IELTS™

5 An empirical investigation of the process of writing Academic Reading test items for the International English Language Testing System

Authors

Anthony Green
Roger Hawkey
University of Bedfordshire, UK

Grant awarded Round 13, 2007

This study compares how trained and untrained item writers select and edit reading texts to make them suitable for a task-based test of reading and how they generate the accompanying items. Both individual and collective test editing processes are investigated.

ABSTRACT

This report describes a study of reading test text selection, item writing and editing processes, with particular reference to these areas of test production for the IELTS Academic Reading test. Based on retrospective reports and direct observation, the report compares how trained and untrained item writers select and edit reading texts to make them suitable for a task-based test of reading and how they generate the accompanying items. Both individual and collective test editing processes are investigated.

For Phase 1 of the study, item writers were invited to respond to a questionnaire on their academic and language teaching and testing background, experience of IELTS and comments on its reading module (see Appendix B). Two groups of participants were selected: four officially-trained IELTS item writers (the experienced group) and three teachers of English for academic purposes who had prepared students to take IELTS, but had no previous experience of item writing for the IELTS Academic Reading module (the non-experienced group). In Phase 2 of the project both groups were asked to select and prepare texts and accompanying items for an IELTS Academic Reading test, and to bring their texts and items to separate interview and focus group sessions. In the first of these sessions, participants were interviewed on how they had selected and edited their texts and how they had generated the items. In a second session, the item writers worked in their two groups to further refine the texts and items to make them more suitable for the test (as the trained item writers would normally do in a test editing meeting).

The analyses of the texts and accompanying items produced by each group, and of the discussions at all the Phase 2 sessions have produced valuable insights into the processes of text selection, adaptation and item writing. The differences observed between the experienced and non-experienced groups help to highlight the skills required for effective item writing for the IELTS Academic Reading test, while at the same time suggesting improvements that could be made to the item production process so that it might more fully operationalise the IELTS reading construct.

AUTHOR BIODATA

DR ANTHONY GREEN

Has a PhD in language assessment. Is the author of *IELTS Washback in Context* (Cambridge University Press) and has published in a number of international peer reviewed journals including *Language Testing*, *Assessment in Education*, *Language Assessment Quarterly* and *Assessing Writing*. Has extensive experience as an ELT teacher and assessor, contributing to test development, administration and validation projects around the world. Previously worked as Cambridge ESOL Validation Officer with responsibility for IELTS and participated as a researcher in IELTS funded projects in 2000/1, 2001/2 and 2005/6. Current research interests include testing academic literacy and test impact.

DR ROGER HAWKEY

Has a PhD in language education and assessment, is the author of two recent language test-related books, *Impact Theory and Practice: Studies of the IELTS test and Progetto Lingue 2000* (2006) and *A Modular Approach to Testing English Language Skills* (2004). Has experience of English language teaching, program design and management posts and consultancy at secondary, teacher training and university levels, in Africa and Asia, Europe and Latin America. Research interests include: language testing, evaluation and impact study; social, cognitive and affective factors in language learning.

CONTENTS

1 AIMS

This research report describes a study of reading, test text selection, item writing and editing processes, areas of test production that have rarely been transparent to those outside testing organisations. Based on retrospective reports, direct observation and analyses of the texts produced, the report compares how trained and untrained item writers select and edit reading texts to make them suitable for a task-based test of reading and how they generate the accompanying items. Both individual and collective editing processes are investigated. The analyses in the study are expected to inform future high-stakes reading test setting and assessment procedures, in particular for examination providers.

2 BACKGROUND AND RELATED RESEARCH

2.1 A socio-cognitive test validation framework

The research is informed by the socio-cognitive test validation framework (Weir 2005), which underpins test design at Cambridge ESOL (Khalifa and ffrench, 2008). The framework, further developed at the Centre for Research in Language Learning and Assessment (CRELLA) at the University of Bedfordshire, is so named because it gives attention both to context and to cognition in relating language test tasks to the target language use domain. As outlined in Khalifa and Weir (2009) and Weir *et al* (2009a and 2009b), in the socio-cognitive approach difficulty in reading is seen to be a function of 1) the complexity of text and 2) the level of processing required to fulfil the reading purpose.

In Weir *et al* (2009a) IELTS texts were analysed against 12 criteria derived from the L2 reading comprehension literature (Freedle and Kostin 1993, Bachman *et al* 1995, Fortus *et al* 1998, Enright *et al* 2000, Alderson *et al*, 2004 and Khalifa and Weir 2009a) These criteria included: Vocabulary, Grammar, Readability, Cohesion, Rhetorical organisation, Genre, Rhetorical task, Pattern of exposition, Subject area, Subject specificity, Cultural specificity and Text abstractness. In the current study, we again employ such criteria to consider the texts produced by item writers and to analyse the decisions they made in shaping their texts.

In Weir *et al* (2009b) the cognitive processes employed by text takers in responding to IELTS reading tasks are analysed, with a particular focus on how test takers might select between expeditious and careful reading and between local and global reading in tackling test tasks.

Local reading involves decoding (word recognition, lexical access and syntactic parsing) and establishing explicit propositional meaning at the phrase, clause and sentence levels while global reading involves the identification of the main idea(s) in a text through reconstruction of its macro-structure in the mind of the reader.

Careful reading involves extracting complete meanings from text, whether at the local or global level. This is based on slow, deliberate, incremental reading for comprehension. Expeditious reading, in contrast, involves quick, selective and efficient reading to access relevant information in a text.

The current study was expected to throw light on how the item writers might take account of the processes engaged by the reader/ test taker in responding to the test tasks and how item writers' conceptions of these processes might relate to reading for academic study.

2.2 Item writing

Item writing has long been seen as a creative art (Ebel 1951, Wesman 1971) requiring mentoring and the flexible interpretation of guidelines. This has been a source of frustration to psychometricians, who would prefer to exert tighter control and to achieve a clearer relationship between item design characteristics and measurement properties. Bormuth (1970) called for scientifically grounded,

algorithmic laws of item writing to counter traditional guidelines that allowed for variation in interpretation. Attempts at standardisation have continued with empirical research into the validity of item writing rules (Haladyna and Downing 1989a and 1989b); the development of item shells – generic items with elements that can be substituted with new facts, concepts or principles to create large numbers of additional items (Haladyna 1999); and efforts to automate item generation (Irvine and Kyllonen 2002). Numerous studies have addressed the effects of item format on difficulty and discrimination (see Haldyna and Downing 1989a, Haladyna, Downing and Rodriguez 2002) and guidelines have been developed to steer test design and to help item writers and editors to identify common pitfalls (Haladyna and Downing, 1989a, Haladyna 1999). For all this, Haladyna, Downing and Rodriguez (2002) conclude that item writing remains essentially creative as many of the guidelines they describe remain tentative, partial or both.

Yet stakeholder expectations of evidence-based, transparently shared validation for high-stakes language exams are increasingly the order of the era (see Bachman, 2005, and Chalhoub-Deville, Chapelle, and Duff (eds), 2006) often specified through codes of practice (eg, ALTE, 1994). Rigour is increasingly expected of item-writer guidelines in the communicative language skills testing sector. The new Pearson Test of English (PTE), due in 2009, aims, like IELTS, to provide language proficiency scores, including reading measures for colleges, universities, professional and government bodies requiring academic-level English. de Jong (2008) proposes an analysis, for PTE item writer training purposes, of item types (14 potentially applicable to the testing of reading) and a schema for item writer training structured around a general guide, item specific instructions, reference materials, codes of practice, an item writer literature review and the Common European Framework of Reference (CEFR). Cambridge ESOL's own framework for the training and development of item writers is referenced in some detail below.

A number of handbooks include guidance on item design and quality assurance issues in language tests (eg, Valette 1967, Carroll and Hall 1985, Heaton 1990, Weir 1993, Norris *et al* 1998, Davidson and Lynch 2002, Hughes 2003). These provide advice on the strengths and weaknesses of various item formats and stress the need for item review and piloting. It is generally taken as axiomatic that trained test item writers are superior to the untrained (Downing and Haladyna 1997).

While the focus of research has been on the characteristics of items, very little attention has been given to the processes that item writers go through in creating test items and the contributions that these may make to the quality of test material. In a rare piece of research focusing on this area, Salisbury (2005) uses verbal protocol methodology and a framework drawn from the study of expertise to explore how text-based tests of listening comprehension are produced by item writers

Salisbury (2005, p 75) describes three phases in the work of the item writer:

- Exploratory Phase: 'searching through possible texts, or, possibly, contexts'
- Concerted Phase: 'working in an intensive and concentrated way to prepare text and items for first submission'
- Refining Phase: 'after either self-, peer- or editor-review, polishing/improving the test paper in an effort to make it conform more closely to domain requirements'

She found that in comparison to novices, more expert item writers, those producing more positively evaluated texts and items that met the requirements of the test developers (UK examining boards offering tests of English as a Foreign Language):

- are more aware of the test specifications and are quickly able to recognise texts that show potential as test material. Where novices tended to devise a listening script from a source text first and then to write the questions, experts were more inclined to start from the questions and then to build a script to fit with these

- are more aware of the needs of candidates for clear contextual information and are better able to provide accessible contextualising information in the form of short, accessible rubrics and co-text
- explore a range of possible task ideas rather than committing immediately to one that might later prove to unworkable
- use many more learned rules or ruses than non-experts including, for example:
- exchanging words in the text and in the question so that the hypernym appears in the text
- adding additional text to the script to introduce distraction and reduce the susceptibility of the questions to guessing strategies

Although more experienced item writers tended to outperform the recently trained, expertise was not simply a function of experience. One writer with no previous experience of test item writing performed better in the judgement of a review panel than two item writers with extensive experience (Salisbury 2005). Salisbury also concludes that expertise in Listening test item writing is collective in nature. Individual writers rarely have sufficient capability to meet institutional requirements at the first attempt and need the feedback they receive from their colleagues to achieve a successful outcome. It might be added that item writer expertise itself is not sufficient to guarantee test quality. Even where items are subject to rigorous review, piloting usually reveals further deficiencies of measurement.

The Cambridge ESOL approach to test development is described in detail by Saville (2003) and by Khalifa and Weir (2009). The IELTS test production process for the reading and listening papers is outlined in a document available from the IELTS website, www.ielts.org. The goal of this test production process is that 'each test [will be] suitable for the test purpose in terms of topics, focus, level of language, length, style and technical measurement properties' (IELTS 2007, 1).

IELTS test material is written by freelance item writers externally commissioned by Cambridge ESOL in a process centrally managed from Cambridge and carried out according to confidential test specifications or item writer guidelines laid down by the test developers (although see Clapham 1996a, 1996b for an account of the role of externally commissioned item writing teams in developing the IELTS academic reading module). These guidelines, periodically modified to reflect feedback from item writers and other stakeholders, detail the characteristics of the IELTS modules (speaking, listening and academic or general training reading and writing), set out the requirements for commissions and guide writers in how to approach the item writing process. The guidelines cover the steps of selecting appropriate material, developing suitable items and submitting material. However, a good deal of the responsibility for test content is devolved to the externally commissioned workers including the item writers and their team leaders or chairs for each of the modules. Khalifa and Weir (2009) describe the chair as having responsibility for the technical aspects of item writing and for ensuring that item writers on their team are fully equipped to generate material of the highest quality.

According to the Cambridge ESOL website (Cambridge ESOL n.d.) the overall network of Cambridge item writers working across the Cambridge ESOL product range includes 30 chairs and 115 item writers. Reflecting the international nature of the examination, Cambridge ESOL employs teams of IELTS item writers in the United Kingdom, Australia, New Zealand and the USA.

There are one or two commissions each year for each item writing team (IELTS 2007). The writers are commissioned to locate and adapt suitable texts 'from publications sourced anywhere in the world' (IELTS 2007, 1). This work is carried out individually by item writers who may adapt their sources to meet the requirements of the test. Khalifa and Weir (2009) list a number of reasons for an item writer to adapt an original text. These are drawn from the Item Writer Guidelines 2006 for general English examinations (KET, PET, FCE, CAE and CPE) produced by Cambridge ESOL (the organisation that is also responsible for producing IELTS) and include:

- cutting to make the text an appropriate length
- removing unsuitable content to make the text inoffensive
- cutting or amending the text to avoid candidates being able to get the correct answer simply by word matching, rather than by understanding the text
- glossing or removing cultural references if appropriate, especially where cultural assumptions might impede understanding
- deleting confusing or redundant references to other parts of the source text
- glossing, amending or removing parts of the text which require experience or detailed understanding of a specific topic

Item writers submit their material in draft form for review at a preliminary pre-editing meeting. This meeting involves the chairs of the item writer teams, experienced item writers and Cambridge ESOL subject officers – members of staff with overall responsibility for the production, delivery and scoring of specific question papers. Green and Jay (2005) describe how 'at this stage, guidance is given to item writers on revising items and altering texts, and feedback is provided on rejected texts and/or unsuitable item types.' This step is identified by the IELTS partners as an important element in item writer training because advice is given by the pre-editing team on reasons for rejecting or refining texts and on the suitability of proposed item types (IELTS 2007).

Pre-edited material is returned to the item writer together with comments from the pre-editing panel. If the text has been evaluated as potentially acceptable for test use, the item writer then prepares an adapted version with accompanying items ready for inclusion in a test form. The modified material is submitted to an editing meeting, which takes place centrally and, in addition to the writer concerned, involves Cambridge ESOL staff and the chair. According to the IELTS partners (IELTS 2007, 2) 'item writers are encouraged to participate in editing meetings dealing with their material.' because this further contributes to their professional development as writers. Khalifa and Weir (2009) describe the aims of editing as follows:

- to check or re-check the quality of material against specifications and item writer guidelines
- to make any changes necessary to submitted materials so that they are of an acceptable standard
- to ensure that the answer key and rubrics are appropriate and comprehensive
- to further develop the skills of item writers in order to improve the quality of materials submitted and the input of item writers to future editing sessions

Following editing, material either passes into the IELTS test bank for inclusion in pre-tests to be trialled with groups of test takers, or is returned to the item writer for further revision and another round of editing. Pretests are administered to groups of students at selected IELTS centres and data is obtained indicating the measurement characteristics of the test items. A further meeting – the pre-test review meeting – is held to consider the item statistics and feedback from candidates and their teachers. Texts are submitted for pretesting with more questions than will appear in the final version and those items that fall outside target difficulty ranges or that have weak discrimination are eliminated. Again at this point any unsatisfactory material may be rejected.

All IELTS item writers are said to receive extensive training. Ingham (2008) describes the standard processes of recruitment and training offered to item writers. This takes place within 'a framework for the training and development of the externals with whom [Cambridge ESOL] works in partnership.

The framework has the acronym RITCME: Recruitment; Induction; Training; Co-ordination; Monitoring and Evaluation'. To be recruited as item writers, individuals must have a university degree, a suitable qualification in English language teaching and five years' teaching experience together with some familiarity with materials production and involvement in preparing students for Cambridge ESOL examinations (Ingham 2008). After completing a screening exercise and preparatory tasks (induction), successful applicants are invited to complete a 'training weekend' (Ingham, 2008, 5) with Cambridge staff and external consultants. The Cambridge item writer trainers work with between 12 and 16 trainees, introducing them, *inter alia*, to item writing techniques, issues specific to the testing of different skills and the technical vocabulary used in the Cambridge ESOL context.

After joining the item writing team for a specific paper such as the IELTS Academic Reading paper, writers 'receive team-specific training before they start to write' (Ingham 2008, 6). They are invited to further training sessions with their team, led by the chair, on an annual basis. In time, successful item writers gain work on additional products to those for which they were originally recruited and may progress in the hierarchy to become chairs themselves. Less successful writers who fail to generate sufficient acceptable material are offered support, but according to Salisbury (2005, 75) may 'gradually lose commissions and eventually drop from the commissioning register'.

Salisbury (2005) points out that the role of the item writer appears, superficially, to be limited to delivering material in line with predetermined requirements. However, it is also widely recognised that formal written specifications can never be fully comprehensive and are always open to interpretation (Clapham 1996a, Fulcher and Davidson 2007). Perhaps inevitably, what Salisbury (2005) describes as 'non-formalised specifications', representing the values and experience of the item writing team and subject officers, emerge to complement the formal set provided by the test developers. These non-formal specifications are less explicit, but more dynamic and open to change than the item writer guidelines. We have already noted that in the Cambridge ESOL model, elements of these non-formal specifications can become formalised as regular feedback from item writers informs revisions to the guidelines. Item writers are therefore central to the IELTS reading construct.

Khalifa and Weir (2009) point to the critical importance of professional cultures or communities of practice (Lave and Wenger, 1991) within a testing body such as Cambridge ESOL. They suggest that question paper production perhaps depends as much on the shared expertise and values of the item production team as on the procedures set out in item writer guidelines. All members of this team, whether they be internal Cambridge ESOL staff or external consultants, bring their own expertise and experience to the process and shape its outcomes at the same time as their own practices are shaped by the norms of the established community that they are joining.

While a number of language test development handbooks offer advice on suitable item types for testing reading and suggest criteria for judging test items (Weir 1993, Alderson 2000, Hughes 2003) the work of the item writer remains under-researched. Studies have been undertaken to investigate the thought processes involved on the part of candidates in responding to IELTS test tasks (Mickan and Slater 2000, Weir *et al* 2009a and 2009b) and on the part of examiners in scoring IELTS performance (Brown 2003, 2006, Furneaux and Rignall, 2007, O'Sullivan and Rignall 2007), but no research is yet available on how IELTS item writers go about constructing test items and translating test specifications into test tasks.

3 RESEARCH METHODOLOGY AND DESIGN

3.1 Deduction and induction

The review of previous research and current theory and practice related to high-stakes test item-writing underlines the complexity of the process. Its investigation is likely to involve qualitative as well as quantitative data collection and analyses, inductive as well as deductive approaches. In the analysis of the reading texts selected and adapted by our participants, for example, models already

established are used deductively to produce theory-based quantitative measures of difficulty, word frequency and readability – for example the Academic Word List (AWL) (Coxhead 2000), word frequency levels based on the British National Corpus (BNC) (Cobb, 2003) and indices of readability (Crossley *et al* 2008).

However, for the participant discussions relating to text search, selection, adaptation, item writing and item editing (audio-recorded with the permission of the participants) a generally inductive approach to data analysis is used. In this process observations are made with the expectation of contributing qualitative insights to a developing theory, seeking processes and patterns that may explain our 'how' and 'why' questions. Patton (1990, p 390) sees such inductive qualitative analysis as permitting patterns, themes, and categories of analysis to 'emerge out of the data rather than being imposed on them prior to data collection and analysis'. Dey (1993, p 99) finds that induction allows a natural creation of categories to occur with 'the process of finding a focus for the analysis, and reading and annotating the data'. As our description of the project's discussion sessions in Section 6 below will indicate, the analysis 'moves back and forth between the logical construction and the actual data in a search for meaningful patterns' (Patton, 1990, p 411). The meaning of a category is 'bound up on the one hand with the bits of data to which it is assigned, and on the other hand with the ideas it expresses' (Dey, 1993, p102).

3.2 Design

The research was undertaken in two phases. In the first, an open-ended questionnaire (see Appendix B) was distributed to the item writers accepting our invitation to participate. Questionnaire respondents included all seven Phase 2 participants and three other experienced item writers from the UK, Australia and New Zealand. The instrument elicited data relating to their background and experience, served to contextualise the second, in-depth focus group phase of the study and informed the analyses of the item writer interview and focus group sessions described below.

Two groups of item writers were involved in these sessions. One group consisted of four trained IELTS item writers. This required the cooperation of Cambridge ESOL in facilitating contact with item writers able to participate in the research, permitting their involvement and in providing the researchers with access to the item writer guidelines for the academic reading paper. As the guidelines are confidential we were asked not to discuss them in detail or to quote from them in this report.

The second group included three teachers of English for academic purposes with a range of experience of the IELTS test and of IELTS preparation but no previous experience of writing reading test items for an examinations board. These teachers were familiar with the appearance of the test, but not with its underlying design.

Data collection took place over two sessions. On the basis of Salisbury's (2005) division of the item writing process into exploratory, concerted and refining phases, the first session concentrated retrospectively on the exploratory phase and prospectively and concurrently on the concerted phase (see above). In the second session the item writers worked as a group to further refine their texts and items to make them more suitable for the test (as the trained item writers would normally do in an actual test editing meeting). In Salisbury's terms, this session may be said to have been concerned retrospectively with the concerted phase and prospectively and concurrently with the refining phase.

In preparation for Phase 2, each participating item writer was sent a commissioning letter (Appendix A), based on a model provided by Cambridge ESOL, inviting them to choose a text that would be suitable for use in IELTS, to edit this text as appropriate and to write 16 or 17 test questions to accompany the text.

In the first session of Phase 2, we sought insights into the strategies that item writers use in selecting and preparing texts and the role that the test specifications, experience and other sources of knowledge might play in this process for experienced and inexperienced writers. Writers were interviewed about

their selection of texts for item writing purposes. Key questions for this session included how item writers select texts, how they adapt the texts to shape them for the purposes of the test and how they generate items. The focus was on the specific text selected by the item writer for this exercise, the features that made it attractive for the purpose of writing IELTS items and the edits that might have been required to shape the text to meet the requirements of the test.

The second session of Phase 2 was similar to an IELTS editing meeting (see above). Item writers brought their texts and items to the focus group to discuss whether these did, as intended, meet the requirements of the test. Again, observation of differences between the experienced and inexperienced writers was intended to provide insights into the practices of those item writers working within the IELTS system for test production. Here the researchers sought to understand the kinds of issues that item writers attend to in texts prepared by others, the changes that they suggest and features of texts and test questions that are given approval or attract criticism. Once again, the analyses of the deliberations linked themes and categories emerging from the recordings and transcripts to the insights provided by the socio-cognitive framework Weir 2005, Khalifa and Weir 2009, Weir *et al* 2009a). It was expected that differences between the experienced and non-experienced groups would highlight the practices of item writers working within the IELTS system for test production and the nature of their expertise. As will be seen below, the study provides insights into how item writers prepare texts and items, and their focus of attention in texts prepared by others; also into the features of texts and test questions that attract approval or criticism in editing.

4 ANALYSIS AND FINDINGS FROM INTERVIEWS AND FOCUS GROUP DISCUSSIONS

4.1 Non-experienced IELTS item writer group

Session 1: participant discussion of their experience with their commission to select an appropriate IELTS academic reading text, edit and adapt for testing purposes and generate test items

This first information collection exercise was organised as a researcher-led discussion session. Here participants discussed their experience with their commission to select an appropriate IELTS academic reading text, edit and adapt it for testing purposes and generate test items. Each of the participants in turn (see Table 10 in Appendix B for cv and other information on them) was first invited to describe the processes through which an 'IELTS' text was selected and adapted, then reading test items created. The intended ethos was participant-centred and informal, with discussion welcomed of each participant's initial account of the experience concerned. Both researchers were present but played a low-key role, intervening infrequently and informally. All proceedings were recorded (see above).

4.1.1 IELTS text search, selection and characterisation

The experiential information provided orally by the three participants on the selection of potential reading texts for IELTS use during the first discussion session of the day is summarised in Table 1, which analyses responses by the three participants according to criteria emerging from the analysis of the transcripts made by the researchers.

Source/ influence?		*Item Writer*	
	Victoria	Mathilda	Mary
Own interest	√	√	
Other's interest		√	
Web	√	√	√
IELTS website	√	√	
Published IELTS papers	√	√	√
Magazines, journals	√	√	
Newspapers			√
Bookshops			√
Environment topics		√	

Table 1: Non-experienced participants: Sources of, and influences on IELTS academic reading module text selection

Table 1 may be read, for example, as indicating that, in their accounts of IELTS text selection, both Victoria and Mathilda (all participant names used throughout this report are pseudonyms) referred in the discussion to using, among other sources, magazines and journals in their selection of suitable texts. For all three participants, it will be noted from the table (and the three non-experienced item writers' own flow-charts of their whole item-writing process, from text search to adapted text and accompanying items in Table 3) that topic interest and web searches are key initiating steps, as is public domain IELTS information accessed via the IELTS website and IELTS test preparation material.

Table 2 below summarises the characteristics of target IELTS-type texts as interpreted by the three participants and the number of mentions of each as counted from the transcript of the discussion. It will be noted from the table that IELTS texts tend to be perceived as likely to be on subjects of popular interest presented in a formal, report-like format, academic in tone, but not so technical that non-specialist readers would be handicapped in understanding them. The three participants differ interestingly across the text criterial characteristics used in Table 2 as potentially significant in this part of the discussion. Mary, for example, is apparently more concerned with the characteristics of IELTS texts from an assessment point of view. Victoria, perhaps influenced by her experience as an IELTS writing paper Assistant Principal Examiner, appears more confident in her interpretation of what IELTS texts are like than the other two non-experienced item writers (see her generally higher criterion counts).

4.1.2 Participant text search treatment and item development: flowcharts and discussions

We now analyse more qualitatively the non-experienced item writers' discussion session of their item writing processes. These deliberations had been recorded, transcribed and coded by topic before the quantitative summary analysis as presented in Tables 1 and 2 above. Table 3 below summarises the more qualitative inductive description here, allowing further inferences to be drawn on the processes involved in efforts by the three non-experienced item writers to locate and select potential IELTS academic reading texts. The submitted materials – texts and accompanying items – are provided in Appendix C.

Perceived IELTS text characteristics	*Item writer*		
	Victoria	Mathilda	Mary
Academic	7	2	
Report	1		
Descriptive /conceptual	2	1	3
Impersonal, hedging	2		1
Pop-scientific/current	1	2	1
Not too specialist	1	2	
Technical but not too	2	1	
Literary	1	2	
Not journalistic / news item	1		1
Avoidance of bias, offence	4		2
Of an assumed difficulty			3
Length			3
Grammar			
Cohesion	1		1
Range/ complexity			2

Table 2: Non-experienced participants: Perceived characteristics of IELTS Academic Reading module texts

The three were asked to sketch flowcharts of the ways they had located, edited and prepared items for their IELTS Reading tests, after which they were invited in turn to explain their flowcharts (see Table 3). It was intended in the design of the study that this activity would provide internal triangulation for the findings of the previous discussion by the participants of their experience in selecting and characterising an appropriate IELTS Academic Reading text, editing and adapting for testing purposes. This proved indeed to be the case. The main points made by the three participants in their discussions of their flowchart are summarised in Table 3 under the headings: text search, editing and item writing, with a final question on their preferred items. The table should be read both for the similarities and for the differences in the processes engaged in across the three participants.

Text search		
Victoria	***Mathilda***	***Mary***
5-6 step flowchart (Victoria thinks now there are more steps than in her flowchart) 1. task familiarisation 2. topic selection (based on knowledge from past papers, website, course books) 3. begin task to determine suitability 4. research topic to test credibility and usefulness of text 5. satisfied with text 6. editing text for cohesion and text type Googled neuro-linguistic programming (NLP) and other potential topics > decided on topic of content of dreams > refining down topic > sub-topics within dreams > other articles > also possible choices? > so settled on the dreams text > tried items out on her EL1 partner; 'apparently NS do really badly on IELTS reading'	5-main steps in flowchart 1. looking at sample IELTS texts 2. browsing for a suitable text 3. selection of text from shortlist 4. text adaptation 5. selecting parts of text to target and writing questions / tasks based on the example of the sample tests Used practice IELTS tests (and her own experience as a candidate) Googled scientific magazines first 'then within the magazines I looked for specific things'… 'you get articles related to it then do a search on words related to it'	6-step flowchart: 1. task assessment, 2. background research 3. text search and rejection 4. text decision and editing 5. text review 6. item writing and text adjusting Used *IELTS Express, Impact IELTS,* past papers, old IELTS copies (Internet) searched under variety of topics, 'try to refine, refine, refine' eg, *science and nature*, down to *robot*s, 'using more and more refined words in order to be able to find an article that would be suitable ' tested text and items on friend

Text editing		
Victoria	***Mathilda***	***Mary***
Believes in significant 'fixing up process' on text Did various things to make the text more academic: took out by-line, added more research-type 'rigour' (eg, evidence-based), more hedging	Mathilda made hardly any changes: about three words	Text editing can mean: 'changing text structure, paragraphing, cohesion' Didn't want to change text too much but one person's 'changing a lot' not another's? Different texts need different amount of changing; editing is relative Is text editing for the sake of the tasks, changing text to fit a task type …. a validity issue?

Item writing		
Victoria	***Mathilda***	***Mary***
Knew the 10 task types, returned to IELTS Website hand out re format and stylistic aspects of task types her 'fixing up' of the text 'summons up the kind of task types there are'; so she could see eg, MCQ, wanted to do a Y?N?NG (students 'have a hard time with NG'; ended up doing another type as well she 'forgot to stop'. text very 'driven by definitions, which lend themselves to 'confusing test-takers'; so a lot of her MCQ definitional; test-takers can be led astray by MCQ text bits adjacent to the term; MCQ items testing whether Cs 'have kept up with the order'; linked items with reading purposes eg, careful reading where you have to 'go back to text and work hard to understand it' MCQ distractors of similar lengths but not necessarily the same style? tried to keep the items in the order of the text as with IELTS wished there were only 3 alternatives; 4th just an 'add on', 'just rubbish', easy for test-taker to spot asks 'can you use words that you know, not in the text'; Must it be ion the text? What's the rule? Victoria not much practice in SAQs; too many alternative responses; hard to generate all possible answers	Looked at task types (IELTS website says 10 different types) checked which would suit the text deciding which bits of info in text or which passages to summarise, making decisions on that in parallel; back and forth at same time decided to use matching paras with short summaries task as …more suitable' for this type of text used true / false / not given task …'put in a few correct ones, made up a few others' eg, collapsing info 'that did not really go together …' to reveal lack of understanding Tested vocab. eg, 'if you don't know that adjacent means next then you don't know whether that info is correct or not…' i MCQ suitable task for text as it has text lots of straightforward info suitable? relatively easy finding distractors: easy to find similar info which could be selected 'if you don't look properly or if you understood it half way' found a fine line between good and bad distractors, and also between distractors 'which could also be correct … because the text might suggest it and also because …. you could actually accept it as a correct answer' marked up text suitable for items ie, that seemed important for overall understanding and 'for local, smaller bits of info where I thought I would be able to ask questions'; then made up items, vocab, others asking for longer stretches as text 'sort of like offered itself'. Adjusting if she felt that they were either too easy (distractors obviously wrong , didn't really test anything or item wording did not make clear what I mean) Regrets not testing items with someone. 'if you …. word them and reword them and go over them again you ….lose touch with it and don't really understand it yourself anymore' Threw away only one or two items but modified about half or her original items Thought the Website said all the items are in the order they are in in the text Short answer questions (SAQs) may be good for definitions, too	Matching task (paras with researcher names) selected to test summary of main text topics summary completion task suited density of description of an experiment short paraphrasal text with candidates to use words from text in new context, to check their understanding didn't just want to test vocab. meaning; tried to elicit specific answers favoured the control offered by multiple choice (MCQ) but now felt she should have been more careful in designing distractors often had difficulty finding the 4th alternative should there be distractors not actually in the text but from test designer's mind? should we actually add to text to get distractors? Mary thinks no as it impairs authenticity never threw any questions away, but did dispense with 'a couple of distractors' IELTS items do not have to be in the order the item topic appears in the text?

Which of your sections you happiest with?		
Victoria	***Mathilda***	***Mary***
likes her T/F NG – it works Stylistically her MCQ wrong because the items are of uneven length, though the questions are 'sort of OK' In her SAQs she is not convinced the answers are the only ones possible	MCQ strongest, not a NS so can 'imagine what it's like' so easier to 'make up the wrong ones'! Task type 7, summary info to match paras, too vague, so her worst	matching (sentences to researcher names) the best. summary completion task the easiest to write so perhaps the worst! MCQ task actually the worst because of her difficulty finding the final distractors summary completion the easiest – so the worst No her first section (the matchings)

Table 3: Non-experienced participants descriptions of the item writing process

Item writer Victoria had begun by visiting the official IELTS website for information and samples of academic reading module topics and task types. She then, like all the three untrained participants, carried out an internet search for potential topics which she had already identified (there were six of these) and selected the one of most interest to her, ie, neuro-linguistic programming. The text on this, however, she rejected as 'too technical, too specialist', as she did her next text, on the Japanese tea ceremony, which though 'a really pretty text', she found too 'instructional', and – a common theme in text selection – biased in favour of particular candidate groups. Victoria's final choice she rated immediately as the kind of 'really studious' topic 'that IELTS uses', namely: 'How the Brain Turns Reality into Dreams' (see Section 7 below for the full description of the text concerned). For Victoria, the search was about 'choosing a text, looking at it, deciding what I can do with it'.

Victoria, as we shall see emphasised in the next section, was from the outset viewing prospective texts in terms of what she could do with them to make them suitable as IELTS texts with appropriate tasks to go with them. The Dreams text she found right because it was 'pseudo-scientific', a view shared by all three in the group as characterising IELTS texts (see below) and, significant for our discussions of test text adaptation in the section below, because it 'lent itself to being fixed up' (Victoria's frequent term for adapting texts).

Mathilda confessed to being initially unsure of the level of difficulty and complexity of IELTS reading texts. Her visit to the IELTS Website suggested to her 'sort of' scientific texts but not too specific, specialist; 'a bit more populist, kind of thing'. She then carried out a search, guided by topics fitting this construct, and which were 'very up-to date' and which 'nowadays should interest most people'. She thus used search terms such as 'environment' and 'future' but rejected several texts as too specialist, too material-intensive given the IELTS reading time limit. Mathilda saved four possible texts and made her final choice, of the one on environmentally friendly cities of the future, which she found engaging, information rich and apparently suitable for test questions.

Mary found the text search time-consuming and quite difficult. She had started by checking with IELTS tests in the Cambridge Practice Tests for IELTS series, focusing in particular on their subject matter. She had then searched in magazines such as the New Statesman, the Economist and the New Scientist, as well as newspaper magazine sections. Articles from these sections she rejected because of their length (Mary 'would have struggled to edit down'), complexity or cultural bias. Mary pursued the topic of robots online after reading a newspaper article on the subject, although this had been much too short for IELTS purposes. Mary then searched the BBC website without finding texts she felt she would not have to edit too heavily – something (see below) Mary expressed particular antipathy towards doing. Finally, through Google News, Mary found an article on robots which she considered at the right level of difficulty, grammar and range: expressing opinions, yet with an appropriate descriptive element. The piece Mary said 'would have been something I would have read at uni. had I studied anything like this!'

4.1.3 Participant focus group discussions

The non-experienced group participated next in a focus group discussion structured around a set of nine semantic differential continua (Osgood, 1957) using the unlabelled scale format (compared with other formats by Garland, 1996) and as seen in Table 4 below. In the table, summaries of the comments made by the participants in their 25 minutes of unmediated discussion are placed in their approximate location on the continua for the nine scales. The adjectives for the continua were selected by the researchers.

clear	choosing texts (Victoria, Mary)	IELTS reading texts supposed to be at three different levels (Victoria) Balancing general vs specific items (Mary)	getting texts the right level (Mathilda) whether items should be in order of the text (Mary) guidelines on the target reading construct?	designing 4 good MCQ distractors (Mary, Victoria, Mathilda) lack of guidelines on how tasks are made and assessed (Mathilda, Mary, Victoria)	confusing
interesting	achieving good text and items (Victoria, Mary) Writing items (Mary) literary, fiction texts would be (Mathilda) but might not be appropriate (Mary, Victoria)	trying to drive the process, not letting the text drive it (Victoria)	finding the text (Mary) informative texts (Mathilda) finding items (Mathilda)		dull
time-consuming	everything! (Mary) looking for texts (Mathilda)	developing items (Mary) editing (Mary, Victoria)	editing (Mathilda)		quick
rewarding	finally finding the right text (Victoria, Mary) finishing everything (Victoria, Mary, Mathilda)	driven by possibility it will be used as a 'real' test (Victoria)	unsure whether doing it right (Mathilda, Mary)	no-one's going to answer the items (Mary, Victoria) no feedback, no knowledge underneath the task they're doing (Mary, Victoria, Mathilda)	unrewarding
worrying	not knowing if they are doing it right (Mathilda, Mary)	worrying about the right level (Mary) not being privy to the process of editing, trialing(Victoria)			pleasing
creative		whole process of creating items, driving the process oneself (Mary) making up credible distractors (Mathilda)	straightforward informational text (Mathilda) forcing in distractors (Mary)		programmatic
	The creative is constrained by the programmatic (Mathilda, Mary, Victoria)				
challenging	Creating a viable 4th distractor in MCQ (Victoria, Mary)	forcing text into particular task types (Victoria) how much to edit (Mary) matching text and task types (Mathilda)	choosing task types (Mary)		straightforward
frustrating	finding the right text (Mary)	making items for the matching tasks (Mary)	completing the matching task (Mary) perfecting answer keys for SAQ task (Victoria)	finishing preparation and editing of a good, cohesive text (Victoria)	satisfying
supported	feedback of friend useful (Mary) Topic checks with friends (Victoria) IELTS materials vital (Mary, Mathilda)		Mathilda didn't know she could seek help Too little help on level of difficulty (Mathilda) needed more samples and guidelines for texts (Mathilda)	Item writer guidelines confidential (Victoria)	unsupported

Table 4: summary of non-experienced participant focus group comments and ratings on semantic differential scales

The points made by the three participants in the focus group discussion certainly served as triangulation for the views they had expressed in the preceding IELTS text search and treatment and item development: flowcharts and discussions already reported. Once again we see strong evidence of time-consuming searching for suitable texts but uncertainty of the target level(s) of such texts and, to some extent, the topic range; major problems with the design of tasks, in particular multiple choice (MCQ) items and, as might be expected of this non-experienced item writer group, frustration caused by lack of item writing guidance.

The research team pursued with the participants certain emerging issues immediately after the end of the participant-led semantic differential discussion, in particular the issue of `the level of English language proficiency associated with IELTS' about which the three participants admitted to being uncertain. Mathilda had learnt from her own experience as an IELTS test-taker but still felt that the IELTS website and other guidance on proficiency levels was 'vague'. Victoria felt that she had had to develop her own proficiency level criteria while selecting her text and making items. She noted how the text 'comprehensibility factor' seemed to dominate her decisions on text and item difficulty. Mathilda felt that her text would not be 'that easy' for candidates whose English 'was not so developed' as her own. Participants were aware that an IELTS Band of 6 or 6.5 was conventionally seen as a cut-off point for students entering BA courses. Mary and Victoria were also informed by the levels of their own IELTS students (IELTS bands 5.0-7.5, and 8.0 respectively), which, for Mary meant that her test might not discriminate effectively at the higher end as she felt that she might not have enough experience of the highest scoring candidates to be able to target items at this group.

The discussion was now focusing on the actual reading construct espoused by IELTS. Victoria and Mary had heard that EL1 users had difficulty with the IELTS academic reading module, and that test performance on this module tended anyway to be weaker than on the other IELTS modules, even for stronger candidates. This is a common perception of IELTS (see Hawkey 2006), although test results published on the IELTS website show that overall mean scores for reading are higher than for the writing and speaking papers. Mathilda wondered whether the IELTS academic reading module was perhaps testing concentration rather than 'reading proficiency'. Victoria recalled that IELTS was described as testing skimming and scanning, but thought that skimming and scanning would also involve careful reading once the information necessary for the response had been located. But Mary was sure that reading and trying to understand every word in an IELTS text would mean not finishing the test. Mary felt that a candidate could not go into an IELTS exam 'not having been taught how to take an IELTS exam' and that a test-taker might not do well on the test just as a 'good reader'. Mary also claimed that she had never, even as a university student, read anything else as she reads an IELTS reading text. When reading a chapter in a book at university, one generally wants one thing, which one skims to locate, then 'goes off' to do the required reading-related task (although, conversely, Mathilda claimed often to 'read the whole thing').

The participants were then asked what other activities the IELTS text selection, editing and item writing processes reminded them of. Victoria recalled her experience working for a publisher and editing other people's reading comprehension passages for the Certificate of Proficiency in English (CPE) examination, which included literary texts (see Appendix B).

Mary had worked on online language courses, where editing other people's work had helped her thinking about the question-setting process (as well as surprising her with how inadequate some people's item-writing could be). The experience had reminded Mary how much easier it was to write grammatical rather than skills-based items. Victoria agreed, based on her own (admittedly rather unrewarding) experience composing objective-format usage of English items which she had prepared during her experience in publishing.

The participants were then asked whether their experience with the research project commission had changed their opinions of the IELTS Academic Reading paper. Victoria had found herself asking more

about the actual process of reading, her answers to this question underlining why IELTS academic reading was such 'a tough exam' for candidates. Mathilda had become more curious about how the test was used actually to measure proficiency, something she feels must be difficult to 'pin down'. Mary feels more tolerant of IELTS texts that may appear boring, given the difficulty she experienced finding her own text for the project. All three participants would welcome further experience with IELTS academic reading item writing, especially the training for it.

4.2 Procedures with and findings from the experienced IELTS Item Writer Group

Session 1: experienced item writer participant discussion of their experience with their commission to select an appropriate IELTS academic reading text, edit and adapt for testing purposes and generate test items

As with the non-experienced group, the four experienced participants discussed this commission to select an appropriate IELTS academic reading text, edit and adapt for testing purposes and generate test items, but this group was also, of course, able to discuss the regular experience of carrying out IELTS item writing commissions. Again this was organised as a researcher-led discussion session. Each participant (see Table 11 in Appendix B for background information) was invited to describe the processes through which an 'IELTS' text was selected and adapted, and then reading test items created. Again, both researchers were present, but intervened only infrequently and informally. All proceedings were recorded (see above).

4.2.1 Participant text search treatment and item development: flowcharts and discussions

The experiential information provided orally by the four participants is summarised in Table 5, which analyses responses on the issue of text sources.

Source/ Influence?	*Item Writer*			
	Jane	Anne	William	Elizabeth
IELTS Guidelines or Commission	√	√	√	√
Own interest			√	
Web	√	√	√	√
Magazines, journals		√		√
Newspapers	√			
Bookshops		√		

Table 5: Experienced participants: Sources and influences re IELTS Academic Reading module text selection

Unlike the non-experienced writers, this group did not mention the IELTS website or published IELTS material as a source of information on text selection. All reported that they referred to the item writer guidelines and to specific recommendations on topics made in the IELTS commissioning process.

Table 6 summarises the characteristics of target IELTS-type texts as interpreted by the four participants. The experienced writers seemed to share with the non-experienced group the perception of IELTS texts: subjects of popular interest presented in a formal, report-like format, academic in tone but not so technical that non-specialist readers would be handicapped in understanding them. As with the non-experienced group, there were differences between participants in the attention given to different text features. William was particularly concerned with issues of bias and cultural sensitivity while Jane seemed to pay most attention initially to the suitability of a text for supporting certain item types.

Perceived IELTS text characteristics	*Item Writer*			
	Jane	Anne	William	Elizabeth
Academic	1	2	2	3
Including a number of ideas/ opinions		2	1	1
Factual		1	1	
Not too specialist	1	1		1
Accessible to the general reader	1	2		2
Not too technical (for item writer to understand)	1		2	
Avoidance of bias, offence	1	2	5	1
Small and specific rather than big and general				1
Cohesion	1	1	1	1
Range/ complexity		1		
Suitability for (multiple) task types	3	1	1	2

Table 6: Experienced participants: Perceived characteristics of IELTS Academic Reading module texts

As with their non-experienced counterparts, the four experienced item-writers were asked to sketch flowcharts of the ways they had located, edited and prepared items for their IELTS Academic Reading tests, after which they were invited in turn to explain their flowcharts. In the following section we analyse the four experienced item writers' discussions. As above, these were transcribed and coded for topic before the semi-quantitative summary analysis as presented in Tables 5 and 6. The discussion is summarised in Table 7.

Three of the four item writers involved were able to use texts that they already had on file, although in William's case, this was because his initial effort to find a new text had failed. Anne reported that in between commissions she would regularly retain promising IELTS texts that she had found and that in this case she had found a suitable text on the topic of laughter (although actually finding that she had a suitable IELTS text on file was rare for her). From the outset, the potential for the text to generate items was a key concern. An ongoing challenge for Anne was to locate texts that included enough discrete points of information or opinions to support enough items to fulfil an IELTS commission: 'with a lot of articles, the problem is they say the same thing in different ways'.

The propositional 'complexity' of the text seemed to be of central concern so that a suitable text 'may not be for the academic reader, it may be for the interested layperson… if the complexity is right'. On the other hand there was a danger with more clearly academic texts of what Anne called 'over-complexity': 'over-complexity is when the research itself or the topic itself needs so much specialist language'. A good IELTS text would be propositionally dense, but not overly technical. Occasionally Anne might add information from a second source to supplement a text – Elizabeth and William (and Victoria of the non-experienced group) had also done this for IELTS, but not Jane.

Initially Anne would carry out 'a form of triage' on the text, forming an impression of which sections she might use as 'often the texts are longer than we might need' and considering 'which tasks would

be suitable'. Once she had settled on a text, she would type it up and it would be at this point that she could arrive at a firmer conclusion concerning its suitability. On occasion she would now find that she needed to take the decision – 'one of the hardest decisions to take' – that 'in fact those tasks aren't going to fit' and so have to reject the text. Anne saw personal interest in a text as being potentially a disadvantage when it came to judging its quality: 'it blinds you the fact that it isn't going to work'.

Elizabeth reported that she asked herself a number of questions in selecting a text: 'is the content appropriate for the candidature? Is the text suitable for a test, rather than for a text book? Will it support a sufficient number of items?' She considered that an ideal IELTS text would include, 'a main idea with a variety of examples rather than just one argument repeated'. Elizabeth reported that she usually selected texts that were considerably longer than required. As she worked with a text, she would highlight points to test and make notes about each paragraph, using these to identify repetitions and to decide on which item type to employ. Passages which were not highlighted as a source for an item could then be cut.

Like Anne, Elizabeth also reported looking for texts between commissions: 'you sort of live searching for texts the whole time'. On this occasion, she too had a suitable text on file. In approaching a text she reported that she considers the candidature for the test (an issue we return to later), the number of items that could be generated and the 'range of ideas'. Although she did not type up the text as Anne did, she made notes on it 'per paragraph' because this 'helps to see if it's the same ideas [being repeated in the text] or different ideas'. An 'ideal [IELTS] text' would 'have a point to it, but then illustrate it by looking at a number of different things; a main idea with examples or experiments or that sort of thing rather than one argument'. On the basis of these notes she would then begin to associate sections of text with task types so that, for example, 'paragraphs one to three might support multiple choice questions… there might be a summary in paragraph five, there's probably a whole text activity like matching paragraphs or identifying paragraph topics'.

At this point Elizabeth would begin cutting the text, initially removing material that could obviously not be used including 'taboo topics, repetitions, that sort of thing' but would still expect to have a longer text than would be required. With the text and the developing items displayed together on a split screen she would then highlight sections of text and produce related items. After completing the items, she might then remove sections of text that had not been highlighted, 'fairly stringently' to end up with a text of the right length.

William had decided to write about a 'particular topic', but 'wasted over two hours' looking for a suitable text on this topic on the internet. He was unable to 'come up with anything that was long enough or varied enough'. Instead he turned to a text that he had previously considered using for a commission, but had not submitted partly because of doubts about the perceived suitability of the topic ('too culturally bound to Britain') and the need to explain the names being discussed (Blake, Wordsworth). The text was somewhat problematic because of its length so that William 'ended up not only cutting it a lot, but rewriting parts of it and moving things around more than [he] would aim to do'. As a result of this rewriting 'there was a risk that it might end up not being as coherent as it ought to be'; a risk that might, in a regular IELTS commission, have led him to reject the text. William reported feeling 'nervous about IELTS in particular because there are so many rules that arise, sometimes unexpectedly' and so he usually sought to 'play safe' with the topics he chose.

William scanned the text from the source book and worked with it on his PC. He reported that he would usually shorten the text by cutting it at this point to 'a little over the maximum'. He would then work on the items and text together with a split screen, adapting the text 'to make sure it fits the tasks'. In choosing the tasks, he would ask himself which tasks 'fit the specifications' and, ideally, 'leap out from the text', but also which are 'worth the effort' and 'pay better'. On this basis 'if I can avoid multiple choice I will' because he found that multiple choice items (in fact the item type with the highest tariff) took much longer to write than other types. He would ensure that the tasks 'work' and

would change the text ‘to fit’ as necessary. The text was not ‘sacrosanct’, but could be adapted as required.

Jane, reported that she did not ‘normally’ store texts on file, but went to certain sources regularly on receiving a commission. On this occasion she looked for a new source. As ‘case studies’ had been requested in a recent IELTS commission, she took this as a starting point and searched for this phrase on the internet. There were ‘quite a few texts’ that she looked at before taking a decision on which to use. Typically, Jane takes an early decision on the task types that would best suit a text: ‘something like multiple choice requires a completely different text to True/False’. As she first scanned it, she identified the text she eventually chose as being suitable for ‘certain task types, not really suitable for others’. She also noticed that it contained too much technical detail, which she would need to cut. She claimed that texts are ‘nearly always three times, if not four times the length that we need’. There was then a process of ‘just cutting it and cutting it and cutting it, deciding which information you can target and which bits of the text will be suitable for particular task types’. Like the others she used a split screen to work on the items and text simultaneously.

Overview of the Item Writing Process			
Jane	Anne	William	Elizabeth
6 step flowchart: 1. Refer to commissioning letter to identify topics to avoid, sections needed (10 mins) 2. Finding possible sources, read quickly to decide whether possible (1hr-2hrs) 3. Collect likely sources and read again – topic suitability, suitable for task types, enough testable material (1hr) 4. Start cutting to appropriate length, identifying information to test and which parts go with which item types (1hr-2hrs) 5. Work on tasks, amending and cutting text as needed to fit tasks (1-2hrs per task type) 6. First draft – check that tasks work, check for overlap between items, cut to word limit (1hr)	11 step flowchart: 1. Text sourcing: check in files, investigate previously fruitful websites, Google a topic suggested in commission or that seems promising (30 mins-1 day) 2. Careful reading (30 mins) 3. Typing up with amendments (1 hr) 4. Length adjustment (to target plus 100-200 words) (15 mins) 5. Work on first (most obvious) task type (30 mins–2hrs [for MCQ]) 6. Mark up further areas of text for suitable items (30 mins) 7. Work on further tasks – amending text as necessary (1hr-2hrs) 8. Print off and attempt tasks (30 mins-1hr) 9. Write answer key (10 mins) 10. Check length and prune if necessary (10 mins-1hr) 11. Review and proof read (10mins-30mins) Found text already in her file (keeps an eye on potential sources) – looking for a Section 1 (relatively easy) task	11 step flowchart: 1. Think of subject – look at own books and articles for inspiration 2. Google possible topics 3. Locate a text and check suitability – how much needs glossing, any taboo subjects? 4. Consider whether text will work with task types 5. Scan or download text 6. Edit text to roughly to required length (or slightly longer), modifying to keep coherence 7. Choose and draft first task, modifying text to fit (abandon task if necessary) 8. Prepare other tasks 9. Revise text for coherence, length, to fit tasks, adapting tasks at the same time as needed 10. Have a break 11. Check and revise text and tasks Timings Steps 1 and 2: 10 mins-2 hrs; Steps 3 to 9: 1hr-2 hrs; Step 9: 20 mins; Step 10: 10 minutes to 1 week; Step 11: 20 mins	10 step flowchart: 1. Keep eyes open for texts 2. Choose from available texts 3. Evaluate selected text 4. Summarise main points and edit out redundant/ inappropriate material 5. Identify possible task types 6. Write items 7. Cut text to required length 8. Tidy up text and items checking keys 9. Leave for a day, print out and amend as needed 10. Send off No timings given

Text search			
Jane	Anne	William	Elizabeth
'I don't normally have texts waiting' 'I have certain sources that I go to regularly'' There were quite a few texts and I made a decision 'Texts are nearly always nearly three or four times the length we will need' If I can't understand a text, I wouldn't use it Opinion texts are more difficult to find 'You can't assume that the candidates are specialists'	Sometimes has to reject texts at her typing stage 'one of the hardest decisions to take' 'I think sometimes being interested in a text is a positive disadvantage' It is a challenge to find articles that have enough discrete information or opinions: 'A lot of articles, the problem is they say the same thing in different ways' It may not be for the academic reader, it may be for the interested layperson… if the complexity is right The over complexity is when the research itself or the topic itself needs so much specialist language	Subject matter is the first thing. 'It's finding the text that takes longest' For this commission 'I decided I would like to write about a particular topic and wasted over two hours on the internet: I couldn't come up with anything that was long enough or varied enough so I gave up' 'You get nervous about IELTS in particular because there are so many rules [restricting topic areas] that arise, sometimes unexpectedly' as a result 'I try to play safe'	'You're looking for texts the whole time' Asks the following questions about the text: 'Is the content appropriate for the candidature*? Will the text support the items? Does it have a range of ideas?' A suitable text, 'has a point to it but then illustrates it by looking at lots of different things' *The candidature I think about places I have worked and people I have know and try and look at it through their eyes You can't assume they are particularly interested in the UK We are going for the academic reader, but it's got to be understood by anyone

Text editing			
Jane	Anne	William	Elizabeth
I have a split screen working on items and text at the same time.	Sometimes we might add a bit from other sources. I cut out the first paragraph from my text because it was journalistic.	This text was long therefore… 'I ended up not only cutting it a lot and moving things around more than I would aim to do usually' Journalistic texts tend to begin from a hook – an example or 'attractive little anecdote' – more academic texts start from the general and move to the specific examples. IELTS texts should reflect the latter an have an academic tone. 'Adapt the text to fit the tasks', don't see the text as 'sacrosanct' 'Rewriting the text and trying out a task, then rewriting the text again and so on' 'Make a task loosely based on the text then make sure the text can fit the task.' Expressing a number of ideas and opinions, which would make it a Section 3 If its fairly factual more Section 1 Genuine academic texts are unsuitable because they assume too much knowledge and would require too much explanation I try and make sure that I understand it and can make it comprehensible	Articles that are not written by a specialist, but by a journalist can misrepresent a subject. To check this, 'I quite often Google stuff or ask people [about the topic]' Need to edit out references to 'amazing', 'surprising' or 'incredible' information in journalistic text

Item writing			
Jane	Anne	William	Elizabeth
I think I make a decision fairly early on about which task type I will use I decided this particular text was suitable for certain task types In other papers you choose a text with one tasks type – IELTS needs a text that will work with three: sometimes this is quite difficult: it doesn't work as easily with the third task With discrete information you can make it work with that Multiple choice questions fit best with an opinion text	Headings are difficult True-false is usually quite easy I don't like dong the diagram type ones or flowcharts Quite often in articles you get a little diagram, but it's too complex or guessable I read a lot oftexts and cut them down before I decide which one to use.	My first main thing is how well the tasks fit that text. Chooses tasks that 'leap out from the text' Not something that could be answered by someone who knows the subject Considers which tasks pay more, which are worth the effort and so avoids MCQ if possible. Factual information you can test with true false not given We need to cover the whole text – every paragraph is tested A text ought to lend itself to having a topic in each paragraph that can be captured in a heading I think the paragraphs overlapped in this case MCQ: coming up with four plausible opinions which are wrong is difficult: the danger is that you are pushed into testing something that is trivial.. they should all be important pieces of information or opinions or functions Flow charts are either a sequence that can be guessable or it's a false way of presenting the information – it's not really a flow chart I work on items at an early stage and will dump a text after ten minutes if I feel it will not work.	I think multiple choice can work across a range of texts including at a more basic factual A diagram or even a flow chart can be more personal than you realise. I made a diagram from one text that failed because it was my idea and it didn't reflect other peoples ideas I often write notes on texts before deciding which one to use.

Which of your sections you happiest with?			
Jane	Anne	William	Elizabeth
		Unusually, I wrote all three tasks simultaneously There were problems of overlap with other tasks. Questions 1, and 16 were all about Blake and Wordsworth: a bit problematic and other people might feel they are not independent of each other Paragraphs F and H each only have one item, which is not ideal Something like a summary of one paragraph can be too easy because the answers are all together. Identifying the paragraph containing information where it's in random order and could be anywhere in the text requires you to scan the whole text for each individual item which seems to me to be far more difficult for candidates	The need to scan the whole text three times for different information seems unfair: 'you wouldn't usually scan [a text] three times for different sorts of information' – we have had advice to cut down on that now. I usually try to focus two of my tasks on specific information and have a third one that is more of an overview. This text does have one basic idea and really the whole text is saying that. I was testing the support for the idea. There is a stage when I think 'this is going to work and I'm not going to dump this. I thought there were enough discrete words that would make a key to support multiple choice. I am very conscious of how much of a text I am exploiting.

Table 7: Experienced participants' descriptions of the item writing process

clear	The guidelines are clear (Anne)			Finding texts can be confusing (Anne) Trying to read pre-editing teams' minds can be confusing (William) Texts can be confusing (Elizabeth) Some tasks confusing for candidates (William) 'We used to fill in a form identifying what each item was testing – it was confusing but also really useful in focussing the mind on what items are actually doing' (Elizabeth)	confusing
interesting	The topic and the texts – I have learnt a lot (William) you get to grips with texts that you might not otherwise read (Anne) Texts must be engaging to keep you interested for a day (Jane)	Final stages of item writing – proof reading (Elizabeth) MCQ can be quite interesting and creative (Anne) Making sure that everything fits together (William)	More interesting than business English texts (Anne)	Proof reading (Jane)	dull
time-consuming	Finding the texts (All)	Editing can be 'deathly' when you are working with other's text that Is problematic (William) Sometimes rewriting is easier alone than by committee (Wiliam)	Depends on time of day (Anne) and team (Elizabeth)	If it's the right text, it can be quick (Anne, Jane)	quick
rewarding	Making it work (William) Pretest review acceptance (William)	Improving the quality of the source text (Anne) Often we are in effect creating a new text – fit for a different purpose (William)			unrewarding
worrying	When you can't find a text (Anne)	You can easily spend half a day and come up with nothing (Elizabeth)		Getting the task to work (Jane)	pleasing

creative	All the writing is creative, even though we are starting with something – rather like putting on a play (William)		Editing problem solving can be creative, but not satisfactory when you seem to be doing another item writer's work for them (William)	Proof reading Techniques for writing enough items – 'in summaries you've got to go for the nouns, which you didn't know when you first started' (Anne)	programmatic
challenging	Finding the texts and shaping them (Anne) Understand a subject you may not be familiar with (William)		Creating the items once a suitable text has been chosen (Elizabeth)		straightforward
frustrating	Feedback that you don't agree with (William) 'There are times when you have to have a quick walk round the garden' (Anne)	Losing a submission altogether (rejection)	Disagreement about issues of bias – William finds Business papers less sensitive: others find Cambridge Main Suite papers more sensitive.		satisfying
supported	Editing and pre-editing is supportive on the whole (Anne) Colleagues are generally helpful and supportive. Rejection of tasks comes when topic not checked in advance (William)	You can ask for elaboration of pre-editing feedback. (Elizabeth) I don't' think I have ever disagreed with pre-editing feedback (Jane)	Some texts accepted when I could answer on basis of topic knowledge, others rejected when answers did not seem guessable to me (William) The whole issue of how guessable items are is difficult (Anne) A collocation can be guessable to a native speaker, but not to NNS (William) but part of reading is the ability to predict(Elizabeth)	Looking for a text is unsupported (Anne)	unsupported

Table 8: Summary of experienced participant focus group comments and ratings on semantic differential scales

4.2.2 Participant focus group discussions

The experienced group participated next in a focus group discussion structured around a set of nine semantic differential continua (Osgood, 1957, using the unlabelled scale format compared with other formats by Garland, 1996) and as seen in Table 8. In the table, summaries of the comments made by the participants in their 20 minutes of unmediated discussion are placed in their approximate location on the continua for the nine scales. The adjectives for the continua were selected by the researchers.

Again, points made by participants in the focus group discussion served to triangulate views expressed in the preceding interview activity concerning IELTS text search and treatment and item development: flowcharts and discussions already reported. Following discussion of the semantic differentials, the research team pursued emerging issues with the group.

The experienced group, like the non-experienced, expressed uncertainty about candidates' level of English language proficiency. The four discussed the need to keep the candidates in mind when writing items, but agreed that it was challenging to do this, given the 'the variety of the situation and [the candidates'] levels of English'. Each participant had their own points of reference for these. Anne

also worked as an examiner for the speaking paper and so met many candidates while both William and Elizabeth had experience of preparing students for the test. However, Elizabeth reminded the group that the candidates they met in the UK would not be representative of the full range of candidates taking the test – especially those from relatively underprivileged backgrounds.

Item writers also received information about candidates from IELTS. An annual report on demographic data is provided by Cambridge ESOL and 'common wrong answers' to open response items are discussed at pretest review meetings. What Anne described as the 'off the wall' nature of some of these wrong answers and the observation that 'some people have been accepted at universities, where I thought their English was totally inadequate' led William to the conclusion that 'you can do reasonably well on IELTS, I think. And still have what seems to be a low level of English'. Elizabeth also questioned whether IELTS candidates would need to arrive at a full understanding of the text in order to succeed on the questions, suspecting that in IELTS 'half the time the candidates don't read the text from beginning to end because they don't have to' because local details in the text were being tested by the items rather than the overall meaning. However, Anne wondered whether William's concern could be justified as success on the test would require adequate levels of performance on the direct speaking and writing papers as well as reading and listening.

There was discussion of how the participants had developed their item writing expertise. For Jane this was not easy to explain: 'It's difficult to say sometimes exactly what you're doing and how you're doing it'. Anne agreed, observing that 'the processes you go through aren't necessarily conscious'.

However, there were item writing skills that could be learnt. Anne had come to appreciate the importance of 'working the task': attempting it as a candidate would. Jane agreed that this was helpful, but admitted she rarely did this prior to submission because of the pressure of deadlines. Elizabeth had found very helpful the advice given to her at her initial training session to focus on what she felt to be the key points of the text, finding that this could help her when she was 'stuck on something'.

Anne felt that her items had improved 'over years of seeing other peoples' and having to mend your own'. William pointed to the value of attending editing meetings to obtain insights and Elizabeth felt that feedback at editing meetings had been one of her main sources of learning about item writing especially where a the chair of the meeting, as an experienced and successful item writer, had been effective at showing how a text or item could be improved.

William spoke of having learnt how to devise plausible distractors for multiple choice items. However, there were limits to how far this could be learnt as an item writing skill and he wondered about the role of background knowledge in eliminating incorrect options: 'I think there's a risk with IELTS because if it's a scientific text, I may not know nearly enough to know what would be a plausible distractor. What seems plausible to me could be instantly rejected by somebody who knows a little more about the subject.'

Testing implicit information was seen to be problematic. There were cases of disagreement between the item writers and their colleagues carrying out pre-editing reviews about 'whether [a point] is implicit, but strongly enough there to be tested or not' (William). For Jane, testing the writer's interpretation against others' was a further argument in favour of the pre-editing and editing processes: 'fresh eyes are invaluable when it comes to evaluating a task'.

Although Jane reported that she tried to keep the level of language in mind as she wrote, the group agreed that the difficulty of items was not easy to predict. None of the writers seemed to have a clear sense of the proportion of items associated with a text that a successful IELTS candidate at band 6.0 or 6.5 might be expected to answer correctly. Pretesting results often revealed items to be easier or more difficult than expected.

5 ANALYSIS AND FINDINGS ON THE TEXTS

The analysis here is applied to the texts as they were submitted by the seven participants, before any changes made during the public editing process reported below. The texts and items submitted by the item writers (in their adapted, but unedited state) are presented in Appendix C. This analysis shows how the texts were shaped by the writers and so serves to contextualise the comments made in the interview and focus group sessions.

In this section, we again begin with the texts submitted by the non-experienced group. Following Weir *et al*. (2009a) we employed automated indices of word frequency and readability to inform and supplement our qualitative text analyses. Outcomes of these procedures are given in Figures 1 to 3 below and are discussed in relation to each submission in the following section.

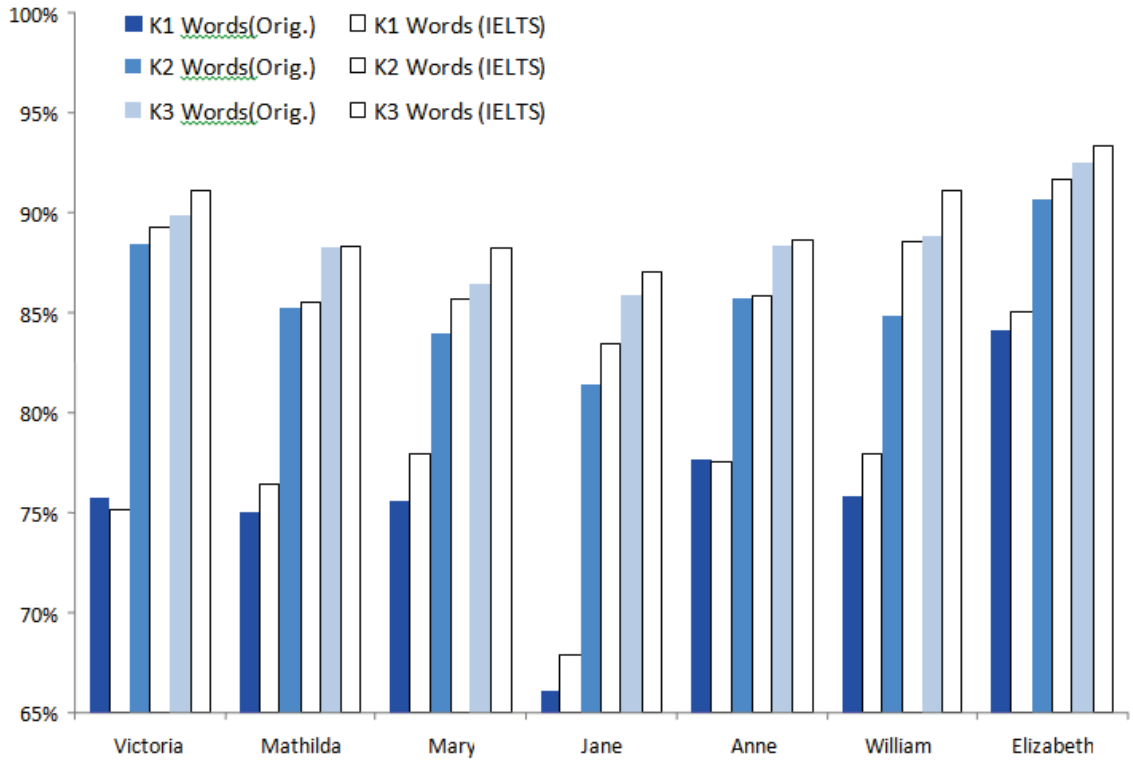

Figure 1: Results of word frequency analyses for original source texts and adapted IELTS text: percentage of very frequent words at the BNC 1000, 2000 and 3000 word frequency levels

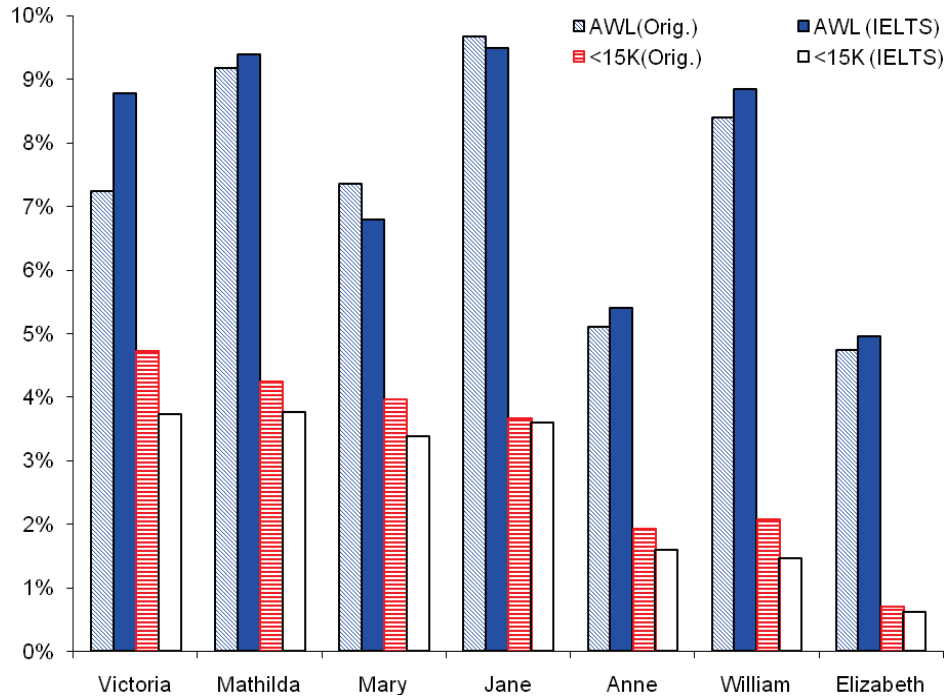

Figure 2: Results of word frequency analyses for original source texts and adapted IELTS text: percentage of sub-technical academic (AWL) and very infrequent words

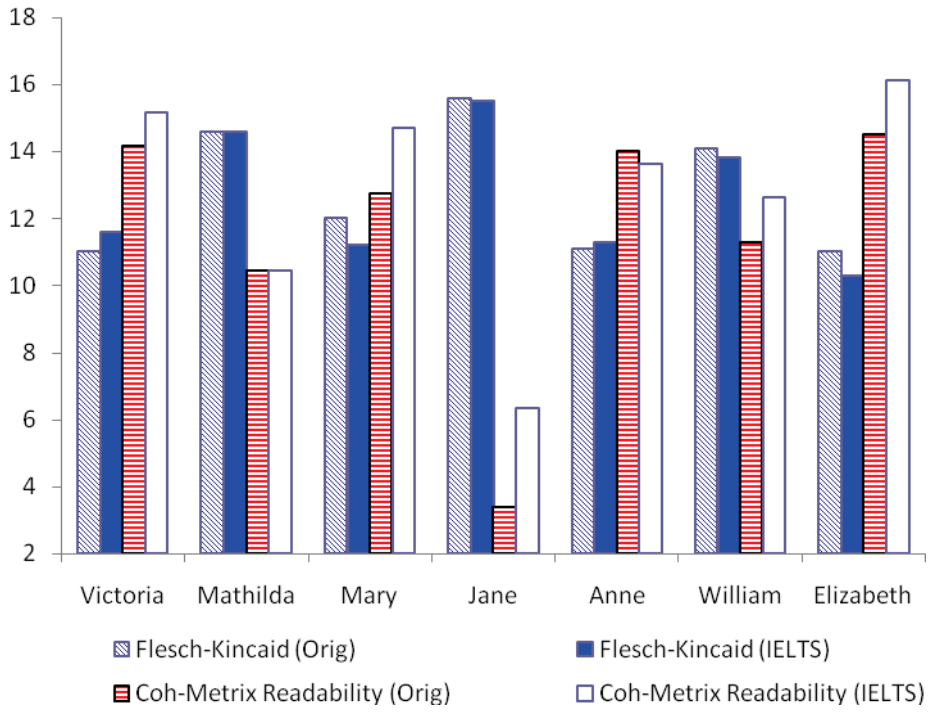

Figure 3: Results for Flesch Kincaid grade level and Coh-Metrix readability estimates for original source texts and adapted IELTS texts
NB lower scores on Flesch Kincaid and higher scores on Coh-Metrix represent greater reading ease.

5.1 The non-experienced group

Victoria's text:

> *How the brain turns reality into dreams: Tests involving Tetris point to the role played by* '*implicit memories*' Kathleen Wren
>
> MSNBC: www.msnbc.msn.com published online 12 October 2001

Victoria's text was a science feature published on the website of online news service MSNBC. It describes research into the nature of dreams recently reported in the journal *Science*. The text is organised around a problem-solution pattern. The problem is that of accounting for how dreams relate to memory. The solution is provided by new research, based on the dreams of amnesiacs, identifying dreams with implicit rather than declarative memories.

Victoria made the most extensive changes of all the untrained writers, making revisions to all but one of the paragraphs in her text with a total of 77 edits. Uniquely, among writers in both groups her adapted text was longer (by 44 words) than her source. It also involved an increase in AWL words and a reduction in the most frequent words (BNC 1,000 word level) in the text (Figure 1 and Figure 2). However, in common with all the writers in the study except Mathilda, the effect of Victoria's adaptations was to increase the proportion of words with a frequency in the BNC of one in 3000 or higher.

Victoria reported that in editing the text she wanted to make it more academic in register and therefore better suited to the context of university study. She had achieved this, she said, by increasing the complexity of sentences, using passive forms and hedges to create academic distance and by adding a methodology section to the article.

There are a number of changes that would seem to be directed at making the text appear less journalistic. A reference to 'Friday's issue of *Science*' in the opening paragraph, which reflects the news value of the article, is removed (although this is the only reference in the article to another text).

These changes include reframing the relationship between writer and reader. The original text addresses the reader as 'you', while the revised version instead employs 'we', passive constructions or, in one case, 'subjects' (in the sense of research subjects). Contractions are replaced with full forms or alternative constructions, as in, 'the hippocampus ~~isn't~~ is not active during REM sleep' or the substitution of 'people with amnesia shouldn't dream' by 'individuals suffering with amnesia should not be capable of dreaming'.

Further changes to the text seem to reflect the intention to achieve a more formal, academic register. These include the use of less frequent vocabulary – 'different parts of the brain' becomes 'a region of the brain'; nominalisation – 'But they can still affect your behavior' becomes 'But they still have the potential to affect behaviour' (note that Victoria changes behavior to behaviour to reflect British spelling conventions); use of reporting verbs – 'said' becomes 'states', 'believes' becomes 'upholds'; references to research procedures – 'therefore' becomes 'from these results', 'the people in the study…' becomes 'The methodology designed for Stickgold's study had two groups of subjects…'; and hedging – 'Much of the fodder for our dreams comes from recent experiences' in the original text is prefixed in the adapted version with 'Such research suggests that…'.

Pronoun references are made more explicit: 'That's called episodic memory' becomes 'To differentiate this information from declarative memory, this particular [form] of recollection is referred to by scientists as episodic memory' and '…the procedural memory system, which stores information…' is expanded to give '…the procedural memory system. This particular system stores information…'

Victoria does not generally choose to replace technical vocabulary with more frequent alternatives, but in one case does add a gloss that does not occur in the source: 'amnesia, *or memory loss*'. She replaces one instance of 'amnesiacs' with 'people suffering from memory loss', but in three other instances she chooses to use 'amnesiacs' directly as it appears in the source text and in a fourth replaces it with 'the amnesiac group'. She also follows the source text in glossing such terms such as 'neocortex', 'hippocampus' and 'hypnogagia', but (again following the source) chooses not to gloss 'REM sleep'. Victoria's changes make the text more difficult to read by the Flesch-Kincaid grade level estimate, which is based on word and sentence length, but easier according to the Coh-Metrix readability formula (Crossley *et al* 2008), which reflects vocabulary frequency, similarity of syntax across sentences and referential cohesion. (Figure 3).

Mathilda's Text

> *How—and Where—Will We Live in 2015? The future is now for sustainable cities in the U.K., China, and U.A.E*. by Andrew Grant, Julianne Pepitone, Stephen Cass
>
> *Discover* Magazine: discovermagazine.com, published online 8 October 2008

Mathilda made the fewest changes of any writer to her source text, which came from *Discover*, a Canadian magazine concerned with developments in science, technology and medicine. This text also has a problem-solution structure, although it is more factual and descriptive and less evaluative than Victoria's. The article portrays three new city developments in diverse locations that are all intended to address ecological problems. The majority of the text is devoted to describing the innovative features of each city in turn: transport, power and irrigation systems.

Mathilda reported that she too had found her text on the internet after looking at examples of IELTS material from the IELTS website. Although she would have preferred a more emotionally engaging literary text, she looked for such popular science topics as 'the environment', 'dreams' and 'the future'

in the belief that these were closer to the topics of the IELTS texts she had seen. After briefly scanning a large number of possible texts, she saved four to her computer for more detailed consideration. She had considered using a text concerning the evolution of the human skeleton, but rejected this as being too technical: 'pure biology'. She made her choice because she felt it was 'easy to read' and had sufficient information to support a large number of questions. In common with both Mary and Victoria, she found choosing the text the most time consuming element in the process.

In editing the text Mathilda cut the attribution and removed the pictures, but left the text itself largely untouched. All four of the textual edits that she made involved replacing relatively infrequent words with more frequent alternatives: 'gas-guzzling cars', which she felt was too idiomatic, became 'gas-consuming cars'. Relatively technical terms were replaced with more frequent words; 'photovoltaic panels' was replaced with 'solar technology'; 'potable water' with 'drinking water' and 'irrigate' with 'water'. These changes somewhat increased the proportion of very frequent and AWL words (panels, technology), and reduced the proportion of very infrequent words, but did not affect the length of the text (748 words) or the readability estimates.

Mary's text

The Rise of the Emotional Robot by Paul Marks

From issue 2650 of *New Scientist* magazine, pages 24-25, published 5 April 2008

As noted in Section 4 above, Mary eventually chose a source text from *New Scientist*, the science and technology magazine noted by Weir *et al*. (2009b) as a popular source for IELTS texts. Unlike both Mathilda and Victoria, Mary chose a source text that, at 1,094 words needed to be pruned to bring it within the maximum IELTS word limit of 950 words. This text, like Victoria's, reports on recent research. The writer reports two studies in some detail and cites the views of other researchers. The situation of human emotional engagement with robots is described and solutions involving making robots appear more human-like are explored. As in Victoria's text, there is an element of evaluation and different points of view are quoted.

Mary was concerned with the authenticity of her text and sought to make as few changes as possible in adapting it for IELTS. Like Mathilda, Mary, who made 30 edits in all, made a number of changes to the vocabulary of her text. These included changing 'careering' to 'moving'; 'resplendent in' to 'wearing'; 'myriad' to 'a multitude of'; 'don' to 'put on' and two instances of 'doppelgänger' to 'computerised double' and 'robotic twin'. As in Mathilda's text, these changes all involved replacing relatively infrequent words with more frequent alternatives, although, reflecting the nature of the text, none of these appear particularly technical to the field of robotics. Mary's changes reduced the proportion of both AWL and infrequent words while increasing the proportion of very frequent words (Figure 1 and Figure 2).

Mary explained that the need to reduce the length of the text led her to remove contextualising points of detail such as the identity of a researcher's university ('…who research human-computer interaction at the Georgia Institute of Technology in Atlanta'), reporting '…presented at the Human-Robot Interaction conference earlier this month in Amsterdam, the Netherlands', or the location of a research facility ('in Germany') and references to other texts '(*New Scientist*, 12 October 2006, p 42)'.

Mary also chose to summarise stretches of text. For example, she reduced 'But Hiroshi Ishiguro of Osaka University in Japan thinks that the sophistication of our interactions with robots will have few constraints. He has built a remote-controlled doppelgänger, which fidgets, blinks, breathes, talks, moves its eyes and looks eerily like him. Recently he has used it to hold classes…' to 'Scientist Hiroshi Ishiguro has used a robotic twin of himself to hold classes…' However, she chose to introduce this section of the text with three sentences of her own composition, 'Whether robots can really form relationships with humans and what these can be is much disputed. Only time will really tell. However, despite the negative criticism there is one scientist with strong evidence for his view.' This

would seem to reflect the focus of her tasks on the identification of views expressed by different experts mentioned in the text.

There is evidence that Mary was aware of the need to avoid potentially sensitive topics in IELTS when choosing her cuts as well as in the initial text selection. Three of the four sentences in a paragraph concerning the emotional attachment formed by American soldiers to robots employed in the Iraq war were deleted from the IELTS text.

Although expressing the most concern for authenticity and favouring a light editorial touch, of all the writers, Mary was the only one to substantially reorder her text. She reported that she had found the original text poorly organised. She wanted to focus in her questions on opinions expressed by different researchers, but found that these were distributed across paragraphs and felt that her questions would be more effective if the paragraphing was addressed.

The first four sentences of the fifth paragraph in her source text, which quotes the views of a named researcher, are cut, and appended to the sixth paragraph. The final sentence is removed altogether. The change, which brings together two quotations from the same expert, reflects Mary's words (see above) concerning the influence of the task type (matching views to protagonists) and the need to avoid diffusing the views of the experts across the text. Taken together, Mary's changes had the effect of making the text easier to read according to both the Flesch-Kincaid grade level estimate and the Coh-Metrix readability formula (Figure 3).

We now turn our attention to the texts submitted by the experienced item writers.

5.2 The experienced group

Jane's text

Wildlife-Spotting Robots by Christine Connolly,

Sensor Review: Volume 27 Number 4 pages 282-287 published in 2007

Uniquely among the writers in this study, Jane chose a text originating in a peer reviewed journal, albeit one directed more towards an industrial than an academic audience (*Sensor Review: The international journal of sensing for industry*). The text concerned the use of remote robotic sensors in wildlife photography exemplified by a secondary report on an application of this technology to capture evidence of a rare bird. The text describes the role of robotic cameras in wildlife observation with examples of the equipment used. There is an extended description of the use of an autonomous robotic camera system in a search for a rare bird, and of a further development of the technology which allows for remote control of the camera over the internet.

Ranging from 1592 to 2518 words, the source texts used by the experienced writers were all very much longer than those of the non-experienced group (748 to 1094 words). At 1870 words the length of Jane's source text was typical for the experienced group. She cut it by 50%, making 43 edits, to give an IELTS text of 937 words.

This was the most technical of all the texts and like other writers Jane cut a number of technical terms. These related both to wildlife and animal behaviour ('hawks', 'herons', 'double knock drummings') and to the technology being used to record it ('RECONYX cameras', 'XBAT software', 'auto-iris'). However, she also retained many such words in her IELTS text including, 'ornithology', 'geese', 'fieldwork', 'vocalisations', 'actuators', 'teleoperation' and 'infrared'. In spite of the changes, Jane's final text included the lowest proportion of high frequency words of any writer. The most frequent 3,000 words of the BNC accounted for just 88.6% of her IELTS text while the 95% coverage said to be required for fluent reading (Laufer 1989) came only at the 8000 word frequency level of the BNC.

Some of Jane's edits appear to be directed at clarification or at improvement of the quality of the writing. Compare the original and edited versions of the following:

> *Original text*: 'More than 20 trained field biologists were recruited to the USFWS/CLO search team, and volunteers also took part'.
> *IELTS text*: 'The project started in 2005 with over 20 trained field biologists taking part in the search team, and volunteers also being recruited'.

> *Original text*: 'The search also made use of… cameras … for monitoring likely sites without the disturbance unavoidable by human observers'
> *IELTS text*: 'The search also made use of… cameras … for monitoring likely sites. This method was ideal since it did not lead to the disturbance that is unavoidable with human observers'

Jane expanded some abbreviations ('50m to 50 metres', '8h per day' to '8 hours per day'), but not others ('10 m to 40 mm' is retained to describe a camera lens focal range, and sound is 'sampled at 20 kHz for up to 4 h per day'). 'UC Berkeley' is expanded to 'University of California, Berkeley' on its first occurrence, but not on its second. Three occurrences of 'Texas A&M' are retained unchanged.

The deletion of the abstract, subheadings and the two citations had the effect of making the final text appear less like a journal article. The removal of a block of 653 words in five paragraphs that described the technical attributes of robotic cameras, together with the cutting of photographs of the equipment and examples of the images captured, had the effect of foregrounding the application to wildlife research (problem-solution) and diminishing the attention given to the attributes of the equipment (description/ elaboration): the central concern of the journal. One paragraph within this block explained why the equipment qualified as 'robotic' and its deletion modifies and diminishes the relationship between the title (Wildlife-spotting robots) and the adapted text. In the IELTS the 'robotic' nature of the cameras is not explicitly explained, although three uses of the term do remain. This became a source of some confusion for the editing team (see Section 7).

Jane's edits had little effect on the Flesch-Kincaid grade level of the original text, but did make it easier to read according to the Coh-Metrix readability formula. However, by both measures her IELTS text was the most difficult of all the edited texts in this study.

Anne's text

> *The Funny Business of Laughter* by Emma Bayley
>
> *BBC Focus*: May 2008, pages 61 to 65

Anne's text was taken from *BBC Focus*, a monthly magazine dedicated to science and technology. This expository text, which draws on a range of research from different disciplines, describes and elaborates the functions and origins of laughter and their implications for our understanding of the human mind. She reported that she had found this text in a file she kept for the purpose of item writing, storing suitable texts between item writing commissions.

Like all the experienced writers, Anne took a relatively lengthy source (1606 words) and cut it extensively (her edited text was 946 words long), making 57 edits altogether. She discarded 15 of the 31 words in the source text that fell outside the 15K frequency level and 31 of 82 from the AWL. This results in a slightly higher proportion of academic words and a lower proportion of very infrequent words in the edited text than in the source (Figure 2).

In common with all the other writers Anne chose to cut a number of technical terms including 'neurological' and 'thorax' (replaced with 'chest') although she retained 'bipedal' and 'quadrupedal' as well as other technical words such as 'neuroscientist', 'primate' and 'stimulus'. She also excised a number of infrequent words including synonyms for laughter (the topic of the text) such as 'chortle',

'yelping' and 'exhalations', replacing this latter word with another infrequent (though more transparent) word borrowed from the deleted opening section of the original: 'outbreath'.

One means of reducing the length of the text that Anne exploits is to cut redundancy in word pairs such as 'rough ~~and tumble~~ play' or restatements such as 'laboured breathing ~~or panting.~~'. Some changes seem to reflect an editor's desire to improve the linguistic quality and accuracy of the text: she inserts the conjunction 'that' in the sentence 'It is clear now *that* it evolved prior to humankind' and replaces 'most apes' with 'great apes', presumably because the text has cited only orang-utan and chimpanzee behaviour.

Anne eliminated references to a 'news' aspect of her story by deleting the first and last paragraphs: the original article opened and closed with references to the forthcoming 'world laughter day'. Another change that makes the text less journalistic, in line with Anne's stated desire to reduce 'journalese', is the increase in formality. The idiomatic 'having a good giggle' is replaced by 'laughing'; some abbreviations and contractions are exchanged for full forms so that 'lab' becomes 'laboratory', 'you've' becomes 'you have' and 'don't' is replaced with 'do not'. However, unlike Victoria, Anne chooses to retain contractions such as 'that's' and 'it's' and even modifies one occurrence of 'it is' in the original to 'it's'. In her final IELTS text, 'it's' occurs three times and 'it is' four times. Whimsical, informal and perhaps culturally specific references to aliens landing on earth and to the 'world's worst sitcom' are also removed.

Through her deletions Anne relegates one of the central themes of her original text – the role of laughter in the evolution of socialisation and the sense of self. As a result, the IELTS text relative to the source, although less journalistic, seems more tightly focussed on laughter as a phenomenon *per se* than on its wider significance for psychology or, as expressed in a sentence that Anne deletes, 'such lofty questions as the perception of self and the evolution of speech, language and social behaviour'. However, elaboration is the primary rhetorical function of the IELTS text as it is for the source. The effect of Anne's changes on the readability of the text is to make it somewhat more difficult according to both the Flesch Kincaid and Coh-Metrix estimates.

William's text

> *Introduction* from *Poor Monkey: The Child in Literature* by Peter Coveney
>
> Published in 1957 by Rockliff

William's source text, the only one taken from a book, was an essay by Peter Coveney (1957). This was the longest chosen by any writer and William cut around 60% of the original, making 65 edits in developing his 909 word IELTS text. The third and eighth paragraphs of the original text are almost entirely discarded, as are lengthy stretches (50 words or more) of every paragraph except the first and fourth.

Much in the rejected passages concerns the original author's informing theory of the relationship between literature and social change. In the third paragraph, he anticipates criticism and defends his approach; 'To suggest a relation between literature and society might seem to imply that too much, perhaps, is to be explained too easily by too little'. This is eliminated from the IELTS text, while in other cases William offers summaries of parts of the original of varying length. The first two sentences of the original text – 'Until the last decades of the eighteenth century, the child did not exist as an important and continuous theme in English literature. Childhood as a major theme came with the generation of Blake and Wordsworth.' – is replaced by a single sentence in the edited text – 'Childhood as an important theme of English literature did not exist before the last decades of the eighteenth century and the poetry of Blake and Wordsworth.', saving nine words. The sentence 'Art was on the run; the ivory tower had become the substitute for the wished-for public arena' substitutes for 169 words on this theme in the original.

References to specific works of literature (*The Chimney Sweeper*, *Ode on Intimations of Immortality*, *The Prelude*, *Hard Times*, *Dombey and Son*, *David Copperfield*, *Huckleberry Finn*, *Essay on Infantile Sexuality*, *Way of All Flesh*, *Peter Pan*) and to a number of writers (Addison, Butler, Carroll, Dryden, James, Johnson, Pope, Prior, Rousseau, Shakespeare, Shaw, Twain) are removed, together with references to other critics (Empson), although the names of Blake, Dickens, Darwin, Freud, Marx and Wordsworth are retained. Some technical literary vocabulary such as 'Augustan', 'ode', 'Romantics' and 'Shakespearian' is cut (although 'lyrics', 'poetry' and 'sensibility' are retained), as are relatively infrequent words such as 'cosmology', 'esoteric', 'moribund', 'congenial' and 'introversion'. As a result, in common with most other writers, the proportion of frequent words is higher and the proportion of very infrequent words lower in the edited text than in the source (Figure 1 and Figure 2).

As was the case for Anne and Jane, one effect of William's changes is to narrow the scope of the essay. The edited version is focussed more closely on the theme of the treatment of childhood at the expense of discussion of specific works and of arguments supporting the thesis of literature as an expression of social change and crisis. As a result, the adapted text takes on more of the characteristics of an historical narrative with a cause/effect structure and loses elements of persuasion and argumentation. The changes to the text had little effect on the Flesch-Kincaid grade level estimate (Figure 3), but made it easier to read according to the Coh-Metrix readability formula.

Elizabeth's text

> *Time to Wake Up to the Facts about Sleep* by Jim Horne
>
> *New Scientist*: published on 16 October 2008, pages 36 to 38

In common with Mary, Elizabeth, chose a source text from the *New Scientist*. As was the case for Anne, this was a text that Elizabeth already held on file. The text questioned popular myths about people's need for more sleep. Resembling the texts chosen by Victoria, Mary, Jane and Anne, this article reports on recent research, although in this case the author of the text is one of the researchers and refers to a study carried out by 'My team' (the IELTS text retains this). The author argues against perceptions that people living in modern societies are deprived of sleep and draws on a range of research evidence, including his own study, to support his view. Like William's, this is a text that involves argumentation and is organised around justifying a point of view. Reflecting the personal tone of the original, Elizabeth retains the attribution by incorporating it into a brief contextualising introduction following the title: 'Claims that we are chronically sleep-deprived are unfounded and irresponsible, says sleep researcher Jim Horne'.

Elizabeth cut the 1592 word source text by 60% to 664 words, making 54 edits. Like Mary, Elizabeth cuts references to other texts – '(*Biology Letters* vol 4, p 402)' – and removes a number of technical terms: she removes the technical 'metabolic syndrome', but retains 'metabolism'. She also chooses to keep 'obesity', 'insomnia', 'precursor', 'glucose' and the very infrequent 'eke'. Elizabeth's source text included relatively few academic and very low frequency words and more high frequency words than the texts chosen any other writer (Figure 1 and Figure 2).

Like Anne and Victoria, Elizabeth replaces informal journalistic touches with more formal alternatives – 'shut eye' becomes 'sleep' (although 'snooze' is retained), 'overcooked' becomes 'exaggerated' (but 'trotted out' is retained).

The most intensively edited section of the text is an extended quotation from a researcher. As was the case for Anne and Jane, clarity and style seem to be important. Compare the following:

> *Original text*: We did this by asking when they usually went to sleep and at what time they woke up, followed by, 'How much sleep do you feel you need each night?'
>
> *IELTS text*: We asked respondents the times when they usually went to bed and woke up, and the amount of sleep they felt they needed each night.

Another change may reflect the need for sensitivity to cultural diversity in IELTS mentioned by Elizabeth in relation to her awareness of candidate background. The author's assumption about the identity of his readers seems to be reflected in one phrase that he uses: 'we in the west'. In the IELTS text this becomes the less positioned 'most people in the west'. Rhetorically, Elizabeth retains the function of the text as an opinion piece organised around justification of a point of view.

The changes made in editing had the effect of making the text easier to read according to both the Flesch-Kincaid grade level estimate and the Coh-Metrix readability formula (Figure 3).

6 ANALYSIS AND FINDINGS ON THE EDITING PROCESS

The participants were mainly left to organise and implement the joint editing session without intervention from the research team. The summary here seeks to identify and quantify the occurrences of key points raised, as informing the investigation of IELTS Academic Reading test item writing processes.

The analysis of the texts as originally submitted by the three non-experienced participants appears in Section 6 above. This section describes the changes made to the texts and items in the process of joint test-editing. We begin with the non-experienced group.

6.1 The non-experienced group

Victoria text editing

As noted in the text analysis below, Victoria's text, 'How the Brain Turns Reality into Dreams', was taken from the online news website MSNBC, describing research into dreams reported in the journal *Science*. Victoria, who, it will be recalled, often referred to her process of 'fixing up' her text, made 77 edits, revised all her paragraphs and actually increased the length of the original text from 897 to 941 words.

At the beginning of the editing session on her text and items, it was suggested by her colleagues, who had just read her text, that Victoria should make the following additional changes to her text:

- the deletion of one or two hedging phrases she had added to give the text a more academic tone
- the shortening of two clauses for compactness.

Victoria item editing

Victoria had chosen True/ False/ Not Given (T/F/NG), Multiple Choice (MCQ) and Short Answer Questions (using not more than three words from the passage) (SAQ) as her task types.

The following were the main issues raised over the tasks and items proposed by Victoria:

- the possibility, especially in the T/F/NG task, that test-takers may infer differently from the item-writer, but plausibly, yet be penalised even when their understanding of the point concerned is not wrong
- the question whether, in actual IELTS item-writing, there were conventions on the distribution of the T/F and NG categories in a set
- the colleagues themselves found Victoria's multiple choice items difficult
- that having two incorrect alternatives which mean the same (though in different words) was in a way increasing the test-taker's chance of selecting the right alternative
- that the SAQ task should be a test of content rather than grammatical structure

Mathilda text editing

As noted above and confirmed in the text analysis below, Mathilda made the fewest changes, only four, of any writer to her source text, 'How – and Where – will we Live in 2015?' which came from *Discover*, a Canadian science and technology magazine. Her text was relatively short at 748 words.

At the beginning of the editing session on her text and items, Mathilda wondered whether her text was perhaps too easy, being straightforward and factual, with no complex argument and a sequential key point structure. Mathilda was reminded by her colleagues that a straightforward text might well be accompanied by difficult questions. In fact, this would not be in accordance with IELTS practice.

Mathilda item editing

The following matters were raised in discussions of the tasks and items proposed by Mathilda:

- whether it was legitimate test practice to include, for example in the multiple choice distractors, information which is not actually in the text
- the 'give-away' factor when a distractor is included that clearly comes from a part of the text distant from the one on which the question set is focusing
- the possible bias of items concerning a project in countries from which some candidates and not others, actually came, and who might know more from personal experience

In the editing discussion of items here, as for all three texts, colleagues were able to point out one or two items which were flawed because of a falsifying point in the text unnoticed by the actual item-writer.

Mary text editing

Mary's text, 'The Rise of the Emotional Robot', had been taken from the *New Scientist*. She had herself reduced the original by 15% to meet the 950 word maximum for an IELTS text. Mary was found (see next section) to have made 30 edits in all, including vocabulary changes – (more changes in fact than Mary herself had indicated, feeling, as she claimed, that texts should not, in the interests of authenticity, be changed too much – see Table 3 above).

At the beginning of the editing session on her text and items, Mary made the following additional points regarding changes to her original text:

- modifications to render the text more academic, 'cohesive' (and 'IELTS-like') through order change
- changes to the final paragraph to add strength and self-containedness to the end of the text
- one deletion from the original had been both to shorten the text to within IELTS limits (950 words) and because the experiment concerned was not one she intended to ask questions about

After discussion with Victoria and Mathilda, who had just read her text, two further modifications were made to Mary's text:

- one sentence was deleted from the text, as repetitive
- reference to the theory of mind was reinstated from the original text
- the order of sentences in the final paragraph was modified for stylistic reasons

Mary item editing

In the context of the research, the discussions of the tasks and items drafted by Mary, Mathilda and Victoria should be informative with regard to both the item writing and editing processes. The following were the main issues raised over the tasks and items proposed by Mary:

On the matching task:

- potential overlap was identified across the source statements leading to some ambiguity in the pairings; modifications were suggested accordingly
- use in the items of the same word(s) as in the text could give away some answers; IELTS-oriented textbooks tend to teach for parallel meanings

On the summary completion task:

- there was some confusion over the difference, if any, between 'passage' and 'text'
- it was clarified that the (not more than three) completing words had to actually appear in the original text but some doubt remained over whether a different form of the same word was eligible for use
- the summary completion passage was modified to allow for this

On the multiple choice task:

- instances of more than one item choice being acceptable because of semantic overlap eg, respect and love, were discussed
- the discussion here raised a multiple choice task issue of whether all alternatives should be similar in function, eg, all four about facts or all four inferences, or whether alternatives can be mixed in terms of function, presence or absence in the text (as in a true / false / not given item) etc? do candidates know such IELTS rules or conventions? in such cases, the test designer has the option of changing the item or changing the distractors
- the test item-writing and editing process here is described by Mary as 'finding the area and going over it with a fine-tooth comb'

It emerged during the editing session that as a part of the editing process both Mary and Victoria had asked friends to take their tests as a check on whether these were successful. Both writers had found this helpful in guiding further improvements.

This part of the session ended after 40 minutes' discussion of the items.

6.1.1 Choosing the text for the exam

The initial choices among the three non-experienced item-writers were as follows:

Mary favoured Mathilda's 'Sustainable Cities' text, finding:

- the robot text (her own) lacked 'meat'
- the dreams text was 'too hard' (for her)
- the cities text, being descriptive, was more easily exploited for items and distractors

Mathilda favoured Mary's 'Robots' text, finding:

- it contained enough meat in the opinions expressed, the tensions described, the hurdles presented
- it was at an appropriate level of difficulty, yet was reader-friendly

Mathilda now considered her own 'Sustainable cities' text:

- too fact-based and argument free
- lacking the challenge or need for deeper understanding of an argumentative text

6.1.2 Change of view caused by the editing process?

Victoria still liked her 'Dreams' text but was now less confident about her tasks. She considered it necessary to do far more analysis of potential texts and tasks. The three in the group still did not know the optimum processes but were, rather, acting on the basis of common sense and their experience as teachers. Mathilda felt the need for a whole range of IELTS tests and tasks to analyse to increase her awareness of suitable texts, tasks and what they are supposed be testing. Mary agreed, not having been trained as an IELTS item writer, it was difficult to know which words you can use in a text, how much you can test inferences.

Victoria would like to know about technical testing matters such as evenness of distractor and response lengths, Mathilda wanted more official information on IELTS to know more about IELTS level of difficulty, mark allocation and analysis. All three participants felt that the 'rules' of IELTS are 'pretty well hidden'. Their own help to their IELTS students in how to deal with IELTS Reading test items was common sense rather than officially informed.

Victoria, who was aware of IELTS Writing paper pre-testing and other validation procedures, wondered whether the reading paper was subject to similar procedures. As will be apparent from our review above, IELTS does publish information on these issues on its website and through other sources such as *Cambridge Research Notes*. The response of this group therefore may indicate lack of awareness rather than lack of availability.

In response to the final question, what the three participants felt they had learnt from their day:

- Victoria, assuming that reading for an IELTS reading test was different from other reading and in spite of having looked at information on task types on the IELTS website, still wished to know how test-takers should read in the test
- Mary, on the issue of what we are meant to be testing and how do we test it, wondered when is it appropriate to be testing vocabulary and when is it not
- Mathilda, wished to know, relatedly, how questions towards the broader and narrower understanding of a text should be balanced

Learning from the activities of the day, the participants noted the following things that they would do differently in future:

- Mary would, in advance, chart her intended items in terms of their intended difficulty and scope, to ensure evenness of coverage
- Mathilda would like to challenge herself using a fictional, more literary text (for example a short story) for her item writing; she still wondered why IELTS reading seemed not to use such texts; her impression was that the range of topics covered by IELTS academic reading modules was somewhat narrow
- Victoria would do a close analysis of more IELTS reading papers before she began setting her own; she was seeking a match between text type and task type; this could

mean a change of direction from choosing text and topic first; as an item writer, she might prefer to feel more responsible for the kinds of task she was going to set

- Victoria did not feel that the experience of the day had clarified why test-takers often found the IELTS reading module more difficult than the other modules (although, as noted above, reading scores are generally higher than for writing and speaking). Perhaps it was less clear with the reading module than with the others what test-takers were supposed to be doing

6.2 The experienced group

With Anne acting taking the role of chair, the participants were asked to organise and implement the joint editing session as they would a routine IELTS editing meeting (without further intervention from the research team). The intention was to prepare at least one text and set of items for the next stage in the test production process: pretesting.

Given the constraints on time, it was anticipated that it might not prove possible to go through the full process with all of the texts. In the event, the group were able to carry out the full editing process with Jane's text and looked closely at the text and one of the three item sets for both William's and Elizabeth's submissions. The group spent an intensive 85 minutes on Jane's text and items – the majority of the time (66 minutes) being devoted to the items. This seemed to the participants to be quite typical of the degree of attention that might usually be given to a submission in an editing meeting, although the point was made that a number of the issues might have been identified in a pre-editing session: a step that was not included in this project.

The 85 minutes spent on Jane's submission compares with a total of 68 minutes spent on the other two submissions considered at the meeting (29 minutes on William's and 39 minutes on Elizabeth's). Because of the time constraints and because is not usual for the chair of an editing meeting to lead the evaluation of her own submission, Anne's was not addressed at the meeting, although her text is considered in Section 6 above. As with the non-experienced writers, the following summary focuses, qualitatively and inductively on key points raised.

In each case, the group began by commenting on a text, suggesting changes which were noted by the chair. They then looked in detail at the related items, agreeing on and noting changes before passing on to the second writer's work.

Jane text editing

There was some discussion about the meaning of the text and the nature of the automated systems described. For example, the use of 'scheduled', 'selective' and 'sift' in the first paragraph caused some confusion with discussion about whether it was the machines or human experts selecting and sifting material. Elizabeth asked whether others shared her understanding that the 'CONE' system was partly and 'ACONE' entirely autonomous. William sought to clarify the roles of the university partners in the study and this question was discussed at some length. Anne queried the ordering of the units used in describing the camera's focal range in the fifth paragraph: 10m to 40mm. William also questioned whether this was accurate as 40mm seemed very short. It was agreed that the figures should be checked.

A number of proof reading errors were identified. For example, William found an intrusive comma in line 3. Problems were also noted with the formatting of the text and the appearance of abbreviations for measures. It was agreed that the names of the universities involved in the research should be standardised and the order of their listing standardised.

Some issues were identified concerning technical vocabulary: Anne suggested glossing 'GPS' in the third paragraph and this was agreed.

A number of changes were intended to improve the coherence of the text:

- There were questions relating to the paragraphing. Elizabeth suggested having the first sentence as a subheading as it seemed not to relate closely to what followed. This was agreed and the change was made. She also questioned whether the last sentence of the second paragraph should be moved to the third paragraph. This was not agreed.
- Elizabeth suggested removing the sentence 'They also use ultra-light aircraft to conduct aerial surveys' in the third paragraph as it seemed to contradict statements about the entirely automated nature of the ACONE system. This was agreed.
- The first sentence of the fourth paragraph was reworded. The original wording was: 'In February 2007, the University of California, Berkeley announced the installation of a high resolution intelligent robotic video system…'.

This was revised to give 'In February 2007 a further stage of the project began when the University of California, Berkeley announced the installation of a high resolution intelligent robotic video system…', This was felt to clarify the relationship between the installation of the video system described in the fourth paragraph and the research described in the third paragraph.

- Elizabeth suggested that, as the acronym had already appeared, CONE in the final paragraph did not need to be expanded again here. This was agreed.

Jane item editing

On the True/ False/ Not Given items:

- **Item 1** Anne questioned whether the first item was sufficiently precise as it could be taken to refer to all wildlife experiments. Elizabeth pointed out that the whole task covered the use of equipment in wildlife experiments and that the necessary information might therefore be difficult to locate. Suggested rewordings were not satisfactory and, following discussion, the item was rejected. As a result, a new item needed to be written.
- **Item 2** Elizabeth suggested that 'only a few occasions' might better reflect the sense of the text than the more negative 'little chance' in the item. Jane wanted to replace 'chance' as this repeated a word in the text. The word 'record' was preferred to 'capture' which seemed ambiguous when discussing wildlife.
- **Item 3** William objected that the text implied that this 'Not Given' statement was true. Elizabeth queried the meaning of 'examine the different species': in the text it was recordings that were being examined. These two objections were dealt with by rewording the item as, 'Those examining the data on target species would benefit from further training' which was felt to be more clearly 'Not Given' in the text.
- **Item 5** Anne queried the order of items 4 and 5. Jane confirmed that these should be reversed. The tense of 'will activate' was changed to 'activates' and 'some' was replaced by 'certain'.
- **Item 4** This was accepted with little discussion.
- **Additional Item** The group agreed that a new item could be generated from the untested material at the end of the second paragraph. The distinction that had been discussed earlier between CONE and ACONE was identified as important information. The group arrived at 'CONE relies entirely on input from non-human sources'. William suggested that the word 'entirely' could be a trigger to test wise candidates, but this objection was overridden. Nonetheless the group was unable to arrive at an agreed wording. After four minutes of discussion, the group failed to arrive at a satisfactory conclusion and decided

to move on. Returning to this section after editing the other items and following discussion of a number of alternative suggestions, finally they settled on: 'CONE uses data from robotic devices instead of input from scientists'.

On the Notes Completion Items:

- **Item 6** The item wording was ambiguous: 'in' could refer to time (1940, the intended key) or place (North America). Adding 'year when' at the beginning was felt to be unnatural for a set of notes. The compromise arrived at involved changing the text provided to the candidate to read: 'when bird was last seen'
- **Item 7** This was accepted
- **Item 8** 'involves' was replaced by 'causes'. In the following line, 'old recordings' was felt to be ambiguous and 'of target species' was added to clarify this
- **Item 9** The item was seen to depend on syntax: 'a' in the item cued 'match' in the text. It could be unclear what the 'match' referred to. The item was revised to give: 'results analysed to identify any [matches] with original recordings'. The text was also edited so that 'a match' became 'matches'
- **Item 10** This was accepted
- **Item 11** This was accepted following discussion
- **Item 12** The use of 'already' was questioned. This was replaced with 'have'. William pointed out that 'birds' would also be correct here. The agreed item read: 'birds that cameras have taken pictures of:'

On the short answer questions:

- **Item 13** Anne had not been able to find the correct answer and suggested replacing 'professional group' with 'professionals'. This was accepted.
- **Item 14** This was accepted without changes.
- **Item 15** William felt this was ambiguous as 'from' could be associated with the events or the monitoring: 'events from the field' or 'monitoring events from the laboratory'. It was decided that 'from' should be replaced with 'in' in the text. The agreed item read 'Where are the biologists in the CONE study located when they are monitoring events': the key was '(in) (their) laboratories'.
- **Item 16** The word 'feature' was replaced with 'function'.

William text editing

The reference to the doctrine of 'original sin' in the second and seventh paragraphs was queried on the grounds that this might be confusing to students from a non-Christian background. 'Christian tradition' was replaced with 'long held belief'. William argued that the term 'sinful' should be acceptable without glossing, but the religious implications were seen to make the text questionable. Alternatives such as 'wickedness' and 'guilt' were considered, but rejected. Anne felt that 'it would be very difficult to get round this, quite frankly' because religion was considered a 'taboo' subject for IELTS. William observed that 'most history seems to be impossible' because of the cultural element. Words such as 'church' or 'mosque' could not, he felt, be used in IELTS. The question of how to eliminate the religious element in the text was put to one side so that editing could proceed.

Elizabeth and Jane both questioned the use of 'ivory tower'. After a number of attempts at rewording, the sentence 'Art was on the run; the ivory tower had become the substitute for the wished for public arena' was eliminated on the grounds that the idea had appeared in the previous sentence.

The 'dense' nature of the text was seen to be a potential shortcoming and there was some confusion over the temporal progression of ideas. Elizabeth asked for clarification of 'late' C19th in Paragraph 7.

William item editing

- The group looked closely at William's second set of questions (matching) and identified certain issues:
- Potential guessability: Jane had been able to guess items 8 and 10, but wondered whether these would be guessable for certain candidates. How far might candidates be expected to know about the history of English literature?
- The stems for items 7 and 11 ('Authors working prior to the late 18th century' and 'In the harsh society of the 19th century, some authors') did not seem to fit well with the stems for items 8, 9 and 10 which (names or lists of names of individual authors)

The conclusion of this session was that the text would probably have been returned to the writer at the pre-editing stage with comments on the cultural elements. The issues identified and communicated to the writer would need to have been resolved before the text could have progressed to editing.

Elizabeth's text editing

All three other writers queried the inclusion, in paragraph 3, of 'eke out the very last quantum of sleepiness', but Anne decided to delay revising this until the group came to address item 2, to which it related. They also questioned 'trotted out' as being too colloquial. The latter was replaced with 'frequently put forward'. These were the only issues raised in relation to Elizabeth's text.

Elizabeth item editing

- **Item 1** Anne had failed to find the correct answer, although William believed it was 'strongly there'. The use of 'accurately reported' in option C was questioned as it might refer to the original reporting of the Stanford study by the researchers rather than to subsequent misrepresentations of it. The use of 'misunderstood' seemed to address this. Anne suggested replacing 'with' in the question stem with 'in'.
- **Item 2** William felt that option B could also be true. The use of 'unrealistic' was identified as problematic and was replaced with Elizabeth's suggestion of 'stressful'. Here the focus moved to finding an appropriate rewording of the problematic element in the text identified earlier. After discussion, 'they are able to eke out the very last quantum of sleepiness' was replaced with 'participants are able to exploit their opportunity to sleep to the full'. As a result of the change, 'unnoticeable' at the end of the sentence became problematic. This had modified 'sleepiness'. The issue was resolved by substituting 'unattainable' for 'unnoticeable'. Elizabeth then suggested reversing the order of options C and D so that the key (originally D) would not come last in the set. This was agreed.
- **Item 3** No suggestions were made and the item was accepted.
- **Item 4** No suggestions were made and the item was accepted.
- **Item 5** All distractors dealt with the issue of the relationship between sleep and obesity and were felt to be acceptable.

- **Item 6** William suggested that confusion that might be caused by using the negatively worded 'underestimating how little sleep'. The alternative 'overestimated the amount of sleep' was preferred.
- **Item 7** The use of the vague 'a particular type of question' in the stem was queried. This was replaced with 'a question like "would you like more sleep?"' which had the advantage of being both more explicit and matching exactly the relevant section of the text (the eighth paragraph). However, the implications of making the relationship between item and text so much more explicit were not discussed. Option B was then felt not to work with the revised stem. This was replaced with 'may give rise to answers on other topics'. The options were then reordered to make D the key, balancing the number of each option that appeared as the key.

Conclusions from the day's activities and discussions

Reflecting at the end of the day, the group felt that the meeting had represented a typical editing meeting, working at what Anne described as a 'realistic pace', at least on Jane's material. However, the point was made that the pre-editing stage would probably have helped to eliminate some of the textual issues that had emerged.

William's submission had highlighted the difficulties of exploiting arts texts, answering, in a sense, the question raised by Mathilda during the inexperienced item writers' deliberations. Arts texts often included a number of culture-specific elements. Elizabeth suggested that it was also an issue that such texts assumed background knowledge: 'they always assume you have read the work or seen the picture'. William was reminded that the editing meeting would always throw up problems that he had failed to find when reviewing the text: 'I always find things at editing that I hadn't noticed before'.

Aspects of Elizabeth's text such as 'trotted out' had highlighted the need to remove journalistic touches from the text to achieve a more neutral academic style. Magazine articles often began with an attention grabbing anecdote or example before moving to a more general point, while, William suggested, academic texts more often started from a generalisation. Anne had cut the first paragraph from her source text for this reason.

There was a contrast between the length of IELTS texts and the length of the texts that students would need to read '900 words versus a book' as Elizabeth put it. Elizabeth defended the use of relatively short tests in IELTS, stating that 'we are not testing what they may be able to do after a few months at university; we are testing whether they will be able to cope, I think'. William pointed to the great variety of texts that could be encountered at university, some of which would be more straightforward for students than IELTS texts. He suggested that 'somebody who struggles with texts like these might be able to cope perfectly well with physics texts' which might contain more technical vocabulary, but less subordination.

Anne felt that IELTS, by moving between topics and by moving from fact based to more discursive texts might 'reflect in miniature what [students] have to do… look at a variety of sources, get key ideas, get attitudes, get opinions' while Elizabeth countered that, given the practical restrictions on what could be covered in a one-hour test 'there is a huge amount we don't do of course: dealing with contents, dealing with indexes, dealing with chapters and all that sort of macro stuff. We can't do it.'

Preparation courses were considered to be helpful in improving reading skills and in building exam technique. Elizabeth reported that students she had taught had found learning for IELTS useful in preparing them to read longer texts. Elizabeth believed that there was a 'core vocabulary' for the test that could be taught and there was general agreement that the strategies used in IELTS would transfer to reading texts for academic purposes.

6.2.1 Analysis and findings on the items

As with the texts in Section 5, the analysis here is applied to the items as they were submitted by the seven participants, before any changes made during the public editing process. Again, links are made with the comments from the participants and the edits made during the meetings.

Table 9 shows the task types selected by the three writers for their commissioned items. No writers chose Types 3 (Sentence Completion), 5 (Labelling a Diagram), 6 (Choosing Headings for Paragraphs or Sections of a Text) or 9 (Classification). This may reflect the difficulty, discussed by the experienced writers, of finding or creating suitable diagrams. It is also of interest, given William's comments on the cost-benefit of multiple choice items and the reservations expressed by Jane that all three non-experienced writers attempted these, but only one of the four experienced writers did so. However, this might also reflect the relative familiarity of this item type for the non-experienced group.

Task Types		*Victoria*	*Mathilda*	*Mary*	*Jane*	*Anne*	*William*	*Elizabeth*
Type 1	Multiple Choice	√	√	√				√
Type 2	Short-answer questions	√			√			
Type 3	Sentence Completion							
Type 4	Notes, Summary or Table/Flow-chart Completion			√	√	√		√
Type 5	Labelling a Diagram							
Type 6	Choosing Headings for Paragraphs or Sections of a Text							
Type 7	Locating Information		√			√	√	√
Type 8	Identification of Writer's Views/Claims or of Information in a Text	√	√		√	√	√	
Type 9	Classification							
Type 10	Matching			√			√	

Table 9: Task types (based on list given at www.ielts.org) selected by each item writer

The most popular task type, chosen by six of the eight writers, was Identification of Writer's Views/Claims or of Information in a Text or True/ False/ Not Given (T/F/NG). It is clear from the focus group discussions that this is seen by the experienced writers as a flexible and relatively straightforward task type to work with. In the following section we analyse the writers' items in some detail, drawing on Macmillan's (2007) typology of lexical relationships between texts and items, to

explore how items are used and how the two groups interpreted the requirements. In this case, we begin with the experienced writers' items.

Jane was the only one of the experienced writers whose T/F/NG items were edited at the meeting. The comments and revisions made provide insight into the experienced writers' conception of this item type.

Jane's only False item (item 1) relies on antonymy. The location of the necessary information is clearly signalled by the repetition of 'programmed' and 'data' in the item while 'random intervals' in the stem is contrasted with 'scheduled intervals' in the text. However, the editing team objected to the open-ended reference of 'wildlife experiments' and were unable to find a satisfactory alternative. As a result they chose to reject the item. A replacement item was written and added to the end of the set during the editing meeting and is discussed below.

Item 2 (True) is also clearly signalled (through the repetition of 'chance' and 'species') and involves lexical repetition: 'the chances are very low' (item): 'there is little chance' (text); 'the target species' (item): 'the species being investigated' (text); and synonymy: 'the equipment used will capture': 'recording an occurrence'. The phrase 'some cameras' in item 5 (True) matches 'some wildlife cameras' in the text and the item paraphrases a single sentence from the text.

In item 3 (Not Given), the location of the necessary information might seem to be most clearly cued by the similarity between 'field studies' in the stem and 'fieldwork' in the text, although this is probably not the intended location as 'fieldwork' occurs in the opening line and items of this type usually follow the sequence of the text. The word 'experts' in the stem repeats 'expert' in the first paragraph, although this word occurs twice more in the text. The repetition of 'species' might serve to limit the search, but also cues the previous item and so might be thought to jeopardise item independence.

Assuming that it is the occurrence of 'expert' in the first paragraph that is intended, the successful test taker would need to recognise either that there is no mention of either experts or cameras examining 'the different species' or that the 'insufficient' number of 'experts' mentioned in the stem is not suggested as a reason for sifting the field data. It may be, however, that this is a plausible inference. For this reason, this does not appear to be a very satisfactory item. The item writers recognised the plausibility of the inference and rewrote the item. The revised item 'Those examining the data on target species would benefit from further training.' includes direct repetition of the phrase 'target species' and 'examining' matches 'examination' in the following line. There is a reference to 'trained biologists' in paragraph 3, which might serve to distract the test taker, but there is no other reference to training. There is no relationship between 'would benefit from further training' and the information given in the text so the new item appears to be unambiguously Not Given.

The location of the necessary information for item 4 (also Not Given) is more clearly signposted: 'cameras for wildlife filming' and 'surveillance cameras' in the stem co-refer with (and repeat words from) the phrases 'wildlife cameras' and 'surveillance systems' in the text. The text compares the operation of the two systems, but does not make any statement about the organisations that manufacture them. Here both the noun phrase 'organisations' and the verb phrases 'produce' and 'make' relate to information that is also entirely absent from the text. This item was accepted by the group without changes.

The task types used by Anne include Type 8: Identification of Writer's Views/Claims or of Information in a Text(True/ False/ Not Given), Type 4: Summary Completion and Type 7: Locating Information. Information required to respond to the first section was located in the first, second, fifth, sixth and seventh paragraphs (of nine). Information required to respond to the second section could be found in paragraphs four and five and that for the final section in paragraphs eight and nine.

The information required to respond to Anne's T/F/NG items can be found in sequence in the first seven paragraphs of her text. Each question consists of a paraphrase of information given in the text. Identifying whether the answer is correct rests on the ability to identify co-reference and hence to map words or phrases in the question onto words in the text. Generally there is a clue to the location of the information provided by use of words or phrases in the question that precisely match words in the text ('six million years ago' in item 3, 'tickling machine' in item 8), are different forms of the same words ('humour' in item 4 matches 'humorous' in the text), or that are close synonyms ('wrote down' in item 1 matches 'noting' in the text; in item 7 'research into tickling' can be matched to 'studies of tickling' in the text).

Inference is also important to finding the correct answers, despite the potential risk of personal differences of interpretation (see above). In item 1, the correct response (False) requires the understanding that 'students' are not equivalent to 'people' in 'public places' while in item 6 'nearly' implies development later than six million years ago, not 'before' as in the item. Antonymy is also used: in item 7 (False) a 'considerable amount' contrasts with 'thin on the ground'.

In Anne's first Not Given item (item 3), there seems to be no clear signal of where the necessary information occurs in the text. There are a number of plausible lexical links to the text: the word 'episode(s)' comes at the beginning of the second paragraph, followed by the division ('sorted') according to the characteristics of research subjects, but this information is targeted by the previous item and there is the identification of three 'facts about laughter' in the following sentence. In either case, the test taker might recognise that the division mentioned is not connected to 'kinds of laughter', as in the stem. Further, there is no mention here of options that Provine (the key researcher in the text) may have 'considered' for his data analysis. Recognising that such information is not available is likely to require reading both more of the text and more careful reading than Jane's item 4 or revised item 3.

Compared with Anne, William includes more direct phrasal matches – lexical repetition – in his items with the relevant information in the text. His items also involve more direct word matches than Jane's. Item 1 has 'Blake and Wordsworth' and 'expressing'; item 2 has 'industrial revolution' ('Industrial Revolutions' in the text) and 'social problems' matching 'social, political and especially intellectual problems' in the text. Item six (False) has the most extensive cues with almost half of the words also occurring in the text including 'the 19th century' and 'the concept of the innocence of children' which repeat elements of their co-referents in the text: 'the nineteenth century' and 'the concept of the child as a symbol of innocence'. As in Anne's items, William's questions make extensive use of paraphrase, synonymy and antonymy. Item 1 (False), for example, would seem to require understanding the contrast between 'adapted a tradition' in the item stem and 'an essentially new phenomenon' in the text.

Perhaps because so many phrases recur in the text, a number of William's items might appear to a candidate plausibly to relate to several locations. Item 6 contrasts 'retained its power' with 'deterioration', although this proposition is echoed in the following sentence: 'only a residue', 'retaining little or nothing'. Similarly, there are several clues to the answer to item 4 (True): the proposition that serious writers were marginalised by the growth of mass literature is effectively repeated three times: 'mature voice… diminished', 'art was on the run' and 'ivory tower… arena' – a fact exploited in the editing meeting when the team decided to eliminate the reference to 'ivory towers'. Item 5 (True) seems to paraphrase a sequence of three related sentences which repeat the idea that nineteenth century authors used the image of the child to express their alienation from industrial society.

William's two Not Given items both repeat lexis from the text to point to the location of the necessary information. In each case one phrase in the text is inaccurately paraphrased in the item stem so that in both cases, substitution of one phrase would yield a True item. For item 2, in the text, it is the author of the text, rather than 'a number of writers' who identifies the industrial revolution as a 'cause of

social problems' while in item 3 the phrase 'was proportionally diminished' – paraphrased in the item by 'featured less often' – relates to the 'creative voice' of the serious writer rather than 'children'.

The variation in the relationship between the items and the text found among these writers is consistent with and so perhaps might help to explain the variation found in the strategies used by test takers responding to four T/F/NG test sections by Weir *et al*. (2009b). In that study, test takers made greater use of word matching strategies and knowledge of vocabulary in some T/F/NG sections of the tests they took than in others. There were also differences in whether the information necessary to answer the question was most often reported as being found within sentences or across sentences. Thus different interpretations of the guidelines appear to lead writers to produce items that target different types of reading on the part of test takers. We note that there was no discussion among the item writers of how changes in the items might affect the reading skills being used by test takers or of the implications of variation in T/F/NG items for the nature of the test.

Of the three non-experienced writers, Victoria and Mathilda employed T/F/NG items. Victoria's T/F/NG items are closer to the experienced item writers' than are Mathilda's in their use of paraphrase and synonymy. She prefers to reorder or rephrase constituents from the text in her items so that in item 1 (False) 'dreams seem to make perfect sense to people…' is rephrased as 'people tend to make the most sense of their dreams…'; in item 4 (True), 'loss of memory' becomes 'memory loss'; in item 6 (True), 'much like [a], [b]' is replaced with 'both [a] and [b]'. There are lexical repetitions between text and item – 'experiences' (Item 3), declarative' (Item 4), 'the hippocampus' (Item 5), but these are usually individual words rather than phrases. Arriving at correct responses to Victoria's items generally involves relating phrases in the items to co-referents in the text. Sometimes, as in the case of item 1 (False), this also involves resolution of referencing within the text – recognising that 'some answers' refers to the question of 'their origins', which in turn refers to 'dreams'.

In comparison to the trained item writers and to Victoria, Mathilda's T/F/NG items make less use of synonymy and paraphrase, instead her strategy involves repeating extensively from sections of the text, paraphrasing only the necessary information. The sequencing of information within the sentence in the text is retained in the item -item 2 (True) item reads 'More than a quarter of carbon emissions in the USA result from burning oil for transportation'. This closely reflects the relevant passage in the text: '28 percent of carbon emissions in the United States result from the burning of 14 million barrels of oil a day for transportation'. Similarly in item 1 (True) the item closely reflects the text and matching the paraphrase 'being built next to' with 'under construction adjacent to' gives the answer. Item 3 (False) is equally explicit, but the paraphrase ends with 2013 (which occurs in the preceding clause) in place of 'next year' from the text. Mathilda's two Not Given items represent rather different approaches to the item type. In the first (item 4), she paraphrases a sentence from the text, here replacing one constituent, 'residential sources' with another, 'motor traffic'. Item 5, in contrast, is inadequate as a paraphrase because it incorporates details from the following clause into the definition of the 'green roof'.

7 COMPARISONS BETWEEN GROUPS

The approach adopted for this study involved asking both experienced and inexperienced writers about their practices based around Salisbury's (2005) phases of the item construction process. The study collected both written (flowchart) and oral (interview and focus group) data on item writer processes and products (draft and edited texts and items) and incorporated both deductive and inductive approaches to analysis.

This approach has proved useful in identifying differences across the item writer groups and between individuals within the groups. These differences highlight both the importance of item writer training and guidelines and suggest changes that might be advisable. In the context of recent research undertaken by the University of Bedfordshire into the IELTS Academic Reading test, the current study

can help to explain some of the characteristics of IELTS texts identified by Weir *et al* (2009a) and the types of reading employed by IELTS test takers (Weir *et al* 2009b).

7.1 Item writing processes

Both the experienced and non-experienced item writers seem to pass through similar steps in constructing their items. They typically begin from a topic, locate texts related to the topic, identify and evaluate potential IELTS texts before selecting one that seems appropriate – this is clearly Salisbury's (2005) exploratory phase. Both groups reported that they found this the most time-consuming stage in the item writing process.

With the exception of Jane, the experienced writers all included more steps in their item writing flow-charts than their non-experienced counterparts. The flow charts include similar attention to text editing in both groups, but there is greater attention to task development among the experienced group: this being broken down into number of steps including revision and re-editing of the text following or in conjunction with item writing.

In the next phase – recognisable as Salisbury's (2005) concerted phase – all of the writers carried out an iterative process of editing the text and developing the items. Unlike the writers in Salisbury's (2005) study, who were devising scripts for tests of listening comprehension, these writers could not be said to have started from their items in writing their texts. However, as observed by Salisbury (2005) in her study, the experienced writers seemed to have a repertoire of gambits for efficiently exploiting their source texts and paid attention to task type in text selection. They also paid attention to potential items during the initial exploratory phase – highlighting or making notes on testable material. While the untrained writers selected material that was already close to the appropriate length, trained writer texts chose much longer pieces then progressively cut out passages that seemed to repeat information or that included elements that would not be tested. The extent of editing and the desire to avoid repetition perhaps explains why the texts analysed in Weir *et al* (2009a) displayed relatively high type: token ratios in comparison with undergraduate textbooks (indicative of a wide range of vocabulary use and rapid progression of ideas).

As a first step in what Salisbury (2005) calls the refining phase, the experienced group favoured attempting the task themselves after an intervening period (although deadlines sometimes limited the opportunities for this). The non-experienced writers also reported attempting their own tasks, but Mary and Victoria additionally asked friends to respond to their tasks and so were able to obtain some further feedback on how well the items were working before coming to the editing session.

7.2 The texts

The non-experienced writers drew on very similar sources to their experienced counterparts. Both Mary and Elizabeth chose articles from *New Scientist* articles while both Mary and Jane selected texts concerning robot technology. Victoria's text was an article from a popular science magazine concerning dreams while Anne's was an article from a popular science magazine concerning sleep. Readability statistics for the two groups were also very similar. The easiest and most difficult texts according to the Flesch Kincaid and Coh-Metrix measures were both produced by experienced writers (Jane and Elizabeth respectively).

Both groups expressed a concern that the selection of topics in the test may be rather narrow. Where the non-experienced group saw this as a constraint imposed by the need to produce IELTS-like texts, the experienced group saw it as a by-product of the need for accessibility and cultural neutrality: arts texts tend to assume or require background knowledge in a way that popular psychology or technology-based texts do not.

Members of both groups edited their (magazine) texts to make them more 'academic' in style and tone and less journalistic. All of the texts involved plausibly academic topics presented for the general

reader. All writers in both groups edited to eliminate (some) vocabulary on the grounds that it was either too technical for the general reader, too colloquial to be appropriate in an academic text or too infrequent and so difficult for IELTS candidates. Both groups included factual texts (Mathilda's text on 'cities' and Jane's 'wildlife cameras' text) and opinion texts (William's essay on literature, Elizabeth's 'sleep' text, Anne's' laughter test' from the experienced group; Mary's 'robots' text and Victoria's 'dreams' text from the untrained group).

Members of both groups also sought to avoid potentially divisive or offensive issues and to eliminate culturally specific knowledge from their texts. Mary removed a paragraph from her text concerning war. The experienced group was concerned to avoid religious issues in William's text.

The trained writers seemed more ready to edit their texts; reshaping them if necessary to meet the requirements of the items. Of the untrained writers Mary seemed to have the strongest objections to revising her text, but in fact made the most substantial changes of this group. These changes included moving material between paragraphs to square her text with the items she wanted to use.

In sum, the effect of editing for both groups, apparent in the analysis of the submitted texts and from the discussions in the editing meetings, was to increase the coherence and information density of the texts and to make them more accessible to readers from non-English speaking backgrounds. The changes also served to reduce technical and cultural specificity, colloquialism, journalistic touches (such as sensationalism, personal engagement of writer etc) and, particularly in the case of the experienced group's texts, to reduce the repetition of ideas.

In devising their items, both groups made use of a range of item types. The True/ False/ Not Given (T/F/NG) item type was chosen most often across groups, but no clear differences in item type selection could be seen from the small sample submitted.

As was to be expected, the experienced item writers submitted items of better quality and were better able to correct the problems that they found. A greater number of shortcomings that would breach the IELTS item writer guidelines could be identified in the untrained writers' submissions. For some untrained writers, items within sets did not consistently follow the order of information in the text where this would usually be expected (as in Mary's MCQ items: 15 and 17 concern the first paragraph, 16 is associated with the sixth paragraph and the necessary information for item 18 is distributed throughout the text). Items within a set were sometimes isolated from each other: Mathilda's item 17, for example, relates to her ninth paragraph while the rest of her T/F/NG items are associated with the first four paragraphs of her text.

The items submitted by the untrained writers sometimes addressed the same parts of the text more than once. Victoria, for example, has three pairs of items that seem to address the same sentences in her text (items 2 and 3; 4 and 5; and 8 and 13). Untrained item writers' texts included stretches of untested material: five of Victoria's 16 paragraphs did not include information required to respond to any of her items.

The non-experienced writers felt that their lack of guidance about the test inhibited their ability to produce adequate items. They felt that they would have benefited from information on devising MCQ distractors and on the skills being targeted by items of different types. It should be noted that these writers had been directed to the Teaching Resources section of the IELTS website, which provides some guidance on this question under the heading of 'What skills are tested in this task type?' However, the information is inexplicit. For Task Type 8 – Identification of Writer's Views/Claims or of Information in a Text, the explanation is as follows:

The first variation of this task type aims to test the candidate's ability to recognise opinions or ideas, and is thus often used with discursive or argumentative texts.

This is not clear enough to guide an item writer. The intended relationship between the items and the text is not made plain and so the type of reading required is not explicit. The lack of guidance is reflected in the very different ways in which Mathilda and Victoria interpreted this task type.

In the editing meeting, the non-experienced group were relatively less critical of each others' work (although it should also be noted that, unlike the experienced group, they had not met each other before the day of the editing meeting). The experienced writers appeared far more efficient in their approach to the editing meeting and worked intensively on improving the texts and items. Each writer contributed numerous suggestions and the chair sought consensus on the proposed changes.

The experienced group were pleased with the guidance they had received from the item writer guidelines and from the experience of training and editing meetings and felt that this had contributed to their expertise. Nonetheless there were clear inconsistencies in the interpretation of task requirements between the experienced writers. The group seemed to share a conception that IELTS tasks should target key, salient facts or opinions expressed in a text and appeared less concerned with the reading skills involved.

The group had discussed at some length the nature of the information that could be targeted using Type 1 MCQ items and the extent to which inferences might be tested using Type 8 T/F/NG items. These discussions left open the possibility that different writers might be targeting different reading skills when using the same item type – as observed in Section 8, each set of T/F/NG items bore a somewhat different relationship to its partner text. This has implications for the comparability of different forms of the test as it makes it more challenging to ensure that every form reflects the required range of reading skills. These issues had not been resolved by the end of the session.

When reviewing and revising items, the writers identified ambiguities and suggested clarifications, but did not generally discuss the implications of changes of wording on the nature of the reading skills that might be required in arriving at a correct response or to the balance of skills being tested in a passage. The three task types in Anne's submission, for example, all appear to involve careful local reading. The items include eight Type 8 T/F/NG items, which involve paraphrase of information in the text, and two Type 7 Locating Information items which are also based on recognising paraphrases of information in the text – in this case distinguishing between the two sentences that paraphrase the information in the text (similar to True items) and the three that do not (similar to False and Not Given items). The item below illustrates how similar this is to a T/F/NG item. There are similar lexical relationships involving repetition (speech), synonymy (develop: evolve) and co-reference (early man: our ancestors).

In the item: Human speech began to develop when early man ceased walking on four legs.

In the text: When our ancestors stood up on two feet, the chest was freed of these mechanical demands making it possible for speech to evolve.

The third item set – Type 4 Summary Completion – involves selecting two-word expressions from the text to complete a summary of paragraphs 3, 4 and 5 – also seems to require understanding at a local level.

8 CONCLUSIONS AND RECOMMENDATIONS

The researchers were favourably impressed by the conscientiousness and professionalism of the IELTS item writers that we interviewed and observed and the quality of the texts and items that they produced. Nonetheless, we would suggest that there are a number of recommendations that could be made on the basis of our study to refine the IELTS academic reading item production process. The inter- and intra- group differences revealed by our research have implications for test preparation that could be addressed through information provided to teachers of IELTS candidates and implications for

the consistency of test material that could be addressed through the guidelines and training given to item writers and the process of text and test review.

Firstly consideration should be given to better informing teachers about the test by increasing the amount of guidance offered concerning the reading skills being targeted and the intentions behind using the variety of item types on the test. The information currently offered on the IELTS website does not seem to be sufficient to inform teachers about the construct of the test. The non-experienced writers felt that, based on the examples they had seen, they had a clear sense of the kinds of texts being used in IELTS, but were less clear on why such texts were being used. This impression was largely borne out by the texts they produced, which resembled the texts produced by the experienced group and those analysed by Weir *et al*. (2009a). However, the untrained writers did not feel well equipped to produce items that would test the same skills as the IELTS Academic Reading test. Although all were familiar with the Academic Reading test, they did not feel well informed about the function of the different items types or the implications of these for the types of reading being tested. More information on these aspects of the test could be communicated through the IELTS handbook, website and other communication channels.

From the testimony of both groups, there seems little doubt that observing the processes of editing and refinement that we report in this study is of great help to item writers in developing their own skills. Indeed, we would suggest that this research could be of use in training new item writers by providing insights into how texts and items are reshaped for the test and might also serve to inform a wider audience about the extent of the work that goes into producing each IELTS item. However, there would seem to be a need for an additional strand of training and guidance that pays greater attention to the construct of academic reading intended to be operationalised through the IELTS academic reading module.

In an earlier study (Weir *et al* 2009a) we recommended that objective text analysis tools could play a valuable role in assisting the evaluation of texts (and perhaps items) as part of the review process. We repeat this recommendation here because, as exemplified in our analyses, such tools can help in the identification of infrequent or technical vocabulary, help to highlight inconsistencies between the texts used across versions of the test and assist in identifying differences in textual genre that might be helpful in better defining the requirements for texts suited to the purposes of the test.

The participant interview and focus groups raised a number of questions that should be addressed in the item writer guidelines or related training packages:

- What are the reading skills that the test as a whole is intended to address? And in what proportion?
- Why these reading skills? And how do they relate to the available task types?
- Within each task, what kinds of linguistic relationships should T/F/NG (and other types of items) have to the text and in what proportion? What are the implications of these for the reading skills being targeted?
- What range of skills should be addressed in each section of the test and what item types should be used to target them?

The experienced item writers speculated on a number of issues including the kinds of information that could legitimately be targeted by MCQ and the extent to which inference could legitimately be targeted: there appeared to be room for individual interpretation in these areas. It is perhaps inevitable that individual writers (and separate teams of writers) will interpret specifications differently (and that there will be some misinterpretation), but questions of this kind should be answerable through the item writer guidelines. To the extent that they are, there should be greater attention to the guidelines during editing. To the extent they are not, the guidelines should be updated to address them.

The test providers should keep item writers informed about relevant assessment issues including current theoretical perspectives on the reading process, the nature of the reading demands made on beginning university students and the implications of these for IELTS. Such meetings, by raising issues of concern to writers, could also serve to direct further research into these questions that will inform the design of the test.

Elizabeth made reference to the discontinued practice of asking item writers to identify the skills being tested by each of their items. Elizabeth had found this difficult, but useful and consideration might be given to re-introducing such a practice as a training exercise if not as a routine requirement. It might also be advisable to introduce clearer controls on the range of task types and the range of skills to be targeted for each text.

Item writers reported that from their perspective some decisions made about test content could appear inconsistent. The fairness review element of the pre-editing and editing process was one area of concern. Items based on factual details in a text might inevitably advantage candidates who are familiar with the subject matter, but the question of which facts should be considered to be widely known and which not was a grey area for our participants. Similarly, these writers, who all worked on other Cambridge ESOL papers as well as IELTS, felt that there might be inconsistencies in the definition of potentially 'offensive' or 'sensitive' material across examinations. It may be that there is a rationale for such differences based in the nature of the candidatures for these different tests, but the implications for test content were not sufficiently clear to the item writing team. If this view is shared more generally by item writers, mechanisms should be found to create greater consistency in the interpretation of the rules, or to better articulate to item writers justified differences across testing programmes within Cambridge ESOL.

Finally, we believe that this study points towards a number of interesting avenues for future research. A comparison between the item writer practices investigated here and test taker strategies of the kind investigated by Weir *et al* (2009b) would provide insights into the extent to which candidate reading behaviours conform to item writer expectations. Similarly, it would be interesting to obtain candidate views on the kinds of editing changes made by item writers or to compare candidate judgements of what constitutes 'key' information in a text with item writer judgements. It would be useful, as a form of evaluation, to carry out a follow-up study after changes to item writer training and guidance have been implemented.

REFERENCES

Alderson, JC, 2000, *Assessing Reading*, Cambridge University Press, Cambridge

Alderson, JC, Figueras, N, Kuijper, H, Nold, G, Takala, S and Tardieu, C, 2004, 'The development of specifications for item development and classification within the Common European Framework of Reference for Languages: learning, teaching, assessment. Reading and listening. Final report of the Dutch CEF construct project' Retrieved from http://eprints.lancs.ac.uk/44/1/final_report.pdf 15 August 2008

Association of Language Testers in Europe, 1994, *Code of Practice*, University of Cambridge Local Examinations Syndicate, Cambridge

Babbie, E, 1992, *The Practice of Social Research*, 6th ed, Wadsworth Publishing Company, Belmont, CA

Bachman, L, 2005, 'Building and supporting a case for test use', *Language Assessment Quarterly*, vol 2, no 1, pp 1-34

Baranowski, RA, 2006, 'Item editing and editorial review', in *Handbook of Test Development*, eds, Downing, SM and Haladyna, TM, Lawrence Erlbaum Associates, New York

Bauman, J, 2005, 'The item writing process for access to ELLS: outline involvement by ESL teachers', presentation to the *Illinois 28th Annual Statewide Conference for Teachers serving Linguistically and Culturally Diverse Students*, 24-27 January, Centre for Applied Linguistics, Washington DC.

Bormuth, JR, 1970, *On the Theory of Achievement Test Items*, University of Chicago Press, Chicago

Brown, A, 2000, 'An investigation of raters' orientation in awarding scores in the IELTS oral interview', in *IELTS Research Reports Volume 3*, ed, Tulloh, R, IELTS Australia, Canberra

Brown, A, 2003, 'Legibility and the rating of second language writing: An investigation of the rating of handwritten and word-processed IELTS Task Two essays', in *IELTS Research Reports Volume 4*, ed, Tulloh, R, IELTS Australia, Canberra

Brown, A, 2006, 'An examination of the rating process in the revised IELTS Speaking Test', in *IELTS Research Reports Volume 3*, eds, McGovern, P and Walsh, S, IELTS Australia and British Council, London, pp 41-70

Cambridge ESOL, n.d., 'Cambridge ESOL exams and the CEFR: Familiarisation with the CEFR – making it work for us', retrieved 20 June 2008 from www.cambridgeesol.org/what-we-do/research/cefr/cef-familiarisation.html

Carroll, BJ and Hall, PJ, 1985, *Make Your Own Language Tests: A practical guide to writing language performance tests*, Pergamon, Oxford

Chalhoub-Deville, M, Chapelle, C and Duff, P, eds, 2006, *Inference and Generalizability in Applied Linguistics: Multiple perspectives*, Benjamins, Amsterdam

Clapham, C, 1996a, 'The Development of IELTS: A study of the effect of background knowledge on reading comprehension', *Studies in Language Testing, Volume 4*, UCLES and Cambridge University Press, Cambridge

Clapham, C, 1996b, 'The academic modules: reading', *IELTS Research Report 3: Constructing and trialling the IELTS test*, eds, Clapham, C and Alderson JC, British Council, UCLES and IDP Education Australia, London

Cobb, T, 2003, 'VocabProfile, The Compleat Lexical Tutor', retrieved on 5 December 2008 from www.lextutor.ca

Coxhead, A, 2000, 'A new academic word list' *TESOL Quarterly* vol 34, no 2, pp 213-238

Crossley, J, Greenfield, J, and McNamara, DS, 2008, 'Assessing text readability using cognitively based indices', *TESOL Quarterly*, vol 42, no 3, pp 475-493.

Davidson, F, and Lynch, BK, 2002, *Testcraft: A teacher's guide to writing and using language test specifications*, Yale University, New Haven, CT

De Jong, J, 2008, 'Procedures for Training Item Writers and Human Raters', presentation, 5 May 2008. http://www.ealta.eu.org/conference/2008/docs/friday/deJong.pdf

Dey, I, 1993, *Qualitative Data Analysis: A user-friendly guide for social scientists*, Routledge, London

Downing, SM and Haladyna, TM, 1997, 'Test Item Development: Validity Evidence from Quality Assurance Procedures', *Applied Measurement in Education* vol 10, no 1, pp 61-82

Downs, PE, 1978, Testing the upgraded semantic differential, *Journal of the Market Research Society* vol 20, pp 99-103

Ebel, RL, 1951, 'Writing the Test Item', in *Educational Measurement*, ed, Lindquist EF, American Council on Education, Washington, DC

Enright, MK, Grabe, W, Koda, K,. Mosenthal, P, Mulcahy-Ernt, P, and Schedl, M, 2000, 'TOEFL 2000 Reading Framework: A working paper', *TOEFL Monograph MS-17*, Educational Testing Service, Princeton, NJ

Fortus, R, Coriat, R and Ford, S., 1998, 'Prediction of Item Difficulty in the English Section of the Israeli Psychometric Entrance Test', in *Validation in Language Assessment*, ed Kunnan, A , Routledge, London

Freedle, R and Kostin, I, 1993, 'The Prediction of TOEFL Reading Comprehension Item Difficulty for Expository Prose Passages for Three Item Types: Main idea, inference, and supporting idea items', *TOEFL Research Reports RR-93-44*, Educational Testing Service, Princeton, NJ

Friedman, HH, Friedman, LW and Gluck, B, 1988, 'The effects of scale-checking styles on responses to a semantic differential scale', *Journal of the Market Research Society* vol 30, pp 477-482.

Fulcher, G and Davidson, F, 2007, *Language Testing and Assessment*, Routledge, London

Furneaux, C and Rignall, M, 2007, 'The effect of standardisation-training on rater judgements for the IELTS Writing Module', in *IELTS Collected Papers: Research in speaking and writing assessment*, Studies in Language Testing, Volume 19, eds Falvey, P and Taylor, L, Cambridge University Press, Cambridge

Garland, R, 1990, 'A Comparison of Three Forms of the Semantic Differential' *Marketing Bulletin* vol 1, pp 19-24, Article 4, Page 1 of 7 http://marketing-bulletin.massey.ac.nz

Green, A and Jay, D, 2005, 'Quality Assurance and Quality Control: Reviewing and pretesting examination material at Cambridge ESOL' *Research Notes* vol 21, pp 5-7

Haladyna, TM, 1999, *Developing and Validating Multiple Choice Test Items* Lawrence Erlbaum Associates, Mahwah, NJ

Haladyna, TM and Downing, SM, 1989a, 'A taxonomy of multiple-choice item-writing rules', *Applied Measurement in Education* vol 2, no 1, pp 37-50

Haladyna, TM and Downing, SM, 1989b, 'Validity of a taxonomy of multiple-choice item-writing rules' *Applied Measurement in Education* vol 2, no 1, pp 51-78

Haladyna, TM, Downing, SM, and Rodriguez, MC, 2002, 'A review of multiple-choice item-writing guidelines for classroom assessment', *Applied Measurement in Education* vol 15, no 3, pp 309-334

Hawkins, DL, Albaum G and Best R, 1974, 'Stapel Scale or semantic differential in marketing research?' *Journal of the Market Research Society* vol 11, pp 318-322.

Heaton, JB, 1990, *Classroom Testing*, Longman, London

Holmes, C, 1974, 'A statistical evaluation of rating scales', *Journal of the Market Research Society* vol 16, pp 87-107

Hughes, A, 1989, *Testing for Language Teachers*, Cambridge University Press, Cambridge

Hughes, GD, 1975, 'Upgrading the semantic differential', *Journal of the Market Research Society* vol 17, pp 41-44

IELTS, 2007, 'The IELTS Question Paper Production Process', retrieved 21 March 2008 from www.ielts.org.

Ingham, K, 2008, 'The Cambridge ESOL approach to Item Writer training: the case of ICFE Listening', *Research Notes* vol 32, pp 5-9.

Irvine, SH and Kyllonen, PC, 2002, *Item Generation for Test Development,* Lawrence Erlbaum Associates, New York

Khalifa, H and ffrench, A, 2008, 'Aligning Cambridge ESOL Examinations to the CEFR: Issues and Practice', paper presented at the *34th Annual Conference of the International Association for Educational Assessment*, Cambridge, UK, 7-12 September 2008.

Khalifa, H and Weir, C, 2009, *Examining Second Language Reading. Studies in Language Testing 29* Cambridge ESOL and Cambridge University Press, Cambridge.

Laufer, B, 1989, 'What percentage of text-lexis is essential for comprehension?', in *Special Language: From Humans Thinking to Thinking Machines*, eds Lauren, C and Nordman, M, Multilingual Matters, Clevedon

Macmillan, F, 2007, 'The role of lexical cohesion in the assessment of EFL reading proficiency'. *Arizona Working Papers in Second Language Acquisition and Teaching* vol 14, pp 75-94.

Menezies, D and Elbert, NF, 1979, 'Alternative semantic scaling formats for measuring store image: an evaluation', *Journal of Marketing Research* vol 16, pp 80-87

Mickan, P and Slater, S, 2000, 'Text analysis and the assessment of academic writing', in *IELTS Research Reports Volume 4,* ed Tulloh, R, IELTS Australia, Canberra:

Norris, JM, Brown, JD, Hudson, T and Yoshioka, J, 1998, *Designing Second Language Performance Assessments*, University of Hawaii, Honolulu, HI

O'Sullivan, B and Rignall, M, 2007, 'Assessing the value of bias analysis feedback to raters for the IELTS Writing Module', in *IELTS Collected Papers: Research in Speaking and Writing Assessment, Studies in Language Testing, Volume 19*, eds Falvey, P and Taylor, L, Cambridge University Press, Cambridge

Osgood EC, Suci, GJ and Tannenbaum, PH, 1957, *The Measurement of Meaning*, University of Illinois Press, Urbana

Patton, MQ, 1990, *Qualitative Evaluation and Research Methods* Sage, Newbury Park, CA

Rodriguez, MC, 2005, 'Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research', *Educational Measurement: Issues and Practice* vol 24, no 2, pp 3-13.

Salisbury, K, 2005, *The Edge of Expertise: Towards an understanding of listening test item writing as professional practice*, unpublished doctoral thesis, King's College London

Saville, N, 2003, 'The process of test development and revision within UCLES EFL', in *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913-2002*, Studies in Language Testing, vol 15*,* eds Weir, C.J. and M. Milanovic, UCLES/Cambridge University Press, Cambridge

Schertzer, CB and Keman, JS, 1985, 'More on the robustness of response scales', *Journal of the Market Research Society* vol 27, pp 261-282.

Valette, RM, 1967, *Modern Language Testing: A Handbook*, Harcourt, Brace and World, New York

Weir, CJ, 1993, *Understanding and Developing Language Tests*, Prentice Hall, Hemel Hempstead

Weir, CJ, 2005, *Language Testing and Validation: An evidence-based approach*, Palgrave Macmillan, New York

Weir, CJ, Hawkey, R, Green, A, Devi, S and Unaldi, A, 2009a, 'The relationship between the academic reading construct as measured by IELTS and the reading experiences of students in their first year of study at a British university', in *Research Reports Volume 9*, ed, P Thompson, British Council/IDP Australia, London

Weir, CJ, Hawkey, R, Green, A and Devi, S, 2009b, 'The cognitive processes underlying the academic reading construct as measured by IELTS', in *Research Reports Volume 9*, ed, P Thompson, British Council/IDP Australia, London

Wesman, AG, 1971, 'Writing the test item', in *Educational Measurement*, ed, Thorndike RL, American Council on Education, Washington, DC, pp 99-111

Wildt, AR and Mazis, W, 1978, Determinants of scale response: label versus position, *Journal of Marketing Research* vol 15, pp 261-267

APPENDIX A: COMMISSIONING LETTER (BASED ON A MODEL LETTER USED IN IELTS COMMISSIONING SUPPLIED BY CAMBRIDGE ESOL)

Date

Address

Dear XXX

IELTS Academic Reading Commission (Item Writer Research Study), September 2008

Thank you for agreeing to produce material for our Item Writer Research Study. I am now writing to confirm that we would like you to produce the following as indicated:

- One IELTS Academic Reading section with 16 or 17 items.
- The text should be of between 750 and 950 words in length.
- Suitable sources include magazines, newspapers, books, academic papers and journals.
- The text may be cut and edited as you see fit to make it more suitable for IELTS.
- You may use 2 or 3 different item types for your questions.
- The items should be arranged in sections according to type – eg, 6 multiple choice items followed by 6 matching questions followed by 5 short answer questions.
- Each item will carry 1 mark.
- The tasks may focus on understanding gist, main ideas/ themes, specific information, making inferences or recognizing opinions/ attitudes.
- The item types used should be based on the list and guidance provided by Cambridge ESOL at http://www.cambridgeesol.org/teach/ielts/academic_reading/index.htm

APPENDIX B: BACKGROUND QUESTIONNAIRES

	Victoria	*Mathilda*	*Mary*
Qualifications	BA English and Linguistics Postgraduate Diploma in Education/MA TESOL	BA Communication and English Language/MA Pragmatics	BA Politics Certificate in English Language Teaching to Adults (CELTA) Diploma in English Language Teaching to Adults (DELTA)
Experience in EL	18 years teaching, 5 years as an examiner, 2 years in publishing	As advanced learner	6 years teaching
Experience of IELTS	IELTS examiner (writing and speaking), IELTS preparation teacher	As test taker and preparation course student	2 years teaching IELTS preparation
Comments on IELTS	Through working as an examiner I have noticed that I sometimes get repeat candidates who score very well: band 8+ on the other three tests and then perform poorly in comparison on reading. Unfortunately for these candidates they have to score well on all four tests. One candidate I have been examining for the last two years. She is a pharmacist originally from Lebanon who in order to practise in the UK needs band 7 or more for all four tests. My current employer used to run its own internal IELTS test for placement purposes. The tests used were past papers. Here too candidates/ students consistently performed badly on reading in relation to the other three tests. Interestingly, native speakers are reputed to not score well on IELTS reading	My impression is that factual texts by far outweigh literary texts (if any). The latter might be more suitable for people intending to study/work with the literary or similar genre rather than in a technical field.	I find the reading test to be very dense which does help the student concentrate on the skills needed for reading, instead of just the reading. I think this is a really positive thing. However, I think some of these skills are quite specific and need to be taught (which could be seen as quite a negative thing. I think the True False and Not Given questions are not always presented well in practise materials. I find the reading topics to be questionable sometimes.

Table 10: Non-experienced item writers

	Jane	*Anne*	*William*
Qualifications	BA (Hons) RSA Diploma in Teaching English as a Foreign Language to Adults (Dip TEFLA)	MA Modern Languages RSA Dip. TEFLA	BA (Hons) in English Language and Literature MA in Modern English Language RSA Dip TEFLA
Experience in EL	12 years teaching EFL 12 years item writing for Cambridge ESOL papers 8 years chairing Cambridge ESOL papers Co-author of preparation book for another Cambridge ESOL exam	20 years teaching EFL In-house teacher training, course design, materials preparation and testing Item writing for other Cambridge ESOL examinations Examiner for IELTS speaking and writing tests Published author of IELTS and other exam preparation books	16 years teaching EFL Writing and marking written university exams, conducting oral university exams, Examiner for BEC Vantage and Higher Writing Translation from a number of languages into English
Experience of IELTS	7 commissions	Can't remember – 4 or 5 years	15+ commissions Published IELTS preparation book
Training as item writer	Editing meetings Formal training session at Cambridge ESOL		Training sessions plus informal training through participation in editing meetings
Background and training you feel have most helped you in your item writing for the academic reading test **(Please explain how this training has helped you)**	Teaching and experience on writing for other papers Approaches to finding a suitable text, and the editing of texts are similar across papers. Also, IELTS shares many task types with other papers and experience of having written these helps considerably.	Item writer training days and editing meetings plus pretest review meetings Seeing what other writers do eg, the texts they choose and the way they craft the items is one of the best ways of learning. Attending pretest review meetings shows you what kind of items work out too hard or too easy and what sort of texts prove too complex for candidates.	Work on BEC Higher reading (the first reading paper I was involved in) provided excellent groundwork and understanding of the demands of item-writing, eg giving awareness of potential pitfalls. My academic background has been helpful for selecting and understanding texts.

Elizabeth	*UK1*	*AUS1*	*NZ1*
BA//MA	Cert Ed B Ed Adv Dip Ed M Ed RSA Cert in TEFLA	BA (Hons) MA RSA Dip in TEFLA	MA, Dip. Tchg., Dip. TESL
34 years teaching EFL 23 years teacher training 30 years course design 32 years materials preparation 32 years testing/examining Published author of IELTS preparation books	13 years teaching EFL EAP course design 4 years as Cambridge ESOL subject Manager 14 years as freelance item Item writer for other IELTS papers and Cambridge ESOL tests Principal Examiner for IELTS Published author of IELTS and other exam preparation material	21 years teaching EFL Designed courses and materials for IELTS preparation 9 years as IELTS examiner	3 years teaching EFL and 7 years as Director of Studies at University English Language Centre 8 years as IELTS Examiner Trainer 13 years as IELTS Examiner
Item writer – 27+ commissions	So many I can't remember	I can't recall – 6 years	10 – 12 commissions
As provided by Cambridge ESOL	'I have trained many item writer teams for IELTS and wrote most versions of the item writer guidelines'	1 day training session Regular update sessions, and visits from Cambridge ESOL staff	One formal 2-day course with a Principal Examiner. A couple of informal sessions with team leader
Ten years' experience of teaching / materials/test writing and course design for courses for medical and science undergraduates overseas provided confidence and interest in science and technology and awareness of approaches to the testing of reading skills	Training I have received for other CESOL papers has helped me with IELTS; eg, production of MCQs and gapfill questions. It is useful to know what type of material can be targeted using these items, how to write good distractors, how to constrain keys, etc. I have used my experience and what I have gained from the literature on testing reading to build on my understanding of other task types that are IELTS-specific. This has helped me construct guidelines for writers.	I don't know how to answer this – all the training we have had has been helpful. The best thing is probably editing meetings.	Everything in my background has been useful.

	Jane	*Anne*	*William*
Do you feel that your item writing for the academic reading test has improved with experience?	I hope so. It is always helpful to see the kinds of texts other writers are using and how they approach the task types.	Definitely. It takes a while to learn how to write for different papers and also where to source appropriate texts from.	Yes – I think I work faster than I used to.
Has the proportion of successful reading commissions increased in that time?	This has remained fairly constant.	Yes	Most have been accepted
IELTS support materials (e.g item writer guidelines) you consult before or while you select texts and write test items:	When I am writing the items, I sometimes refer to the item writer guidelines Commissioning letter from Cambridge ESOL		Item writer guidelines Commissioning letter from Cambridge ESOL Item writer feedback forms from pre-editing review
What additional support materials might assist you with text selection, editing and item writing?	I'm not sure. There is considerable support in the item writer guidelines – especially for writing the items Text selection is probably one of the hardest aspects and more guidance with this is always welcome.	Samples of good practice ie, texts and tasks that have worked well	None
Do you see your IELTS academic reading material in the meetings below? If so, how much of your material do you see at each? (eg, If you see all your material edited at editing meetings, you should put 100%)			
pre-editing meetings:	Writers don't always attend these		No
editing meetings:	I have attended one pre-editing meeting since I started	100%	No
pretest review meetings:	Every commission	It depends – sometimes quite a lot. On other occasions none of your material may come up at a pretest review meeting	Yes – generally 100%, occasionally 66% or 75%
How would you describe the experience of attending these IELTS meetings?	I have attended a few PTR meetings – perhaps once a year	Helpful	It depends very much on the other people involved. IELTS academic reading is positive.

Elizabeth	*UK1*	*AUS1*	*NZ1*
Not necessarily just due to experience – guidelines have become more precise	Yes – writing well for any paper is an ongoing, developmental experience.	Sometimes I wonder – I still have end up with a lot of work to do on my items.	Yes, definitely.
Most have been accepted	Yes.	Most have been accepted	Yes, somewhat, though has always been fairly high.
Item writer guidelines I might use Google to check up on details of input text content where this seems unclear I may also use a Thesaurus to provide synonyms eg, in multiple choice questions.	I always refer to the guidelines when I am writing. The Specimen materials are dated so I don't use these.	What materials are there apart from the Guidelines?	Item writer guidelines
Access to appropriate published materials on the internet eg, if I had access to undergraduate/ postgraduate sites	It might be useful to have more good 'models', particularly for some of the trickier task types. A better insight into how candidates take the test might be helpful.		
Item writers do not normally see their own material at pre-editing for IELTS or other Cambridge ESOL examinations	None.	nil	100%
usually 100%	Most, say 85%.	100%	100%
Variable	Ditto.	100%	NO
Productive	I am usually chairing these meetings. From that perspective, they run smoothly, writers are very experienced and make very useful contributions.	A tremendous struggle, but great. We have a really terrific team.	Very positive and useful.

	Jane	*Anne*	*William*
What do you think are the key characteristics of more/less successful IELTS academic reading items?	PTR meetings are always informative. As a writer, it helps to see which items candidates have found too difficult and to think about the reasons for this.	Successful items require candidates to process the text carefully Less successful items are - too easy (eg, numbers that stand out - answerable without understanding surrounding text) - convoluted - based on complex extracts which candidates cannot process	More successful: - clear focus on a specific piece of text (fact, opinion, etc), clearly distinguishable from text tested by surrounding items - unambiguous phrasing
What do you think are the characteristics of more/less successful IELTS academic reading item writers?	A range of things including: - giving candidates support for where to look in the text - clarity in phrasing of items//awareness of language level of candidates	Successful item writers find appropriate texts which suit the item types and which do not need to be altered very much	More successful: - eye for detail - ability to identify ambiguity - being pedantic - being intelligent enough to understand not only the text but interrelations between items, and options in different items - having a wide enough general knowledge to understand a range of texts - understanding how different types of items work and how they can go wrong
What do you enjoy/ dislike about IELTS academic reading item writing work?	Experience of writing other test material and attention to detail.	I like writing for IELTS because you can use texts which are challenging and interesting. I often find that I learn a lot while I am working with the texts.	Enjoy: selecting texts on a variety of subjects, intellectual satisfaction of adapting text to suit the constraints and creating successful items Dislike: what can sometimes seem like oversensitivity to candidates' sensibilities
Which aspects do you find easiest/ most challenging?	The source material that I read through to find suitable texts is different from other papers that I work on and is sometimes interesting.		Easiest: finding interesting texts Most challenging: making text fit certain task types, eg, flow-chart, 5-option multiple choice that isn't too easy

Elizabeth	*UK1*	*AUS1*	*NZ1*
Successful items focus on key points of the text and involve the strategies that would be used by a good reader of that text.	Clear keys; appropriate level; well-sequenced (if appropriate); targeting salient information, well cued without giving the answer away; written according to item specific guidelines; independent; well written surrounding text (when appropriate).	Successful items are ones that don't need to have a lot of work done on them after they have been submitted.	-
Ability to identify and evaluate texts from a variety of academic subject areas and to exploit texts using the specified item types.	More successful writers follow the guidelines, ensure that their texts will yield sufficient items before they begin, have time to do the job, pay meticulous attention to detail, repeatedly proof their work, are good writers of English themselves, listen well at meetings and have clear heads Less successful writers do not satisfy some of the criteria above and may have their own agenda about what IELTS should test.	Having the knack of choosing the right parts of the text to test; being helpful to other writers; knowing a good text when they see one	
I enjoy most aspects of this work, particularly the variety of content.	The texts are generally more interesting than other EL texts, which makes it more enjoyable looking for them. The task types are more varied and there is some choice; writing questions is less mechanical than it is for other papers. There isn't much I dislike about this paper!	I enjoy editing texts; attending meetings; finding a good fit between text and items. I dislike the fact that I am not as good at writing items as I would like to be.	I enjoy the challenge of finding a suitable text, editing or adapting it, writing the best items possible, and working with a small and committed team of like-minded professionals who are very good at what they do.
Finding the texts is most challenging. The rest is relatively straightforward though teamwork is necessary to produce good working items.	Some task types are more difficult to write than others; namely: summary completion, paragraph headings and sentence completion with a box. Other task types are more straightforward to write.	Finding and editing texts is the easy part.	After the creative part is over it always feels tedious when the feedback comes back after a considerable time lapse, to have to re-visit the commission and input all the recommended changes.

	Jane	*Anne*	*William*
In what respects do you think the IELTS Academic Reading test reflects (or fails to reflect) the reading skills needed by students entering English-medium higher education?	Finding suitable texts is usually easier than for general English papers.	I think it does reflect the skills required quite well – above all the ability to read a quite complex text at quite a fast pace and get the main points from it without being thrown by items of vocabulary that are completely unfamiliar.	Reflects: requirement for speed; distinguishing between main points and detail; interpreting texts that include unfamiliar vocabulary; need for a broad vocabulary and understanding of a wide range of structures; need for intellectual curiosity
What changes, if any, would you like to see in the IELTS Academic Reading test and why?	As far as I know, the test reflects the reading skills needed by students reasonably well. //I think the test is different, however, from independent reading. The texts are short and most of the less significant information has already been cut by the item writer. Also, many of the items, for example note-taking or tables, provide the reader with a framework.		None
Other comments on IELTS academic reading module item writing.		It's certainly a challenge at times but enjoyable on the whole.	None

Elizabeth	UK1	AUS1	NZ1
My students found preparation for IELTS useful in developing awareness of text type and structure as well as academic vocabulary and reading skills such as understanding main idea and detail. The relatively heavy reading load also encouraged them to increase their reading speed. The exam does not test reference skills (eg, use of contents/index or reading at chapter/whole book level) or text evaluation skills. Appropriate humanities texts (eg, English literature) are hard to find due to cultural and stylistic reasons. When looking for a well-written text at the target level, it is much easier to find texts on science and technology than on the humanities.	IELTS differs from other reading papers in that there is a wider range of task types to reflect the range of skills that students might need at university; eg, reading for detail, main idea, gist, etc. It is possible to match certain tasks to skills and to feel assured that the test covers the main ones. The texts are selected for inclusion according to the density of ideas, vocabulary level and text type (descriptive, argument-based, etc.) There may be other requisite reading skills for academic study that are subject specific or task specific but these may not be 'assessable' in a global language test of this kind. (Candidates may attempt every question using strategies that evade 'real' reading but this is not to say that they will get the answers correct and/or get the band score they are capable of getting had they used more appropriate skills.)	I think the texts are often far away from the level of difficulty encountered in real academic texts. Things (eg lexical items) which are in the least challenging are glossed or removed. Many of the students who enter our higher education institutions should not be doing so anyway – IELTS 6 is in no way adequate to allow someone to undertake an English-medium degree.	I think it's a good reflection.
It would be interesting to look at the possibility of varying text length – eg, one much shorter text with items focusing on detail and one much longer one, to test outlining / summary skills. However, given the size and nature of the candidature, any changes of this nature would need to be very well researched before implementation.	Perhaps a greater variety of text types and fewer potential short cuts for candidates. (The latter is really an editing issue).		-
	I have faith in the ACR paper and believe that candidates get a good deal. It is rigorously produced and meets many of the pedagogical/ theoretical requirements of an EAP reading test.		-

TABLE 11: EXPERIENCED ITEM WRITERS

APPENDIX C: ITEM WRITER SUBMISSIONS

Non-experienced group: Victoria

Text

How the Brain Turns Reality into Dreams

Dreams seem to make perfect sense as we are having them. Yet, on awakening they typically leave us feeling befuddled; without any clear idea about their origins. Research, however, investigating the dreams of individuals with amnesia may provide some answers.

Such research suggests that much of the fodder for our dreams comes from recent experiences. For this reason, scientists have tentatively supposed that the dreaming brain draws from its "declarative memory" system. This system stores newly learned information.

The declarative memory stores the type of information that can be "declared" to be known by subjects; the name of one's dog, for example. Often, subjects can even remember when or where they learned something – for example, the day you discovered the harsh truth about Santa Claus. To differentiate this information from declarative memory this particular of recollection is referred to by scientists as episodic memory.

It seems subjects who permanently suffer from amnesia or loss of memory are unable to add new declarative or episodic memories. The part of the brain involved in storing this type of information, a region called the hippocampus, has been damaged. Although, subjects who suffer from memory loss are able to retain new information temporarily, they are unable to permanently retain it. Studies have shown that new information for such individuals is lost sometimes within minutes.

If such is the case, that dreams originate from declarative memories, then individuals suffering with amnesia should not be capable of dreaming at all. Current research directed by Robert Stickgold of Harvard Medical School, however, suggests quite the opposite.

Stickgold's study shows that, much like individuals with normal memory, amnesiacs also replay recent experiences whilst asleep. The only difference seems to be that the amnesiacs are unable to recognize what they are dreaming about.

The methodology designed for Stickgold's study had two groups of subjects playing several hours of the computer game Tetris, which requires users to direct falling blocks into the correct positions as they reach the bottom of the screen. At night, the amnesiac group did not remember playing the game but, they did describe seeing falling, rotating blocks while they were falling asleep.

A second group of players with normal memories reported seeing the same images.

From these results, Stickgold's research team felt reassured in making the claim that dreams come from the types of memory amnesiacs do have, defined as implicit memories. Such memories can be measured even when individuals have no conscious awareness of them. One class of implicit memories is found in the procedural memory system. This particular system stores information that is used, but is somehow beyond the individuals' ability to state how they know, to perform actions. A pertinent example being, when one rides a bicycle for the first time in years, a reliance on procedural memory has come into play.

Another type of implicit memory uses semantic knowledge, and resides in a region of the brain called the neocortex. One aspect of semantic knowledge involves general, abstract concepts. Both groups of Tetris players, for example, only described seeing blocks, falling and rotating, and evidently did not see a desk, room, or computer screen, or feel their fingers on the keyboard.

Without help from the hippocampus, new semantic memories are too weak to be intentionally recalled. But they still have the potential to affect behaviour. In contrast, the information in episodic memories is associated with specific times, places or events thus providing "anchors" to reality. In contrast, implicit memories based on semantic knowledge do not possess such grounding and it is for this reason the study's authors say that dreams are so illogical and full of discontinuity.

We have to enquire as to the benefit to the individual of being able to dream. Stickgold upholds that dreams serve a purpose for the brain, allowing it to make necessary emotional connections among new pieces of information.

> *Dreams let us consolidate and integrate … experiences, without conflict with other input from real life. Dreaming is like saying, 'I'm going home, disconnecting the phone, and nobody talk to me. I have to do work Stickgold.*

Because the hippocampus seems to be inaccessible for this "off-line" memory processing, the brain may use the abstract information in the neocortex instead.

According to Stickgold's theory, dreaming is like choosing an outfit by reaching into bins labelled 'shirts, trousers' and so on. You'll happen upon something to wear, but it won't be a perfectly matching ensemble.

The period of sleep that Stickgold's team studied is called hypnagogia: an in-between state between being fully awake and fully asleep. Many people who have just had an intense new experience of some kind, either mental or physical, often report replays of that experience during this stage.

In his poem, 'After Apple Picking', for example, Robert Frost describes seeing the apples and apple blossoms, and feeling the ladder sway as he nods off to sleep. Stickgold's first encounter with this phenomenon occurred after a day of mountain climbing, when he felt the sensation of rocks under his fingertips as he fell asleep.

Hypnagogic sleep is different from REM sleep, the period marked by rapid eye movement, when standard dreams most often occur. According to Stickgold, other studies suggest that the hippocampus is not active during REM sleep either. Therefore, he proposes, the brain activity responsible for the Tetris images is probably similar to the dreaming that occurs in REM sleep.

Interpreting REM sleep dreams, however, is a highly subjective process. Stickgold states, 'what is so nice about the images in our experiments is that they are so accurately re-creating the Tetris experience- no interpretation is necessary.'

Non-experienced group: Victoria

Items

Type 8

Do the following statements agree with the information given in the Reading Passage?

True	if the statement agrees with the information
False	if the statement contradicts the information
Not Given	if there is no information on this

1 People tend to make the most sense of their dreams while waking up.

2 Dream research may one day help people with memory loss.

3 The content informing dreams comes from experiences had by an individual in the last few days.

4 Permanent memory loss is associated with the declarative memory storing section of the brain.

5 One way in which to restore memory of amnesiacs is to repair the hippocampus.

6 Both amnesia sufferers and people with normal memory function go over recent activities during sleep.

Type 1

Multiple Choice

Choose the correct letter, A, B, C or D.

7 Declarative memory can be best understood as

A memory that is similar to episodic memory.

B memory that entails episodic memory.

C memory that is distinct from episodic memory.

D memory that is identical to episodic memory.

8 The research team used the following evidence to state that dreams come from implicit memories because

A the normal group could both recall playing Tetris and dreaming about it.

B the amnesiac group could both recall playing Tetris and dreaming about it.

C the normal group could not recall playing Tetris but could describe having dreamt about tumbling shapes.

D the amnesiac group could not recall playing Tetris but could describe having dreamt about tumbling shapes.

9 Implicit memory differs from episodic memory in that

A it does not relate to definite events.

B it only applies to non specific events.

C it only applies to semantic memories.

D it is completely distinct from episodic memory.

Task Type 2 – Short-answer Questions

Choose **NO MORE THAN THREE WORDS** from the passage for each answer.

10 The writer describes several different types of memory. The type of information stored by declarative memory *can be known/ declared.*

11 The writer describe the kind of condition during the hypnagogia stage of sleep *an in - between state*.

12 The writer uses the Frost poem as an example of *intense physical experience*.

13 The research benefit the Tetris experience provides is that it *makes interpretation unnecessary/ redundant*.

Words in italics are the answers

Non-experienced group: Mathilda

Text

How-and Where-Will We Live in 2015?

The future is now for sustainable cities in the U.K., China, and U.A.E.

Future City Madsar

No cars in the land of oil.

In an ironic twist, the first city to fully turn its back on petroleum is likely to spring up in the United Arab Emirates, the oil-producing giant in the Middle East. Madsar, a carbon-neutral, zero-waste, walled metropolis now under construction adjacent to the Abu Dhabi airport, will have many innovative green technologies, but it may be most noteworthy for one thing it won't have: gas-consuming cars.

Nearly all of the world's motor vehicles run on petroleum, and the environmental consequences are obvious. For example, 28 percent of carbon emissions in the United States result from the burning of 14 million barrels of oil a day for transportation, primarily in cars and small trucks. Madsar will do away with this problem. Urbanites will walk along shaded sidewalks, and if the sweltering desert heat gets to them, they will never be more than 500 feet from a public transportation network that puts traditional buses and subways to shame. Small electric vehicles, guided in part by magnets embedded in the road, will act as driverless taxicabs serving 83 stations situated throughout the roughly 2.5-square-mile city. Meanwhile, two electric rail systems will connect Madsar to the outside, carbon-polluting world.

The Madsar project was announced in 2006, and development is already in full swing; much of the financing is coming from the emirate of Abu Dhabi, which committed $15 billion. Developers have set a goal of sustaining 55,000 residents and visitors by 2013, with the first section of the city scheduled to open next year.

Future City London

An old industrial site gets a green makeover

A. In 2006 London produced eight percent of the United Kingdom's 560.6 million tons of carbon emissions, 70 percent of it from residential sources. In response, the city has developed an ambitious long-term plan known as the Mayor's Energy Strategy, which calls for, among other things, the establishment of one zero-carbon community in each of the city's 32 boroughs by 2010.

B. A prototype is planned for a three-acre area on the Royal Albert Dock. Called Gallions Park, it will be a sustainable community with at least 200 residential units. What makes this site important for other cities attempting to shrink their carbon footprint is that the dock area is land that was previously used by industry. Many upcoming eco-cities are being built on virgin land; success at Gallions Park would open other abandoned industrial sites to similar development possibilities.

C. While the Gallions Park development includes several earth-friendly features, such as community greenhouses, a key element of the zero-carbon strategy will be a combined heat and power (CHP) plant to generate electricity and provide hot water. The CHP plant will use biomass, such as wood, for fuel. The community's buildings will also create renewable energy through roof-mounted wind turbines and solar technology that converts light into electricity.

D. A budget has not yet been released by the developer, but the planning application for Gallions Park was filed in July, and construction is expected to begin by early 2009.

Future City Dongtan

China watches over every drop of water

E. On a small, thinly populated island about 14 miles off the coast of Shanghai, a city is rising that could spell salvation for the 1 billion people expected to live in China's urban areas by 2045. Like several other planned cities, Dongtan will showcase an array of eco-friendly technologies such as wind power and zero-emission vehicles, but its most important innovation may be that it is designed to consume 43 percent less water than a conventional city. If Dongtan succeeds, many of its technologies will be employed in other cities in China.

F. Access to water has become a critical issue for much of the world. The United Nations estimates that by 2025, 1.8 billion people will live in regions where drinking water is scarce. The problem is particularly urgent in China, where major rivers (including the Yangtze, which feeds into Shanghai) are heavily polluted. Dongtan aims to reduce its water needs by using technologies such as green roofs-building tops covered with plants-to capture and filter rainwater and by recycling sewage and other waste to fertilize and water nearby farms.

Although Dongtan is in the earliest stages of construction, Arup, the U.K. design and engineering firm hired by the Chinese government to oversee its development, says that as many as 5,000 people will be living there by 2010. There have been delays and setbacks-originally Arup anticipated up to 10,000 settlers by 2010-but the firm says the city is still on track to have as many as 500,000 residents by 2050.

Non-experienced group: Mathilda

Items

Task Type 1 – Multiple Choice

1. *Choose the correct letter, **A**, **B**, **C** or **D**.*

What is going to be the most special feature of the future city Madsar?

A It is going to play a major role in oil production.

B There will be no cars that run on petrol.

C The city will pioneer in carbon waste recycling.

D There will be no airport in the city's vicinity.

2. *Choose the correct letter, **A**, **B**, **C** or **D**.*

Madsar will do away with the problem of

A overcrowding on traditional buses and subways.

B oil consumption of cars and small trucks in the United States.

C vehicles causing environmental damage due to carbon emissions.

D people walking along the sidewalks.

3. *Choose the correct letter, **A**, **B**, **C** or **D**.*

Which country is contributing considerably to the financing of the Madsar project?

A Emirate of Abu Dhabi.

B China.

C USA.

D United Kingdom.

4. *Choose the correct letter, **A**, **B**, **C** or **D**.*

What makes Gallions Park a particularly important example for other environmental projects?

A It will have residential units.

B It is a three-acre area.

C It is not clear yet what the budget for the project is going to be.

D It was previously used by industry and is not built on virgin land.

5. *Choose the correct letter, **A**, **B**, **C** or **D**.*

The CHP plant will generate electricity and hot water by using

A wind turbines.

B solar technology.

C biomass.

D fossil fuel.

6. *Choose the correct letter, **A**, **B**, **C** or **D**.*

Who has the job of overseeing the development of the city Dongtan?

A The Chinese government.

B The future residents of the city.

C The United Nations.

D A design and engineering firm from the UK.

Answers:

1. **B**
2. **C**
3. **A**
4. **D**
5. **C**
6. **D**

Task Type 7 – Locating Information

Questions 7-11

The reading passage has 6 paragraphs, A-F

Which paragraph contains the following information?

Write the correct letter A-F in boxes 7-11 on your answer sheet.

NB You may use any letter more than once.

7 Information about the planning stages of the London project

8 A description of the technologies employed in environmentally-friendly developments in order to minimise water use

9 An example of an area that is being developed into a zero-carbon community

10 Mention of a lack of drinking water caused by polluted rivers

11 The introduction of a future city that will use less water than a conventional city

Answers:

7. D

8. F

9. B

10. F

11. E

Task Type 8 – Identification of Information in a Text

Questions 12-17

Do the following statements agree with the information given in Reading Passage 1?

In boxes 12-17 on your answer sheet write

TRUE if the statement agrees with the information

FALSE if the statement contradicts the information

NOT GIVEN if there is no information on this.

12 The city of Madsar is being built next to the Abu Dhabi airport.

13 More than a quarter of carbon emissions in the USA result from burning oil for transportation.

14 The first section of the city of Madsar is going to open in 2013.

15 In London, a large part of carbon emissions is caused by motor traffic.

16 The long term plan for London is to develop 32 zero-carbon communities by 2010.

17 A green roof is a building top covered with plants used to fertilize crops.

Answers:

12. TRUE

13. TRUE

14. FALSE

15. NOT GIVEN

16. TRUE

17. FALSE

Non-experienced group: Mary

Text

The rise of the emotional robot

- 05 April 2008
- From *New Scientist* Print Edition.
- Paul Marks

Duke is moving noisily across a living room floor wearing the dark blue and white colours of Duke University in Durham, North Carolina. He's no student but a disc-shaped robotic vacuum cleaner called the Roomba. Not only have his owners dressed him up, they have also given him a name and gender. Duke is not alone. Such behaviour is common, and takes a multitude of forms according to a survey of almost 400 Roomba owners, conducted by Ja-Young Sung and Rebecca Grinter. "Dressing up Roomba happens in many ways," Sung says "and people also often gave their robots a name and gender". Kathy Morgan, an engineer based in Atlanta, said that her robot wore a sticker saying "Our Baby", indicating that she viewed it almost as part of the family.

Until recently, robots have been designed for what the robotics industry dubs "dull, dirty and dangerous" jobs, like welding cars, defusing bombs or mowing lawns. Even the name robot comes from robota, the Czech word for drudgery. But Sung's observations suggest that we have moved on. "I have not seen a single family who treats Roomba like a machine if they clothe it," she says. "With skins or costumes on, people tend to treat Roomba with more respect." Sung believes that the notion of humans relating to their robots almost as if they were family members or friends is more than just a curiosity. "People want their Roomba to look unique because it has evolved into something that's much more than a gadget," she says.

These changing relationships with robots are something which is particularly in the minds of roboticists at present. Figuring out just how far humans are willing to go in shifting the boundaries towards accepting robots as partners rather than mere machines will help designers decide what tasks and functions are appropriate for robots. Meanwhile, working out whether it's the robot or the person who determines the boundary shift might mean designers can deliberately create robots that elicit more feeling from humans. "Engineers will need to identify the positive robot design factors that yield good emotions and not bad ones – and try to design robots that promote them," says Sung.

To work out which kinds of robots are more likely to coax social responses from humans, researchers led by Frank Heger at Bielefeld University are scanning the brains of people as they interact with robots. The team starts by getting humans to "meet" four different "opponents": a computer program running on a laptop, a pair of robotic arms that tap the keys of a laptop, a robot with a human-shaped body and rubbery human-like head, which also taps at a laptop, and a human. Then the volunteers put on video goggles and enter an MRI machine. While inside the machine, a picture of the opponent they must play against flashes up inside their goggles. The volunteers then must choose between cooperating with their opponent or betraying them. As they can't tell what their opponent will do, it requires them to predict what their opponent is thinking. The volunteers then indicate their choice from inside the scanner.

Heger's team have carried out the experiment on 32 volunteers, who each played all four opponents. Then they compared the brain scans for each opponent, paying particular attention to the parts of the brain associated with assessing someone else's mental state. This ability is considered a vital part of

successful social interactions. Unsurprisingly, the team found that this part of the volunteers' brains were active to some extent when playing all opponents. However, it was more active the more human-like their opponent was, with the human triggering the most activity in this region, followed by the robot with the human-like body and head. Heger says that this shows that the way a robot looks affects the sophistication of an interaction.

Not surprisingly, though there are similarities between the way people view robots and other human beings, there are also differences. Daniel Levin and colleagues at Vanderbilt University showed people videos of robots in action and then interviewed them. He says that people are unwilling to attribute intentions to robots, no matter how sophisticated they appear to be. Further complicating the matter, researchers have also shown that the degree to which someone socialises with and trusts a robot depends on their gender and nationality.

These uncertainties haven't stopped some researchers from forming strong opinions. Herbert Clark, a psychologist at Stanford University in California, is sceptical about humans ever having sophisticated relationships with robots. "Roboticists should admit that robots will never approach human-like interaction levels – and the sooner they do the sooner we'll get a realistic idea of what people can expect from robots." He says that robots' lack of desire and free will is always going to limit the way humans view them.

Whether robots can really form relationships with humans and what these can be is much disputed. Only time will really tell. However, despite the negative criticism there is one scientist with strong evidence for his view. Scientist Hiroshi Ishiguro has used a robotic twin of himself to hold classes at his university while he controls it remotely. He says that people's reactions to his computerised double suggest that they are engaging with the robot emotionally. "People treat my copy completely naturally and say hello to it as they walk past," he says. "Robots can be people's partners and they will be."

Non-experienced group: Mary

Items

Questions 1-5: Matching

Look at the following list of statements based on research into robots and their emotional relationships to humans.

Match each statement (1-5) with the correct person A-E.

1. People are reluctant to think that robots can have intentions.

2. People's opinion of robots will always be limited.

3. People have moved on from thinking of robots as only useful for risky and repetitive tasks.

4. People react more to robots that are physically similar to humans.

5. People can and will interact with humans in a completely natural way.

Researchers:

A: Herbert Clarke

B: Hiroshi Ishiguro

C: Ja-Young Sung

D: Daniel Levin

E: Frank Heger

Questions 6-13: Summary Completion

Complete the summary below.

Choose NO MORE THAN TWO WORDS from the passage for each answer.

A recent study by Frank Heger concluded that how a robot 6. _____ is essential in determining the 7. _____ it can have with humans. In his experiment, volunteers had to assess four opponents ranging from very machine-like equipment to real 8. _____. Volunteers were put in a MRI scanner, which measured the activity of their 9. _____, wearing 10. _____ showing images of their opponents. They then had to decide whether to work with or 11. _____ the

opponent. Their scans showed that their assessment of their opponents' 12. _____ was more active when their opponent appeared more 13. _____.

Questions 14-17: Multiple Choice

Choose the correct letter (A,B,C or D)

14. People dress up their Roombas and give them names because...

A... they want their robots to be different from all the others.

B... they love their robots.

C.... they give their robots names and genders.

D... their robots are respected.

15. Daniel Levin indicated that levels of social interaction with robots can depend on...

A... the age of the human.

B.... their intentions.

C...the sex and nationality of the human.

D... the way they view humans.

16. Roomba is...

A... a baby

B... a cleaning device.

C... a robot in human form.

D... a student.

17. Experts' views on the extent that robots can be humanised...

A... are varied.

B... are critical.

C... are positive.

D... are consistent.

Answer Key

Questions 1-5: Matching

1. People are reluctant to think that robots can have intentions. **D Daniel Levin**
2. People's opinion of robots will always be limited. **A Herbert Clarke**
3. People have moved on from thinking of robots as only useful for risky and repetitive tasks. **C Ja Young Sung**
4. People react more to robots that are physically similar to humans. **E Frank Heger**
5. People can and will interact with humans in a completely natural way. **B Hiroshi Ishiguro**

Questions 6-13: Summary Completion

A recent study by Frank Heger concluded that how a robot **6. looks** is essential in determining the **7. (social) interaction** it can have with humans. In his experiment, volunteers had to assess four opponents ranging from very machine-like equipment to real **8. humans**. Volunteers were put in a MRI scanner, which measured the activity of their **9. brains**, wearing **10. video goggles** showing images of their opponents. They then had to decide whether to work with or **11. betray** the opponent. Their scans showed that their assessment of **12. mental state** was more active when their opponent appeared more **13. human-like**.

Questions 14-17: Multiple Choice

Choose the correct letter (A,B,C or D)

14. People dress up their Roombas and give them names because...

    A... they want their robots to be different from all the others. **Correct**

15. Daniel Levin indicated that levels of social interaction with robots can depend on...

    C...the sex and nationality of the human. **Correct**

16. Roomba is...

    B... a cleaning device **Correct**

16. Experts views on the extent that robots can be humanised...

    A... are varied. **Correct**

Experienced group: Jane

Text

Wildlife-spotting robots

Conservationists are using robotic cameras to help search for rare wildlife. Biology fieldwork is very labour intensive, so there is increasing use of technology to collect data in an unobtrusive way. Autonomous sound recording units and video imaging equipment can be programmed to collect data at scheduled times in remote areas, and it is often possible, via long-range wireless communications, to access the data from a distant location. However, the chances of recording an occurrence of the target species are very low, so it is important that the equipment should be selective in what it records, or have the ability to sift the data and pick out likely candidates for expert examination.

Some wildlife cameras are controlled by passive infrared motion sensors which detect a nearby animal, point the camera towards it and trigger an image-capture sequence. Some are directly controlled by remote observers, in a similar way to web cameras or surveillance systems. A project, led by University of California, Berkeley and Texas A & M aims to integrate signals from both sensors and humans in the control of a hybrid teleoperated/ autonomous robotic device called the collaborative observatory for natural environments (CONE). The developers are building an "observatory" that will enable scientists to study animals in their natural habitat via the Internet. A purely automatic version of this, the ACONE, is aiding researchers at Cornell University in their systematic search for the North American ivory-billed woodpecker.

The ivory-billed woodpecker is a large, vividly coloured bird that was widely thought to be extinct. There has been no officially confirmed sighting of it since 1940, but a spate of recently reported glimpses inspired a determined search effort by the United States Fish and Wildlife Service and Cornell Lab of Ornithology (CLO). The project started in 2005 with over 20 trained field biologists taking part in the search team, and volunteers also being recruited. They had to become familiar with the use of GPS, digital video cameras and microphones, and cell phone technologies. The search also made use of time-lapse video recording cameras for monitoring likely sites. This method was ideal since it did not lead to the disturbance that is unavoidable with human observers. They played back a 1935 recording of the ivory-billed woodpecker to try to attract the bird. They also used ultra-light aircraft to conduct aerial surveys. Autonomous sound-recording units recorded the ambient sounds at selected sites, and sampled at 20 kHz for up to 4h per day, and the results were analysed at CLO to pick out any new recordings that were a match to the known vocalisations of the ivory-billed woodpecker. Despite the 6,347 field-hours of the Cornell-coordinated search, no definitive sighting was made in the 2005-2006 season.

In February 2007, the University of California, Berkeley announced the installation of a high-resolution intelligent robotic video system developed in collaboration with Texas A&M University. Mounted on an electrical transformer and positioned to view the birds flying through the narrow corridor surrounding an existing power line, two cameras collect video images, and software examines them in real time, discarding any images that have no relevance. The software looks for a large bird flying at 20 to 40 mph. The images saved are then examined by human experts.

The camera lens has a focal range of 10 m to 40 mm, giving a horizontal field of view varying from 32° to 8°. The cameras are positioned so that they look along the corridor in opposite directions. They are mounted 3 metres the water in the marsh, and angled upwards to observe birds flying between the

tree-tops through the 50 metre corridor. With a 20° horizontal field of view, each camera observes the full width of the corridor at a point 366 metres away, and can switch to a wider field of view as the bird approaches. Each camera captures 20 frames per second, and the software continually examines the images, carrying out statistical background subtraction to allow for changing weather conditions, and looking for large connected foreground components on every seventh image. In this way, the system has already collected clear images of geese. In its first 120 days, operating for 8 hours per day, it captured a total of 76 million images, but the software selected only 5,825 of these for storage. This autonomous system has already demonstrated its success in capturing images of large birds, and its powerful selectivity avoids wasting experts' time.

A more general project to develop "Collaborative Observatories for Natural Environments" (CONE) is underway at the Universities of Texas A & M and UC Berkeley, with input from natural scientists and documentary filmmakers, funded by the National Science Foundation. The observatories are located in the field and are run on solar energy. They incorporate sensors and actuators and software to carry out a periodic systematic observation of the surrounding scene, or automatically point the camera in the direction of a sensed event, and also respond to the signals of biologists who are making observations and inputting instructions from their laboratories. This combination of teleoperation and autonomy builds upon years of work by Professor Ken Goldberg, an engineering professor, and Dr Dehzen Song, assistant professor of computer science. Song's work involved the development of a system that allows multiple users to share simultaneous live control of a camera via the internet. The system collects the requests from users and calculates, for example, which zoom setting would best satisfy them all. This advance in intelligent software has helped CONE biologists to improve the selectivity of images captured and stored, and improved the responsiveness of remote systems to their live imaging needs.

Experienced group: Jane

Items

Questions 1-5

Do the following statements agree with the information in Reading Passage I?

In boxes 1-5 on your answer sheet, write

TRUE	*if the statement agrees with the information*
FALSE	*if the statement contradicts the information*
NOT GIVEN	*if there is no information on this*

1 The equipment in wildlife experiments is programmed to obtain data at random intervals.

2 There is little chance that the equipment used will capture the species being investigated.

3 There are insufficient experts to examine the different species that are recorded in field studies.

4 The organisations that produce surveillance cameras also make cameras for wildlife filming.

5 The movement of animals through an area will activate some cameras.

Questions 6-12

Complete the notes below.

*Choose **NO MORE THAN ONE WORD AND / OR A NUMBER** from the passage for each answer.*

*Write your answers in boxes **6-12** on your answer sheet.*

ACONE study of the ivory-billed woodpecker

<u>background to the study</u>

- bird last seen in: **6**……..
- most people believed the bird to be: **7**…….

<u>2005 -2006 season</u>

- camera study preferable since observation by people involves: **8**……
- old recordings played to attract the bird

- results analysed to identify sounds that were a: **9**…….
- no definitive sighting
- cameras positioned to view birds flying through narrow corridor
- images analysed to assess their : **10**…….
- cameras fixed at a height of: **11**……
- cameras already taken pictures of: **12**……...

Questions 13-16

*Answer the questions below. Choose **NO MORE THAN TWO WORDS AND/OR A NUMBER** from the passage below for each answer.*

Write your answer in boxes 13-16 on your answer sheet.

13 Which professional group is working with scientists on the CONE study?

14 What are the CONE observatories powered by?

15 Where are biologists in the CONE study monitoring events from?

16 Which camera feature can be controlled by the software developed by Dehzen Song ?

Key

1 F

2 T

3 NG

4 NG

5 T

6 1940

7 extinct

8 disruption

9 match

10 relevance

11 3 metres

12 geese

13 (documentary) filmmakers

14 solar energy

15 (their) laboratories

16 zoom (setting)

Experienced group: Anne

Text

What makes us laugh?

If we ask ourselves what triggers a good laugh, the obvious answer is that it is a response to something we find funny. But one scientist, Robert Provine, who has spent nearly two decades studying laughter, says that humour has surprisingly little to do with it. Provine, a neuroscientist at the University of Maryland in the US and author of Laughter: A Scientific Investigation, realised early on in his research that you cannot capture real-life laughter in the laboratory because as soon as you place it under scrutiny, it vanishes. So, instead, he gathered his data by standing around in public places, eavesdropping on people's conversations, secretly noting when they laughed.

Over the course of a year he collected 1200 laugh episodes – an episode being defined as the comment immediately preceding the laughter and the laughter itself – which he sorted by speaker (the person talking), audience (the person being addressed), gender and pre-laugh comment. His analysis of this data revealed three important facts about laughter. Firstly, that it is all about relationships. Secondly, that it occurs during phrase breaks in speech. And thirdly, that it is not consciously controlled. 'It's a message we send to other people – it practically disappears when we are by ourselves,' he says. Perhaps most surprising of all is Provine's finding that only 15-20 per cent of everyday comments preceding laughter are remotely humorous. 'Laughter usually follows comments like "I've got to go now" or "Here's John." '

The fact that we don't have control over when we laugh suggests that it must be deeply embedded in our nature. Indeed, studies of the play behaviour of great apes suggest that laughing has been around a lot longer than we have. Chimpanzees laugh while they are having play fights although the sound is quite different to that made by humans due to their different vocal apparatus. Instead of chopping a single outbreath into the 'ha-ha' sound that characterises our laughter, chimps' laughter sounds like panting.

A recent study of orangutans reveals a deeper similarity with humans. A team of researchers watched the play behaviour of 25 individuals aged between two and twelve at four primate centres. 'In particular we analysed the facial expressions that they produce during social play,' says Dr Marina Davila-Ross of the University of Portsmouth. 'It's a relaxed expression where they open their mouth and show their upper teeth. It's very similar to the human expression of laughter.'

The team discovered that when one orangutan displayed this expression, its playmate would often produce the same expression less than half a second later. The speed with which this mimicry occurred indicated that the orangutan had not had time to decide on the response – in other words the laughter was contagious. 'In humans, mimicking is a mechanism that enables us to understand our social partner better, and this helps us to cooperate and form social bonds. It is clear now that it evolved prior to humankind,' says Davila-Ross.

The fact that we share laughter with great apes suggests that it emerged in our ancestors sometime before the split with chimpanzees six million years ago. But it may have evolved even earlier than that. Research conducted at Bowling Green State University in Ohio, US, found that even rats produce chirping sounds comparable to laughter when playing or when tickled and the common ancestor of rats and humans lived 75 million years ago. The fact that laughter is triggered by tickling suggests a strong link to humans, because, as Provine puts it, 'tickle is the most ancient and reliable stimulus of

laughter.' One of the earliest games parents and children play is when the parent tickles the baby on the stomach or under the arms until it laughs.

Studies of tickling, although thin on the ground, should therefore be able to tell us a lot more about laughter. For example, we all know that we cannot make ourselves laugh by tickling ourselves. But could a machine tickle us? One team of researchers at the University of California at San Diego built a mechanical tickling machine to look at this very question. They discovered that their subjects laughed just as much in response to the machine as to the experimenter. This tells us that a successful tickle does not depend on another person, but merely on another entity, something that's not you.

Discovering that laughter can be used as a tool to explore other aspects of our behaviour has, for Provine, been one of the most rewarding aspects of his research. Perhaps his most important insight concerns the evolution of speech.

Provine believes that the evolution of speech and bipedal locomotion are causally related. He came to this conclusion after analysing the difference between chimp and human laughter. 'It occurred to me that basically the human 'ha-ha' came about as a result of the evolution of breath control that chimps lack,' he explains. We hold our breath to lift heavy objects and quadrupedal animals must do the same when moving in order to support their body when their four limbs hit the ground. When our ancestors stood up on two feet, the chest was freed of these mechanical demands making it possible for speech to evolve.

By breaking away from traditional models of laughter and discovering its links to deep elements of human nature such as speech and sociality, Provine has reinforced just how ancient laughter is. It has been around for as long as rough play, an activity that you see in mammals, from rats and squirrels to chimpanzees and humans, and has most likely evolved from the laboured breathing that accompanies such play.

(946 words)

Experienced group: Anne

Items

Questions 1-8

Do the following statements agree with the information given in Reading Passage X?

In boxes 1-8 on your answer sheet, write

TRUE	*if the statement agrees with the information*
FALSE	*if the statement contradicts the information*
NOT GIVEN	*if there is no information on this*

1 Provine wrote down more than a thousand examples of what made some of his students laugh.

2 Provine classified his research material into male and female subjects.

3 Provine considered dividing the laugh episodes into the kind of laughter generated.

4 Provine observed that laughter is mostly stimulated by remarks that are without humour.

5 Copying another person's gestures or behaviour is believed to assist in the creation of communal attachments.

6 It is clear that laughter developed in man nearly six million years ago.

7 There has been a considerable amount of research into tickling.

8 The tickling machine is to be tried out on a range of different mammals.

Questions 9-14

Complete the summary below.

*Choose **TWO WORDS ONLY** from the passage for each answer.*

Write your answers in boxes 9-14 on your answer sheet.

Laughter in Great Apes

When observing chimpanzees, researchers have noted that laughter occurs when the animals are involved in **9**…….. The chimpanzees make a noise similar to **10**…….and this is because their internal **11**…….is not the same as that of humans.

Other researchers have studied orangutans in captivity and focused on the common **12**…….that they exhibit when relaxing together. The researchers were especially interested in the fact that the top **13**……of the orangutans were visible when they were 'laughing' . When observing one animal 'laughing', researchers frequently noted that another orangutan immediately copied this behaviour, suggesting that the laughter could be described as **14**…..

Questions 15 and 16

*Choose **TWO** letters, A-E.*

Write the correct letters in boxes 15 and 16 on your answer sheet.

Which **TWO** of the following statements are mentioned in the passage?

A It is thought that laughter in apes is related to their ability to stand upright at times.

B Laughter in chimpanzees probably originated when they learned to hold their breath.

C Human speech began to develop when early man ceased walking on four legs.

D All mammals demonstrate some kind of laughter when playing.

E Laughter may originate in the physical response to the exertion of play.

KEY

1 False

2 True

3 Not Given

4 True

5 True

6 False

7 False

8 Not Given

9 play fights

10 panting

11 vocal apparatus

12 facial expressions

13 teeth

14 contagious

15 /16 C/E (in either order)

Experienced group: William

Text

The changing image of childhood in English literature

A Childhood as an important theme of English literature did not exist before the last decades of the eighteenth century and the poetry of Blake and Wordsworth. There were of course children in English literature before then, as the subject of lyrics and complimentary verses. But in drama, the main body of poetry and the novel, the child is virtually or entirely absent.

B With Blake and Wordsworth we are confronted with an essentially new phenomenon, that of major poets expressing something they considered of great significance through the image of the child. In time, the child became the central figure of an increasingly significant proportion of English literature. The concept of the child's nature which informed the work of Blake and Wordsworth was that children were naturally innocent, and were slowly corrupted by the society in which they lived – in contradiction to the long Christian tradition that everyone, child and adult alike, is sinful.

C The nineteenth century saw the beginnings of a spiritual crisis. The securities of the eighteenth-century peace dissolved in the era of revolution, leading to social and political ferment. The social, political, and, more especially, the intellectual problems arising from the French and Industrial Revolutions found no resolution. In a rapidly dissolving culture, the nineteenth-century artist faced alienation. The concern of the modern European intellect has been, in part, the maintenance of individual integrity within the search for the security of universal order. At no time has that maintenance and search been so pressing in its demand as in the nineteenth century, when long-accepted ideas were challenged not only by the upheavals mentioned above, but also by the revolutionary thinking of Darwin, Marx and Freud.

D The society created by the industrial developments of the late eighteenth and nineteenth centuries was increasingly unconcerned with and often hostile to art. The novelist Charles Dickens was the last major English writer to have a really successful public voice, in the mid 1800s. By the end of the century, there was a new literate public who were unresponsive to the best creative work. A new mass literature supplied the demands of uninformed literacy; and the relative influence of the mature creative voice was proportionally diminished. Art was on the run; the ivory tower had become the substitute for the wished-for public arena.

E In this context of isolation, alienation, doubt and intellectual conflict, it is not difficult to see the attraction of the child as a literary theme. The child could serve as a symbol of the artist's dissatisfaction with the society which was in process of such harsh development about him or her. In a world given increasingly to utilitarian values and the machine, the child could become the symbol of imagination and sensibility, of nature set against the forces in society actively de-naturing humanity. Through the child the artist could express awareness of the conflict between human innocence and the cumulative pressures of social experience, and protest against the horrors of that experience.

F If the central problem of the artist was in fact one of adjustment to the realities of society, one can see the possibilities for identification between the artist and the consciousness of the child, whose difficulty and chief source of pain often lie in adjustment and accommodation to his or her environment. In childhood lay the perfect image of insecurity and isolation, of fear and bewilderment, of vulnerability and potential violation. Some authors took the identification to an extreme, turning to this image as a means of escape, a way of withdrawal from spiritual and emotional confusion in a tired culture. They could be said to have taken refuge in a world of fantasy and nostalgia for childhood.

G The nineteenth century saw the deterioration of the concept of the child as a symbol of innocence. The symbol which had such strength and richness in the poetry of Blake and Wordsworth and some later novels became in time the static and sentimentalised child-figure of the popular nineteenth-century imagination; only a residue of a literary theme, retaining little or nothing of the significance it had earlier borne. It was against this conventionally innocent child that a revolution was effected at the turn of the nineteenth century. Just as the eighteenth century had turned from the Christian doctrine of original sin to the cult of original virtue in the child, so the nineteenth century turned from the assumption of original innocence to the scientific investigation of the infant and child consciousness and its importance to the development of the adult mind.

H A distinction can be made between those late eighteenth- and nineteenth-century authors who went to the child to express their involvement with life, and those who approached the symbol as a retreat from 'life's decay'. In writing of childhood, we find that in a very exact and significant sense the modern author is writing of life. In the literature of the child in the nineteenth and twentieth centuries we have a reflection of the situation of certain artists in modern times; their response, at a deep and significant level, to the condition in which they found themselves; and, if their feelings could achieve the projection, the condition in which they found humanity. Considering the nature of that condition, it is perhaps not remarkable that through writing of childhood there should be those who wanted to go back to the beginning to begin again, and others who wanted just to go back.

Experienced group: William

Items

Questions 1-6

Do the following statements agree with the information given in Reading Passage 0?

In boxes 1-6 on your answer sheet, write

	TRUE	if the statement agrees with the information
	FALSE	if the statement contradicts the information
	NOT GIVEN	if there is no information on this
1	Blake and Wordsworth adapted a tradition of expressing ideas through children.	
2	A number of writers identified the industrial revolution as a major cause of social problems.	
3	Children featured less often in 19th century literature for the masses than in serious literature.	
4	During the 19th century, serious writers found themselves increasingly marginalised by the popularity of mass literature.	
5	Some 19th century authors saw in childhood a reflection of their own difficulties in adjusting to society.	
6	The concept of the innocence of children retained its power as a symbol throughout the 19th century.	

Questions 7-11

Complete each sentence with the correct ending, A-G below.

Write the correct letter, A-G in boxes 7-11 on your answer sheet.

7	Authors working prior to the late 18th century
8	Blake and Wordsworth
9	Darwin, Marx and Freud
10	Dickens
11	In the harsh society of the 19th century, some authors

A	wrote about the relationship between children and their parents.
B	added to the difficulty of reconciling individual needs with those of society.
C	recognised the damage that children could inflict on each other.
D	used children as a vehicle for protest.

E	rarely included children in any significant role.
F	researched the effects of revolution on children.
G	used children as symbols of innocence.
H	gained a degree of popularity that later 19th century writers failed to equal.

Questions 12-17

Reading Passage 0 has eight paragraphs, A-H.

Which paragraph contains the following information?

Write the correct letter, A-H, in boxes 12-17 on your answer sheet.

12	a comparison between changes in concepts of children in two distinct periods
13	a reference to the impact of new ideas during the 19th century
14	a contrast between two psychological motives for writing about children
15	a reference to an increase in the number of people able to read
16	how Blake's and Wordsworth's view of the child differed from the prevailing concept
17	a contrast between qualities symbolised by children and the realities of society

Key

1	False	10	H
2	Not given	11	D
3	Not given	12	G
4	True	13	C
5	True	14	H
6	False	15	D
7	E	16	B
8	G	17	E
9	B		

Experienced group: Elizabeth

Text

Time to wake up to the facts about sleep

Claims that we are chronically sleep-deprived are unfounded and irresponsible, says sleep researcher Jim Horne

A

Ask people whether they would like more sleep, and most will say yes. Does that mean they are not sleeping enough? The apparent desire for more sleep, together with oft-repeated assertions that our grandparents slept longer, all too easily leads to the conclusion that most people in the west are chronically sleep-deprived. It has also been claimed that inadequate sleep causes obesity and related disorders such as diabetes.

Claims of widespread sleep deprivation in western society are nothing new – in 1894, the British Medical Journal warned that the 'hurry and excitement' of modern life was leading to an epidemic of insomnia. But even then it probably wasn't true. The fact is that most adults get enough sleep, and our sleep debt, if it exists at all, has not worsened in recent times.

B

The well-known 'fact' that people used to sleep around 9 hours a night is a myth. The figure originates from a 1913 study by researchers at Stanford University, which did find that average daily sleep was 9 hours – but this applied to children aged 8 to 17, not adults. Even today, children continue to average this amount. Over the past 40 years, there have been several large studies of how much sleep people actually get, and the findings have consistently shown that healthy adults sleep 7 to 7½ hours a night.

More support for today's epidemic of sleep debt supposedly comes from laboratory studies using very sensitive tests of sleepiness, such as the multiple sleep latency test, in which participants are sent to a quiet, dimly lit bedroom and instructed to 'relax, close your eyes and try to go to sleep'. These tests claim to reveal high levels of sleepiness in the general population, but as they are performed under relaxing conditions they are able to eke out the very last quantum of sleepiness which, under everyday conditions, is largely unnoticeable.

Another line of evidence trotted out for chronic sleep deprivation is that we typically sleep longer on vacation and at weekends, often up to 9 or 10 hours a night. It is often assumed that we do this to pay off a sleep debt built up during the week. However, just because we can easily sleep beyond our usual daily norm – the Saturday morning lie-in, the Sunday afternoon snooze – it doesn't follow that we really need the extra sleep. Why shouldn't we be able to sleep to excess, for indulgence? After all, we enthusiastically eat and drink well beyond our biological needs.

C

What of the risk of a sleep shortage causing obesity? Several studies have found a link, including the Nurses' Health Study, which tracked 68,000 women for 16 years. The hazard, though real, is hardly anything to worry about. It only becomes apparent when habitual sleep is below 5 hours a day, which

applies to only 5 per cent of the population, and even then the problem is minimal. Somebody sleeping 5 hours every night would only gain a kilogram of fat per year.

The link between short sleep and diabetes has also been exaggerated. It's true that healthy young adults who are restricted to 4 hours' sleep a night for several nights show the beginnings of glucose intolerance, which can be a precursor to type 2 diabetes. However, that doesn't mean it happens in the real world. For one thing, the effect quickly reverses after one night of recovery sleep. Moreover, 4 hours' sleep is highly artificial and the vast majority of people cannot sustain it for more than a few days. Our very lowest natural limit seems to be 5 hours, yet the researchers did not test the effect of 5 hours' sleep on metabolism, and many have just assumed that what is found with 4 hours' sleep applies to short sleep in general.

D

Not only have chronic sleep deprivation and its consequences been overstated, I also believe that our apparent desire for more sleep isn't all it seems. Do we really mean it when we say 'yes' to the question, 'Would you like more sleep?' It's a leading question that invites a positive response, in the same way as asking whether you would like more money, a bigger house or more holiday. Who, in all honesty, would say no? The real test of inadequate sleep is whether people feel excessively sleepy during the daytime.

E

My team recently investigated sleep deprivation by giving around 11,000 adults a questionnaire asking about it in an indirect way. We asked respondents the times when they usually went to bed and woke up, and the amount of sleep they felt they needed each night. The responses to these two questions allowed us to get an estimate of the shortfall. They also completed another questionnaire to assess daytime sleepiness. Half the respondents turned out to have a sleep shortfall and around 20 per cent had daytime sleepiness.

We then asked, 'If you had an extra hour a day, how would you prefer to spend it?' The alternatives were exercising, socialising, reading or relaxing, working or sleeping. Few people opted to use their extra hour for sleep. It seems that people may want more sleep, but they may not actually need it, and they will happily forego extra sleep in favour of other leisure activities.

F

Does any of this matter? I believe it does. Propagating the myth of a sleep-deprived society adds to the anxieties of people who wrongly believe they are not getting enough sleep, leading to a greater demand for sleeping pills. Rather than attempting to increase our sleep, maybe we should spend those 'extra' hours of wakefulness doing something more productive.

New Scientist 18 October 2008

Experienced group: Elizabeth

Items

Questions 1-6

*Choose the correct letter, **A**, **B**, **C** or **D**.*

Write the correct letter in boxes 1-6 on your answer sheet.

1 What problem does the writer identify with the study done at Stanford University in 1913?

A The research was based on a false assumption.

B The findings conflict with those of later studies.

C The conclusion has not been accurately reported .

D The researchers did not clearly identify age groups.

2 The writer claims tests such as the multiple sleep latency test may not have valid results because

A they do not use a representative sample of the population.

B they require the subjects to try to sleep in unrealistic conditions.

C they do not make precise enough measurements of the time slept.

D they encourage the subjects to sleep more than they would normally.

3 The writer mentions the 'Saturday morning lie-in' as an example of

A a treat that may actually be harmful to health.

B something unnecessary that is done for pleasure.

C a time when we can catch up on the sleep we need.

D something that may not actually lead to extra sleep.

4 What is the writer's conclusion about the link between sleep and obesity?

A A good way to lose weight is to sleep less.

B The risk of lack of sleep causing obesity is insignificant.

C Too much sleep leads to obesity in only 5% of cases.

D There is no relationship between lack of sleep and obesity.

5 The writer criticises a study linking lack of sleep with diabetes because

A it was not based on a natural situation.

B it underestimated how little sleep people really need.

C it only concentrated on recovery sleep.

D it did not test the effect of lack of sleep on metabolism.

6 The writer suggests that when researchers use a particular type of question, this

A may provide data that is inaccurate.

B may show how materialistic people are.

C may elicit information that is surprising.

D may make people afraid of answering honestly.

Questions 7-12

Complete the summary below.

*Choose **NO MORE THAN TWO WORDS** from the passage for each answer.*

Write your answers in boxes 7-12 on your answer sheet.

The writer's team carried out a study on 11,000 adults. Perceptions of sleep deprivation were estimated by comparing the answers to two **7**…… questions, and the team found that half the respondents had sleep deprivation. **8**…… was also assessed, and found to be less common. The team also found that if they were given an extra hour a day, few people would use this for sleeping.

The writer concludes that people do not **9**…… more sleep. He says his findings are important because false beliefs about sleep deprivation are creating **10**…… which have no basis in reality, and encouraging people to ask for **11**…… . People should therefore not try to **12**…… the number of hours they sleep.

Questions 13-17

Reading passage x has six sections A-F.

Which paragraph contains the following information?

*Write the correct letter, **A-F**, in boxes 13-17 on your answer sheet.*

***NB** You may use any letter more than once.*

13 a mention of a medical condition which may precede a more serious illness

14 a reference to sleep deprivation in a specific academic publication

15 some examples of things people could do instead of sleeping

16 a statement of the amount of sleep the writer believes is needed by an adult

17 a summary of the reasons why sleep deprivation is seen as a problem today

Answer key

1	C	4	B
2	D	5	A
3	B	6	A

7 indirect

8 Daytime sleepiness

9 need

10 anxieties

11 sleeping pills

12 increase

13 C

14 A

15 F

16 B

17 A