

Centre for English Language Studies
University of Birmingham

Masters in Teaching English as a Foreign or Second Language

The Validity and Reliability of the Cambridge

First Certificate in English

By: Deborah Grossmann

September 2010

Question:

Describe an English language test with which you are familiar and discuss how *valid* and *reliable* the test appears to be. (If possible, include illustrative examples from the test itself.) Describe any procedures you would use to establish its validity and reliability. (You should not carry out these procedures unless they are quick and simple to complete.)

TABLE OF CONTENTS

1	INTRODUCTION.....	4
2	THE FIRST CERTIFICATE IN ENGLISH.....	5
3	LITERATURE REVIEW.....	7
3.1	TYPES OF TEST	7
3.2	TYPES OF TESTING	8
3.3	THE TEST DEVELOPMENT PROCESS.....	9
4	VALIDITY AND RELIABILITY OF THE FCE TEST.....	14
4.1	CONSTRUCT VALIDITY	14
4.1.1	<i>Reading (Section 8.1, pp. 38-32)</i>	<i>17</i>
4.1.2	<i>Writing (Section 8.2, pp. 43-44)</i>	<i>20</i>
4.1.3	<i>Use of English (Section 8.3, pp. 45-50)</i>	<i>21</i>
4.1.4	<i>Listening (Section 8.4, pp. 51-56)</i>	<i>21</i>
4.1.5	<i>Speaking (Section 8.5, pp. 57-62).....</i>	<i>23</i>
4.1.6	<i>Evaluation of Construct Validity.....</i>	<i>24</i>
4.2	CONTEXT VALIDITY	25
4.2.1	<i>Demands</i>	<i>25</i>
4.2.2	<i>Setting</i>	<i>26</i>
4.3	SCORING VALIDITY	26
4.3.1	<i>Level</i>	<i>26</i>
4.3.2	<i>Reliability.....</i>	<i>27</i>
5	PROCEDURES TO ESTABLISH VALIDITY.....	30
5.1	RESEARCHING THE EFFECT KNOWLEDGE OF TOPIC ON PERFORMANCE	31
5.2	RESEARCH INTO THE CORRELATION BETWEEN PAPERS	31

5.3	RESEARCH COMPARING FCE AND ANOTHER EXAMINATION.....	32
6	CONCLUSION	32
7	BIBLIOGRAPHY	33
8	APPENDIX ONE	36
8.1	SAMPLE READING PAPER	36
8.2	SAMPLE WRITING PAPER	41
8.3	SAMPLE USE OF ENGLISH PAPER	43
8.4	SAMPLE LISTENING PAPER	49
8.5	SAMPLE SPEAKING PAPER	55

1 Introduction

According to Hughes (2003, p. 1), many ELT professionals “harbour a deep mistrust of tests and testers” and, he claims, this view is often justifiable due to the poor quality of many language tests. Furthermore, Bachman and Palmer (1996, p. 6) suggest that “there is no such thing as the one ‘best’ test, even for a specific situation”. Nevertheless, a cursory glance through ELT publishers’ catalogues shows that “high stakes” language tests are big business. The “high stakes” label refers to the fact that candidates’ futures are often determined by achieving a satisfactory grade in one or other of the tests. Examples include IELTS examinations, frequently used as an entry requirement for UK universities, or the TOEFL test, often required to enter tertiary education in the US. Given these tests can affect the future of candidates around the globe, it is vital that providers develop examinations where candidates can perform to the best of their ability. Similarly, the education departments, employers or universities using these tests to measure the language ability of individuals are reliant on the results truly reflecting how language is used in context. For these stakeholders, therefore, consistent accuracy and valid measurement is paramount. Therefore, a high-stakes test-developer’s aim should be to create tests that clearly state their purpose and offer results that can be empirically proven to be reliable; in other words, tests that are valid and reliable.

This paper considers Cambridge ESOL’s First Certificate in English (FCE), which, at B2 level, is formally recognised by the Department of Education in Switzerland and can be taken as part of the “Berufsmatura” qualification (equivalent of A level) (Randall, 2010, p. 3). Therefore, in a Swiss context, the FCE constitutes a high-stakes test, as it

determines whether candidates can attend a Swiss university. In this paper, I will describe the FCE test, review the existing literature to outline the theories underpinning foreign language testing, and offer an analysis of the apparent reliability and validity of the FCE test. Finally, I will consider research which could be undertaken in the classroom to test both the reliability and validity of the FCE examination.

2 The First Certificate in English

The Cambridge ESOL's FCE was introduced in 1939 as the Lower Certificate in English. It has since undergone 4 major reviews the most recent in 2008. According to Hawkey (2009, p. 7), these modifications were made to reflect the changes in language testing "with the continuing ascendancy of communicative approaches to language teaching". A sample of the 2008 version is given in Appendix One.

The FCE consists of five papers, Reading, Writing, Use of English, Listening and Speaking, (table 2.1) with each paper contributing 20% to the total mark. Two of the papers (writing and speaking) are marked subjectively, using both global and specific criterion-referenced rating scales, while the remaining papers are objectively marked.

Paper	Timing	Content
1: Reading	1 hour	Part 1: A text followed by 8 multiple-choice questions. Part 2: A text from which seven sentences have been removed and placed in a jumbled order, together with an additional sentence, after the text. Part 3: A text or several short texts preceded by 15 multiple-matching questions.
2: Writing	1 hour 20 minutes	Part 1: One compulsory question. Part 2: Candidates choose one task from a choice of five questions (including a set text option).
3: Use of English	45 minutes	Part 1: A modified cloze test containing 12 gaps and followed by 12 multiple-choice items.

		<p>Part 2: A modified open cloze test containing 12 gaps.</p> <p>Part 3: A text containing 10 gaps. Each gap corresponds to a word. The stems of the missing words are given beside the text and must be changed to form the missing word.</p> <p>Part 4: Eight separate questions, each with a lead-in sentence and a gapped second sentence to be completed in two to five words, one of which is a given 'key word'.</p>
4: Listening	45minutes	<p>Part 1: A series of eight, short unrelated extracts from monologues or exchanges between interacting speakers. There is one multiple-choice question per extract.</p> <p>Part 2: A monologue or text involving interacting speakers, with a sentence completion task with 10 questions.</p> <p>Part 3: Five short related monologues, with five multiple matching questions.</p> <p>Part 4: A monologue or text involving interacting speakers, with seven multiple-choice questions.</p>
5: Speaking	14 minutes	<p>Part 1: A conversation between the interlocutor and each candidate (spoken questions).</p> <p>Part 2: An individual 'long turn' for each candidate, with a brief response from the second candidate (visual and written stimuli with spoken instructions).</p> <p>Part 3: A two-way conversation between the candidates (visual and written stimuli with spoken instructions).</p> <p>Part 4: A discussion on topics related to part 3 (spoken questions).</p>

Table 2.1: FCE Content (Cambridge, 2008a, p. inside front cover)

According to Cambridge ESOL (2008a, pp. 3-4), the examination is set at the Association of Language Testers in Europe (ALTE) level 3, and level B2 of the Council of Europe's Common European Framework of Reference for Languages (CEFR). Therefore, as the handbook states (2008a, p. 4),

[a]t 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. Their understanding of spoken language and written texts should go beyond being able to pick out items of factual information, and they should be able to distinguish between main and subsidiary points and between the gist of a text

and specific detail. They should be able to produce written texts of various types, showing the ability to develop an argument as well as describe or recount events.

Having outlined the current format and level of the FCE, I will now address the theoretical basis of language testing in an attempt to provide a framework for validity and reliability.

3 Literature Review

This section outlines types of test, types of testing and the test development process and identifies how these theories relate to the FCE.

3.1 Types of Test

Tests can be categorised by the type of information they provide. Hughes (2003) identifies four types of tests: proficiency, achievement, diagnostic and placement.

Figure 3.1 provides explanations of these terms.

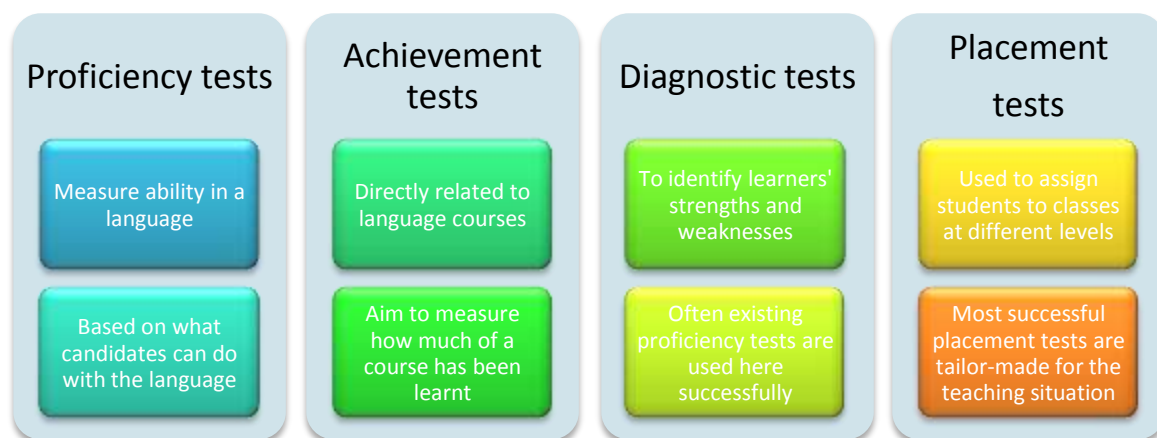


Figure 3.1: Types of tests (adapted from Hughes, 2003, pp. 11-17)

Bachman (1990, p. 114) argues that tests such as the FCE are achievement tests as the majority of candidates complete a preparation course. In contrast, Hughes (2003, p. 12)

points out that the FCE is a general type of proficiency test. This corresponds with Cambridge's (2008a, p. 3) view that the test measures "overall communicative ability".

3.2 Types of testing

Within each type of test, examining boards choose between a variety of testing elements, combining them to meet particular testing needs. The possibilities are shown in figure 3.2.

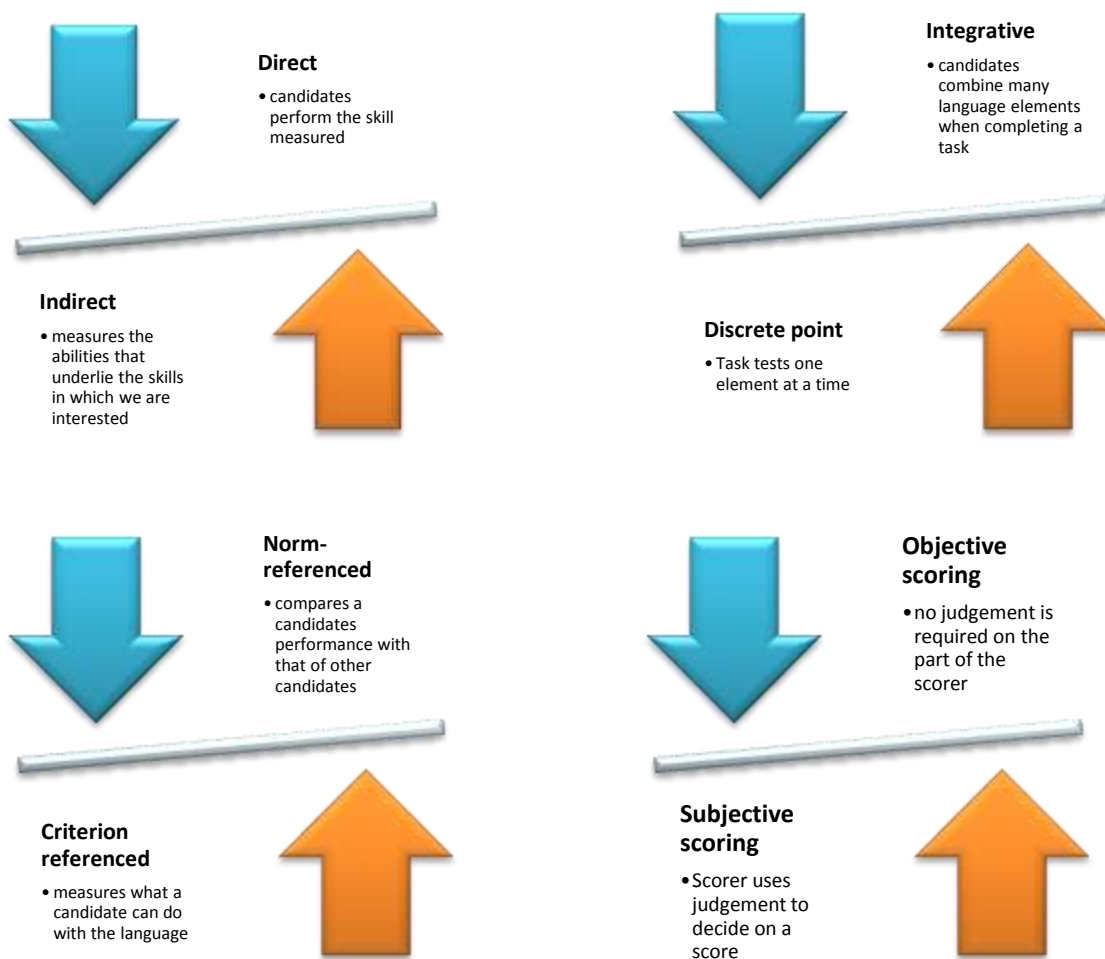


Figure 3.2 Types of testing methods (adapted from Hughes, 2003, pp. 17-22)

The FCE examination uses both objective (papers 1, 3 and 4) and subjective scoring (papers 2 and 5) and is criterion referenced, measuring what candidates can do with the

language. In addition, the FCE combines both direct and indirect testing methods, which are mainly integrative; however, some elements of the grammar paper could, in my opinion, be considered discrete point, particularly part 3, as it tests knowledge of word building. Having outlined the elements that constitute a test, I will now consider the test development process.

3.3 The Test Development Process

Test development, according to Bachman and Palmer (1996, p. 85) incorporates the “entire process of creating and using a test, beginning with its initial conceptualization and design, and culminating in one or more archived tests and the results of their use”. They provide a conceptual framework for the development process which is organised into three stages (see figure 3.3). Accordingly, test usefulness is the most important quality of any test and should be taken into account at all stages of the development process; usefulness includes reliability, construct validity, authenticity, interactiveness, impact and practicality, which test developers should balance appropriately to optimise the usefulness of their test (Bachman & Palmer, 1996, pp. 17-18).

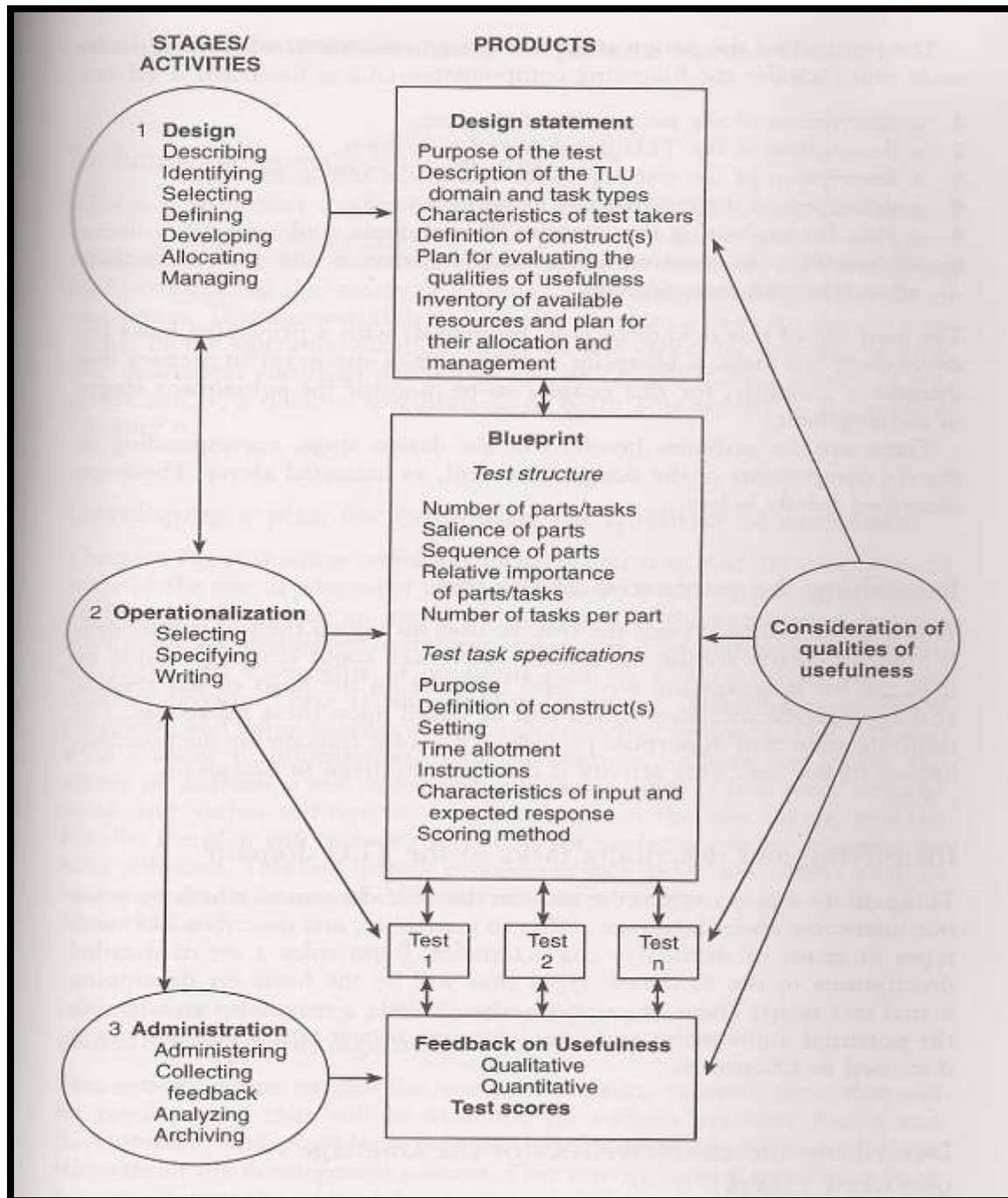


Figure 3.3 Stages of test development (Bachman & Palmer, 1996, p. 87)

Although the Bachman and Palmer model influenced Cambridge ESOL's current framework, it did not reflect the socio-cognitive view of test validation, central to Cambridge's approach, and therefore, in Taylor's view (2006 cited in Hawkey 2009

p.173), Bachman and Palmer’s framework “ha[s] not proved as useful for practical test design and operation as [it] perhaps once promised”. Consequently, Cambridge ESOL have provided an alternative model of test development (Figure 3.4) as well as a framework for the 2008 FCE review project (Figure 3.5).

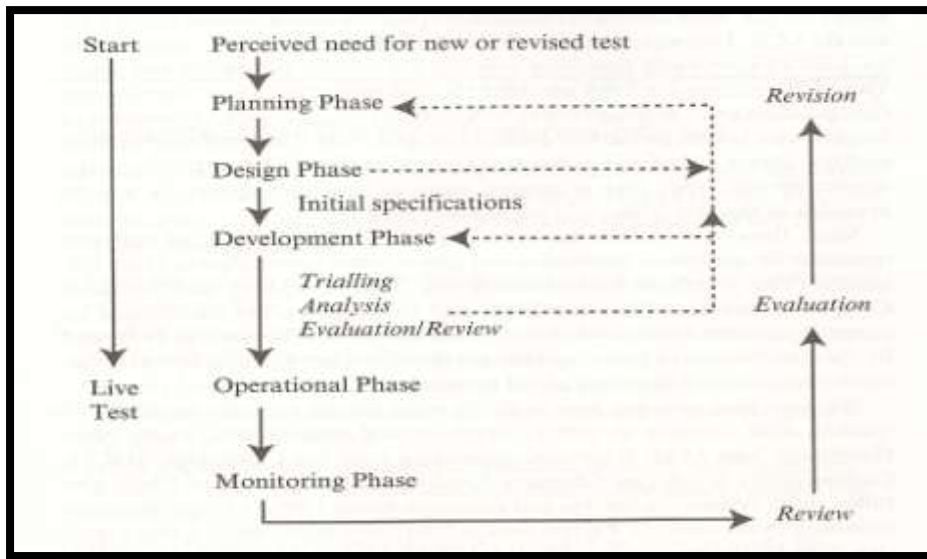


Figure 3.4 Cambridge ESOL model of the test development process (Hawkey, 2009, p. 138)

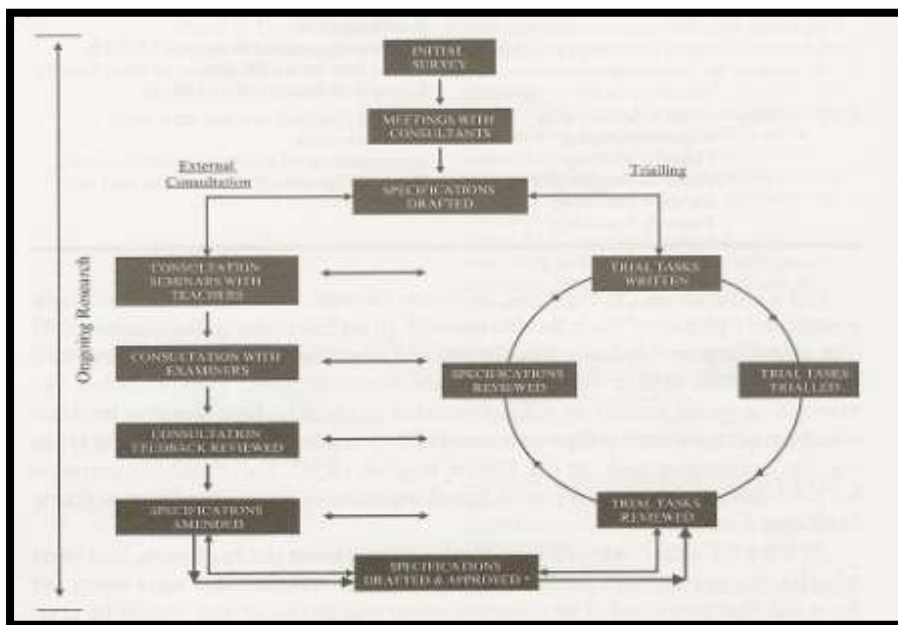


Figure 3.5 Model of the FCE review project (Hawkey, 2009, p. 142)

In addition, Cambridge ESOL have provided a basis for assessing usefulness consisting of four key examination qualities, validation, reliability, impact and practicality (Hawkey, 2009, p. 127). This process of validation is operationalised using Weir's (2005) cited in Hawkey p.173) socio-cognitive approach to test validation. It views any testing activity as "a triangular relationship between three critical components: the test taker's cognitive abilities, the task and context, and the scoring process". Weir's framework sees construct validity as consisting of three symbiotic elements: cognitive, context and scoring validity. By separating context validity from scoring and cognitive validity the framework allows for adjustments to be made depending on the skill being tested. An example of Weir's framework is shown in Figure 3.6. The processes illustrated in Figures 3.4 – 3.6 show the comprehensive nature of the systems used in the validation of the FCE, which, in my opinion, adds to the test's validity.

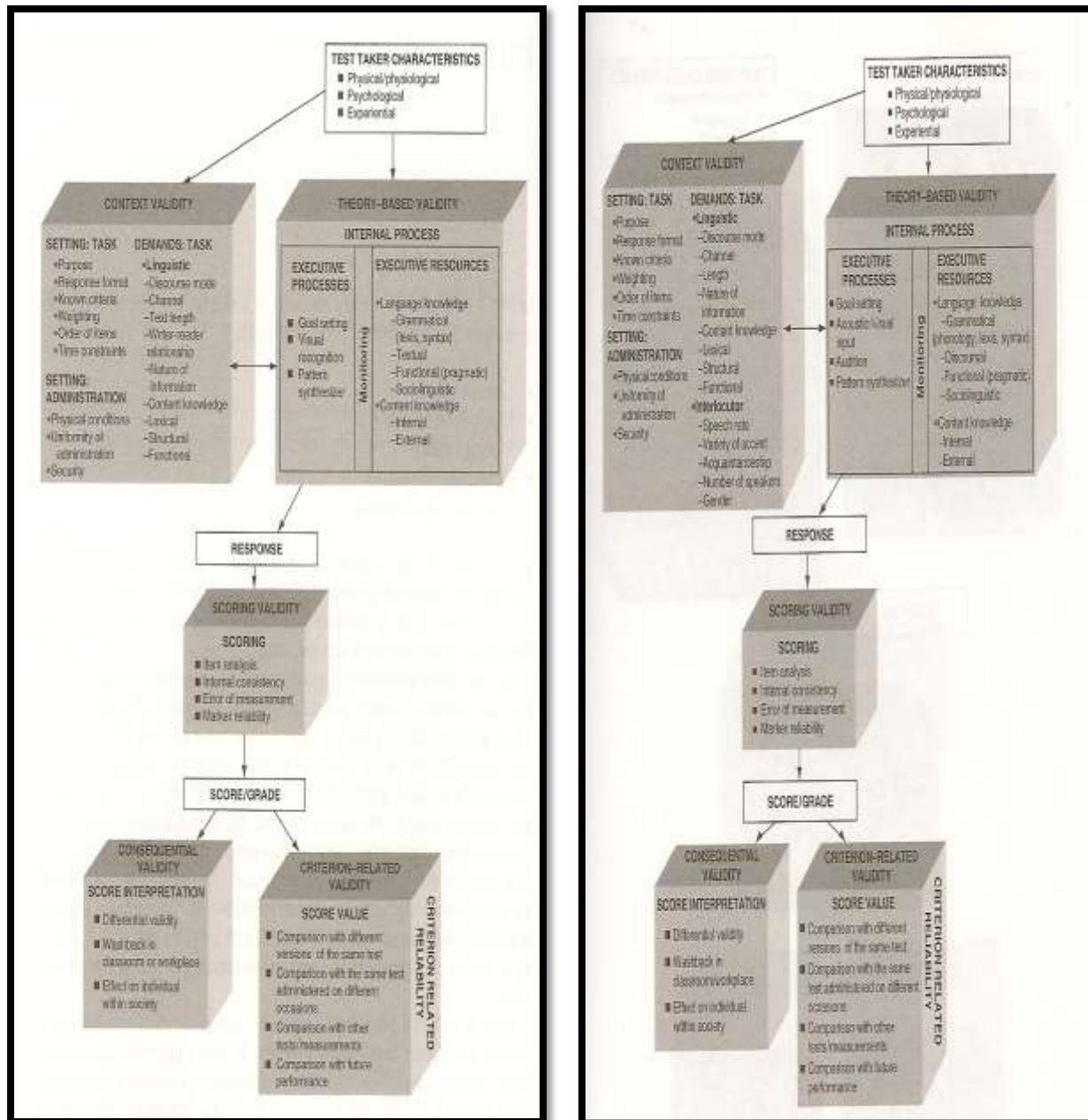


Figure 3.6: A socio-cognitive framework for a reading (left) and a listening (right) test (Weir C. J., 2005)

Having considered how validity can be evaluated in theoretical terms, the next section will discuss how valid the FCE appears to be in reality.

4 Validity and reliability of the FCE test

Given the processes discussed in section 3.3 and the amount of research undertaken as part of the FCE review project, Cambridge ESOL appears to provide a great deal of evidence of the FCE's validity and reliability. However, I will look at varying aspects of validity and consider each in more detail to determine whether the wealth of research does in fact prove validity. Due to the constraints of this assignment it is not possible to consider all elements of validity for all of the FCE papers, so in this section I will discuss the most relevant elements and relate them to the varying papers, so that conclusion can be drawn regarding the validity of the FCE examination as a whole.

4.1 Construct Validity

The first thing to consider when addressing the question of validity is the construct, that is what the test intends to measure. A test that demonstrates construct validity can provide evidence that the scores obtained accurately reflect the construct being measured. According to Hawkey (2009, p. 171) the reference to construct was "implicit" prior to the 1996 review. Research by Bachman et al (1995) into the comparability of the TOEFL and the FCE examinations led to a call for a clearer and more explicit focus on examination constructs and construct validation, which were adopted for the 1996 version of the test. The 1996 constructs are detailed in Table 4.1.

FCE 1996 Paper	Constructs
Reading	Understanding of propositional, functional and sociolinguistic meanings at word, phrase, sentence or discourse levels and of reading outcomes relevant to FCE takers (i.e gist, specific information, detail, main idea, deduced information)
Writing	Ability to demonstrate range of vocabulary and structure; accuracy of vocabulary, structure, spelling and punctuation; appropriacy; organisation and cohesion; task achievement through text types

	relevant to FCE test takers i.e. transactional and personal letters, articles, reports, compositions
Use of English	Lexico-grammatical competence, including components of meaning, word formation, collocations, lexical relationships, lexical cohesions, modality, complementation, phrase structuring, information focus; morphology, phrase structure, clause structure, clause combining, grammatical cohesion.
Listening	Understanding of propositional, functional and sociolinguistic meanings at word, phrase, sentence or discourse levels and of understanding/outcomes (i.e gist, specific information, detail, main idea, deduced information).
Speaking	Ability to demonstrate use of grammar, vocabulary, pronunciation, interactive communication, task achievement in those interaction routines relevant to FCE target users and most amenable to implementation in an examination context i.e. social encounters, social interaction, informal discussion, informal planning and decision making.

Table 4.1: Test constructs of the 1996 FCE examination (Hawkey, 2009, p. 82)

However, in the 2008 handbook and specifications there is no mention of the 1996 constructs. Instead, it states that the FCE examination measures “overall communicative ability”, which for practical purposes is subdivided into the four main skills of reading, writing, listening and speaking (Cambridge, 2008a, p. 3). The 2008 specifications provide more explicit detail for each paper.

Paper	Test focus
1: Reading	<p>General: Candidates are expected to show understanding of specific information, text organisation features, tone, and text structure.</p> <p>Part 1: detail, opinion, gist attitude, deducing meaning, text organisation features (exemplification, comparison, reference), tone, purpose, main idea.</p> <p>Part 2: text structure, cohesion, coherence.</p> <p>Part 3: specific information, detail, opinion, attitude.</p>
2: Writing	<p>General: Candidates are expected to be able to write non-specialised text types such as article, essay, letter, email, report, review, short story, with focus on advising, apologising, comparing, describing, explaining, expressing opinions, justifying, persuading,</p>

	<p>recommending and suggesting.</p> <p>Part 1: focus on expressing opinions, justifying, persuading, comparing, recommending, suggesting, advising, apologising, describing and explaining.</p> <p>Part 2: varying focuses according to task, including: expressing opinions, justifying, comparing, recommending, advising, describing and explaining.</p>
3: Use of English	<p>General: Candidates are expected to demonstrate the ability to apply their knowledge of the language system by completing a number of tasks.</p> <p>Part 1: lexical/lexico-grammatical</p> <p>Part 2: grammatical/lexico-grammatical</p> <p>Part 3: lexical/lexico-grammatical</p> <p>Part 4: lexical and grammatical</p>
4: Listening	<p>General: Candidates are expected to be able to show understanding of attitude, detail, function, genre, gist, main idea, opinion, place, purpose, situation, specific information, relationship, topic and agreement.</p> <p>Part 1: general gist, detail, function, purpose, attitude, opinion relationship, topic, place, situation, genre, agreement.</p> <p>Part 2: detail, specific information, stated opinion.</p> <p>Part 3: same as for Part 1</p> <p>Part 4: opinion, attitude, gist, main idea, specific information.</p>
5: Speaking	<p>General: Exchanging personal and factual information; expressing and finding out about attitudes and opinions.</p> <p>Part 1: general interactional and social language</p> <p>Part 2: organising a larger unit of discourse, comparing, describing, and expressing opinions.</p> <p>Part 3: sustaining an interaction; exchanging ideas, expressing and justifying opinions, agreeing and/or disagreeing, suggesting, speculating, evaluating, reaching a decision through negotiation, etc.</p> <p>Part 4: expressing and justifying opinions, agreeing and or disagreeing.</p>

Table 4.2: FCE Test Focus (adapted from (Hawkey, 2009, pp. 222-223) and (Cambridge, 2008, p. 4)

As Weir and Shaw (2006, p. 9) point out “adequate construct definition for purposes of test validation is a vital principle in language testing”. In my opinion, given that the test focus (table 4.2) details much of the information shown in the test construct document

(table 4.1) and, in fact, in greater detail, taking away the reference to constructs provides critics with ammunition to discredit the examination. It may, therefore, have been advisable to maintain the use of the word constructs, as this would have ensured transparency. Having identified the construct for each of the papers of the 2008 examinations, I will now consider how the papers shown in Appendix 1 reflect the constructs for each test.

4.1.1 Reading (Section 8.1, pp. 38-32)

Table 4.3 itemises the part one reading questions and answers, as well as the sentence in the text that allows the candidates to identify the correct answer. These first two columns have provided the basis for my assessment of the construct. It should be noted at this stage that identifying the focus of each questions was, at times, extremely difficult as it was hard to differentiate between some elements identified in the construct.

Part one question and answer	Sentence from text	Construct
1. As he travelled the writer regretted his choice of ... a) seat	I had taken a seat on the wrong side where the summer sun beat on the window.	testing ability to find detailed information from text
2. What had surprised the writer about the job? d) He had been selected for an interview	It hadn't seemed possible when a letter came	testing ability to find detailed information from text
3. The writer uses the phrase 'I had grabbed the lifeline' to show that he felt b) ready to consider any offer	Not applicable	testing ability to deduce meaning from context.
4. What impression had the writer previously had of Yorkshire? b) It was a boring place	I was prepared for a place of solid respectability and dullness	testing ability to find detailed information from text
5. What did the writer find unusual about Darrowby? d) the lack of activity	there was not another sound or movement anywhere	testing ability to find detailed information from text
6. What did the writer feel that the guidebooks had missed about Darrowby? c) the lovely views from the town	Everywhere from the windows of houses in Darrowby you could see the hills	testing ability to find detailed information from text

7. How did the writer recognise Skeldale House? c) There was a certain plant outside	I knew by the ivy that grew untidily over the red brick	testing ability to find detailed information from text
8. How did the writer's attitude change during the passage? a) He began to feel he might like living in Darrowby	The pressure of the city ... seemed to be falling away from me and this could be the place I would work as a veterinary surgeon	testing ability to determine gist attitude from the text

Table 4.3 Construct represented by Reading paper part one questions

As can be seen in Table 4.3, part one of the Reading paper mainly requires the candidates to consider details with just question 3 testing deducing meaning and question 8 gist attitude. This could be considered an inadequate representation of the construct as part one should cover nine functions, rather than the three tested.

Part two, on the other hand, appears to clearly test the construct i.e. the candidates understanding of cohesion, coherence and text structure, as, in my view, candidates would experience problems completing the task successfully without an understanding of how the main text and the missing sentences link together (see table 4.4.).

Question	Part of main text providing link to answer	Answer
9	It wasn't an amazing success	F) I finished last , but it didn't matter as I enjoyed it.
10	In fact, there's quite a lot of putting up tents in muddy fields	H) It's not all stardom and glamour , though.
11	Both events were completely different from the UK race scene	C) the courses were twice as long and the crowds were twice as big .
12	You quickly learn how to do it so as not to injure yourself .	A) I've fallen off more times than I care to remember.
13	downhill racing wasn't taken seriously as a mountain biking discipline	E) The attitude was: how much skill do you need to sit on a saddle and point a bike in the same direction for a few minutes?
14	Your legs hurt so much they burn ... but, in a race you switch off to the pain until you have finished .	B) I usually have to stop during practice .
15	people think you need to spend	G) Nothing could be further from

	thousands of pounds a reasonable beginner's downhill bike will cost you around £400	the truth.
--	---	------------

Table 4.4 Links in main text helping candidates identify correct answer in part two Reading paper

Table 4.5 shows the sentences in the text that assist candidates in choosing the correct answers for part three of the Reading paper, and from this I have concluded the element of the construct represented by each question. As can be seen, part three seems to test specific information and detail exclusively without, in my view, any reference to opinion and attitude.

Question and Answer	Sentence from text	Construct
16. had to restart their collection? B	He had to sell his valuable collection He took up the interest again	testing ability to find specific information from text
17. has provided useful advice on their subject? C	Her book gives simple and safe home tests for identification	testing ability to find specific information from text
18. was misled by an early success? A	a dealer came and bought everything I'd brought along. I thought 'Great! This is my future life'. But after that I never sold another one.	testing ability to find specific information from text
19. received an unexpected gift? D	Later, to her astonishment, he went round to her flat and presented them to her.	testing ability to find detailed information from text
20. admits to making little practical use of the collection? A	Apart from making sure they work, he rarely touches them.	testing ability to find specific information from text
21. regrets the rapid disappearance of certain items? C	because it is vital to keep examples	testing ability to find specific information from text
22. is aware that a fuller collection of items exists elsewhere? B	Whilst acknowledging that the Royal Camera Collection in Bath is probably more extensive than his own	testing ability to find detailed information from text
23. has a history of collecting different items? D	already had twenty years of collecting one thing or another	testing ability to find specific information from text
24. performed a favour for	so she asked Barton to	testing ability to find

someone they knew? A	look at it for her	specific information from text
25. is a national expert on their subject? C	one of the foremost authorities on plastics in Britain	testing ability to find specific information from text
26. is aware that they form part of a growing group? C	the band of collectors is constantly expanding	testing ability to find specific information from text
27. insists on purchasing top quality items? D	she will only buy a fan if it is in excellent condition	testing ability to find specific information from text
28. noticed items while looking for something else? A	he was searching for bits of second-hand furniture and kept seeing beautiful old sewing machines	testing ability to find detailed information from text
29. has to protect their collection from damage? D	the fans are on show but are kept behind glass	testing ability to find specific information from text
30. would like to create a hands-on display of their collection? B	hopes to open his own photographic museum where members of the public will be able to touch and fiddle around with the cameras	testing ability to find detailed information from text

Table 4.5 Construct represented by Reading paper part three questions

4.1.2 Writing (Section 8.2, pp. 43-44)

In part one of the Writing paper, candidates write an informal letter which recommends, suggests, advises, expresses an opinion and gives an explanation. This covers 50% of the items included in the construct which, I would argue, is a reasonable expectation, given the word limit. In addition, even though many informal letters may not require the use of so many functions, when replying to a letter from a friend, it is quite likely that many functions are utilised in real life communicative situations, providing excellent construct validity. However, as there is only one question in part one, there is no opportunity for candidates to demonstrate their ability in more formal registers, which must reduce the validity of the paper. Nevertheless, constraints on practicality mean that an inevitable balance has been reached. In part two, five of the six options test

ability to write different genres, whereas, question 5a involves writing an informal letter. This could mean that candidates write only informal letters, which would not be a fair representation of their writing ability, across a wider range of genres. This, in my opinion, affects the validity of the paper. Having said that, only candidates who have elected to read the set-book, will benefit from the situation. The remaining questions do reflect the construct accurately and can be considered valid.

4.1.3 Use of English (Section 8.3, pp. 45-50)

All four parts of this paper give candidates an opportunity to demonstrate their understanding of the system of English grammar, both lexically and structurally and therefore provides a valid representation of the construct for the paper.

4.1.4 Listening (Section 8.4, pp. 51-56)

Table 4.6 details the part one Listening questions and answers, as well as the sentence in the listening script that allows candidates to correctly identify answers. This information has provided the basis for my assessment of the construct being covered. Once more, deciding which element of the construct is being tested proved extremely difficult to determine.

Question and Answer	Sentence in transcript	Construct
1. Why did he go back to college? C: He needed new skills	My new job involved managing staff, which I had no experience of	testing candidates' ability to understand detailed information from a dialogue
2. What is he? A: an inventor	protect their ideas ... people who create mechanical objects	testing ability to understand the general gist of a text
3. How does the artist feel about his work? C: He is happy to see his work destroyed	the highest compliment he can hope to be paid is to have his work sliced up and swallowed	testing ability to deduce someone's attitude of something from the text
4. Why is she talking to him?	Contact the mountain guide service... get as much	testing candidates' understanding of functions

C: to make a suggestion	information as you can ...	
5. What is the lecturer describing? A: reasons why something is changing	Whole text	testing candidates' ability to deduce the purpose of something
6. Why did he decide to become a chef? B: to develop a natural talent	I knew how to make things taste good and that's what I wanted to build on when I went to college	testing ability to understand detailed information from the dialogue
7. How does she feel when she is playing the sport? C: confident	I know exactly what I'm doing	testing ability to deduce someone's opinion of something from the text
8. How will he travel once he is across the river? B: on horseback	use horses rather than trucks for the rest of the trip	testing ability to understand detailed information from dialogue

Table 4.6 Construct represented by Listening part one tasks

As is evident from Table 4.6, the questions test a range of construct elements, making this part a good representation of the test construct. Equally, in part two the construct is well represented, although there is only one question from ten (question 15) relating to stated opinion. As is evident from table 4.7, in part three, only three of the construct elements are tested, however, as there are only five questions this is might be due to the constraints of practicality.

Question and answer	Sentence from transcript	Construct
19. mistaking someone's identity	we didn't recognise each other at all	testing ability to understand detailed information from dialogue
20. getting a particular date wrong	turning up at the airport to find your flight had left the day before	testing ability to understand detailed information from dialogue
21. ignoring someone's advice	Len had recommended that we wait ... so I suggested that we got on with it	testing candidates' understanding of functions
22. losing something important	I'd no idea where I'd put it	testing a candidates' understanding of the situation
23. failing to inform someone about something	If he'd have known earlier he could have taken me to see the sights of Paris	testing a candidates' understanding of the situation

Table 4.7 Construct represented by Listening part three tasks

Finally, table 4.8 shows the constructs covered by part 4, which, in my opinion, also offers a reasonable representation of the Listening test constructs.

Question and answer	Sentence from transcript	Construct
24. What feature of the cable car makes it particularly good for seeing wildlife in the rainforest? B: the height at which it travels	you get to see some of the most beautiful wildlife on our planet living in ... the tall rainforest trees ... many [animals] only live in the treetops	testing a candidates' ability to understand specific information in the text
25. What is the main aim of the cable car project? B: to persuade people to save the rainforest	one of the things we need to do ... is to provide education ... you focus on conservation ... and the other thing	testing candidates' ability to understand the gist of the message
26. What is the advantage of the project for local people? C: More work is available for them	This project keeps 50 people in work, far more than would be employed on this land otherwise.	testing ability to deduce someone's opinion of something from the text
27. Why was the cable car redesigned? B: to avoid cutting down too much forest	redesigning it so that it wouldn't damage the beauty of the site	testing ability to understand detailed information from dialogue
28. How does Donald react to the suggestion that he has disturbed the wildlife? A: He explained what happened in the past	the property we have was purchased from private individuals who hunted here and cut done trees commercially	testing ability to deduce someone's attitude of something from the text
29. Why is Donald sure his project is a success? A: This piece of forest has survived	If we hadn't been here, this area would have been cut down now.	testing candidates' ability to understand the gist of the message
30. Donald thinks the future survival of the rainforest will depend on B: the attitude of the people towards it	only if they can be made to appreciate it will people value the rainforest and so want to protect it	testing ability to deduce someone's attitude of something from the text

Table 4.8 Construct represented by Listening part 4

4.1.5 Speaking (Section 8.5, pp. 57-62)

The construct is well reflected in all four parts of the paper, giving candidates the opportunity to show the examiner all of the functions listed.

4.1.6 Evaluation of Construct Validity

Although not all elements of the test construct are present in this paper, the majority are represented. The constraint of limiting the number of questions for practical reasons can account for the elements of the construct not covered. It is important, however, that the elements omitted in this test should be present in other versions of the test for the construct to remain valid in the future. Research comparing the elements of the construct which have been covered in the live versions of the exam since its launch in December 2008, would provide useful confirmation of the on-going validity of the FCE.

Hawkey argues that research by Geranpayeh and Somers has provided strong evidence for the skills approach to language testing (Hawkey, 2009, p. 321). However, this research has not been published, as it is contained within a Cambridge ESOL internal report. In addition, if one examines the reference list in the *Studies in Language Testing (SILT)* volume 28, there are several references to Cambridge ESOL internal reports. One reference in particular *Taylor et al (2006) Defining the construct(s) underpinning the Cambridge ESOL Upper Main Suite (UMS) tests: a socio-cognitive perspective on overall language proficiency and the four language skills* equates with a piece of published research by Weir and Shaw (2006), entitled *Defining the constructs underpinning the Main Suite Writing Tests: a socio-cognitive perspective*. The fact that data relating to the writing papers alone was published, rather than the whole internal report could, once again, suggest that there is a lack of transparency, and that only data that supports Cambridge's approach to testing is published. This would, if true, provide strong grounds to claim that the Cambridge examinations are invalid measurement tools. An explanation for this could be that the time and effort required to transform all internal reports into published research is considered excessive, and therefore, not

undertaken. Moreover, I find it hard to believe that Cambridge ESOL would consider a policy which would threaten not only the credibility of the ESOL division but also of the University as a whole.

4.2 Context Validity

In this section, I will consider context. I believe an examining board's failure to adequately control this element can have considerable influence on validity, as it affects the candidates' ability to give their best performance. Cambridge ESOL (2009, p. 6) views a fair test as "one in which the ability being tested is the primary focus and where all irrelevant barriers to candidate performance have been removed". In other words, a fair test is one that has a high context validity. In Weir's (2005) socio-cognitive model of test validation the context validity consists of two elements: the demands and the setting. I will consider each of these in turn.

4.2.1 Demands

The handbook for teachers (Cambridge, 2008a), which can be ordered from local Exam Managers, gives full details of all the tasks in each paper, offers advice on preparing students for the examination, and provides two full sample papers with answer key. In addition, Cambridge ESOL offers on-line resources for both teachers and candidates. Furthermore, in many countries, seminars are organised to help teachers maximise the potential of their candidates. This means that if teachers support the use of the examination, which is generally the case, the candidates know what to expect in the examination, increasing their chances of performing well and context validity is achieved.

4.2.2 Setting

There are two elements of setting identified in the Weir model: task and administration. As mentioned above, not only do Cambridge ESOL provide a wealth of information about the tasks, but they also produce a report on each examination, which offers an analysis of the results and responses given by candidates. These elements together add validity to the examination.

The second element of the setting, administration, is also carefully controlled by Cambridge ESOL, through employment of staff to independently monitor test centres. This team is commissioned to undertake spot checks on examinations to ensure that the centre adheres to the regulations, including suitability of venue, security measures, distance between candidates, noise levels etc. From my experience of the examination as a teacher, and from the candidates using our local centre, this system is extremely effective in Switzerland, ensuring a valid examination.

4.3 Scoring Validity

Within scoring validity, there are two elements of significance that can affect the validity of a test. The first is the reliability of the score itself and the second the level that the score should represent. I will first consider the level.

4.3.1 Level

The FCE is set at ALTE level 3, which is equivalent to the B2 level of the CEFR. Cambridge ESOL has undertaken numerous research projects to align their examinations to the CEFR and, as Khalifa and French (2008, p. 4) point out,

Cambridge ESOL views the alignment of its tests to the CEFR as a key aspect of their validity; it views alignment as an imbedded and integrated feature of its test development and validation model. The model has an ongoing intera[c]tive cycle

from perceived need through test design, trialling and administration to post exam review.

In addition, Khalifia and French (2008, p. Abstract) document the provision by the Council of Europe of “a toolkit of resources, including a draft pilot *Manual* for relating language examinations to the CEFR and a technical reference supplement”. The manual itemises the information required to align examinations to the framework, thus providing external reliability, and recommends that examining boards collect the required information. Nevertheless, there appears to be no independent process to review the information collected. This does not mean that the information collected is invalid but, in my opinion, if an independent body were established to review and verify the evidence of alignment, transparency could be achieved.

Another system to ensure level consistency is item-banking, which was adopted in the 1990s by Cambridge ESOL (2009a). This involves collecting examination questions whose level of difficulty is known, and from this constructing a single measurement scale against which all items can be compared. This system allows all new material to be calibrated to the level of a particular examination. This is achieved by

common person linking, where a group of learners might for example take test papers at two different levels, and *common item* linking, where different tests contain some items in common. This is the basic approach used in pretesting, where each pretest is administered together with an anchor test of already calibrated material (Cambridge, 2009a).

Item banking, which is applied to all three objective papers, therefore, ensures internal reliability of the FCE examination.

4.3.2 Reliability

Another area of scoring which can create problems with validity is that of reliability.

According to Bachman (1990, p. 160),

[t]he investigation of reliability is concerned with answering the question, 'How much of an individual's test performance is due to measurement error, or to factors other than the language ability we want to measure?' and with minimizing the effect of these factors on test scores.

In other words, the more accurately a score represents a candidates' true language ability the more valid a test will be. Geranpayeh (2004, p. 22) states that composite reliability for the entire FCE test has consistently measured 0.92 since 2000. This means that the correlation between all five papers is high and reflects a reliable examination, adding to its validity. With regard to the individual papers, the average reliability figures (Cronbach's Alpha) for the period 2002-3 were Reading - 0.84; Listening – 0.85; Use of English – 0.91. Geranpayeh points out that the higher reliability figure for the Use of English paper is probably the result of the higher number of items in that paper, which directly affects the Alpha measurement. This data shows that the objective tests do indeed reflect the trait being measured, increasing the test's validity. However, I have been unable to obtain reliability figures for the 2008 version, but one would expect them to be lower for the Reading and Use of English papers, as the number of items has been reduced.

The other two papers are marked subjectively, using rating scales, and this presents its own concerns for the reliability of the examination. The first point to note in this area, as McNamara (1996, p. 127) cited in (Galaczi & ffrench, 2007, p. 29)) points out, is that inter-rater variance is "a fact of life". Cooze and Shaw (2007, p. 17) argued that the inter-rater reliability coefficients of 0.70 for prototype tasks in FCE part one Writing were encouraging. In addition, Galaczi (2005) found inter-rater reliability coefficients of between 0.79 and 0.85 for the speaking test in 2003. However, Orr (2002) found inter-rater reliability to be poor and that non-criteria elements were affecting the judgements

of oral examiners. He recommended that examiners be offered more training and that the use of assessment scales be questioned. Given the conflict of opinion, I will now examine some of the systems in place to ensure the reliability of the speaking test.

The 1996 revision of the FCE saw the introduction of the paired format for speaking, new rating scales, and a set of procedures, using the acronym RITCME (recruitment, induction, training, co-ordination, monitoring and evaluation) to maximise the accurate utilisation of the scales. This, combined with the system of team leaders (employed by the local Exam Manager) and senior team leaders (employed by Cambridge ESOL), aims to ensure the consistent and accurate performance of oral examiners (Saville & Hargreaves, 1999).

The 2008 version made few changes to the FCE speaking test itself, but did provide new rating scales, which used the CEFR as a starting point, and were developed based on four principles: positiveness, definiteness and clarity, brevity and independence (Galaczi & ffrench, 2007, p. 28). Research (Galaczi & ffrench, 2007, p. 29) was then undertaken to investigate the validity of the new rating scales and results showed “high levels of agreement between raters involved”. The changes to the wording of the band descriptors has been welcomed by oral examiners many, myself included, finding them easier to “process when marking” (Galaczi & ffrench, 2007, p. 30). However, the research also showed oral examiners tended to cluster marks around the mid-range score of 3. It was recommended that this problem be addressed through training. In addition to the supplementary training, a Professional Support Network was introduced for oral examiners. This new on-line system, which complements the existing co-ordination procedures, not only helps team leaders identify examiners who are

consistently over- or under-marking, even slightly, but it also provides examiners with opportunities to view recorded tests just before each examining session.

Furthermore, Cambridge ESOL (2010, p. 4) undertakes extensive post-test analysis of the marks given by oral examiners to identify any tendencies for error that may exist as a centre or by individual examiners. They conclude

The results of our programmes of post-exam analysis indicate that the vast majority of centres and oral examiners are well within the very high standards that we set. However, Cambridge ESOL's commitment to producing the highest quality assessment means that we will continue to scrutinise how our exams, examiners and centres perform, and continually seek new ways to improve the service we offer candidates.

From the above discussion it is possible to conclude that the FCE is currently a valid measure of language ability at B2 level. However, again, an increase in transparency would prove beneficial for all concerned.

5 Procedures to establish validity

Cambridge ESOL undertook an extensive range of research as part of the 2008 FCE review. As Barker et al (2007, p. 32) point out

[t]he research was undertaken in relation to three areas: the construct models of FCE ..., reviewing the mark schemes and assessment criteria, and investigating tasks, topics and general content within the exam...

This research covered all areas of the construct, as well as the effects shortening the examination and changing the rating scales for the Speaking test would have on validity and reliability. In addition, research was undertaken into each of the specific skill areas. Rather than detail this research base, this section will identify some procedures that could be undertaken in the classroom so that teachers using the examination could investigate areas that might improve candidates' performance.

5.1 Researching the effect knowledge of topic on performance

Part two of the Use of English paper is an open cloze text, which causes a great deal of problems, especially for weaker candidates. In my opinion, it would be interesting to see the effect that knowledge of a particular topic might have on candidates' ability to successfully complete this part. This could be achieved by selecting two tests on different topics, preferably ones that candidates are unlikely to know much about. Before administering the tests under examination conditions, a third of the class should be asked to find out as much information as possible, in English, about the topic of one of the tests, a third should do the same for the other topic, and a third should not be asked to find out about either topic. This latter group would act as a control to ensure that the two tests are equivalent. When the test has been administered and scored, the control group, who have not investigated either topic, should demonstrate similar rank orders and scores. If this is the case, then the two tests can be deemed equivalent. The rank order and scores of the two groups with knowledge of one of the topics should also be determined and compared. If the scores for the test with prior knowledge of the topic are higher than those without, then it would be possible to conclude that knowledge of topic can improve a candidate's performance.

5.2 Research into the correlation between papers

In addition, it would be interesting to see if any two of the Reading, Use of English or Listening papers are equivalent. This would involve administering two tests and comparing the scores achieved. If the two tests were equivalent, one would expect to see that the rank order and actual scores were similar. If the rank order were similar but the actual scores were consistently higher or lower one could argue that one test was easier or more difficult than the other.

5.3 Research comparing FCE and another examination

A further possibility would be to compare the results of the FCE test to another test. One interesting comparison for Switzerland would be to compare the rank order of FCE results against the rank order of “Berufsmatura” results as well as the correlation coefficients of the two examinations. Having this information would allow students and future employers to firstly see if the examinations measure the same thing and also to see whether one test is more difficult than the other.

6 Conclusion

In this paper I have considered, the First Certificate in English, which in the Swiss context, is used as an entrance requirement for tertiary education, earning it the high stakes label. I have reviewed the systems in place as part of the test’s validation process, suggested additional means for testing validity and reliability, and found that the FCE currently seems to be a valid and reliable test. I have also argued, however, that an independent review of the evidence aligning the examination to the CEFR and the publication of all internal reports detailing research would be beneficial as it would silence critics, like Hughes (2003, p. 1) who “harbour a deep mistrust of tests and testers”.

7 Bibliography

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Cambridge: CUP.

Bachman, L., & Palmer, A. (1996). *Language Testing in Practice*. Cambridge: CUP.

Bachman, L., Davidson, F., Ryan, K., et al (1995). *An investigation into the comparability of two tests of English as a Foreign Language*. Cambridge: CUP.

Barker, F., McKenna, S., Murray, S., et al (2007, November). Overview of the FCE and CAE Review Project research activity. *Cambridge ESOL Research Notes 30* , pp. 31-34.

Cambridge, ESOL (2009, October 30). Cambridge ESOL and fairness. *Cambridge First* , p. 6.

Cambridge, ESOL (2009a). *Cambridge ESOL exams and the CEFR*. Retrieved August 6th, 2010, from Cambridge ESOL: What we do?: <http://www.cambridgeesol.org/what-we-do/research/cefr/item-banking.html>

Cambridge, ESOL (2010, March 31). Ensuring the quality and accuracy of speaking test assessment. *Cambridge First* , p. 4.

Cambridge, ESOL (2008, November). *FCE and CAE exam specifications 2008*. Retrieved August 4, 2010, from Cambridge ESOL Bulletin: http://www.cambridgeesol.org/assets/pdf/fcecae_review5.pdf

Cambridge, ESOL (2008a). *First Certificate in English Handbook for Teachers*. Cambridge: University of Cambridge.

Cooze, M., & Shaw, S. (2007, November). Establishing the impact of reduced input and output length in FCE and CAE Writing. *Cambridge ESOL Research Notes 30* , pp. 15-16.

Galaczi, E. (2005, May). Upper Main Suite speaking assessment: Towards an understanding of assessment criteria and oral examiner behaviour. *Cambridge ESOL Research Notes 20* , pp. 16-19.

Galaczi, E., & French, A. (2007, November). Developing a revised assessment scales for Main Suite and BEC Speaking tests. *Cambridge ESOL Research Notes* , pp. 28-31.

Geranpayeh, A. (2004, February). Reliability in First Certificate in English objective papers. *Cambridge ESOL Research Notes 15* , pp. 21-33.

Hawkey, R. (2009). *Studies in Language Testing 28: Examining FCE and CAE*. Cambridge: CUP.

Hughes, A. (2003). *Testing for Language Teachers 2nd Edition*. Cambridge: CUP.

Khalifa, H., & French, A. (2008). *Aligning Cambridge ESOL Examinations to the CEFR: Issues and Practice*. Cambridge: Cambridge ESOL.

McNamara, T. (1996). *Measuring Second Language Performance*. London: Longman.

Orr, M. (2002). The FCE speaking test: Using rater reports to help interpret test scores. *System Volume 30.2* , 143-154.

Randall, S. (2010, May). Cambridge ESOL's growing impact on English language teaching and learning in national education projects. *Cambridge ESOL Research Notes 40* , pp. 2-3.

Saville, N., & Hargreaves, P. (1999). Assessing speaking in the revised FCE. *ELT Journal Volume 53/1* , 42-51.

Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Oxford: Palgrave.

Weir, C., & Shaw, S. (2006, November). Defining the constructs underpinning the Main Suite Writing Test: a socio-cognitive perspective. *Cambridge ESOL Research Notes 26* , pp. 9-14.

8 Appendix One

8.1 Sample Reading paper

Part 1

You are going to read an extract from a novel. For questions 1 – 8, choose the answer (A, B, C or D) which you think fits best according to the text.

Mark your answers on the separate answer sheet.

I shifted uncomfortably inside my best suit and eased a finger inside the tight white collar. It was hot in the little bus and I had taken a seat on the wrong side where the summer sun beat on the windows. It was a strange outfit for the weather, but a few miles ahead my future employer might be waiting for me and I had to make a good impression.

There was a lot depending on this interview. Many friends who had qualified with me were unemployed or working in shops or as labourers in the shipyards. So many that I had almost given up hope of any future for myself as a veterinary surgeon.

There were usually two or three jobs advertised in the *Veterinary Record* each week and an average of eighty applicants for each one. It hadn't seemed possible when the letter came from Darrowby in Yorkshire. Mr S. Farnon would like to see me on the Friday afternoon; I was to come to tea and, if we were suited to each other, I could stay on as his assistant. Most young people emerging from the colleges after five years of hard work were faced by a world unimpressed by their enthusiasm and bursting knowledge. So I had grabbed the lifeline unbelievably.

line 15

The driver crashed his gears again as we went into another steep bend. We had been climbing steadily now for the last fifteen miles or so, moving closer to the distant blue of the Pennine Hills. I had never been in Yorkshire before, but the name had always raised a picture of a region as heavy and unromantic as the pudding of the same name; I was prepared for solid respectability, dullness and a total lack of charm. But as the bus made its way higher, I began to wonder. There were high grassy hills and wide valleys. In the valley bottoms, rivers twisted among the trees and solid grey stone farmhouses lay among islands of cultivated land which pushed up the wild, dark hillsides.

Suddenly, I realised the bus was clattering along a narrow street which opened onto a square where we stopped. Above the window of a small grocer's shop I read 'Darrowby Co-operative Society'. We had arrived. I got out and stood beside my battered suitcase, looking about me. There was something unusual and I didn't know what it was at first. Then it came to me. The other passengers had dispersed, the driver had switched off the engine and there was not a sound or a movement anywhere. The only visible sign of life was a group of old men sitting round the clock tower in the centre of the square, but they might have been carved of stone.

Darrowby didn't get much space in the guidebooks, but where it was mentioned it was described as a grey little town on the River Arrow with a market place and little of interest except its two ancient bridges. But when you looked at it, its setting was beautiful. Everywhere from the windows of houses in Darrowby you could see the hills. There was a clearness in the air, a sense of space and airiness that made me feel I had left something behind. The pressure of the city, the noise, the smoke – already they seemed to be falling away from me.

Trengate Street was a quiet road leading off the square and from there I had my first sight of Skeldale House. I knew it was the right place before I was near enough to read *S. Farnon, Veterinary Surgeon* on the old-fashioned brass nameplate. I knew by the ivy which grew untidily over the red brick, climbing up to the topmost windows. It was what the letter had said – the only house with ivy; and this could be where I would work for the first time as a veterinary surgeon. I rang the doorbell.

- 1 As he travelled, the writer regretted his choice of
- A seat.
 - B clothes.
 - C career.
 - D means of transport.
- 2 What had surprised the writer about the job?
- A There had been no advertisement.
 - B He had been contacted by letter.
 - C There was an invitation to tea.
 - D He had been selected for interview.
- 3 The writer uses the phrase 'I had grabbed the lifeline' (line 15) to show that he felt
- A confident of his ability.
 - B ready to consider any offer.
 - C cautious about accepting the invitation.
 - D forced to make a decision unwillingly.
- 4 What impression had the writer previously had of Yorkshire?
- A It was a beautiful place.
 - B It was a boring place.
 - C It was a charming place.
 - D It was an unhappy place.
- 5 What did the writer find unusual about Darrowby?
- A the location of the bus stop
 - B the small number of shops
 - C the design of the square
 - D the lack of activity
- 6 What did the writer feel the guidebooks had missed about Darrowby?
- A the beauty of the houses
 - B the importance of the bridges
 - C the lovely views from the town
 - D the impressive public spaces
- 7 How did the writer recognise Skeldale House?
- A The name was on the door.
 - B It had red bricks.
 - C There was a certain plant outside.
 - D It stood alone.
- 8 How did the writer's attitude change during the passage?
- A He began to feel he might like living in Darrowby.
 - B He became less enthusiastic about the job.
 - C He realised his journey was likely to have been a waste of time.
 - D He started to look forward to having the interview.

Turn Over ►

Part 2

You are going to read an article about a woman who is a downhill mountain-bike racer. Seven sentences have been removed from the article. Choose from the sentences A – H the one which fits each gap (9 – 15). There is one extra sentence which you do not need to use.

Mark your answers on the separate answer sheet.

Downhill racer

Anna Jones tells of her move from skiing to downhill mountain biking and her rapid rise up the ranks to her current position as one of the top five downhill racers in the country.

At the age of seven I had learnt to ski and by fourteen I was competing internationally. When I was eighteen a close friend was injured in a ski race, and as a result, I gave up competitive skiing. To fill the gap that skiing had left I decided to swap two planks of wood for two wheels with big tyres.

My first race was a cross-country race in 1995. It wasn't an amazing success. After entering a few more cross-country races, a local bike shop gave me a downhill bike to try. I entered a downhill race, fell off, but did reasonably well in the end, so I switched to downhill racing.

I think my skiing helped a lot as I was able to transfer several skills such as cornering and weight-balance to mountain biking. This year I'm riding for a famous British team and there are races almost every weekend from March through to September. In fact, there's quite a lot of putting up tents in muddy fields.

Last season I was selected to represent Great Britain at both the European and World Championships. Both events were completely different from the UK race scene. I was totally in awe, racing with the riders I had been following in magazines. The atmosphere was electric and I finished about mid-pack.

Mountain biking is a great sport to be in. People ask me if downhill racing is really scary. I say, 'Yes it is, and I love it.' Every time I race I scare myself silly and then say, 'Yeah let's do it again.'

When you're riding well, you are right on the edge, as close as you can be to being out of control. However, you quickly learn how to do it so as not to injure yourself. And it's part of the learning process as you have to push yourself and try new skills to improve.

Initially, downhill racing wasn't taken seriously as a mountain-biking discipline. But things are changing and riders are now realising that they need to train just as hard for downhill racing as they would do for cross-country.

The races are run over ground which is generally closer to vertical than horizontal, with jumps, drop-offs, holes, corners and nasty rocks and trees to test your nerves as well as technical skill. At the end of a run, which is between two and three minutes in this country your legs hurt so much they burn. But in a race, you're so excited that you switch off to the pain until you've finished.

A lot of people think that you need to spend thousands of pounds to give downhill mountain biking a go. A reasonable beginner's downhill bike will cost you around £400 and the basic equipment, of a cycle helmet, cycle shorts and gloves, around £150. Later on you may want to upgrade your bike and get a full-face crash helmet, since riders are now achieving speeds of up to 80 kilometres per hour.

- | | |
|--|---|
| A I've fallen off more times than I care to remember. | E The attitude was: how much skill do you need to sit on a saddle and point a bike in the same direction for a few minutes? |
| B I usually have to stop during practice sessions. | F I finished last, but it didn't matter as I really enjoyed it. |
| C The courses were twice as long and the crowds were twice as big. | G Nothing could be further from the truth. |
| D I'm not strong enough in my arms, so I've been doing a lot of upper-body training this year. | H It's not all stardom and glamour, though. |

Part 3

You are going to read a magazine article about people who collect things. For questions 16 – 30, choose from the people (A – D). The people may be chosen more than once.

Mark your answers on the separate answer sheet.

Which person

- | | | |
|--|----|----------------------|
| had to re-start their collection? | 16 | <input type="text"/> |
| has provided useful advice on their subject? | 17 | <input type="text"/> |
| was misled by an early success? | 18 | <input type="text"/> |
| received an unexpected gift? | 19 | <input type="text"/> |
| admits to making little practical use of their collection? | 20 | <input type="text"/> |
| regrets the rapid disappearance of certain items? | 21 | <input type="text"/> |
| is aware that a fuller collection of items exists elsewhere? | 22 | <input type="text"/> |
| has a history of collecting different items? | 23 | <input type="text"/> |
| performed a favour for someone they knew? | 24 | <input type="text"/> |
| is a national expert on their subject? | 25 | <input type="text"/> |
| is aware that they form part of a growing group? | 26 | <input type="text"/> |
| insists on purchasing top-quality items? | 27 | <input type="text"/> |
| noticed items while looking for something else? | 28 | <input type="text"/> |
| has to protect their collection from damage? | 29 | <input type="text"/> |
| would like to create a hands-on display of their collection? | 30 | <input type="text"/> |

The World of Collecting

A **Ron Barton** shares his home with about 200 sewing machines. His passion began when he was searching for bits of second-hand furniture and kept seeing 'beautiful old sewing machines that were next to nothing to buy'. He couldn't resist them. Then a friend had a machine that wouldn't work, so she asked Barton to look at it for her. At that stage he was not an authority on the subject, but he worked on it for three days and eventually got it going.

Later he opened up a small stand in a London market. 'Most people seemed uninterested. Then a dealer came and bought everything I'd taken along. I thought, "Great! This is my future life." But after that I never sold another one there and ended up with a stall in another market which was only moderately successful.'

Nowadays, he concentrates on domestic machines in their original box containers with their handbooks. He is often asked if he does any sewing with them. The answer is that, apart from making sure that they work, he rarely touches them.

B As a boy, **Chris Peters** collected hundreds of vintage cameras, mostly from jumble sales and dustbins. Later, when the time came to buy his first house, he had to sell his valuable collection in order to put down a deposit. A few years after, he took up the interest again and now has over a thousand cameras, the earliest dating from 1860.

Now Peters 'just cannot stop collecting' and hopes to open his own photographic museum where members of the public will be able to touch and fiddle around with the cameras. Whilst acknowledging that the Royal Camera Collection in Bath is probably more extensive than his own, he points out that 'so few of the items are on show there at the same time that I think my own personal collection will easily rival it.'

C **Sylvia King** is one of the foremost authorities on plastics in Britain. She has, in every corner of her house, a striking collection of plastic objects of every kind, dating from the middle of the last century and illustrating the complex uses of plastic over the years.

King's interest started when she was commissioned to write her first book. In order to do this, she had to start from scratch; so she attended a course on work machinery, maintaining that if she didn't understand plastics manufacture then nobody else would.

As she gathered information for her book, she also began to collect pieces of plastic from every imaginable source: junk shops, arcades, and the cupboards of friends. She also collects 'because it is vital to keep examples. We live in an age of throw-away items: tape-recorders, cassettes, hair dryers – they are all replaced so quickly.'

King's second book, *Classic Plastics: from Bakelite to High Tech*, is the first published guide to plastics collecting. It describes collections that can be visited and gives simple and safe home tests for identification.

King admits that 'plastic is a mysterious substance and many people are frightened of it. Even so, the band of collectors is constantly expanding.'

D **Janet Pontin** already had twenty years of collecting one thing or another behind her when she started collecting 'art deco' fans in 1966. It happened when she went to an auction sale and saw a shoe-box filled with them. Someone else got them by offering a higher price and she was very cross. Later, to her astonishment, he went round to her flat and presented them to her. 'That was how it all started.' There were about five fans in the shoe-box and since then they've been exhibited in the first really big exhibition of 'art deco' in America. The fans are not normally on show, however, but are kept behind glass. They are extremely fragile and people are tempted to handle them. The idea is to have, one day, a black-lacquered room where they can be more easily seen.

Pontin doesn't restrict herself to fans of a particular period, but she will only buy a fan if it is in excellent condition. The same rule applies to everything in her house.

8.2 Sample Writing paper

Part 1

You must answer this question. Write your answer in 120 – 150 words in an appropriate style on the opposite page.

- 1 You have received an email from your English-speaking friend, Sara, who is planning to open a restaurant. Read Sara's email and the notes you have made. Then write an email to Sara, using all your notes.

email

From: Sara Martins
Sent: 15th March 2006
Subject: Restaurant

You remember how Alex and I have always wanted to open a restaurant – well, we're going to do it!

We want to serve food from different countries in our restaurant so we're planning to travel around to collect some ideas. We want to come to your country. When is the best time to come?

Say when and why —————

We want to find out what people cook at home every day. What's the best way for us to do that?

Suggest ... —————

We'd also like to go to some local restaurants which serve traditional food. Can you recommend one? ————— Yes, give details

When we open the restaurant in July, we'd like you to come. Will you be free? ————— No, because ...

Reply soon.

Sara

Write your email. You must use grammatically correct sentences with accurate spelling and punctuation in a style appropriate for the situation.

Part 2

Write an answer to **one** of the questions 2 – 5 in this part. Write your answer in **120 – 180** words in an appropriate style on the opposite page. Put the question number in the box at the top of the page.

- 2 You have seen this announcement in an international magazine.

MY FAVOURITE TEACHER

Tell us about a favourite teacher of yours and say what you remember about him or her.

We will publish the most interesting articles next month.

Write your **article**.

- 3 You recently saw this notice in an English-language magazine called *Theatre World*.

Reviews needed!

Have you been to the theatre recently? If so, could you write us a review of the play you saw? Include information on the characters, costumes and story and say whether you would recommend the play to other people.

The best reviews will be published next month.

Write your **review**.

- 4 Your teacher has asked you to write a story for an international magazine. The story must **begin** with the following words:

Anna had a very special reason for getting up early the next day, so she set the alarm for 5 am.

Write your **story**.

- 5 Answer **one** of the following two questions based on one of the titles below. Write the letter (a) or (b) as well as the number 5 in the question box on the opposite page.

- (a) *The Citadel* by A.J.Cronin
This is part of a letter from your English-speaking penfriend.

*We are reading *The Citadel* in class. Didn't you say you've seen the film? What do you think of the main character, Andrew Manson?*

Write a **letter** to your penfriend, giving your opinion. Do not write any postal addresses.

Write your **letter**.

- (b) *Round the world in 80 days* by Jules Verne
Phileas Fogg and Passepartout are very different characters. Which one do you think enjoys the journey most? Write an **essay** saying who you think enjoys the journey most and why.

Write your **essay**.

8.3 Sample Use of English paper

Part 1

For questions 1 – 12, read the text below and decide which answer (A, B, C or D) best fits each gap. There is an example at the beginning (0).

Mark your answers on the separate answer sheet.

Example:

0 A called B named C referred D known

0	<u> A </u>	<u> B </u>	<u> C </u>	<u> D </u>
---	--------------	--------------	--------------	--------------

A love of travelling

For Nigel Portman, a love of travelling began with what's (0) a 'gap year'. In common with many other British teenagers, he chose to take a year out before (1) to study for his degree. After doing various jobs to (2) some money, he left home to gain some experience of life in different cultures, visiting America and Asia. The more adventurous the young person, the (3) the challenge they are likely to (4) themselves for the gap year, and for some, like Nigel, it can (5) in a thirst for adventure.

Now that his university course has (6) to an end, Nigel is just about to leave on a three-year trip that will take him (7) around the world. What's more, he plans to make the whole journey using only means of transport which are (8) by natural energy. In other words, he'll be (9) mostly on bicycles and his own legs; and when there's an ocean to cross, he won't be taking a (10) cut by climbing aboard a plane, he'll be joining the crew of a sailing ship (11)

As well as doing some mountain climbing and other outdoor pursuits along the way, Nigel hopes to (12) on to the people he meets the environmental message that lies behind the whole idea.

- | | | | | |
|----|-----------------|--------------|---------------|----------------|
| 1 | A settling down | B getting up | C taking over | D holding back |
| 2 | A achieve | B raise | C advance | D win |
| 3 | A stronger | B wider | C greater | D deeper |
| 4 | A put | B set | C aim | D place |
| 5 | A result | B lead | C cause | D create |
| 6 | A come | B turned | C reached | D brought |
| 7 | A just | B complete | C whole | D right |
| 8 | A pulled | B charged | C forced | D powered |
| 9 | A relying | B using | C attempting | D trying |
| 10 | A quick | B short | C brief | D swift |
| 11 | A anyway | B alike | C instead | D otherwise |
| 12 | A leave | B keep | C pass | D give |

Part 2

For questions **13 – 24**, read the text below and think of the word which best fits each gap. Use only **one** word in each gap. There is an example at the beginning **(0)**.

Write your answers **IN CAPITAL LETTERS** on the separate answer sheet.

Example:

0	A	S																			
---	---	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

The temple in the lake

Lake Titicaca, often known **(0)** the 'holy lake', is situated in South America on the border between Bolivia and Peru. The lives of the people **(13)** tools and pottery have **(14)** found on its shores have long remained a mystery. However, scientists taking **(15)** in an exploration project at the lake have found what they believe to **(16)** a 1000-year-old temple under the water.

Divers from the expedition have discovered a 200-metre-long, 50-metre-wide building surrounded by a terrace for crops, a road and a wall. It is thought that the remains **(17)** those of a temple built by the Tihuanacu people who lived beside Lake Titicaca before it became a part **(18)** the much later Incan empire.

'The scientists have not yet had time to analyse the material sufficiently,' says project director, Soraya Aubi. 'But some have **(19)** forward the idea that the remains date from this period **(20)** to the fact that there are very similar ones elsewhere.'

The expedition has so **(21)** this year made more than 200 dives into water 30 metres deep **(22)** order to record the ancient remains on film. The film, **(23)** will later be studied in detail, **(24)** will provide important information about the region.

Part 3

For questions 25 – 34, read the text below. Use the word given in capitals at the end of some of the lines to form a word that fits in the gap in the same line. There is an example at the beginning (0).

Write your answers IN CAPITAL LETTERS on the separate answer sheet.

Example: 0 S E L E C T I O N

Walking holidays

<p>The Real Walkers Company offers a (0) of small group walking holidays which explore some delightful hidden corners of Europe, the Americas and Australasia. There is something for everyone to enjoy on these holidays, (25) of age or level of (26) The brochure includes various destinations and a range of itineraries. These range from sightseeing tours of (27) cities to undemanding walking trips in unspoilt coastal and country regions and, for the more (28) traveller, challenging mountain or hill-walking expeditions.</p>	<p>SELECT</p>
<p>But it would be (29) to give the impression that these holidays are just about walking. According to the brochure, an (30) of walking is often the thing that brings together a group of like-minded people, who share the (31) of good companionship in (32) surroundings.</p>	<p>REGARD</p> <p>FIT</p> <p>HISTORY</p> <p>ADVENTURE</p>
<p>The company believes that its tour leaders are the key to its success. These people are (33) trained and are particularly keen to (34) that each individual traveller makes the most of their trip.</p>	<p>FAIR</p> <p>ENJOY</p> <p>PLEASE</p> <p>ATTRACT</p> <p>FULL</p> <p>SURE</p>

Part 4

For questions 35 – 42, complete the second sentence so that it has a similar meaning to the first sentence, using the word given. **Do not change the word given.** You must use between **two** and **five** words, including the word given. Here is an example (0).

Example:

0 A very friendly taxi driver drove us into town.

DRIVEN

We a very friendly taxi driver.

The gap can be filled by the words 'were driven into town by', so you write:

Example: 0

WERE DRIVEN INTO TOWN BY

Write **only** the missing words **IN CAPITAL LETTERS** on the separate answer sheet.

35 The two boys were sitting by themselves in the classroom.

OWN

The two boys were sitting in the classroom.

36 'I have an interview tomorrow, so I ought to leave soon,' Yannis said.

BETTER

'I have an interview tomorrow, so I soon,' Yannis said.

37 The film will have started, so it's not worth going to the cinema now.

POINT

The film will have started, so in going to the cinema now.

38 Roberto arrived late this morning because his train was delayed.

TIME

If the train Roberto would not have arrived late this morning.

39 I had never met Pia's husband before.

FIRST

It I had ever met Pia's husband.

40 Abdul's mother didn't let him play on the computer until he had done his homework.

MADE

Abdul's mother his homework before he played on the computer.

41 Although the police chased them, the thieves didn't get caught.

EVEN

The thieves managed to get the police chased them.

42 Considering that Luke is so young, you must admit he's making excellent progress as a musician.

ACCOUNT

If you young Luke is, you must admit he's making excellent progress as a musician.

8.4 Sample Listening paper

Part 1

You will hear people talking in eight different situations. For questions 1 – 8, choose the best answer, (A, B or C).

- 1 You hear a young man talking.
Why did he go back to college?
A He needed a better job.
B He needed an evening activity.
C He needed new skills.

- 2 You hear a man talking on the radio.
What is he?
A an inventor
B a company employee
C a writer

- 3 You hear someone talking on the radio about an artist.
How does the artist feel about his work?
A He would like to exhibit it in an art gallery.
B He wants to make his creations last longer.
C He is happy to see his work destroyed.

- 4 You hear a woman talking to her son.
Why is she talking to him?
A to give him a warning
B to refuse permission
C to make a suggestion

5 You hear part of a lecture about the role of retired people in the economy.

What is the lecturer describing?

- A reasons why something is changing
- B errors in statistical information
- C disagreements between researchers

6 You hear a chef being interviewed on the radio.

Why did he decide to become a chef?

- A to follow a family tradition
- B to develop a natural talent
- C to pursue his love of cooking

7 You hear a teenager talking about the sport she plays.

How does she feel while she is playing the sport?

- A uncomfortable
- B embarrassed
- C confident

8 You hear an explorer talking about a journey he is making.

How will he travel once he is across the river?

- A by motor vehicle
- B on horseback
- C on foot

Part 2

You will hear an interview with a woman called Helen Hunter who runs a summer camp for teenagers. For questions 9 – 18, complete the sentences.

SUMMER CAMPS

Helen says that people taking part in the summer camp usually sleep in a

	9
--	---

The summer camp is a chance for teenagers to meet people and learn

	10
--	----

As an example of a practical activity, Helen tells us about a team which built a

	11
--	----

In the next camp, teams will work out problem-solving activities such as a

	12	with clues.
--	----	-------------

Helen gives the example of

	13	as the only typical sporting activity at the camp.
--	----	--

The day when teams can choose their own activities is called

	14
--	----

The summer camp is good for people who don't have opportunities or have little

	15
--	----

On 'Battle of the Bands' day, the teams make a pop record and a

	16
--	----

For the teenagers taking part, the camp lasts for

	17
--	----

You can book for a summer camp that will be held in the month of

	18
--	----

Part 3

You will hear five different people talking about a mistake they recently made. For questions 19 – 23, choose from the list (A – F) the type of mistake that each person made. Use the letters only once. There is one extra letter which you do not need to use.

A ignoring someone's advice	Speaker 1	<input type="text"/>	19
B failing to inform someone about something	Speaker 2	<input type="text"/>	20
C mistaking someone's identity	Speaker 3	<input type="text"/>	21
D arriving somewhere too early	Speaker 4	<input type="text"/>	22
E getting a particular date wrong	Speaker 5	<input type="text"/>	23
F losing something important			

Part 4

You will hear an interview with a conservationist who has built a cable car in the rainforest. For questions 24 – 30, choose the best answer (A, B or C).

- 24 What feature of the cable car makes it particularly good for seeing wildlife in the rainforest?
- A the speed at which it moves
 - B the height at which it travels
 - C the distance that it covers
- 25 What is the main aim of the cable car project?
- A to educate local people
 - B to persuade people to save the rainforest
 - C to raise money for other conservation projects
- 26 What is the advantage of the project for the local people?
- A They can use the land if they want.
 - B They can sell forest products to the visitors.
 - C More work is available to them.
- 27 Why was the cable car redesigned?
- A so that people could touch the trees
 - B to avoid cutting down too much forest
 - C because it had to be brought in by air
- 28 How does Donald react to the suggestion that he has disturbed the wildlife?
- A He explains what happened in the past.
 - B He criticises what happens elsewhere.
 - C He denies that there's been any disturbance.

- 29 Why is Donald sure his project is a success?
- A This piece of forest has survived.
 - B Animals have returned to the area.
 - C Other projects have copied his ideas.
- 30 Donald thinks the future survival of the rainforest will depend on
- A the size of the world's population.
 - B the attitude of people towards it.
 - C the size of the areas left as forest.

8.5 Sample Speaking paper

Part 1

3 minutes (5 minutes for groups of three)

Good morning/afternoon/evening. My name is and this is my colleague

And your names are?

Can I have your mark sheets, please?

Thank you.

First of all we'd like to know something about you.

- Where are you from (*Candidate A*)?
- And you (*Candidate B*)?
- What do you like about living (*here / name of candidate's home town*)?
- And what about you (*Candidate A/B*)?

Select one or more questions from any of the following categories, as appropriate.

Likes and dislikes

- Do you prefer to spend time on your own or with other people? (Why?)
- Do you like cooking? (What sort of things do you cook?)
- What's your favourite food? (Why do you like it?)
- Do you like going to parties? (Tell us about a good party you've been to.)
- Tell us about a day you've really enjoyed recently.

Education and Work

- Do you find it easy to study where you live? (Why? / Why not?)
- Is there something new you'd really like to learn about? (Why?)
- Are you happier doing mental or physical work? (Why?)
- Do you prefer working on your own or with other people? (Why?)
- What kind of work would you really like to do in the future? (Why?)

Part 1

3 minutes (5 minutes for groups of three)

Free Time

- Who do you spend your free time with? (What sort of things do you do together?)
- How much time do you spend at home? (What do you enjoy doing?)
- What sort of music do you listen to? (Why do you enjoy it?)
- Do you ever go to concerts? (What kind of concerts do you enjoy?)
..... (What do you like about them?)
- Where do you like listening to music? (Why?)
- Does anyone you know have an interesting hobby? (What does he/she do?)
- Have you got any plans for this weekend? (What are you going to do?)

Holidays and Travel

- Which area of your country would you like to get to know better? (Why?)
- What's the most interesting place you've visited near here? (Tell us about it.)
- Have you ever used your English on holiday? (Where were you?) (What did you use it for?)
- Do you like to plan your holidays carefully or do you prefer to just go? (Why?)
- Where would you really like to go on holiday in the future? (Why?)

Media

- How much TV do you watch in a week? (Would you prefer to watch more TV than that or less?) (Why?)
- Tell us about a TV programme you've seen recently.
- Do you have a favourite newspaper or magazine? (Why do you like it?)
- Do you use the Internet to learn new things? (What sort of things do you look for?)

1 Making music
2 Educational visits

Part 2

4 minutes (6 minutes for groups of three)

Interlocutor In this part of the test, I'm going to give each of you two photographs. I'd like you to talk about your photographs on your own for about a minute, and also to answer a short question about your partner's photographs.

(Candidate A), it's your turn first. Here are your photographs. They show **people making music in different ways**.

Place **Part 2** booklet, open at **Task 1**, in front of *Candidate A*.

I'd like you to compare the photographs, and say **why you think the music is important to the different groups of people**.

All right?

Candidate A

⌚ 1 minute
.....

Interlocutor Thank you. (Can I have the booklet, please?) Retrieve **Part 2** booklet.

(Candidate B), **which type of music would you prefer to listen to?**

Candidate B

⌚ approximately
20 seconds

Interlocutor Thank you.

Now, *(Candidate B)*, here are your photographs. They show **people of different ages on educational visits**.

Place **Part 2** booklet, open at **Task 2**, in front of *Candidate B*.

I'd like you to compare the photographs, and say **what you think the people will learn on their visits**.

All right?

Candidate B

⌚ 1 minute
.....

Interlocutor Thank you. (Can I have the booklet, please?) Retrieve **Part 2** booklet.

(Candidate A), **which of these things would you like to learn about?**

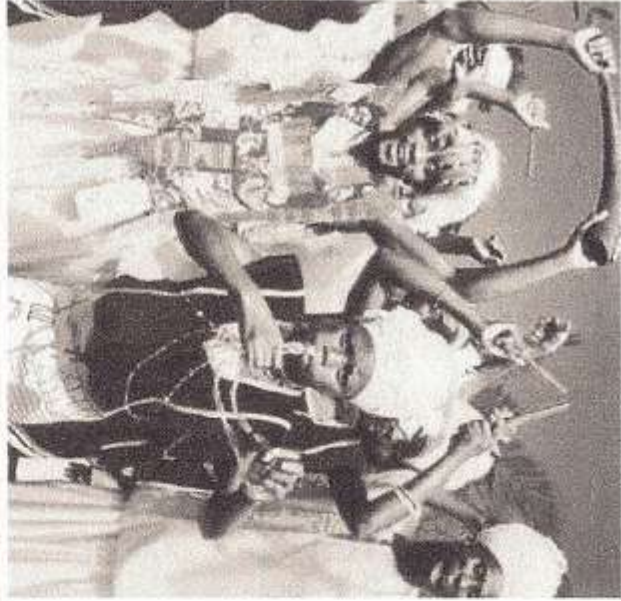
Candidate A

⌚ approximately
20 seconds

Interlocutor Thank you.

Why is the music important to the different groups of peoples?

1



What will the people learn on their visits?

2



21 Jon's café

Parts 3 and 4

7 minutes (9 minutes for groups of three)

Part 3

Interlocutor Now, I'd like you to talk about something together for about three minutes.
(4 minutes for groups of three)

I'd like you to imagine that a local café wants to attract more people. Here are some of the suggestions they are considering.

Place Part 3 booklet, open at Task 21, in front of the candidates.

First, talk to each other about how **successful these suggestions might be**. Then decide **which two would attract most people**.

All right?

Candidates

🕒 3 minutes
(4 minutes for groups of three)

.....

Interlocutor Thank you. (Can I have the booklet, please?) *Retrieve Part 3 booklet.*

Part 4

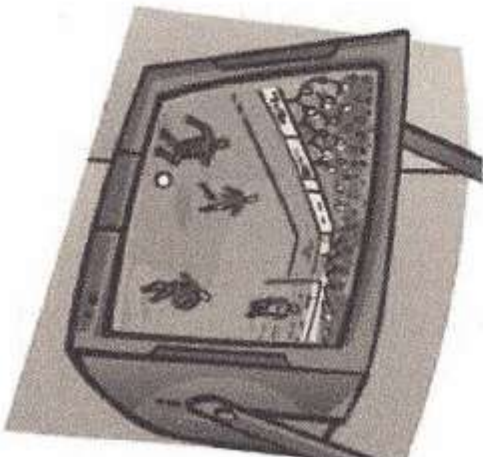
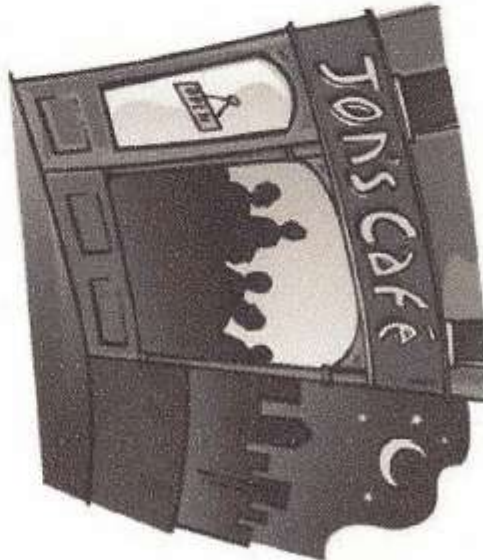
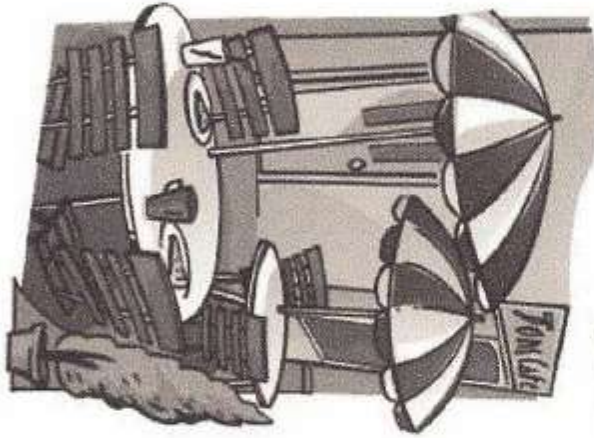
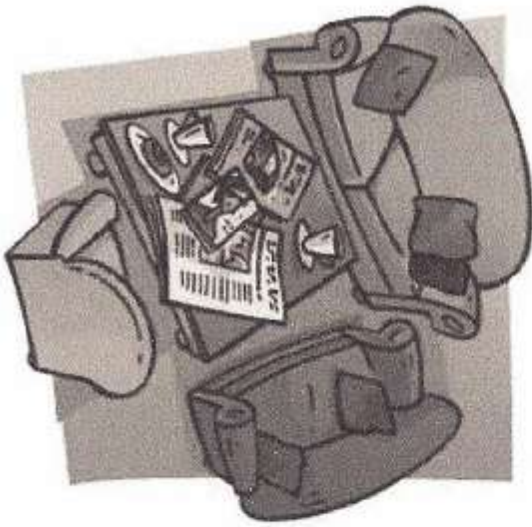
Interlocutor *Select any of the following questions, as appropriate:*

- **Would you like to spend time in a café like this?**
..... (Why? / Why not?)
- **Would you like to work in a café?**
..... (Why? / Why not?)
- **What sort of restaurants are most popular with visitors in your country?**
(Why?)
- **What sort of things do people complain about in cafés and restaurants?**
- **Young people usually go to different places to relax than older people. Why do you think that is?**
- **Some people say that going out to relax is a waste of time and money. Do you agree?** (Why? / Why not?)

Select any of the following prompts, as appropriate:

- **What do you think?**
- **Do you agree?**
- **And you?**

Thank you. That is the end of the test.



- How successful might these suggestions be?
- Which two would attract most people?