulty. The General Training Reading module contains texts that are

based on the type of material candidates would be expected to encounter on a daily basis in an English speaking country. They are taken from sources such as newspapers, advertisements, instruction manuals and books, and test the candidate’s ability to understand and use information. The test includes one longer text, which is descriptive rather than argumentative. (Cambridge ESOL *et al.* 2007: 3)

Cambridge ESOL *et al.* (2007: 7) also specifies that the first section of the General Training module is called ‘social survival’ and contains texts relevant to basic linguistic survival with “tasks mainly about retrieving and providing general factual information.” The second section, ‘training survival’, “involves a text or texts of more complex language with some precise or elaborated expressions”. Finally, the third section, i.e. ‘general reading’, “involves reading more extended prose with a more complex structure but with the emphasis on descriptive and instructive rather than argumentative texts” (Cambridge ESOL *et al.* 2007: 7).

By contrast, in the IELTS Academic Reading module

texts are taken from magazines, journals and newspapers, all written for a non-specialist audience. At least one of the texts contains a detailed logical argument. One text may contain non-verbal materials such as diagrams, graphs or illustrations. If texts contain technical terms then a simple glossary is provided. (Cambridge ESOL *et al.* 2007: 7)

Alderson (2000: 131) makes reference to de Witt’s (1997) account of the IELTS construct and says that “the major skills students need to know, and the things they need to do are: ‘(Know) how to understand main ideas and how to find specific information; (Do) survey the text;

analyse the questions; go back to the text to find answers; check your answers.”” Figure 2.3 gives information about the four texts chosen: the version of the Reading module and the section they belong to, the number of items accompanying each input text and the test method(s) used to elicit the candidates’ responses.

<i>IELTS</i>	<i>Version; Section</i>	<i>N. of items; Test methods</i>
<i>14. Use of University Ground</i>		
	General Training Reading; Section 1	5; true/false/not given
<i>15. West Thames College</i>		
	General Training Reading; Section 2	7; true/false/not given
<i>16. Central Library</i>		
	General Training Reading; Section 2	7; true/false/not given
<i>17. Green Wave</i>		
	Academic Reading; Section 1	13 questions;
	17.01-17.06	yes/no/not given
	17.07-17.09	multiple choice
	17.10-17.13	summary completion by choosing words from a list

Figure 2.3. The IELTS texts chosen for the corpus, the version of the Reading module and the section they belong to, the number of items accompanying each text, and the test methods used.

## 2.4 The PET Sub-corpus

Texts 18, 19, 20, and 21 and their tasks are all representative of parts 3 and 4 of the Reading Paper of University of Cambridge ESOL Examinations’ Preliminary English Test (PET). PET is a test at level B1 of the CEF (Council of Europe 2001), which is defined as the level at which “learners should be able to cope linguistically in a range of everyday situations which require a largely predictable use of language”. Furthermore, language users at level B1 “will be able to use English in their own or a foreign country in contact with native and non-native speakers of English for general purposes” (Cambridge

ESOL 2008: 3). The test consists of Paper 1, i.e. Reading/Writing, Paper 2, i.e. Listening, and Paper 3, i.e. Speaking. The timing for Paper 1 is 1 hour 30 minutes, and its Reading component is made up of 5 parts, “which test a range of reading skills with a variety of texts, ranging from very short notices to longer continuous texts” (Cambridge ESOL 2008: i). In part 1 candidates have to prove they can read and understand various kinds of short texts, and think about the situations in which they would be used in real life. The texts to be read in this part are “authentic notices and signs, packaging information (for example, instructions on a food package or a label on a medicine bottle), and communicative messages (notes, e-mails, cards and postcards)” (Cambridge ESOL 2008: 11). In part 2 test-takers are presented with five short texts which they have to match to five out of eight ‘adapted-authentic’ factual texts on a particular topic. Part 3 tests the ability to read longer ‘adapted-authentic’ factual texts and to find specific information in them.

Frequently these texts take the form of brochure extracts, advertisements in magazines and website information. There are ten questions, which are single-sentence statements about the text. The task is made more authentic by putting these questions before the text, in order to encourage candidates to read them first and then scan the text to find each answer. The information given in the text follows the same order as the content of the questions. (Cambridge ESOL 2008: 11)

In part 4 candidates have to read an ‘adapted-authentic’ long text,

which goes beyond the provision of factual information, and expresses an opinion or attitude. There are five multiple-choice questions with four options [...] . In answering these questions, candidates will demonstrate whether they have understood the writer’s purpose, the writer’s attitude and opinion, or an opinion quoted by the writer, and both the detailed and global meaning of the text. This part requires candidates to read the text very carefully indeed. [...] It may be more practical for candidates to consider the first and last questions together, in that the first focuses on writer purpose and the last on global meaning. The other three questions follow the order of information given in the text and one of the three will focus on attitude or opinion. (Cambridge ESOL 2008: 11)

Finally, in part 5 test participants are presented with an “adaptive-authentic text drawn from a variety of sources [...] of a factual or

narrative nature” (Cambridge ESOL 2008: 10). The text in this part contains ten numbered spaces that test-takers have to fill in. In order to do so, they have to answer a four-option multiple choice question for each space. Texts 18, 20 and 21 were chosen from those available in Cambridge ESOL (2008a: 16-17, 25), a collection of Cambridge ESOL’s Preliminary English Test (PET) sample papers, while text 19 was chosen from the preparation papers for PET in UCLES (2001b: 10-11). Information about these texts and accompanying tasks is summarised in Figure 2.4.

<i>PET</i>	<i>Part</i>	<i>Sub-skills tested</i>	<i>N. of items; Test methods</i>
<i>18. Exploring the Arctic AND 19. Globewise</i>			
	3	processing a factual text; scanning for specific information while disregarding redundant material	10; true/false
<i>20. Ainsley Harriot AND 21. Doug Allan</i>			
	4	reading for detailed comprehension; understanding attitude, opinion and writer purpose; reading for gist, inference and global meaning	5; four-option multiple choice

Figure 2.4. The PET texts chosen for the corpus, the Part of the Reading paper they belong to, the (sub-)skills tested by each item, the number of items accompanying each text, and the test method(s) used.

## 2.5 The FCE Sub-corpus

Both tests 22 and 23 were obtained from UCLES (2000: 4-5, 6-7), a collection of past Reading papers from University of Cambridge ESOL Examinations’ First Certificate in English (FCE). Recently, both FCE and CAE papers and specifications have undergone an updating process and the changes and improvements made will be introduced in December 2008<sup>12</sup>.

---

12 For more information on this review project see *Research Note 30* at <[http://www.cambridgeesol.org/rs\\_notes/](http://www.cambridgeesol.org/rs_notes/)>, last visited on 25<sup>th</sup> April 2008.

University of Cambridge ESOL Examinations' First Certificate in English (FCE) is claimed to be a test at level B2 of the CEF (Council of Europe, 2001). At this level "a learner should be able to handle the main structures of the language with some confidence, demonstrate knowledge of a wide range of vocabulary and use appropriate communicative strategies in a variety of social situations" (Cambridge ESOL 2008b: 3). The current test consists of five papers: Reading, Writing, Use of English, Listening and Speaking. The Reading paper takes 1 hour 15 minutes and consists of 4 parts and 35 questions. The input texts are chosen from "newspaper and magazine articles, reports, fiction, advertisements, correspondence, messages, informational material (e.g. brochures, guides, manuals, etc.)" and vary in length from 350 to 700 words (Cambridge ESOL 2008b: 6). In part 1 of the Reading paper candidates have to prove their ability to "identify the main points in a text at paragraph level. One of two different tasks may appear on the paper: headings or summary sentences" (Cambridge ESOL 2008b: 7). The tasks candidates are asked to perform are concerned with reading a list of especially written headings and decide which heading or summary sentence best fits each paragraph. "The headings or summary sentences are printed in a list before the text to encourage candidates to form an impression of the main points they are looking for before they start reading a text" (Cambridge ESOL 2008b: 7). In part 2 what is tested is the candidates' detailed understanding of a text, including the opinions and attitudes expressed in it. The text and each question about them need to be read carefully,

in order to distinguish between apparently similar viewpoints, outcomes and reasons. [...] The multiple-choice questions appear after the text. They are presented in the same order as the information in the text so that candidates can follow the development of the text as they work through the questions. The final question may require candidates to interpret an aspect of the whole text, e.g. the writer's purpose, attitude or opinion. (Cambridge ESOL 2008b: 7)

In part 3 the "candidates' understanding of how texts are structured and their ability to follow the development of the text" are at issue (Cambridge ESOL 2008b: 7). The task requires the selection from a number of options of the correct extract (either a sentence or a paragraph) to fit in each of six or seven gaps in a text. The last part of