IELTS™
English for International Opportunity

3. The Misinterpretation of Directions for the Questions in the Academic Reading and Listening Sub-Tests of the IELTS Test

Gayle Coleman and Stephen Heap
Institute of Continuing and TESOL Education
The University of Queensland

Publishing details

International English Language Testing System (IELTS)

Research Reports 1998

Volume 1

Editor: Sandra Wood

IELTS Australia Pty Limited
ACN 008 664 766
Incorporated in the Australian Capital Territory
Web: www.ielts.org

ELICOS Association Limited
ACN 003 959 037

National Library of Australia
Cataloguing-in-Publication Data
1999 ed
IELTS Research Reports 1999 Volume 2
ISBN 0 86403 021 5

3 The Misinterpretation of Directions for the Questions in the Academic Reading and Listening Sub-Tests of the IELTS Test

Gayle Coleman and Stephen Heap
Institute of Continuing and TESOL Education
The University of Queensland

Abstract

"I have to choose for answers no more than three words... one or two or three ... I have to get the answers from the box ..... I was a little bit confused, so I have to get the answers from the reading passage."
(Student 7, Academic Reading)

This research project seeks to address the areas of perceived confusion or misunderstanding that candidates sometimes have when attempting to comprehend the rubrics of the IELTS Academic Reading and Listening sub-tests. In the project a sample of actual test question papers and candidate answer papers are analysed and students interviewed in order to determine the extent and nature of the problems of misunderstanding that candidates may have.

Whilst the IELTS specifications are now quite clear, and the rubrics standardised, concise and as simple as possible, the tests that were analysed demonstrate that great care has to be taken over the wording of the rubrics. However, it was found that relatively few students misjudged the requirements of the rubrics; often it was the wording and style of the questions themselves that caused the greatest difficulties. In terms of the rubrics, IELTS can be applauded for the reasonably low level of misunderstandings generated by the rubrics.

1.0 Introduction

This research project was conducted between February 1996 and February 1997 in the Institute of Continuing and TESOL Education at The University of Queensland, and funded by IELTS Australia.

The project arose from concerns expressed by EAP language instructors, IELTS preparation language instructors, and IELTS invigilators. These concerns include:

- Language instructors are often quite bewildered by the scores some of their ELICOS and EAP students receive in the IELTS Reading and Listening sub-tests. It is believed that some candidates may be achieving scores which do not reflect their ability in a certain skill due to their lack of attention to understand exactly what is required of them in answering the questions in those sub-tests.

- When invigilators are objectively marking the Reading and Listening sub-tests, it has become apparent that some students have responded incorrectly to a question because they have misinterpreted the rubric of the question.

- It is commonly the case that Reading scores are often lower than writing and speaking scores. Upon further investigation it appears that some students are misinterpreting the rubric of the questions. Therefore, their sub-band scores are not a true reflection of their ability in Reading and Listening.

- Some IELTS candidates have expressed difficulty in interpreting the requirements of certain questions in the Reading and Listening sub-tests.

The primary area of concern in this project, then, was to identify areas of difficulty for candidates in interpreting the rubrics to the questions for the Listening and Academic Reading sub-tests of the IELTS test. This was to be followed by suggestions for any improvements that might be made either in the wording of some of the actual questions or in the guidelines for item writers.

However, other areas of interest arose as the data were studied, notably question types causing most/least difficulty, the progressive difficulty of test items, the number and frequency of different test items, and incorrect marking.

The area of *rubric design* for listening and reading tests in foreign language learning and teaching is not one that has received a great deal of attention in the literature, and yet it is clearly of considerable importance in the construction of such tests. Candidates can be unfairly penalised if rubrics are misleading or ambiguous. This is highlighted in an article by Ana Huerta-Macias (1995,8) which cites the example of a student who "knew the concept being tested, but was unfamiliar with the language and format of the test". Although this reference is to a fourth-grade (American) science test, the same situation can still potentially apply to tests in English as a second or foreign language. Although Huerta-Macias goes on to list other problems that also adversely affect the test-taker's performance, from our research in this project we would not see the potential problem of rubrics as being sufficient grounds alone for developing alternative assessment procedures. It is alternative assessment that Huerta-Macias is advocating. There are probably other more pertinent grounds (which she names) such as

"norming, linguistic and cultural biases" (p.8) and the desire to give the candidate the widest opportunity to display his/her abilities, that may demand alternative assessment.

The problem of rubrics can be overcome by ensuring, in the case of classroom tests, that candidates have had sufficient practice in the kinds of test items they will experience, and in the case of public tests, that rubrics are designed in the least misleading and least ambiguous way including examples in cases where the procedure may be unfamiliar to candidates. They should also be recurring and limited in number so that candidates have the opportunity of familiarising themselves with the rubrics before taking the tests. From our research, we are able to say that the IELTS Reading sub-test, and to a lesser extent the Listening sub-test, contain rubrics that are seldom misinterpreted.

2.0 Data Collection

2.1 Candidate Answer Papers

Firstly, data, in the form of candidate answer papers from the Academic Reading and Listening sub-tests, were collected after the IELTS tests were conducted at the University of Queensland between November 1995 and February 1996. Table 1 is a breakdown of the versions, dates conducted and the numbers of candidates:

READING	Date Conducted	Number of Candidates
Version 1	11 Nov 1995	177
Version 2	10 Feb 1996	49
Version 3	25 Nov 1995	115

LISTENING	Date Conducted	Number of Candidates
Version 1	11 Nov 1995	73
Version 3	25 Nov 1995	115
	13 Jan 1996	47
Version 4	10 Feb 1996	40

Note: The number of candidates for the Reading and Listening sub-tests differ for the same test dates, as some answer sheets used (photocopies) were sometimes illegible.

Table 1 *Overview of Tests*

The results of these answers, that is, whether a candidate had answered a question correctly, incorrectly, or not at all, were entered into a database. Microsoft Access was used for this purpose. From these data we were able to extract information pertaining to candidates who had answered each question in one of three ways: incorrectly, correctly, or no response. Hard copies showing candidate numbers who answered incorrectly or who had given no response were then printed (Appendix 3.1). This bank of data provided the initial impetus for the analysis of candidate answers, but it was necessary, too, to study actual answer papers.

2.2 Student Interviews

Student interviews were conducted to determine the extent of student understanding of the requirements of the question rubrics. Students from a variety of backgrounds, language levels and IELTS experience were interviewed from Higher Intermediate ELICOS and Higher Intermediate and Advanced EAP classes. These were not the students whose test papers were analysed in the sections above, as it was deemed appropriate to have data from a separate cohort to compare with the data from the actual test candidates. (To differentiate these two groups, the term *candidate* is used for those whose actual test papers were analysed and the term *student* for those who were interviewed.)

The students were instructed to read the rubric and question in front of them, but not to try to answer the question. It was explained that they should read the instructions for the question carefully, and they were then asked: "What must you do to answer the question?" A summary of candidates' responses is included, where applicable, in the data analysis sections for the different item types below. At the time the interviews were conducted the question papers had all been withdrawn from circulation at IELTS test centres, with the exception of Listening test Version 4.

2.2.1 Reading Sample

The students who participated in the Academic Reading sub-test interviews and whose comments are recorded below, included six Upper Intermediate ELICOS candidates who had not previously taken the IELTS test but were planning to do so within the following four months, and seven Advanced level EAP candidates who had all taken the IELTS test once only in the preceding six months. The students who were interviewed about their understanding of the Reading sub-test rubrics were as follows:

Student	Nationality	Gender	Level	IELTS test taken
1.	Thai	Female	Upper Intermediate	No IELTS test taken
2.	Korean	Male	Upper Intermediate	No IELTS test taken
3.	Japanese	Male	Upper Intermediate	No IELTS test taken
4.	Thai	Male	Upper Intermediate	No IELTS test taken
5.	Taiwanese	Male	Upper Intermediate	No IELTS test taken
6.	PRC	Female	Upper Intermediate	No IELTS test taken
7.	Korean	Female	Advanced	IELTS Reading sub-test band score: 6
8.	Japanese	Male	Advanced	IELTS Reading sub-test band score: 4.5
9.	Korean	Female	Advanced	IELTS Reading sub-test band score: 6
10.	PRC	Female	Advanced	IELTS Reading sub-test band score: 5
11.	Japanese	Female	Advanced	IELTS Reading sub-test band score: 6
12.	Korean	Female	Advanced	IELTS Reading sub-test band score: 6
13.	Korean	Female	Advanced	No IELTS test taken

Table 2 Students interviewed about the Reading sub-test

Questions selected were those in which candidates demonstrated possible misunderstanding of the rubrics as evidenced in the data analysis in Table 3. The questions for the Reading sub-test can be found in Appendix 3.2.

Version	Questions	Types of Questions
1:	10-14	Notes/Summary with a choice of possible answers.
	29-32	Notes/ Summary without a choice of answers
2:	1-7	Notes/Summary without a choice of answers.
	24-26	Sentence Completion with a choice of answers.
3:	2 & 3	Multiple Choice Items with one possible answer.
	10-14	Notes/Summary without a choice of answers.
	32-40	Matching.

Table 3 Reading sub-test questions

2.2.2 Listening Sample

The students interviewed for the Listening sub-test were 11 Higher Intermediate EAP students who had already completed the IELTS test and were therefore familiar with the rubrics. All of these students were intending to do the IELTS test again in the near future. The students were as follows:

Student	Nationality	Gender	IELTS test taken
1.	Korean	Female	IELTS Listening sub-test band score: 5.5
2.	Thai	Female	IELTS Listening sub-test band score: 6
3.	Japanese	Female	IELTS Listening sub-test band score: 6
4.	Indonesian	Female	IELTS Listening sub-test band score: 5
5.	PRC	Female	IELTS Listening sub-test band score: 5.5
6.	Taiwanese	Female	IELTS Listening sub-test band score: 5
7.	Japanese	Male	IELTS Listening sub-test band score: 5
8.	Japanese	Male	IELTS Listening sub-test band score: 6.5
9.	Taiwanese	Male	IELTS Listening sub-test band score: 4.5
10.	Taiwanese	Female	IELTS Listening sub-test band score: 4.5
11.	Japanese	Female	IELTS Listening sub-test band score: 6

Table 4 Students interviewed for the Listening sub-test

The questions chosen for the interviews were those which had a high percentage of incorrect responses or no response. The question types for the Listening sub-test are listed in Table 5 and in detail in Appendix 3.3.

Version	Questions	Types of Questions
1:	5 to 7	Short answer questions - requiring a list.
	12	Sentence completion
	13	Short answer questions - discrete questions.
	15 & 16 and 19 & 20	Multiple Choice - multiple answers and 1 mark for each.
	18	Multiple Choice - multiple answers for 1 mark.
	21 to 25	Table Completion - without a choice of possible answers
	32 to 41	Summary Completion - without a choice of possible answers.
3:	11 to 17	Notes Completion - without a choice of possible answers.
	18 to 20	Labelling a diagram which has numbered parts.
	27	Multiple Choice - multiple answers for 1 mark.
	31 to 33	Notes Completion - without a choice of possible answers.
	34 to 38	Summary Completion - without a choice of possible answers.

Table 5 *Listening sub-test questions*

3.0 Data Analysis

3.1 Academic Reading

An analysis of the following tables helps identify and clarify certain aspects of the sub-tests as these relate to the particular candidates who sat the tests at those particular times.

The purpose of extracting this information was to identify particular item types and their frequency of occurrence. In these three versions of the Academic Reading sub-test some item types clearly stand out as being more frequent with item writers, for example, Notes/Summary Completion, whilst others are notable for their absence, for example, Short-Answer Questions.

It is interesting to note that Version 1 has only four different item types. Whilst Version 2 has the widest variety of item types, one of these, *notes/summary completion*, accounts for almost one third of all questions. The IELTS Specifications (February 1996) for the Academic

Reading module state that "If too many item types appeared this might lead to processing problems". This study has not sought to delve into this particular issue but we would suggest that a high number of item types would lead to an increase in rubric misunderstanding, as candidates have to switch from one rubric form to another; candidates may well carry over an aspect of the requirements of one item type to the next question.

Item Type	Version 1	Version 2	Version 3
Multiple choice items			
a. One possible answer	Q. 28	Q. 13-14	Q.2,3,9,21,22
b. Multiple answers for 1 mark	-	-	-
c. Multiple answers for 1 mark each	Q. 24-27	Q. 8-12	-
Short-answer questions			
a. Discrete questions	-	-	-
b. Requiring a list	-	-	Q. 7,8
Sentence completion			
a. With a choice of answers	-	Q. 24-26	-
b. Without a choice of answers	-	-	-
Notes/summary/diagram/ flow-chart/table completion			
a. With a choice of possible answers	Q. 10-14,33-39	Q. 34-38	-
b. Without a choice of answers	Q. 29-32	Q. 1-7	Q.10-14,23-27
Choosing from a selection of headings for identified paragraphs/ sections of the text			
a. For whole passage	-	-	Q. 1
b. For all/some paragraphs/sections	Q. 1-4, 15-19	Q. 27-30	Q.15-20,28-31
Identification of writer's views / attitudes/ claims - yes, no, not given	Q. 5-9, 20-23	Q. 15-19,31-33	Q. 4-6
Classification		Q. 20-23	
Matching			Q. 32-40

Table 6 Classification of Academic Reading questions according to item type

It should be stated here that the test data that was analysed - the question papers and the student answer papers - were selected prior to the publication of the Draft Specifications (February, 1997).

Although there is quite a wide range in incorrect candidate responses for some of the item types, eg Notes/Summary Completion (Version 3): 18.7 % - 68.3%, and Choosing from a Selection of Headings (Version 1): 29.6 % - 82.2 %, this is not consistent among all three versions. Looking at the averages, we can observe that there is little difference between versions (with the exception of Writer's Views). Item types do not of themselves appear to pose

any particular difficulties, with the exception, possibly, of Sentence Completion, but only one set of questions has been considered.

Core item type	Version	No. of questions	Average %	Range %
Multiple choice items	1	1	52.1	N/A
One possible answer	2	2	55.1	49-61.2
	3	5	37.9	21.1-58.5
Multiple choice items	1	4	46.2	27.8-59.2
Multiple answers - 1 mark each	2	5	50.6	24.5-71.4
Short answer questions	3	2	36.6	N/A
Sentence completion	2	3	71.4	63.3-83.7
Notes/summary completion	1	12	48.3	34.3-75.7
Choice of answers	2	5	61.2	40.8-81.6
Notes/summary completion	1	4	39.8	21.3-47.3
No choice of answers	2	7	38.5	26.5-51
	3	10	38.2	18.7-68.3
Choosing from selection of headings for whole passage	3	1	13.8	N/A
Choosing from selection of headings for paragraphs	1	9	53	29.6-81.7
	2	4	49	30.6-73.5
	3	10	49.6	32.5-65.9
Writer's views/claims	1	9	46.4	29.6-82.2
	2	8	47.2	30.6-63.5
	3	3	22.5	13-27.6
Matching	2	4	48	36.7-59.2
	3	9	53.1	39-68.3

Table 7 *Summary of Candidate Responses to Core Item Types*

Question		Core Item Type	% Incorrect	% No Response
TEXT 1	1	Choosing from selection of headings	33.9	0.6
	2	"	48	--
	3	"	39	--
	4	"	42.4	0.6
	5	Identification of writer's claims	39	1.7
	6	"	51.4	2.3
	7	"	29.9	--
	8	"	82.5	0.6
	9	"	43.5	0.6
	10	Notes completion with choice of answers	52	--
	11	"	50.8	2.3
	12	"	43.5	3.4
	13	"	67.8	2.3
	14	"	38.4	2.8
TEXT 2	15	Choosing from selection of headings	29.4	0.6
	16	"	66.7	--
	17	"	76.3	--
	18	"	81.4	1.7
	19	"	58.2	1.1
	20	Identification of writer's claims	35	0.6
	21	"	32.8	--
	22	"	46.3	0.6
	23	"	61.6	0.6
	24	MCQ - multiple answers: 1 mark each answer	27.1	3.4
	25	"	59.9	4
	26	"	54.8	3.4
	27	"	41.8	4
TEXT 3	28	MCQ - one answer only	52.5	4
	29	Table completion without choice of answers	46.3	14.1
	30	"	48	14.1
	31	"	22	18.1
	32	"	45.8	23.2
	33	Summary completion with choice of answers	36.2	14.7
	34	"	53.7	19.2
	35	"	36.2	20.3
	36	"	35	15.3
	37	"	33.3	20.9
	38	"	57.1	18.6
	39	"	75.1	19.2

Table 8 *Summary of candidate responses to questions: Academic Reading sub-test Version 1*

Question		Core Item Type	% Incorrect	% No Response
TEXT 1	1	Summary completion without choice of answers	28.6	--
	2	"	26.5	2
	3	"	49	6.1
	4	"	49	6.1
	5	"	34.7	4.1
	6	"	30.6	4.1
	7	"	51	4.1
	8	MCQ - multiple answers: 1 mark each answer	57.1	2
	9	"	46.9	2
	10	"	71.4	2
	11	"	53.1	2
	12	"	24.5	2
	13	MCQ - one answer only	61.2	2
	14	"	49	2
TEXT 2	15	Identification of writer's claims	30.6	--
	16	"	63.5	--
	17	"	46.9	--
	18	"	49	--
	19	"	49	--
	20	Matching	57.1	--
	21	"	36.7	--
	22	"	38.8	--
	23	"	59.2	--
	24	Sentence completion with choice of answers	63.3	--
	25	"	83.7	--
	26	"	67.3	--
TEXT 3	27	Choosing from selection of headings	51	--
	28	"	40.8	--
	29	"	73.5	--
	30	"	30.6	--
	31	Identification of writer's claims	30.6	--
	32	"	46.9	2
	33	"	61.2	--
	34	Summary completion with choice of answers	59.2	6.1
	35	"	77.6	8.2
	36	"	40.8	12.2
	37	"	46.9	22.4
	38	"	81.6	12.2

Table 9 *Summary of candidate responses to questions: Academic Reading sub-test Version 2*

Question		Core Item Type	% Incorrect	% No Response
TEXT 1	1	Choosing - selection of headings - whole Text	*13*	*3.5*
	2	MCQ - one answer only	20.9	--
	3	"	36.5	--
	4	Identification of writer's claims	26.1	3.5
	5	"	12.2	2.6
	6	"	27	4.3
	7	Short answer questions requiring a list	36.5	5.2
	8	"	33	4.3
	9	MCQ - one answer only	32.2	0.9
	10	Summary completion without choice of answers	15.7	1.7
	11	"	14.8	3.5
	12	"	27	5.2
	13	"	67.8	10.
	14	"	47.8	19.1
TEXT 2	15	Choosing from selection of headings	41.7	5.2
	16	"	29.6	1.7
	17	"	57.4	2.6
	18	"	65.2	1.7
	19	"	49.6	0.9
	20	"	56.5	3.5
	21	MCQ - one answer only	40	1.7
	22	"	57.4	2.6
	23	Summary completion without choice of answers	41.7	10.4
	24	"	28.7	14.8
	25	"	40	16.5
	26	"	33	13.9
	27	"	60.9	9.6
TEXT 3	28	Choosing from selection of headings	30.4	6.1
	29	"	59.1	7
	30	"	52.2	4.3
	31	"	48.7	2.6
	32	Matching	43.5	7.8
	33	"	63.5	13
	34	"	52.2	13.9
	35	"	53.9	6.1
	36	"	39.1	7.8
	37	"	50.4	8.7
	38	"	61.7	8.7
	39	"	68.7	9.6
	40	"	52.2	10.4

Table 10 *Summary of Candidate Responses to Questions Academic Reading sub-test Version 3*

The overriding aim of the tabulated information above was to give some pointers to potential problem areas relating to the rubrics for the questions. It was assumed that if a rubric was genuinely and generally misleading, confusing or constructed in such a way as to cause the candidate to answer the question incorrectly, this would be reflected in the percentage of candidates either answering incorrectly or omitting an answer altogether. The first specific question, then, that was looked at in this way was Version 1, question 8. This question had the highest percentage of incorrect answers (82.5%) of any question in all three versions. However, on looking at candidates actual answer papers for this question, it is clear that it is not a question of misunderstanding of the rubric - students have done what they were asked to do - but that candidates found this a difficult question.

Tables 8, 9 and 10 demonstrate the claim made in the IELTS Specifications (1996, 66) that "the test consists of three sections of increasing difficulty, across the Intermediate to Advanced range", and again "the texts are graded with the easiest appearing as Section 1 and the most difficult as Section 3". If we look at average percentages of Incorrect/No Response answers for each text, this gradation in difficulty can be seen to be operating.

	Text 1	Text 2	Text 3
Version 1	48.5%	53.2%	62%
Version 2	48%	53.8%	58.6%
Version 3	33.9%	52.8%	60.1%

Table 11 Summary of Incorrect/No Response Answers - Academic Reading sub-test

The figures shown above are the average percentages of Incorrect and No Response answers for each separate text within the tests.

As different groups of candidates were used for each version it is not possible to look across the versions comparatively. It would not be reasonable, for example, to say that Version 3 is an easier test than Versions 1 and 2, although on the face of it that would appear to be true. However, it is interesting to note that Version 2 shows very little difference in the level of difficulty between the three texts, whilst Version 3 demonstrates a considerable difference. This, however, masks the fact that within Text 1, Version 1, which is a relatively easier text overall, one of the questions (question 8) has the highest percentage of incorrect responses (82.5%) for the whole test.

3.2 Listening

Firstly an analysis was conducted of all item types used in Versions 1, 3 and 4 of the listening sub-test. The classification is listed in Table 12. There were no core item types involving classification and matching.

Item Type	Version 1	Version 3	Version 4
Multiple choice items			
a. One possible answer	Q.14 & 17	Q.1, 2, 5-8, 9, 10, 25, 28 & 29	Q.7-9 & 14
b. Multiple answers for 1 mark	Q.18	Q.22 & 27	-
c. Multiple answers and 1 mark for each	Q.15,16,19 & 20	-	Q.15, 16, 18 & 19
Short-answer questions			
a. Discrete questions	Q.1-4, & 13	Q.3, 4, 21 & 26	Q17
b. Requiring a list	Q.5-7	-	-
Sentence completion	Q12	Q.30	Q.1-3
Notes/table/summary/flow-chart/diagram completion			
a. Using a box of possible answers	Q.8-11	-	Q.29-35
b. Without a choice of possible answers	Q.21-22, 23.....41	Q.11-17, 23, 24, 31, 32....38	Q.4-6, 10-13, 20-28, 36-40
Labelling a diagram which has numbered parts	-	Q.18-20	-

Table 12 Classification of questions according to item type - Listening sub-test

This analysis was necessary to identify the item types and their frequency of occurrence. With Version 1 more than half of the item types involved completion of a table, notes or summary. Version 3 appeared very problematic when observing the variety of item types included, especially section three where the item type changed six times within nine questions. Version 4 contained an amazingly high proportion of note and table completion item types - 28 from a possible 40 questions. Overall there were six different item types for Versions 1 and 4, and eight different item types for Version 3.

After identifying the different item types, information was obtained on percentages of candidates who answered each question correctly, incorrectly or with no response. Those questions marked incorrectly were then analysed for misinterpretation.

Question		Core Item Type	% Incorrect	% No Response
Section 1	1	Short Answer	12.3%	
	2	Notes Completion	24.7%	
	3	"	23.3%	
	4	"	19.2%	
	5	Short Answer	57.5%	
	6	"	74.0%	
	7	"	75.3%	
	8	Notes/Table Completion	47.9%	
	9	"	34.2%	
	10	"	37.0%	
	11	"	34.2%	
Section 2	12	Sentence Completion	49.3%	4.1%
	13	Short Answer	58.9%	1.4%
	14	Multiple Choice	23.3%	1.4%
	15	"	42.5%	1.4%
	16	"	35.6%	1.4%
	17	"	30.1%	2.7%
	18	"	11.0%	
	19	"	17.8%	
	20	"	52.1%	
Section 3	21	Table Completion	35.6%	4.1%
	22	"	54.8%	1.4%
	23	"	45.2%	2.7%
	24	"	43.8%	1.4%
	25	"	82.2%	2.7%
	26	Notes Completion	42.5%	13.7%
	27	"	37.0%	11.0%
	28	"	34.2%	16.4%
	29	"	41.1%	39.7%
	30	"	60.3%	2.7%
	31	"	69.9%	15.1%
Section 4	32	Summary Completion	67.1%	13.7%
	33	"	56.2%	9.6%
	34	"	63.0%	13.7%
	35	"	37.0%	2.7%
	36	"	41.1%	13.7%
	37	"	39.7%	12.3%
	38	"	34.2%	11.0%
	39	"	58.9%	5.5%
	40	"	23.3%	.4%
	41	"	27.4%	.5%

Table 13 *Summary of Candidate Responses to Questions Listening sub-test Version 1*

Question		Core Item Type	% Incorrect	% No Response
Section 1	1	Multiple Choice	4.3%	
	2	"	17.4%	0.6%
	3	Short Answer	17.4%	
	4	"	23.0%	
	5	Multiple Choice	26.1%	1.2%
	6	"	14.3%	
	7	"	1.2%	
	8	"	27.3%	
	9	"	16.1%	
	10	"	13.7%	
Section 2	11	Notes Completion	10.6%	4.3%
	12	"	24.8%	5.6%
	13	"	18.6%	2.5%
	14	"	35.4%	1.2%
	15	"	21.7%	3.1%
	16	"	44.7%	3.7%
	17	"	64.6%	11.8%
	18	Labelling a diagram	50.3%	2.5%
	19	"	49.7%	1.9%
	20	"	47.8%	2.5%
Section 3	21	Short Answers	13.7%	1.9%
	22	Multiple Choice	34.2%	0.6%
	23	Table Completion	18.0%	0.6%
	24	"	20.5%	1.2%
	25	Multiple Choice	32.3%	0.6%
	26	Short Answers	22.4%	5.0%
	27	Multiple Choice	51.6%	0.6%
	28	"	31.7%	0.6%
	29	"	14.9%	0.6%
Section 4	30	Sentence Completion	4.2%	5.6%
	31	Note Completion	62.7%	12.4%
	32	"	54.0%	11.2%
	33	"	47.2%	12.4%
	34	Summary Completion	3.3%	17.4%
	35	"	70.2%	23.6%
	36	"	69.6%	15.5%
	37	"	36.6%	19.3%
	38	"	63.4%	11.2%

Table 14 *Summary of Candidate Responses to Questions Listening sub-test Version 3*

Question		Core Item Type	% incorrect	% no resp.
Section 1	1	Sentence Completion	27.5%	
	2	"	27.5%	10.0%
	3	"	0.0%	10.0%
	4	Table Completion	52.5%	2.5%
	5	"	25.0%	10.0%
	6	"	27.5%	5.0%
	7	Multiple Choice	12.5%	
	8	"	7.5%	
	9	"	37.55%	
Section 2	10	Notes Completion	32.5%	10.0%
	11	"	40.0%	35.0%
	12	"	40.0%	40.0%
	13	"	35.0%	12.5%
	14	Multiple Choice	40.0%	
	15	"	40.0%	
	16	"	22.5%	
	17	Short Answers	67.55	2.5%
	18	Multiple Choice	35.0%	
	19	"	60.0%	
Section 3	20	Note Completion	60.0%	10.0%
	21	"	42.5%	5.0%
	22	"	37.5%	7.5%
	23	"	30.0%	22.5%
	24	"	47.5%	27.5%
	25	Table Completion	57.5%	30.0%
	26	"	47.5%	40.0%
	27	"	47.5%	47.5%
	28	"	50.0%	47.5%
	29	"	62.5%	
	30	"	50.0%	
	31	"	50.0%	
Section 4	32	Table Completion	27.5%	2.5%
	33	"	47.5%	
	34	"	62.5%	
	35	"	50.0%	
	36	Note Completion	35.0%	7.5%
	37	"	37.5%	22.5%
	38	"	42.5%	15.0%
	39	"	60.0%	12.5%
	40	"	47.5%	15.0%

Table 15 *Summary of Candidate Responses to Questions Listening sub-test Version 4*

It is also worth noting the overall number of incorrect or no responses given for each section of the listening test. This is useful as the test is supposed to be graded from least difficult to more difficult. The IELTS Specifications (1996, 40) state: "The test consists of four sections of increasing difficulty, across the intermediate to advanced range".

	Section 1	Section 2	Section 3	Section 4
Version 1	39.9%	37%	59.7%	54.5%
Version 3	23.7%	40.7%	28%	71%
Version 4	36.1%	51.2%	68.5%	53.8%

Table 16 *Summary of Incorrect/No Response Answers - Listening sub-test*

Version 1 shows that more candidates were incorrect in section 3 than section 4, suggesting that either the listening text or the questions for section 3 may have been more difficult. There is some inconsistency with Version 3. Section 3 appears much easier than both sections 2 and 4. Section 3 also contained a variety of item types, such as short answers, multiple choice, and table completion. Once again in Version 4, section 3 has a higher percentage incorrect than section 4. Overall it appears that Version 4 may be a more difficult test than Versions 1 and 3.

Version 1 shows that more candidates were incorrect in section 3 than section 4, suggesting that either the listening text or the questions for section 3 may have been more difficult. There is some inconsistency with Version 3. Section 3 appears much easier than both sections 2 and 4. Section 3 also contained a variety of item types, such as short answers, multiple choice, and table completion. Once again in Version 4, section 3 has a higher percentage incorrect than section 4. Overall it appears that Version 4 may be a more difficult test than Versions 1 and 3.

4.0 Discussion of Findings

4.1 Reading

Core item types and inappropriate student answers

Because we had expected to see evidence of rubric misunderstanding in the way candidates had answered the questions, certain questions were isolated as being potentially problematic, in that they might lead candidates to write answers in a manner different from that demanded of them. The following section looks at each item type and identifies those rubrics which appear to have been misinterpreted.

4.1.1 Multiple Choice Items

4.1.1a One possible answer

In Version 2, questions 13 and 14, and again in Version 3, questions 2 and 3, candidates are instructed to:

Choose the appropriate letters A-D and write them in boxes 2 and 3 on your answer sheet.

This follows the exact wording given in the IELTS Specifications. However, it is potentially problematic for candidates as the plural words: *letters* and *them* may suggest that more than one answer is correct for each question. However, from the data only one candidate [155] in fact did give more than one answer: for Version 3, question 2, B and C were both given as answers. Although a relatively high percentage of candidates gave incorrect responses to Version 2, questions 13 and 14 (61.2% and 49% respectively - this was Text 1), none of these consisted of two answers. This would be because of the wording of the actual questions, which clearly indicate that single item answers are required:

13. What is the main topic of the passage?
14. What is the author's main purpose in writing this passage?

From the interviews, it was clear that the students were paying more attention to the actual questions than to the preceding rubric.
Interestingly, the wording for Version 3, questions 21 and 22 is in the singular:

Choose the best phrase to complete each statement below and write the appropriate letter A-D in boxes 21-22 on your answer sheet.

This would appear to be a more appropriate wording for such questions. (See recommendations in Section 5.)

Where there is only one question under that particular rubric (as in Version 3, question 9) this is not a problem, as the question is worded in the singular. However, one candidate [253] wrote 2 answers (A and B) for Version 3, question 21 even though this question clearly asked for a single answer.

In the interviews students were asked about their understanding of Version 3, questions 2 & 3. Although the number of incorrect responses for these questions was not particularly high (20.9% and 36.5% respectively), they are the most ambiguous of the multiple-choice questions. All the interview responses, except two, indicated that students were quite sure that only one answer should be selected, for example:

Student 1:	"I have to choose only one answer from A to D."
Student 7:	"Just one I can choose, then I have to write down the capital letter."
Student 10:	"This question asks me to choose one answer."
Interviewer:	"Can you choose two answers?"
Student 10:	"No, no I can't. Just one answer."

This is indicative of the fact that a number of students, on seeing the familiar format of a multiple choice question, did not pay attention to the rubric but instead went directly to the body of the question. Two students interviewed, however, did in their responses demonstrate the ambiguity of the question:

Student 6:	"I think from this topic not sure if you can. I can choose one or two but it depends if they're right or not."
Student 13:	"I will try to choose one or two."
Interviewer:	"What do you mean, one or two?"

Student 13: "I mean Number 2 question I will try to find the answer. I will try to find one or two...more than one answers I will try to find, because 'write them'. It's plural."

It is clear that student 13 has interpreted the plural items: *letters* and *them* as applying to each of the questions rather than to both questions together.

4.1.1b Multiple answers

Questions 24-27 of Version 1 and questions 8-12 of Version 2, which have similar requirements, could potentially cause candidates to write more than one letter against a question number on their answer sheet. However, there were only two instances of this in the Version 1 question and none in Version 2. One candidate [63] wrote two answers for questions 24 and 27, and another [93] wrote C-H, D-H, etc as answers. A further candidate [103] also misunderstood the requirements of the question and wrote single words for two of the questions. These three candidates achieved Reading sub-test band scores of 4.0, 5.0 and 4.5 respectively, indicating fairly poor performance in the other question types too.

Version 1, questions 24-27 also had the highest number of No Responses for Text 2: between 3.4% and 4%. However, this is commensurate with an increasing level of difficulty in the texts and the text questions. The Reading band scores of those candidates were also generally low: two with 4.0, four with 4.5, and one with 5.5.

Interviews with students on this question type were not conducted.

4.1.2 Short Answer Questions Requiring a List

Again, low band scores (4.5 and lower) were evident among candidates who misunderstood Version 3, questions 7 and 8 or did not answer the questions. However, these were not misunderstandings of the italicised rubric but rather of the question:

Give two examples each of types of snakes .....

4.1.3 Sentence Completion with a Choice of Answers

A very high percentage of candidates had incorrect answers for Version 2, questions 24-26, as high as 83.7% for question 25. However, there were not many candidates who appeared to have misunderstood the rubric. There is a slight difference in the wording of the rubric from that given in the Specifications, ie *answer* is used rather than *ending*. The latter is more appropriate as it indicates more precisely the nature of what is required.
The NB in this item type:

NB There are more possible endings than statements.
You may use any ending more than once.

as given also in the Specifications can be misleading. 8.2% of all candidates in this particular cohort appear to have misunderstood the rubric. For example, candidate [179] gave two answers for question 24 and one for questions 25 and 26. Candidate [193] gave three answers for question 24, two for question 25 and one for question 26. Two other candidates, [205] and [222], put either two or three answers for one or more of the questions.

Interviews with students revealed that all except one believed that only one answer was permissible for each question. There was greater uncertainty over whether one answer could be used more than once. Examples of student comments that demonstrate understanding of the rubric are:

Student 1: "This question I find answer from table and I write only one, two, three, four. If I choose number three I just write number three."
Interviewer: "Can you write more than one answer?"
Student 1: "No."
Student 3: "I will write answer number two is twenty-four. I think just one answer."
Student 9: "This answer I try to find from the text. Answer from the box. Different number for different questions and cannot put two numbers for one question."
Student 13: "If I find the answer to question twenty-four, in the case of twenty-five I will find another answer."

The student who misunderstood the rubric said the following:

Student 4: "Maybe I can put two answers for this question."

There was greater uncertainty over whether one answer could be used more than once:

Student 5: "Maybe I can use the same answer for two questions."

Student 12: "I don't think so."

Two students, 9 and 12, did not see or read the Note.

4.1.4 Notes/Summary Completion

4.1.4a With a choice of possible answers

This question type occurs three times in the data: once in Version 2 and twice in Version 1. Although there was a high percentage of incorrect answers for Version 2, questions 34-38 (ranging from 40.8% to 81.6%), all candidates who answered the questions did so as instructed. These were simply difficult questions, attested to by the high percentage of No Response answers (as high as 22.4%). In fact, for Question 38, if the Incorrect and No Response answers are added, we have a total of 93.8% of students getting a zero mark for the answer. A similar situation applies, too, to Version 1, questions 33-39.

However, in Version 1, questions 10-14 (see Appendix 3.2 for the full item) there is quite a different story. 4.5% of all candidates in this test appear to have misunderstood precisely what they should do. But this may not be because of the wording of the rubric (bearing in mind that such problems did not arise with the other questions mentioned in the previous paragraph) but have more to do with the way the question is worded and set out on the page.

Questions 10-14 are potentially confusing if candidates do not recognise, for example, the groups of three words as phrases but as two possible answers, or, do not recognise that the single words are separate answers from the phrases.

This latter problem was evident in the answers of candidate [15] who wrote:

10. content formal and informal
11. theory process and theory

and so on for the remaining questions in that section.

A number of other candidates chose words that were not from the box of possible words, although not for all the questions. For example:

Candidate [305]:	question 11: 'involves the'
	question 12: 'the process'
Candidate [282]:	question 13: 'formal approach'
Candidate [12]:	question 13: 'fixed body of knowledge'

A further problem with questions that demand a written word/phrase for an answer arises in the marking. For example, candidate [345] used two single words for one answer, omitting the conjunction and was marked correct. Candidate [333] wrote 'theory and teaching practice', which is not given as one of the alternatives. This was marked correct, perhaps because it embodies the essential meaning.

Because of its potentially problematic nature, this question featured in student interviews. Whilst a few students were hesitant about how the question should be answered, only one (student 5) misunderstood more than one aspect of the rubric. Only the first sentence of the rubric had been read. Three students misunderstood one part of it, for example, in thinking that words could be guessed ("I can choose from my head"), being unsure whether the same answer could be used for more than one question, or that two answers could be chosen for one question.

Some student responses that indicated clear understanding of this question were:

Student 11: "I have to choose words from this box. I can choose from the right side something and something."

Student 12: "I will look at paragraph C and if I see 'between' maybe I'll choose from here (pointing to the column on the right). Of course I have to choose 'formal and informal' " (when asked if just one of the words from the phrase could be used).

4.1.4b Without a choice of possible answers

Although the rubric for these item types could probably not be clearer, many candidates wrote more than the required number of words. (Occasionally these were given correct, for example, candidate [234] gave five and four words respectively for questions 7 and 8 of Version 2 when the required number was *ONE or TWO*. Perhaps structure words such as 'and' and 'of' are not considered, but in one instance, candidate [288], an extra content word was given.

In Version 1, questions 29-32, *NO MORE THAN THREE WORDS* is asked for. However, on twelve occasions 4 words were given, on four occasions 5 words, on 3 occasions 6 words and on one occasion 7 words.

In Version 2, questions 1-7, candidates are asked to **Choose ONE or TWO WORDS** and again, on six occasions 3 words were given, and on six occasions 4 words.
In Version 3, questions 10-14 and 23-27 candidates are asked to **Choose ONE or TWO WORDS**. On eight occasions 3 words were given, and on two occasions each 4 and 5 words were given.

All these examples point to a clear misunderstanding of the rubric, or at the least a disregard for the rubric even though the important words are capitalised and in bold font.

No misunderstandings other than required number of words were evident in the test data. However, student interviews revealed that the words 'from Reading Passage 3' or 'from the passage' did not always register with students. For example:

Version 1: questions 29-32

Student 1: "I have to guess some words is same meaning between right side and left side..... not more than three words ..... I guess some words."
Interviewer: "You guess your own words?"
Student 1 "Yes."

Version 2: questions 1-7

Student 13: "Some answer will occur to me."

But overall the students had no significant problems understanding the rubric for any of these questions:

Version 1: questions 29-32

Student 4: "Three is maximum."
Interviewer: "Do the words come from your head?"
Student 4: "Oh no no no. I have to find from the article."
Student 9: "I just choose three or two words ..... from the article, paragraph."

Version 2: questions 1-7

Student 2: "This question we must answer the word we can refer to text ... one or two words."
Student 10: "I can use just one or two words to fix the blanks."
Student 12: "Of course from the reading passage."

Version 3: questions 10-14

Interviewer: "Can you choose an answer from your head?"
Student 7: "No, no way."

Two of the Advanced students suggested that a different word form could be used:

Student 12: "See grammar structure and change word."

4.1.5 Choosing from a Selection of Headings

4.1.5a For the whole passage

The only example of a question in this category was in Version 3, question 1. As would be expected for the first question of the first text, there was a very high success rate: 82.9%. However, candidate [124] put DEABC as the answer when asked *to choose the most suitable title* and to *write the appropriate letter A-E.* Candidate [176] wrote: E-D-C-A and candidate [253] D, E.

Interviews with students on this question type were not conducted.

4.1.5b For paragraphs/sections

Again, a very small number of candidates wrote more than one answer. They may have misunderstood that part of the rubric that says they may use any heading more than once. For Version 1, one candidate [63] gave two answers for each of questions 1-3 and questions 15-17.

In Version 2, one candidate [205] gave two answers for question 29, and in Version 3 three candidates gave two answers for no more than two questions each.

Interviews with students on this question type were not conducted.

4.1.6 Identification of Writer's Views/Claims

The only problem with this item type was in not writing the words as specified on the question paper, viz: **YES, NO, NOT GIVEN**. Candidates wrote a mixture of NG, No Given, Y, N, but these are all permissible and given correct. However, in Version 2 candidate [212] was marked correct for writing 'NOT'. Candidate [201] wrote 'N' for question 31 and was given correct but elsewhere the same candidate used 'NO'. This can lead to inconsistency in marking, for example, Version 1, question 9: candidate [1] wrote 'Not Given' and was marked correct, when the answer should have been 'Yes'. In the Reading sub-test Version 1, question 8, candidate [64] was marked incorrect for 'Not Given'; but this is the correct answer.

The use of N it seems could be used equally by candidates to mean either NO or NOT GIVEN.

Interviews with students on this question type were not conducted.

4.1.7 Classification

This item type only occurred in Version 2, and all candidates answered according to the instructions. Interviews with students on this question type were not conducted.

4.1.8 Matching

Version 3, questions 32-40 are the only examples of this item type. The only instances of incorrect answers that may represent a misunderstanding of, or non-adherence to, the rubric are questions for which candidates have given more than one answer. If we look at the band scores for all the students who have given more than the required number of answers, it can be seen that it is more likely that candidates who make this mistake will be weaker in reading overall. Out of a total of 22 candidate questions in which more than one answer was given, 15 questions were answered by students with a Reading band score of 5.0 or lower; two were answered by a candidate with 7.5 and one by a candidate with 8.0.

From the interviews only student 3 thought that it was possible to write more than one answer for each question. The others understood that only one answer was to be given. A number of others were unclear whether the same answer could be used with more than one question. However, the following examples are representative of the students' responses:

Student 7:	"I have to match one by one."
Student 12:	"So I have to choose from this box .... I will choose one answer .... just one .... I can use again, same answer."
Interviewer:	"Can you use two answers for one question?"
Student 12:	"No I don't think so."

4.2 Listening

The classification of item types for each version and the database of incorrect and no responses formed the basis for analysing the various rubrics for each item type in the Listening sub-test. The findings of that analysis are given under the headings for the various item types.

4.2.1 Multiple Choice Items

4.2.1a Where there is one possible answer

For this item type the rubric in Versions 1 and 3 read:

Circle the correct answer

In Version 4 the rubric for this item type is:

Circle the correct letter A-C for each answer.

There were no problems with these instructions. Candidates responded well on all versions and the percentage correct was very high. The percentage incorrect ranged from 1.2% to 40%. Even candidates whose overall band for listening was 4 or lower appeared to understand these rubrics. Students interviewed supported this viewpoint with all students explaining that they were required to circle one letter.

4.2.1b Where there are multiple answers for only one mark

The rubric for this item type was clearly written as follows:

Circle TWO letters OR Circle THREE letters

Listening test Version 1 - question 18 showed no problems with only 10% of candidates responding incorrectly. However, this was not the case with Version 3 - questions 22 and 27. The results were 34.2% and 51.6% respectively. Analysis of these two questions demonstrated that several candidates responded with only one letter rather than two and therefore were incorrect for this question. With question 27 the item type changed constantly within this section of the listening test. Questions 23 and 24 were table completion item types; question 25 was multiple choice; question 26 was short answer and then questions 27 to 29 were multiple choice. This constant changing of item type may have caused problems for some candidates. The IELTS Specifications (1996) also suggest that too many item types in any one section may lead to problems in processing the information.

During the interviews all students clearly understood the rubric. For example:

Student 7: "I choose three answers". (Version 1 - question 18)

Similarly, with Version 3 - question 27 all students interviewed understood the rubric.

Student 1: "so I must choose the materials about the modern bicycle ..I must choose two".

Student 2: "Choose two correct answers".

All students gave similar responses to this question. This suggests that those candidates who only responded with one letter may not have known the other answer.

4.2.1c Where there are multiple answers and one mark for each.

The rubric for this item type in Version 1 states:

Circle TWO letters

There are no examples of this item type in Version 3, but Version 4 has two examples written as illustrated below:

Circle TWO letters A-E

This rubric caused problems for many candidates. As there is one mark for each answer selected it is worth looking at the incorrect responses together. The incorrect response rate for Version 1, questions 15 and 16 was 42.5% and 35.6% respectively. This pattern was more noticeable with the remaining questions. Questions 19 and 20 were 17.8% and 52.1%. Similarly, in Version 4, questions 15 and 16 the incorrect response rate was 40% and 22.5%, and for questions 18 and 19 it was 35% and 60%. The rubric for the latter questions is very wordy, and candidates are required to tick two boxes and then copy the letter to the spaces underneath. They are not concise and simple, and actually require two steps to obtain the answer.

Identify (☑) TWO stated benefits of the Melbourne Olympics for Australia from the list below.

Copy your answers (A-F) against questions 18 and 19 in any order.

Although candidates appeared to understand the rubric, they often transferred their answers incorrectly. For example, Listening Version 1 - questions 15 and 16. The correct response was:

Q15 - B and
Q16 - D in any order.

Many candidates had one of the letters incorrect such as by responding with C, D for both question 15 and 16, that is,

Q15 - C, D
Q16 - C, D

Even though the candidates had selected D, they were not marked correct for it because they did not select both D and B.

Student responses though indicated that this rubric was not problematic. Student responses included:
Version 1 - question 15 and 16

Student 1: "I have to choose two answers" and "I have to ...two answers".
Student 2: "choose the correct answer - choose two correct answer"
Student 4: "I have to circle two letter...is it correct?"

Similarly in the Listening Version 4 - questions 15 and 16 two letters were required for two marks. Candidates scoring 5.5 or less for the listening component of the test selected only one letter. This had ramifications for the following questions in this version in that they transferred the answer for question 17 into the place of the answer for question 16.

There were also a small number of candidates who responded with both letters and the complete answer from the question sheet. Likewise a small percentage wrote the whole answer rather than the letter. This occurred twice in Version 4.

4.2.2 Short-Answer Questions

4.2.2a Discrete questions

The rubrics for this item type were not consistent. In Version 1, question 1 there was no rubric and for question 13 it read:

> Answer questions 12 to 16 according to the information given in the talk.

Version 3 was more explicit with the rubric stating:

> Write NO MORE THAN THREE WORDS for each answer

or:

> Write ONE word.

Version 4, question 17 also excluded any rubric.

Candidate responses were quite accurate for these questions with the exception of Version 1, question 13 at 58.9% incorrect and Version 4, question 17 which was 67.5% incorrect. The former may be accounted for by the changing of item types from sentence completion to short answer to multiple choice. For the latter, candidates were required to write two countries for one mark. A small percentage only wrote one country. This could be due to the candidate not knowing the other country. A second group of candidates transferred their responses as two separate answers for question 17 and 18 rather than the two answers for question 17.

Interviews with students indicated that this rubric was fully understood.

Student 3: "I must write less than three words"
Student 10: "No more three words".

However, student responses indicated they were unsure of how many words to write. When asked how many words, student 5 responded, "I don't know". Three other students responded with similar answers. This is an area where the rubrics seriously need to be formalised to avoid confusing candidates.

4.2.2b Requiring a list

There was only one example of this rubric in Listening Test Version 1 - Questions 5, 6 and 7. The majority of candidates responded incorrectly to these questions, even those whose overall listening band was 7.5 and 8.0. The percentage of incorrect responses was 57.5%, 74% and 75.3% respectively. I believe the inclusion of a negative question confused the candidates. The question reads:

> List three things which the owner will not allow in her flat. Write NO MORE THAN THREE WORDS for each answer.

Analysis of candidate responses demonstrated the subjective nature of marking by invigilators. Incorrect responses included pets, dog, a dog. While some invigilators marked candidates incorrect for some responses, others marked them correct for the same response.

During the interview students responded well to this question. They appeared to understand the rubric.

Student 3: "I have to listen three things which is not allowed in the flat (pause) and I have to write no more than three words".

However when prompted as to what the answers might be, students in the interview responded with negative statements. For example,

Student five "No more than three words for each one...I have to write not speak loudly".

This was the response of most interviewees.

Student 7: "The thing that the owner of flat don't allow me to bring".

Interviewer prompts students.

Student 7: "Pets."

Interviewer: "But how many words can you write"

Student 7: "One or two."

The rubrics of this item type, listing short answers is obviously clearly comprehended, but I believe the inclusion of a negative question caused the high failure rate for this particular question.

4.2.3 Sentence Completion

Within the three versions of the listening test, there were only five sentence completion questions. In Versions 1 and 3 this item type was a solitary question which could pose problems for candidates in that they need to read and change item type quickly whilst still listening to the tape. The response was as follows:

In Version 1, question 12 the rubric included:

> Answer questions 12 to 16 according to the information in the talk.

The incorrect response rate for this question was 49.3% and 4.1% had no response. This high rate could be attributed to it being the first question in this section. Although an example was included with this item type during the interview students rarely looked at this, except to clarify how many words to write.

In Version 3, question 30 the success rate was higher with only 34.2% of candidates answering incorrectly and 5.6% giving no response. The rubric here is more explicit:

> Write NO MORE THAN THREE WORDS for each answer.
> Complete the information below.

Likewise, the rubric for Version 4 was clearly stated:

> Complete the sentences below.
> Write NO MORE THAN THREE WORDS for each answer.

The percentage of incorrect responses was 27.5% for questions 1 and 2, with a further 10% giving no response to question 2. This percentage increased dramatically to 70% incorrect and 10% no response for question 3. Although the tapes were not available, this may be a case of the information being presented too close together on the tape. The IELTS Specifications (1996, 46) clearly state, "Sufficient redundant material should follow an item to allow candidates time to write down their answer".

Analysis of candidate responses indicated that the marking for Version 4 question 3 was inconsistent. Candidates were marked incorrectly if they responded with words whose meaning and content was the same as the answer. It is possible that because these exact words were not included on the answer key they were marked incorrectly.

There were no obvious problems with the rubrics according to the interviews. For Version 1, questions 12 and 13 students responded:

Student 6: "No...I think...you must you must looking carefully before you do this question..so in here you can just use no more than three words."

4.2.4 Notes/Table/Summary/Flow-chart/Diagram completion

4.2.4a Using a box of possible answers

Version 1, questions 8 to 11 included the following rubric for this item type:

> Write the appropriate occupations under the correct flat numbers (I-VIII).
> NB Sam does NOT mention all of the flats. You will have to leave some empty.
> Copy the flat number and the occupation of the people mentioned against 8 to 11 in any order.

The percentage of incorrect responses was 47.9%, 34.2%, 37% and 34.2% respectively. All candidates wrote a response for each question. Only two students misinterpreted the rubric and wrote only the occupation rather than the flat number and the occupation. These candidates were marked incorrect for all four questions.

Version 3 contained no examples of this item type. Version 4, questions 29 to 31 contained the following rubric:

> Complete the table by circling the appropriate course.
> NB Circle only one course for each student.

Despite the fact that all candidates responded to these questions the incorrect response rate was very high, 62.5% for question 29 and 50% for questions 30 and 31. There was no evidence of misinterpretation of this rubric.

Version 4, questions 32 to 35 contained the following rubric:

> Complete the following table. Choose CAUSES A-F from the box below.
> Write letters A-F in the numbered spaces according to the information on the tape.

Nearly all candidates responded to these questions. The percentage of incorrect responses was 27.5%, 47.5%, 62.5% and 50% respectively. Analysis of responses indicated that the rubric was generally understood by all candidates. However, there were a few candidates who wrote the whole answer rather than the letter. Noticeably this occurred only with candidates whose overall listening band score was 5 or lower.

4.2.4b Without a choice of possible answers

The rubric varied considerably for this item type. In Version 1, questions 21 to 25 the rubric was quite lengthy as shown below.

> Listen to the discussion about the numbers of people Stanley originally wanted to include in his research and the numbers Dr Long and he finally

> decide on. Complete the table by writing in the numbers against questions 21 to 25.

The percentage of incorrect responses ranged from 35.6% to 82.2%. These high percentages could be attributed to the fact that the answers are numbers. During student interviews students spent a considerable time reading this question. Responses included:

Student 3: "Listen to the discussion (pause) and complete the table (pause) I have to write a numbers".
Student 10: "Write a number".

But other students clearly did not realise a number was required for each question.

Student 6: "You listen you catch some words to write in this one (pause) How many words (pause) no [Looking at previous page] no it doesn't say".
Student 8: "Listen to the discussion and fill in the gaps. It doesn't say how many words".

From the interviews there was some doubt as to what was required. The fact that a number was necessary for each question should be clearly stated, possibly in bold.

The rubric for Version 1, questions 26 to 31 was clearly stated:

> Complete the notes. Write NO MORE THAN THREE WORDS for each answer.

Although the percentage incorrect was high, ranging from 34.2% to 69.9%, this may be in keeping within the guidelines that the items become more difficult as candidates proceed through the test. From the candidate answer papers they appeared to understand the question. The percentage who gave no response was disparate, ranging from 2.7% to 39.7%.

Similarly, for Version 1, questions 32 to 41 the rubric was clearly stated. Student interviews illustrated that all students knew what was expected.

Student 3: "Fill in the gap and no more than three word again".
Student 6: "Also write no more than three words".

In Version 3, questions 11 to 17 a similar rubric to those above was used. The incorrect response rate varied from 10.6% to 64.6%. Once again, analysis of candidate answer papers indicated that they comprehended the rubric.

Version 3, questions 31 to 33 contained a simple rubric.

> Complete the notes.

However, there is a lot of reading involved and this may have caused problems when processing information from the tape. The incorrect response rate ranged from 47.2% to 62.7%. In total, a third of candidates gave no response. Analysis of candidate answer papers showed that candidates had written one, or up to three words, and student interviews also

suggest that they understood the rubric. However, there was uncertainty about how many words to use.

Student 3: "Fill in the gap...you can use many words as much as we can. It doesn't mention".

Student 6: "I think I can take some note in here, but I don't know how many words".

While two students looked to the top of the page and responded with no more than three words, the remainder clearly thought that it did not state how many words or that it did not matter. The inclusion of a rubric at the top of the page for all questions on that page is problematic, especially where the item type changes.

In Version 4, questions 4, 5 and 6, the rubric clearly stated:

> Look at the diary for the conference. Fill in the numbered spaces.
> Write NO MORE THAN THREE WORDS for each answer.

The incorrect responses were 52.5%, 25% and 27.5% respectively. The change in item type may have caused the high incorrect response rate for question 4. All candidates appeared to understand the rubric.

In Version 4, questions 10 to 13 the rubric was clearly written. It is possible that the lexis used may have caused some problems rather than the listening text. In particular, questions 11 and 12 caused problems. Even candidates with an overall listening band of 7.5 were incorrect for these questions. Upon analysis various forms of spelling were marked correct by some markers, but the same spelling was marked incorrect by others. Hughes (1989, 139) states, "in scoring a test of a receptive skill there is no reason to deduct points for errors of grammar or spelling, provided that it is clear that the correct response was intended". Both of these particular answers were place names. This could be a potential problem for students and may need to be given careful consideration in future item type answers.

While the rubric did not appear to pose any problems, some candidates had problems when they transferred their responses at the conclusion of the listening test. For example, Listening Version 1 - Question 3. The correct response is 27 Freke St, but students sometimes transferred their responses as follows:

Q3 27
Q4 Freke St

Candidates were marked incorrectly for this question and for future questions.

Student interviews support the claim that candidates understood this rubric.

Student 1: "Maybe I have to write the summary about which I heard before from the tape, but I must write down no more than three words...um and I must fill the blanks between the sentence".

Student 10: "Catch the word from the talking... no more than three words".

4.2.5 Labelling a diagram which has numbered parts

The only example of this item type was on the Listening Test Version 3, questions 18 to 20. The rubric stated:

> Look at questions 18 - 20. Identify where the buildings in the questions are located on the map below. Write the appropriate letter A-G in the box next to each question.

An example was also included with the rubric. There was a small number of candidates who gave no response and the overall percentage of incorrect responses was considerably high at 50.3%, 49.7% and 47.8% respectively. All candidates who responded answered with a letter to identify the building illustrating that there was no misinterpretation of the rubric.

Students interviewed also supported this claim.

Student 1: "I must write with a character A from G".

Student 2: "Find out the building in the question (pause) and choose which one is the building (pause) and fill in the table. Just put A, B, C, D, E".

All students understood that a letter was required for the answer.

4.2.6 Classification and Matching

There were no questions of this item type on any of these three versions of the listening test.

5.0 Conclusions and Recommendations

5.1 The first point to make about both the Reading and Listening sub-tests is that the rubrics are generally clear and concise with, in almost all cases, little ambiguity. The data, as presented in this research project highlights the fact that the rubrics are in fact rarely misunderstood. There are only minor areas which need tightening up, as outlined in the findings.

5.2 For multiple choice questions in the Reading sub-test, in which only one answer is required, the wording should be less ambiguous. The IELTS Specifications Handbook, February 1996, needs to take account of this. We would suggest the following:

> Choose the best answer for each question and write the appropriate letter, A-D, in boxes.....

Also in the listening sub-test the multiple-choice rubrics should be kept as concise as possible, so that they require minimal reading time.

5.3 Questions should follow the rubric guidelines, as stated in the IELTS Specifications, and be consistent. Generally, this is the case, but there are a few minor differences, notably in the short answer questions and notes/table/summary completion.

5.4 The Reading Specifications document states that "If too many item types appeared this might lead to processing problems". A large number of item types is less of a problem for Reading than for Listening, the latter demanding very quick mental processing of both the written and spoken word, with no second chance given. The IELTS Specifications (1996, 33) clearly state, "A variety of item types are used but items are grouped according to item type". This is an area that needs to be monitored carefully.

5.5 Markers need to be made more accountable. There are a number of instances in both the listening and academic reading sub-tests in which Markers have not followed the demands of the question, for example, in the number of words required, and instances in which the Marker has either marked an incorrect response correct or a correct response is marked incorrect. We would suggest that at the very least the Marker's signature and name should be written on the answer paper.

5.6 As mentioned in 5.3, short answer questions need to contain at least some rubric. Even if the question may clearly demonstrate how many items are required, we believe it is necessary to include a rubric such as:

Write ONE word.
Write TWO WORDS for 1 mark.

The candidates must be given simple, concise instructions so the requirements of this item type are explicit.

5.7 Negative and double negative questions should not be included in writing particular item types. Item writers should not include these as part of the wording for particular questions. These are tests of reading and listening not tests of grammar.

BIBLIOGRAPHY

Cohen, A.D. (1994) *Assessing Language Ability in the Classroom*, Massachusetts: Heinle and Heinle.

Huerta-Macias, A (1995) *Alternative Assessment: Responses to Commonly Asked Questions* in TESOL Journal, Autumn 1995 p8-11.

Hughes, A. (1989) *Testing for Language Teachers*, Cambridge: Cambridge University Press.

IELTS Specifications, Draft February 1996.

Madsen, H.S. (1983) *Techniques in Testing* New York: Oxford University Press.