

**IELTS™**

# IELTS Research Reports

**Volume 11**



[www.ielts.org](http://www.ielts.org)



# Introducing IELTS

IELTS stands for the International English Language Testing System. It is a test for people who intend to study or work where English is the primary language of communication.

Since we introduced it in 1989, IELTS has been widely regarded by both academics and students as the model of English language testing. In four central areas – experience, authenticity, quality and global reach – it has proved itself time and again. For millions of colleges, universities and companies around the world it is, quite simply, the best way to gauge the English skills of your students and staff.

## EXPERIENCE

IELTS is tried and tested: over the past 20 years, more than six million candidates have taken the test. Yet its development goes back to a decade before when a group of British academics asked themselves how they could create a fair, accessible and effective test of English. Today's exam, then, is the product of 30 years of our investment in development and fundamental research.

The fact that IELTS is so well-established means that test administrators and candidates can rest assured that IELTS contains a huge and continually refreshed bank of resources, so avoiding unwanted repetition of materials. Ours is the test that is always innovating but – crucially – it is the test that is always trusted.

## AUTHENTICITY

IELTS tests authentic skills for success. We recognise the central skill our stakeholders are looking for: the ability to communicate in real-life situations.

IELTS has always used face-to-face interaction in the speaking test because it's by far the most realistic option. It generates the most natural performance from the candidate and, not least, the candidate feels more comfortable speaking to a real person rather than a computer.

This attention to authenticity extends to all four parts of the IELTS test: listening, speaking, reading and writing. IELTS tasks reflect real language proficiency for academic use and everyday experiences.

## QUALITY

We pride ourselves on being the best at what we do. Each IELTS test is based on an exhaustive, ISO-compliant development process ensuring that the test material is of world-leading quality. We take every possible measure to guard against fraudulent test-takers and documentation while all IELTS test centres meet strictly audited standards of quality, security and customer service.

This expertise has not gone unnoticed: IELTS test scores are trusted and recognised by more than 6,000 universities and colleges around the globe. The rigorous way in which we recruit and train our examiners very much contributes to this. All of our examiners – of which there are more than 4,000 – mark to the highest quality and are constantly monitored both electronically and by the Professional Support Network which, overseen by chief examiners, manages and standardises recertification.

## GLOBAL REACH

IELTS is the international test for international mobility. The test is taken by well over a million candidates each year in over 130 countries with each test centre offering unparalleled customer service and support.

This global reach means that IELTS understands both its clients and its clients' markets: real-life questions are sourced by writers based in the UK, Australia, Canada, the United States and New Zealand. This internationalism also applies to the managing partnership of IELTS, which combines the missions of four international organisations dedicated to academic excellence, cultural understanding, student recruitment and creating success worldwide.

[www.ielts.org](http://www.ielts.org)

**IELTS**<sup>™</sup>



## FOREWORD

It is with pleasure that the IELTS partners bring you Volume 11 of the IELTS Research Reports. The studies reported in this volume were funded by the joint-funded research programme, sponsored by British Council and IELTS Australia. The third IELTS partner, Cambridge ESOL, supports the programme by supplying sample materials and assistance to approved researchers.

The programme reflects the IELTS partners' longstanding commitment to be at the forefront of developments in English language testing, as well as their desire to maintain productive relationships with the worldwide academic research community for language education and assessment. The programme represents a major part of the Research and Validation framework of the IELTS test.

IELTS is at the cutting edge of English language testing and is one of the world's most researched English language tests. This research informs the constant review of test design and content which are refined to incorporate advances in applied linguistics, language pedagogy, language assessment and technology. The steady evolution of IELTS in response to research outcomes ensures its contemporary relevance and continued usefulness for test stakeholders. IELTS continues to be the language test that sets the standard through its high level of quality, validity, security and overall integrity.

As with the previous ten volumes of IELTS research reports, Volume 11 brings together several empirical studies. The contents deal with contemporary themes of direct relevance not just to the IELTS test producers and other IELTS stakeholders but to the wider language teaching and testing community.

The annual call for research proposals is widely publicised, with research proposals of all types being carefully considered. A Joint Research Committee, comprising representatives of the three IELTS partners, agrees on research priorities and oversees the tendering process. Committee members, in collaboration with experts in the field of applied linguistics and language testing, review and evaluate the research proposals according to the following criteria:

- relevance and benefit of outcomes to IELTS
- clarity and coherence of the proposal's rationale, objectives and methodology
- feasibility of outcomes, timelines, and budget
- qualifications and experience of proposed project staff
- potential of the project to be reported in a form which would be both useful to IELTS and of interest to an international audience.

As of this year, all the published IELTS research reports are available for download on the IELTS website, [www.ielts.org](http://www.ielts.org) where further information can also be found.



Martin Davidson  
Chief Executive  
British Council



Anthony Pollock  
Chief Executive  
IELTS Australia



## Contents

<b>Introducing IELTS</b> .....	<b>1</b>
<b>Introduction, Dr Lynda Taylor</b> .....	<b>7</b>
<b>1      An impact study into the use of IELTS by professional associations and registration entities: The United Kingdom, Ireland and Canada</b>	
<i>Examines the history and rationale for selection of IELTS as a language benchmarking system for professional associations, and explores the main competitors to IELTS in global testing for professionals, in Canada, the UK and Ireland.</i>	
<b>Glenys Merrifield, GBM Associates</b> .....	<b>21</b>
<b>2      Learning to play the ‘classroom tennis’ well: IELTS and international students in teacher education</b>	
<i>Addresses the question of an appropriate IELTS score for graduate entry teacher education courses by investigating the extent to which current IELTS scores into graduate entry teacher education courses are considered adequate: by the lecturers of such students, by prospective teacher registration authorities and by the students themselves.</i>	
<b>Wayne Sawyer and Michael Singh, University of Western Sydney</b> .....	<b>73</b>
<b>3      A multiple case study of the relationship between the indicators of students’ English language competence on entry and students’ academic progress at an international postgraduate university</b>	
<i>Investigates selection practices and decision making rationales of admissions personnel in an international, postgraduate UK setting and the consequences for borderline non-native English speaking students’ academic progress.</i>	
<b>Gaynor Lloyd-Jones, Charles Neame and Simon Medaney, Cranfield University</b> .....	<b>129</b>
<b>4      Construct validity in the IELTS Academic Reading test: a comparison of reading requirements in IELTS test items and in university study</b>	
<i>Investigates suitability of items on the IELTS Academic Reading test in relation to the reading and general literacy requirements of university study, through a survey of reading tasks in both domains, and interviews with academic staff from a range of disciplines.</i>	
<b>Tim Moore, Swinburne University, Janne Morton, Melbourne University and Steve Price, Swinburne University</b> .....	<b>185</b>
<b>5      An investigation of the process of writing IELTS Academic Reading test items</b>	
<i>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.</i>	
<b>Anthony Green and Roger Hawkey, University of Bedfordshire</b> .....	<b>269</b>
<b>6      The cognitive processes of taking IELTS Academic Writing Task 1</b>	
<i>Investigates cognitive processes of candidates taking IELTS Academic Writing Task One with different graphic prompts at two time points: before AWT1 task training, and after.</i>	
<b>Guoxing Yu, Pauline Rea-Dickins and Richard Kiely, University of Bristol</b> .....	<b>373</b>





## Introduction

The British Council / IELTS Australia joint-funded research program continues to make an important contribution to the ongoing development of the International English Language Testing System (IELTS). Studies funded directly by the British Council and IELTS Australia complement internal research and validation projects conducted or commissioned by Cambridge ESOL, the third IELTS partner. Together they form an essential component of the overall quality management system for IELTS. In particular, the joint-funded program reflects the IELTS partners' long-standing commitment to a process of continuous improvement to the test as well as to ongoing engagement with the wider academic research community for language education and assessment.

The body of published research on IELTS continues to grow steadily. Volume 1 of the *IELTS Research Reports* first appeared in 1998 and has been followed by nine further volumes (some available in a CD-ROM version). These publications were initially produced by IELTS Australia and later in collaboration with the British Council. Copies can be sourced through the IELTS website: [www.ielts.org](http://www.ielts.org). Volumes 1-10 present in the public domain 55 of the empirical studies to have received grant funding from the IELTS partners since 1995. This growing library of IELTS research publications has been enhanced over the past 5 years with volumes in the *Studies in Language Testing* series, published jointly by Cambridge ESOL and Cambridge University Press. They include academic volumes authored or edited by Roger Hawkey (2006), Tony Green (2007), Lynda Taylor and Peter Falvey (2007), and Alan Davies (2008). Details of all these titles can be found on the Cambridge ESOL website: [www.cambridgeesol.org/what-we-do/research/silt](http://www.cambridgeesol.org/what-we-do/research/silt). In addition, many IELTS funded researchers have successfully published articles and papers in peer-review journals or presented academic papers at applied linguistics and language testing conferences worldwide.

*IELTS Research Reports* Volume 11 brings together a set of six empirical studies sponsored under Round 13 of the funded research program (2007). As with much other academic research, there can be a lengthy passage from initial acceptance of a research proposal to completion of the study and subsequent publication of its findings in the public domain. IELTS-funded researchers submit their final project report to the grant-awarding body who then sends it out for peer review by one or more external specialists in the field; it is also forwarded to the IELTS research team at Cambridge ESOL for review and comment. Following a comprehensive cycle of review, feedback and revision, outcomes from the funded studies are finally released into the public domain via a number of channels, including the IELTS website and publications such as this one.

Volume 11 contains studies on varying themes that continue to attract attention and provoke debate among the language testing community and beyond. Three of the six studies in this volume address themes of washback and impact which were also the focus of *IELTS Research Reports* Volume 10 (2009), and of the earlier Volume 8 (2008), but the work presented here breaks some new ground in the area of IELTS washback and impact. One study extends earlier research conducted among professional associations and registration entities using IELTS, while the other two studies explore the use and role of IELTS in contexts that have so far received relatively little research attention, namely teacher education and other postgraduate entry university courses.

Interest in exploring issues of test washback and impact has grown considerably over the past 15 years as awareness has developed of the impact that tests and test results have at a 'macro' as well as a 'micro' level; testing and assessment influence not only the career and life chances of individuals but also social policy and practice in areas such as language and educational planning, migration and citizenship, as well as professional life and employment in an increasingly globalised world. Today many language testers locate concepts of washback and impact within the theoretical notion of 'consequential validity' in which the social consequences of testing are part of a broader, unified concept of test validity. Consequential validity continues to be a core topic for theoretical debate and empirical research within the language testing and assessment community alongside related themes of fairness and ethics.

The notion of consequential validity that developed during the 1990s found partial expression in the emergence of ethical standards and codes of practice. Examples of these include the Association of Language Testers in Europe (ALTE) *Code of Practice* (1994), the AERA/APA/NCME *Standards* in the US (1999), the International Language Testing Association's (ILTA) *Code of Ethics* (2000), and the European Association of Language Testing and Assessment's (EALTA) *Guidelines for Good Practice* (2006) – all of which are available online. All these associations seek to make explicit the obligations of test producers with respect to the users of their tests; at the same time, they advocate a shared responsibility on the part of test users (i.e. learners, teachers, receiving institutions, employers, government agencies, etc) for the appropriate use of tests and test results. The overall aim is to maximise *positive* test washback and impact, and to minimise any negative effects for individuals and for wider society.

It was against this background that, following the last major IELTS revision in 1995, Cambridge ESOL acknowledged the importance of conducting impact studies on IELTS, a high-stakes test with an international gate-keeping function. With support from the other IELTS partners, and with expert input from specialists at the University of Lancaster, Cambridge initiated a long-term research agenda - the IELTS Impact Study – which established the investigation of impact as a priority for internally conducted and commissioned work as well as for externally funded studies. Attention focused on researching various impact-related issues, including: stakeholder attitudes among key user groups; the use of test scores; the nature of score gains; and the provision of test preparation courses and materials. Outcomes from some of this research were published in *Impact Theory and Practice* by Roger Hawkey (2006), Volume 24 in the UCLES/CUP *Studies in Language Testing* series.

Since 1995 a substantial number of impact-related research proposals have received grant-funding and other types of support from the IELTS partners; these too have contributed to a growing body of research evidence that provides valuable insights into IELTS and consequential validity. Given the test's large-scale nature and its high-stakes gate-keeping role worldwide, accumulation of this body of evidence is essential to confirm the extent to which the test is functioning as intended and to identify any issues which may need to be addressed. With the worldwide expansion of higher education and employment opportunities for students and professionals, the IELTS candidature has grown exponentially since 1995; in consequence, IELTS test preparation courses, teaching materials and other resources – once quite limited in nature – have mushroomed beyond all expectations. A complex educational, social and economic infrastructure has thus built up around the test and so the study of consequential validity remains a high priority for the IELTS partners.

The remaining studies in Volume 11 explore issues of **construct validity** within specific components of the IELTS test – Academic Reading and Writing. Two of the three studies investigate the cognitive processes engaged in by test takers when completing the reading and writing test tasks, while the third study reports insights into the process of creating test items and tasks for the Academic Reading module. This work builds on earlier cognitive validity studies for IELTS (see Volumes 6, 7 and 9) but it also breaks new ground by exploring the very nature of the test-writing process itself and the complex factors that shape the creation of appropriate tasks and test items for the IELTS Reading module.

The Academic Reading and Writing variants of IELTS have their roots in the earlier ELTS test (the English Language Testing Service) of the 1980s. From 1989 onwards IELTS continued the tradition of subdividing the overall language proficiency construct for operational test administration purposes into the 4 main components of Listening, Reading, Writing and Speaking, maintaining variants of the Reading and Writing subtests oriented towards academic, university-based education. Over the past 20 years, the individual skills-based modules have evolved in light of findings from the ongoing validation and research activities sponsored or undertaken by the IELTS partners (see, for example, Taylor and Falvey's 2007 account of changes to the IELTS Speaking and Writing components). The modules have also been revised to take account of advances in technology (e.g. computer-based testing), measurement

theory (e.g. the application of IRT and item-banking), test administration (e.g. production of multiple versions for fixed date sessions), and the currency and recognition of IELTS in an increasingly globalised world and among a wider set of test stakeholders (e.g. the use of the test for professional registration and migration purposes).

The evolutionary development of ELTS and IELTS is well documented by Alan Davies in his 2008 volume *Assessing Academic English*, Volume 23 in the UCLES/CUP *Studies in Language Testing* series. Davies analyses the many and various factors that shape the development of any test within a given historical and socio-cultural context, including our understanding of language and of language learning processes and outcomes, as well as of the nature of assessment. What is important is that any test task (or items) should be consistent with a likely focus for responding to the input material and should encourage test takers to engage in appropriate cognitive processes that bear some relation to the processing that would typically take place in the world beyond the test. Validation studies such as those in this volume help the test producers to confirm an appropriate match between task input, cognitive processing and task output. In recent years the joint-funded program has specifically encouraged research into the nature of the cognitive processes underlying the tasks in IELTS, and the relationship of these to activities beyond the test in the world of academic study (see, for example, Volumes 6 and 9). Insights from funded studies feed directly into the ongoing production, review and validation of test materials for Academic Reading and Writing, and will naturally inform future revision projects for these modules.

In summary then, the contents of Volume 11 deal with contemporary themes of direct relevance not just to the IELTS test producers and other IELTS stakeholders but also to the wider language teaching and testing community. An overview of each contribution in this volume is given below, together with some discussion of its implications for the IELTS partners and for the future development of the test.

## **1 AN IMPACT STUDY INTO THE USE OF IELTS BY PROFESSIONAL ASSOCIATIONS AND REGISTRATION ENTITIES: THE UNITED KINGDOM, IRELAND AND CANADA**

**Glenys Merrifield** focuses on an important area for investigation to have emerged in recent years – the increasing use of IELTS for professionals seeking registration or membership of professional associations. This study was designed to examine language testing for professionals in three global markets: the two traditional markets of the United Kingdom and Ireland, and a relatively new market, Canada. Merrifield’s study helpfully builds upon her earlier research investigating similar issues in Australia, New Zealand and the United States of America (see Merrifield 2008 in *IELTS Research Reports*, Volume 8; see also Read and Wette in *IELTS Research Reports*, Volume 10, 2009).

The objectives of Merrifield’s follow-up study were: to examine the history and rationale for selecting IELTS as a language benchmarking system for professional associations; to explore the main competitors to IELTS in global testing for professionals; to determine the minimum band scores being used as criteria for professional registration or membership, the decision-making policies and the perceptions of its fitness for the purpose; and to identify potential risk factors for the test in future. The methodological approach adopted involved desk research and semi-structured, interviews conducted face-to-face, by telephone and by email with key staff in over 20 professional associations across the three countries (UK, Ireland and Canada).

At the outset, Merrifield describes the global context at the start of the 21st century in which broadening patterns of communication and the removal of borders between countries and cultures have led to increased mobility for professionally trained individuals wishing to migrate and seek work in their chosen profession in English speaking countries. She notes how English language competency in the

professions, particularly high stakes community roles such as those in nursing, medicine and veterinary care, has emerged as a critical criterion for entry to English speaking countries as an immigrant to take up a professional role. Against this background, it is not surprising that the IELTS partners have seen a marked growth in interest from professional associations worldwide wishing to use IELTS as a reliable, secure and widely recognised English language proficiency measure.

Merrifield's research provides a useful and informative survey of how IELTS is currently being used by a range of professional bodies in the UK, Ireland and Canada, in particular the entry criterion scores set and the way these can change with time and experience; in addition, she reports the perceptions of the test among test score users within these bodies. Information is also provided on some of the other internationally available tests that are used in a similar way. The report highlights the perceived advantages of IELTS reported by the stakeholders consulted, in particular: worldwide accessibility across a broad testing network; testing date frequency; prompt reporting of results; inclusion of a face-to-face speaking test component; an online verification service for test score users; support for test candidates; a strong research base underpinning the test; and extensive quality assurance measures. It is worth commenting here that several of these features are generally regarded by academic language testers as 'non-technical' features of a test; for this reason they receive little attention or discussion in the research literature. However, they are clearly high priority considerations for test score users and stakeholders, and they are features that can help to explain a stakeholder's preference for IELTS over an alternative test, even over a test that may have been developed in-country or a test with profession-specific content. The significant factors that shape test choice and use need to be better understood and discussed within the professional language testing community. As the Merrifield report makes clear, most of those consulted in this study expressed general satisfaction with the reliability of IELTS and the support received from the test administration when needed.

The study's findings highlight how difficult it can be for institutional users of IELTS (and similar tests) to maintain their knowledge base and familiarity with the test and the meaning of its scores. Staff turnover and changes in responsibilities, especially in smaller organisations, can mean that accumulated knowledge and experience of the test is quickly lost. Merrifield's point about the advisability of the IELTS partners maintaining personal contact with responsible registration staff in the professional associations is well taken. It is encouraging to see that several professional associations have taken the step of reviewing their existing standards and revising their minimum levels overall and for individual skills as appropriate to the needs of the profession. To do so, they have made use of the user-oriented band descriptors for Speaking and Writing available on the IELTS website and the 'IELTS Scores Explained DVD', both of which were developed several years ago to assist stakeholders in IELTS score interpretation and in setting appropriate band score thresholds for the context of use. These and other resources form part of the IELTS partners' established and ongoing strategy to respond appropriately to stakeholder needs.

As Merrifield notes, these benchmark review and revision projects suggest a growing understanding of IELTS and its band scores and they reflect the advice the IELTS partners have always given, i.e. that score users should set their standards in an informed and, where possible, evidence-based manner. As a rule, the IELTS partners work closely with any professional body considering adoption of IELTS as its language proficiency measure, encouraging careful content analysis of the test against the skills and competences of the relevant profession as well as consideration of key pragmatic requirements for the association such as test security, recognition, accessibility and cost prior to any decision to adopt; if the evidence from this exploratory stage suggests the test to be appropriate for the context of use, then the IELTS partners support the carrying out of standard setting studies to set appropriate and safe criterion thresholds for the purpose and context of use. Examples of this type of collaboration with test users include two standard setting studies conducted during 2004: one in the US in collaboration with the National Council of State Boards of Nursing (NCSBN), the BUROS Institute of Measurement and the IELTS partners; and a second in collaboration with the UK General Medical Council, Lancaster

University, and the IELTS partners. Outcomes of both studies were reported in research papers presented at the Language Testing Research Colloquium (LTRC) in Ottawa in June 2005; a research paper on the first by O'Neill, Buckendahl, Plake and Taylor was published in the peer review journal *Language Assessment Quarterly* in late 2007. As Merrifield comments, it would be good to see professional associations commissioning more test validation studies of this type, including analysis of workplace language skills for the various professions.

On the sensitive issue of European legislation that currently exempts European nationals from language assessment, the IELTS partners will continue to work with the relevant departments and agencies in the UK and Ireland to provide appropriate advice. And on the policy issue over whether or not to accept IELTS scores from more than one test sitting or administration, it may be helpful to clarify that this has always been a matter for the user organisation to determine, rather than a stipulation from the IELTS partners. If a professional association wishes to accept an individual's best score outcomes from two or more IELTS tests, then they are at liberty to do so.

Merrifield highlights several perceived risk factors for the future of IELTS including the challenge posed to IELTS' market share by iB TOEFL and by newer tests such as Pearson's Academic Test of English. Market competition may be of less concern, however, than pressure from user organisations to 'tailor' the test content more closely to their professional domain of interest. IELTS (and its predecessor ELTS) was originally designed to be appropriate to the academic context and has served well in that capacity for over 25 years. More recently, some professions have found that the test can also serve their needs when they require practitioners to demonstrate appropriate evidence of general English proficiency. In the mid-1990s IELTS moved away from discipline-specific modules largely because of the impossibility of ensuring fair, relevant and comparable test versions across differing disciplinary fields. Prior to April 1995 IELTS included three discipline-specific subtests for both Reading and Writing - Academic Modules A, B and C – which were designed to meet the needs of candidates in three broad discipline areas: Physical Sciences and Technology (PST), Life and Medical Sciences (LMS), and Arts and Social Sciences (ASS). Despite its attractiveness, this sub-division of the test into three discipline-specific subtests caused some administrative concern because test centres and receiving institutions were often unclear about the appropriate subtests for different higher education courses. Furthermore, it was not always clear whether it would be better to match a candidate to a discipline-specific subtest on the basis of their previous or their intended disciplinary area; even within a broad disciplinary area, the sub-specialities could be so varied, e.g. geology and aeronautical engineering within Physical Sciences and Technology, that candidates taking the same module would not necessarily encounter texts and topics related to their own disciplinary field. Feedback from IELTS administrators and examiners supported a reduction in the number of discipline-specific subtests. At the same time, monitoring of usage showed that the majority of IELTS candidates (around 75%) were taking Module C. In light of these findings, the International Editing Committee recommended research to investigate the effectiveness of a one-module approach and to ensure that Academic candidates would not be disadvantaged if they were to take a single module for each of the Academic Reading and writing subtests. The results of this project, together with results from important research into second language reading and ESP testing by Caroline Clapham (1993, 1995, 1996), showed that a single test for all academic disciplines did not discriminate for or against candidates of any discipline area. For this reason, the three discipline-specific subtests in IELTS were replaced in 1995 by one Academic Reading Module and one Academic Writing Module. Despite the desire for a more 'tailored' approach, it is doubtful that subject-specific content could be reintroduced into IELTS in such a way that the test could meet the expectations of the many different academic and professional domains that are now using it; moreover, the production of such multiple and comparable test versions across diverse disciplines would be neither viable nor sustainable. Merrifield's point is well made, nonetheless, about the need to ensure that existing and future stakeholders clearly understand the difference between the Academic and General Training variants of IELTS, and the IELTS partners will continue to draw attention to the important distinction between the two versions

of the test.

Finally, it is worth noting that a particularly valuable outcome of this and the previous study is that together they enable a cross-national overview and breakdown of the English language proficiency standards set by a broad range of professional associations and regulatory authorities in some key international employment domains. As such, it is hoped that the two studies will prove a useful and up-to-date resource for policy discussion and decision-making in professional bodies more widely.

## 2 LEARNING TO PLAY THE 'CLASSROOM TENNIS' WELL: IELTS AND INTERNATIONAL STUDENTS IN TEACHER EDUCATION

**Wayne Sawyer** and **Michael Singh's** research study focuses on the role of IELTS in selecting international students for teacher education courses in the Australian context. It has an interesting synergy with Merrifield's research since the latter study also surveyed the use of IELTS by two professional teacher associations within the Canadian context.

Sawyer and Singh set out to address the question of an appropriate IELTS score for graduate entry teacher education courses. They did so by investigating the extent to which current IELTS scores for entry into such programs are considered adequate by three key stakeholder groups – course lecturers, prospective teacher registration authorities, and students themselves. This was followed up with exploration of what an appropriate score might be for entry to such courses. The methodology adopted for the study involved analysis of relevant policy documents on course entry requirements complemented with semi-structured interviews with a range of stakeholder constituencies. Academics from four Faculties of Education and one student cohort were interviewed, along with representatives of one state teacher registration authority.

The report of the study begins with an overview of the available research literature relating to international students, English language proficiency, testing instruments and teacher education. Though the focus is here on its relevance to Australian higher education, the review should prove useful and generalisable to other English-speaking contexts since it identifies a range of language skills required by teachers in the classroom. The focus for concern is that teacher education students whose language background is not English not only need to perform adequately in English for the purposes of their academic study, but they also need to be fluent in the public situation of teaching their own classes on practicum and in their future careers. For teacher education students from international backgrounds, successful course completion means that the language issues they must address encompass not just listening and reading comprehension or success in academic writing, but 'performance' of oral and written English in classrooms, along with issues such as the differences in school cultures when compared with their own backgrounds.

Not surprisingly, perhaps, the study's findings suggest that the situation is complex and multifaceted, partly because the issue of language testing for students is not just one of entry-level, i.e. ability to cope with the linguistic demands of teacher training; it is also one of exit-level competence at the end of the course. (Readers may like to note that the issue of IELTS exit score gains was the focus of a study by O'Loughlin and Arkoudis in *IELTS Research Reports*, Volume 10, 2009.) As we have seen so often in studies investigating the setting of language proficiency requirements, there exists no magic IELTS band score formula which can suit multiple contexts of use. Instead, a process of careful analysis and considered judgement is necessary by each institutional user of IELTS scores, taking into account factors such as course length, level of language and other support available post entry, as well as cultural experience and expectations. Sawyer and Singh suggest that the extent of opportunity for language development during a course may be an important consideration in this regard: thus higher IELTS scores may be deemed advisable for accessing shorter courses (e.g. one-year Diplomas), lower scores

more appropriate for longer courses which offer students more time to develop the critical language skills they need to embark on their professional career. The researchers are cautious, nonetheless, and avoid presenting this as a simple solution. Of course setting higher IELTS scores is unlikely to address the issue of familiarisation with the local schooling culture facing international students wishing to train as teachers; such acculturation can really only be addressed during the course itself. To address this angle Sawyer and Singh note the value of post-entry support programs with both a linguistic **and** a cultural focus. Their findings reinforce the view that language proficiency alone is no guarantee of either academic success or failure and their study provides further empirical evidence of the complex factors that interact in teacher education as in any other academic discipline or professional domain, arguing for a richer screening regime for applicants. It is clearly unreasonable and unwise to lay a burden of expectation on IELTS scores as an entry requirement beyond what the test is capable of delivering (that is, a measure of English language proficiency), and the IELTS partners continue to recommend that stakeholders take note, where possible, of additional sources of evidence alongside IELTS scores when making selection decisions, e.g. through interviews, or an in-house, subject-specific diagnostic instrument. However attractive the notion may be, it is unlikely that the IELTS partners could create a teacher-education-specific version of IELTS as suggested by the researchers, with varying course-entry and exit components; as discussed above, technical issues combined with matters of practicality and fairness continue to mitigate against the development of specific IELTS versions to match the multiple domains of medicine, nursing, pharmacy, teacher education, etc.

One particularly useful outcome of this study is the inventory it offers of the English language skills needed by teachers, drawn from the available literature on teacher education; to these can be added the additional language skills identified by respondents in the study as being needed by teachers operating in English as a second or other language, including the need to become familiar with colloquial idiom in the school context and with discipline-specific discourse. Interestingly, the researchers also noted the importance attached to listening comprehension by both student and academic respondents, since it is this that enables them to play ‘classroom tennis’ effectively, ie, being comfortable enough in English so as to respond reasonably quickly and appropriately in spoken classroom exchanges. This has important implications for how listening skills should be tested, as well as taught, at the higher proficiency levels; such insights are valuable for the IELTS test producers because they can help inform the future development of the listening test module in IELTS.

### **3 A MULTIPLE CASE STUDY OF THE RELATIONSHIP BETWEEN THE INDICATORS OF STUDENTS’ ENGLISH LANGUAGE COMPETENCE ON ENTRY AND STUDENTS’ ACADEMIC PROGRESS AT AN INTERNATIONAL POSTGRADUATE UNIVERSITY**

The study by **Gaynor Lloyd-Jones, Charles Neame and Simon Medaney** sets out to investigate the selection processes and decision-making rationales of admissions personnel in an international, postgraduate UK setting and the consequences for the academic progress of borderline non-native English speaking students at their institution. The project was contextualised within UK concerns over declining degree standards due to the impact of internationalisation initiatives on the expanded taught Masters postgraduate sector. This study is particularly welcome because few studies to date have researched the selection practices associated with this area, even though much of the recent expansion in UK international student numbers has taken place in the postgraduate sector involving taught Masters programs. The researchers adopted a case study approach to examine the phenomenon of student selection in a postgraduate UK higher education institution specialising in engineering and management through a variety of different Masters programs.

Their introduction and brief literature review conveniently highlight two important points regarding current decision-making practices in this context. First, the literature appears to indicate two different models of student selection. In one, the language requirement is treated as independent of other selection criteria, a simple accept-or-refuse decision with little attention paid to borderline cases. In the other, decision-making is richly complex with the final decision based upon multiple, interacting criteria. Which model was used appeared to depend partly on the context of use, perhaps dictated by the numbers of student applicants involved and the personnel resources available for the selection process. Secondly, the researchers identified a considerable variation in the test scores on English language proficiency measures accepted for university entry, both across and within higher education institutions. They suggest this may be due to contextual sensitivity to disciplinary, programme and institutional diversity, as recommended by the IELTS partners, though they also acknowledge that it may result from the pressure academic institutions find themselves under to recruit international students, sometimes leading to a lowering of entry scores as a result.

In terms of their research findings, Lloyd-Jones, Neame and Medaney found that the extensive variety of selection procedures highlighted in the relevant literature was reflected within their own postgraduate institution. They suggest this variation derived from school affiliation and possibly former organisational arrangements when there existed two separate campuses. While some parts of the institution depended heavily on the outcome of formal English proficiency testing, others relied upon internal assessment by linguist staff, e.g. through interviews, and still others operated a blend of the two approaches. Regardless of school affiliation or programme, the researchers found evidence of complex decision-making in selection rationales built upon multiple and sometimes competing criteria, involving the integration of language and academic abilities. In most cases, however, selection decisions were made on the basis of several sources of evidence; they represented balanced judgements encompassing a variety of criteria considered in the round rather than singly or in isolation. The researchers explain their findings in relation to several factors, including: contrasting approaches to selection between academic and administration personnel, the former preferring complexity of judgement and the latter preferring a simpler and transparent system; the specific nature of the postgraduate applicant population, often older and with richer background experience; the highly competitive nature of courses at this level meaning that applicants were more likely to comply with entry requirements and borderline cases less frequent.

The study comments that since postgraduate applicants differ substantially from the typical applicants for UK undergraduate courses, this, combined with questions of scale (i.e. fewer postgraduate applicants), may necessitate differing selection procedures and practices at the postgraduate level. Such insights are useful for the IELTS partners when called upon to advise higher education institutions about how best to use IELTS scores when setting benchmark entry levels. While the study resonates with earlier findings from joint-funded studies by Coleman et al (see Volume 5), Rea-Dickins et al (see Volume 7) and O'Loughlin (see Volume 8), it also questions the simple assumption that greater awareness or knowledge about English language proficiency assessment measures among the institution's Course Directors will necessarily result in better selection decisions and judgements.

A key finding of this study was that test scores, even when in line with entry requirements, were no guarantee that a student **could** write satisfactorily in an academic genre, particularly for extended texts such as the thesis. In this regard, it may be worth reflecting here that IELTS is primarily intended to test 'readiness-to-enter' the world of academic study; and though previous studies have identified a healthy correspondence between features of the tasks of the IELTS Writing module and the writing tasks encountered by students during their academic studies (see Moore and Morton in *IELTS Research Reports*, Volume 2, 1999), it might be unrealistic to expect the IELTS Writing test score to be a good predictor of writing ability at the highest level in a postgraduate context. Dissertation and thesis-writing skills will clearly need to be developed and refined during the course, according to the demands of the discipline area. Postgraduate students' oral skills, on the other hand, did not appear to be a cause for concern.



Finally, it is interesting to note that the original proposal for this study had to be modified when the researchers found that their early findings conflicted with initial assumptions and when fewer respondents with English test scores were subsequently available than originally anticipated; their reflections on dealing with this unforeseen development are a valuable aspect of this report and may be instructive for others embarking on research of a similar nature.

#### **4 CONSTRUCT VALIDITY IN THE IELTS ACADEMIC READING TEST: A COMPARISON OF READING REQUIREMENTS IN IELTS TEST ITEMS AND IN UNIVERSITY STUDY**

**Tim Moore, Janne Morton and Steve Price** set out to investigate the suitability of IELTS Academic Reading test items in relation to the reading and general literacy requirements of university study in an Australian context. They approached this through a survey of reading tasks in the two domains of the IELTS test and university study, as well as via interviews with academic staff across a range of academic disciplines. A taxonomic framework was constructed to analyse IELTS and university-based reading tasks, with a focus on two dimensions of potential difference: *level of engagement*, referring to the level of text with which a reader needs to engage to respond to a task (local versus global); and *type of engagement*, referring to the way (or ways) a reader needs to engage with texts on the task (literal versus interpretative). The study sought to make explicit the task demands of reading items in the IELTS Reading test so as to understand the types of interaction being provoked between text and reader, and the extent to which these reflect the types of reading tasks and activities required of students on university programs. This study therefore has a strong construct validity focus, and it adds to similar research undertaken in the UK on reading and IELTS by Professor Cyril Weir and his colleagues at the University of Bedfordshire (see *IELTS Research Reports*, Volume 9, 2009).

The findings of this study are encouraging for the IELTS test producers inasmuch as they provide further empirical evidence of a clear correspondence between the reading requirements in the IELTS Academic Reading test and some of those needed for academic study in the world beyond the test. Similarity was observed in those types of reading requiring a mainly local and literal engagement with material, i.e. a basic comprehension of relatively small textual units. Most of the IELTS reading test items were observed to reflect features of reading tasks found in the corpus of academic texts gathered for the study, texts which had as their focus the need for students to understand certain discipline-based concepts. At the same time, however, there was evidence of some divergence between the two domains, with a variety of reading tasks in the academic corpus appearing to require a more critical engagement with material or interaction with multiple sources and viewpoints. These task types and demands were noticeably less evident in the IELTS task corpus under scrutiny.

The patterns of similarity and difference between the IELTS reading tasks and the academic task corpus were confirmed in the interviews with academic staff, though, interestingly, perceptions varied among subject staff from differing disciplines about the degree of congruence between the type of reading they expected their students to do on courses and the apparent demands of the IELTS test. Moore, Morton and Price reported a broad division between the ‘harder’ technical disciplines on the one hand (e.g. Engineering, Architecture, Physics, Biology), where reading expectations seem to be characterised more narrowly, e.g. as requiring the assimilation of information, and the ‘softer’ humanities-oriented disciplines on the other hand (e.g. Media Studies, Linguistics, History, Management), where academic reading requirements seem to be more complex, often comparative or evaluative in nature. This would suggest that the types of materials students need to read on their courses, and the ways they need to go about reading these, can vary markedly depending upon the disciplinary field.

In discussing their research findings the researchers offer us some valuable insights into key features that appear to differentiate the reading demands of IELTS from the demands of academic reading in the university study context, including specific features relating to epistemic entities, interpretative readings, readings of multiple texts, the contextual nature of reading, the reading-writing nexus, information literacy and genre readings of texts. The researchers' discussion touches upon the central issue in language assessment of *construct under-representation*, i.e. the extent to which a test does, or does not, sample adequately from the universe of tasks and activities linked to the construct of interest, in this case the construct of academic reading. Moore, Morton and Price advocate strengthening the link between the two domains (of the IELTS test and the world of study) by including more test items requiring global/interpretative rather than just local/literal reading. This, they suggest, would help bring the cognitive demands of the test more into line with the type of reading required on students' courses.

It is encouraging to see that the researchers are not at all naïve about the practical considerations involved in test production and they readily acknowledge the challenges associated with modifying the IELTS test in order to improve construct representation along the lines they propose. We clearly need to recognise the limits to which a test such as IELTS can (or should be expected to) simulate language use in the target use situation in its entirety. The testing of reading in IELTS is premised upon a generalist construct of academic reading and the researchers are right to highlight the inevitable challenge that disciplinary variation in reading requirements at university raises for a test such as IELTS. Furthermore, as previously mentioned, IELTS is designed principally to test *readiness to enter* the world of university-level study in the English language and does not assume test takers have already mastered the high-level academic literacy skills they are likely to require for their future studies. Such skills may well need to be developed **during** their studies, perhaps even during the early months of their first year, and within a specific disciplinary context which enjoys its own specialist discourse and approach to literacy.

Despite these caveats, however, the IELTS test producers are committed to maintaining a cycle of systematic monitoring and continuous improvement of the test and they recognise their responsibility to enhance test content and delivery in the light of ongoing research and as conditions and circumstances allow. Thus the researchers' practical suggestions for how IELTS reading tasks might be extended to reflect a greater degree of global and interpretative reading are immediately relevant to the test-writing process. The sample tasks offered at the end of the report should offer valuable input to the IELTS item-writing teams currently working on material for the Academic Reading module. It may be that closer attention can be given by the test writers to ensuring a larger proportion of reading test items that function at the global and interpretative levels. In the longer term, it is interesting to speculate whether future computer-based development of IELTS might permit a greater inclusion in the reading test of some of the features that characterise academic reading, and thus a broader representation of the construct of interest. Innovative computer-based testing techniques, for example, might enable the test-taker to do one or more of the following: engage with larger quantities of text; interact with multiple texts; exercise skills related to the searching and selecting of sources, including electronic media; and even undertake more sophisticated reading-into-writing tasks.

It is worth noting that recent theoretical and empirical work discussed by Weir in the field of L2 reading and reading assessment (see *Examining Reading* by Khalifa and Weir, 2009) highlights the critical parameters that appear to characterise the higher levels of second language reading ability: in particular, the ability to go beyond simply establishing propositional meaning at the clause, sentence and paragraph level in order to build complex mental models, creating a text level representation based on a single text and an inter-textual representation drawing upon multiple textual sources. This is the sort of reading, argues Weir, that characterises the C1 and C2 levels of the Common European Framework of Reference (CEFR); it is the type of high-level academic reading that students typically need to undertake in their university courses. Although Weir et al (*IELTS Research Reports*, Volume 9, 2009) accept that full contextual authenticity is generally unrealistic for language assessments, our growing understanding of the nature of high-level L2 reading proficiency, combined with the evidence from empirical studies

such as those by Moore, Morton and Price in the Australian context, and by Weir and his colleagues in the UK, undoubtedly have important implications for the future development of the IELTS Academic Reading test.

## 5 AN INVESTIGATION OF THE PROCESS OF WRITING IELTS ACADEMIC READING TEST ITEMS

The study by **Anthony Green** and **Roger Hawkey** explores an aspect of the IELTS test that has so far received relatively little attention under the joint-funded research program. While a few previous studies have focused on the characteristics of texts and test items, there has been little investigation of the actual processes that item writers go through in choosing texts and creating items, and the way these contribute to the quality of test material. This study thus breaks new ground for IELTS and is a welcome addition to the growing body of research relating to the Academic Reading test, complementing previous funded research studies that have explored test content and design. Furthermore, this study helps to explain some of the characteristics of the IELTS texts and the types of reading identified by the Weir et al studies published in *IELTS Research Reports*, Volume 9 (2009).

Green and Hawkey investigated the text selection, item writing and editing processes involved in the development and production of material for the IELTS Academic Reading test. Using the methodology of retrospective reports and direct observation, they set out to compare how trained and untrained item writers, both individually and collectively, select and edit reading texts to make them suitable for a task-based test of reading and how they generate the accompanying items. Both written (flowchart) and oral (interview and focus group) data were gathered on item writer processes and products (draft and edited reading texts and items), and both deductive and inductive approaches to analysis were employed. The investigation was useful in identifying differences across the item writer groups and also between individuals within the groups. Both the experienced and non-experienced writers seemed to pass through similar stages when selecting texts and constructing items, though the researchers noted that those in the experienced group were able to articulate their experience more explicitly and in greater detail, and also generated higher-quality test material. The latter group also manifested a repertoire of gambits for efficiently exploiting source texts and task types, including the willingness to confidently edit texts for reasons of accessibility or cultural neutrality, reshaping them as necessary to meet the requirements of the test items. The expertise of the experienced test writing group appears to have been significantly influenced by their item writer training, by the item writer guidelines which guided their activity and by their collaborative approach during editing, which involved not only being able to freely critique each other's material but also make constructive proposals for improving another's work.

This study provides the field with some valuable insights into the processes of text selection, adaptation and item writing for a test of reading comprehension ability, as well as more generally into the nature of expertise. The differences observed between the experienced and non-experienced groups help to highlight the skills that are required for effective item-writing. Overall, the researchers report being favourably impressed by the conscientiousness and professionalism of the trained IELTS item writers that they interviewed and observed and by the quality of the texts and the items that they produced. This should be a source of encouragement for the IELTS test producers who have undertaken extensive investment over the years to develop rigorous policy and procedures for item writer selection, training and monitoring; it also strengthens the view that such expertise is collective in nature, rather than residing in individuals, and it supports the IELTS partners' decision to have IELTS item-writing teams based in different parts of the English-speaking world.

The researchers make some useful recommendations for refining and strengthening the current approach and procedures for IELTS test material production. One recommendation suggests making the principles and processes of test production more transparent and accessible to external stakeholders such as teachers and test preparation material publishers, particularly concerning the types of reading skill being targeted and the intention behind use of certain task-types. This could be done relatively easily by enhancing the public information already available on the IELTS website or through other communication channels, such as stakeholder seminars. Such an initiative would be consistent with the now well-established policy of the IELTS partners to communicate as much useful information as possible to test stakeholders and it would assist those who prepare candidates for IELTS in ensuring the match, in terms of construct validity, between test preparation activities and what candidates actually encounter in the test. Perhaps more important for the IELTS test developers is the recommendation offered in this study to extend and deepen the training of the item writing teams. The insights gained through this study have undoubtedly application in the initial training of new item writers when they join the team, helping them to understand how texts and items can be reshaped for the test and to develop their own skills in this regard. They also have relevance for more experienced item writers who may benefit from additional training and guidance on the detailed nature of the academic reading construct and how this is best operationalised through the IELTS Academic Reading module. The suggestion of using electronic tools for objective text analysis is certainly worthy of consideration by the IELTS item writing teams; software such as Compleat Lexical Tutor or Coh-Metrix could prove valuable practical tools for identifying or confirming key features of academic text genres and helping to ensure comparability across test versions. The point is also well made that 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 on beginning university students and the implications of these for assessment. Articulating the ability construct and approaches to operationalising it for assessment, especially across different proficiency levels and domains, is the underlying rationale for the series of skills-related volumes currently being published by Cambridge ESOL and Cambridge University Press in the *Studies in Language Testing* series. Khalifa and Weir's *Examining Reading* (2009) focuses on the assessment of second language reading ability, including the nature of reading at higher proficiency levels in academic and professional contexts. The hope is that volumes such as these will increasingly be used in practical ways to develop item writers' understanding of the constructs that are the focus of assessment, thus enabling them to more fully operationalise the academic reading construct in IELTS and other tests.

## 6 THE COGNITIVE PROCESSES OF TAKING IELTS ACADEMIC WRITING TASK 1

Like the previous two studies, the final study to appear in this volume, by **Guoxing Yu, Pauline Rea-Dickins** and **Richard Kiely**, maintains a focus on construct validity, this time in relation to the IELTS Academic Writing test rather than the Academic Reading module. Their research aimed to understand the cognitive processes of candidates taking Task 1 of the Academic Writing module with different graphic prompts at two different time points – before a short training session for completion of the task, and post such training. In particular, the study set out to explore the extent to which candidates' cognitive processes are affected by the use of different graphic prompts, by their graphic skills (level of graphicacy) and by their English writing abilities, as well as by a short training exercise. A grounded and multi-layered case study approach was employed to capture data on the cognitive processing of 18 intending IELTS candidates recruited from a large Chinese university. Subjects were asked to complete eight Academic Writing Task 1s under examination conditions and to 'think aloud' their processes while doing so. Four tasks were completed prior to training, and four after training. Data on subjects' English writing abilities and graphic skills were also collected, along with post-task interviews. The think-aloud protocols were analysed using the qualitative data analysis program winMAX and the cognitive

processing patterns that emerged were then interpreted with reference to the other data sources. This enabled the researchers to develop a cognitive processing model which guided their analyses in relation to four main areas for investigation: the nature of processing involved in graphic comprehension and its interpretation in written form; the role of graph familiarity on task performance; the correlation between writing performance on a graph-based task and on a topic-based argumentative essay; and the impact of training on task performance.

Although Task 2 of the IELTS Academic Writing module - an essay-based writing task - has been the subject of several joint-funded studies in recent years, Task 1 has received relatively little research attention by comparison so this study is particularly welcome. Task 1 is an 'integrated' task in that it requires test takers to describe information given to them in the form of a graph, chart, table or diagram and to present the description in their own words. This type of task raises interesting questions about the nature of test input presented in graphical mode (often non-verbal as a result) and the way that such visual input is processed by test takers in order to generate written output. This study therefore contributes to our understanding of the complex features associated with such task prompts and it builds upon two earlier joint-funded studies that investigated Task 1, by Mickan, Slater and Gibson and by O'Loughlin and Wigglesworth (see *IELTS Research Reports*, Volumes 3 and 4 respectively).

The insights resulting from this study are useful for the IELTS test producers because they have the potential to improve our understanding of how graphic prompts can function not only in the writing test but also in tests of listening, reading and speaking. The finding that different graphic prompts (e.g. line graphs, pie charts, tables of statistics, etc) apparently activate different forms of cognitive processing and can stimulate candidates to produce different sorts of written output, differentiated by their lexical and structural choices, is an important one to bear in mind since it touches upon questions of task and test comparability. There is a fine balance to be struck here on the part of test developers between standardising the task prompt in order to achieve task comparability from test version to test version, while at the same time sampling as widely as possible from the target language use domain and thus including a range of representative task types. The findings also raise the question of whether graphicacy should be regarded as an integral component of academic literacy, i.e. whether it is *construct-relevant*; is it reasonable to expect that L2 test takers should have some graphic familiarity so that they can cope with reading-into-writing tasks based on a range of graphic prompts, or do such tasks make unreasonable demands beyond the purely linguistic? As previously mentioned, there is growing interest now in understanding and describing the nature of the L2 learner's language and communication skills at higher proficiency levels, i.e. the C1 and C2 levels of the Common European Framework of Reference (CEFR), where the interface between language ability and academic literacy may be increasingly blurred.

The observation of a strong correlation between written performance on Academic Writing Task 1 and Academic Writing Task 2 is naturally encouraging for the test producers. Nonetheless, concern expressed over variable interpretation of the word 'describe' in the task rubric for Academic Writing Task 1, and how teachers therefore train students to approach this task in the test, highlights the importance of clarity of instruction in the actual test as well as in supporting materials for test takers, such as sample materials, exemplar written performances and accompanying examiner comments. As the researchers suggest, the working model of cognitive processes they propose could help to address this, offering a useful framework for test writers when designing Task 1s and ensuring they take account of the three inter-related stages involved: comprehending the non-graphically presented task instructions; comprehending the graphic information; and reproducing graph comprehension in written discourse in English.

## CONCLUSION

Once again, the contents of this latest volume of *IELTS Research Reports* illustrate the extent to which IELTS continues to offer researchers fruitful ground for investigations of different types. The six studies reflect a range of situated language assessment contexts and address diverse research questions. In all cases, the studies offer valuable reviews and summaries of the recent literature in their area of investigation and they demonstrate a variety of methodological approaches - all of which should be of interest and relevance to the wider research community. The methodological challenges and practical constraints encountered by the researchers, and the ways in which they resolved these, are particularly enlightening. The discussion and results from each study help to broaden and enhance our understanding of the IELTS test and each report highlights avenues for further investigation and exploration. Researchers who are considering applying for future grant funding from IELTS Australia and The British Council may like to note some of the suggestions made, perhaps drawing on the methodologies and instrumentation described in this volume to undertake similar studies in their own context.

Finally, the studies published in Volume 11 are a good example of the collaboration which can be achieved between the IELTS test producers and members of the wider applied linguistics, language education and language testing research communities as part of a joint commitment to sharing knowledge and understanding not only about the quality and usefulness of IELTS but also about the nature and business of language proficiency assessment more broadly.

Dr Lynda Taylor

**IELTS Editor-in-Chief – University of Cambridge ESOL Examinations**

*November 2010*

## 1 An impact study into the use of IELTS by professional associations and registration entities: Canada, the United Kingdom and Ireland

**Author**

**Glenys Merrifield**

GBM & Associates

Grant awarded Round 13, 2007

This study examines the history and rationale for selection of IELTS as a language benchmarking system for professional associations, and explores the main competitors to IELTS in global testing for professionals, in Canada, the UK and Ireland.

### **ABSTRACT**

IELTS was originally designed to assess English language skills for entry to courses of academic study. However, the use of IELTS for professionals seeking registration or membership of professional association has been growing over the past decade. This study was designed to examine language testing for professionals in three global markets: one relatively new market, Canada, and two traditional markets, the United Kingdom and Ireland.

The key objectives of the study were to examine the history and rationale for selection of IELTS as a language benchmarking system for professional associations, to explore the main competitors to IELTS in global testing for professionals, to determine the minimum band scores being used as criteria for registration or membership of professional associations, the decision-making policies and the perception of its fitness for the purpose. This was a qualitative study which included desk research of alternative examinations and assessment systems and the targeted associations, and identification of key contacts, followed by semi-structured interviews conducted face-to-face, by telephone and by email.

The study found that the range of assessments accepted by professional associations varied according to the country. In the UK, eight out of ten associations listed IELTS as the only acceptable assessment system of English language competency, and association staff tended to be well-informed about IELTS and familiar with the band scores.

In Canada, the range of testing systems used was more diverse, with some nationally-produced benchmarking systems also accepted. The majority of registering bodies, however, reported that most applicants presented with either the IELTS or the Educational Testing Service's (ETS) tests. The main challenge to IELTS' market share is the roll-out of iB TOEFL, which tests integrated skills and has largely replaced TOEFL, particularly in Canada where ETS is highly respected.

The UK and Ireland, as members of the EU, are bound by legislation which prevents them from asking for a language assessment from professionals of the EU. Accordingly, the responsibility for ensuring that individuals have the appropriate language skills to operate in an English-speaking context has

been shifted from the regulatory body to employers. This may be a potential growth market for IELTS in those countries.

## **AUTHOR BIODATA**

### **GLENYS MERRIFIELD**

Glenys Merrifield has been involved in international education, and in particular the TESOL sector, since 1975, initially in the United Kingdom and then in Australia. She has owned her own business in the international English language teaching sector, is an experienced manager and has been involved in lecturing and training in universities and private vocational education and training for a number of years. She holds postgraduate qualifications in management. From 1992 to 2004 she managed the National ELT Accreditation Scheme (NEAS), the national accreditation and quality monitoring service for Australia's international student sector and Adult Migrant English Program. She has a consultancy business and has conducted research and project management for the international education industry in Australia, New Zealand, the UK, the USA and Canada related to professional development, quality assurance in ELT, language assessment and other aspects of international education.



## CONTENTS

<b>1</b>	<b>Introduction .....</b>	<b>25</b>
<b>2</b>	<b>Objectives, Scope and Approach .....</b>	<b>26</b>
2.1	Objectives .....	26
2.2	Approach.....	26
2.3	Scope.....	27
2.3.1	Canada .....	27
2.3.2	The United Kingdom .....	27
2.3.3	Ireland.....	28
2.4	Responses.....	28
<b>3</b>	<b>National and International Tests.....</b>	<b>29</b>
3.1	International English Language Testing System (IELTS).....	29
3.2	Test of English as a Foreign Language /Test of Spoken English (TSE) .....	30
3.3	Cambridge ESOL Tests.....	31
3.4	Trinity College London Integrated Skills in English (ISE).....	31
3.5	Canadian English Language Benchmark Assessment for Nurses (CELBAN).....	32
3.6	Canadian Test of English for Scholars and Trainees (CanTEST).....	33
3.7	Michigan English Language Assessment Battery (MELAB) .....	33
3.8	Test of English for International Communication (TOEIC) .....	34
<b>4</b>	<b>Profiles of Professional Associations.....</b>	<b>34</b>
4.1	Profiles of Professional Associations – Canada.....	34
4.1.1	Nursing organisations .....	35
4.1.2	Pharmacy Organisations.....	37
4.1.3	Associations of Teachers .....	40
4.2	Profiles of Professional Associations - The United Kingdom .....	41
4.2.1	The British Acupuncture Council.....	41
4.2.2	Chartered Institute of Marketing .....	42
4.2.3	Faculty of Public Health .....	42
4.2.4	General Dental Council.....	43
4.2.5	General Medical Council and the Professional Linguistic Assessment Board .....	44
4.2.6	General Optical Council.....	44
4.2.7	Health Professions Council.....	44
4.2.8	Nursing and Midwifery Council (NMC).....	45
4.2.9	Royal College of Veterinary Surgeons.....	46
4.2.10	Royal Pharmaceutical Society of Great Britain .....	46
4.3	Profiles of Professional Associations – Ireland .....	47
4.3.1	Irish Nursing Board (An Bord Altranais).....	47
4.3.2	Irish Medical Council.....	47
4.3.3	Pre-Hospital Emergency Care Council .....	48

<b>5</b>	<b>Outcomes of the Study .....</b>	<b>48</b>
5.1	Summarising comments - Canada.....	48
5.1.1	Minimum IELTS band scores.....	49
5.1.2	Alternative language assessment systems accepted - Canada .....	50
5.2	Summarising comments - the United Kingdom.....	52
5.2.1	Minimum IELTS band scores.....	52
5.2.2	Alternative language assessment systems accepted - United Kingdom .....	53
5.3	Summarising comments - Ireland .....	54
5.3.1	Minimum IELTS band scores.....	55
5.3.2	Alternative language assessment systems accepted - Ireland.....	56
5.4	General outcomes.....	57
5.4.1	Main competitors to IELTS.....	57
5.4.2	Rationale for selection of IELTS.....	58
5.4.3	Review of minimum standards of IELTS .....	59
5.4.4	Consideration of IELTS as fit for the purpose .....	60
5.4.5	General feedback.....	60
5.4.6	Risk management and opportunities .....	61
<b>6</b>	<b>Conclusions.....</b>	<b>61</b>
	References.....	63
	Appendix 1: Professional associations and websites.....	64
	Appendix 2: Acknowledgements .....	66
	Appendix 3: Guiding questions for interviews conducted with professional associations .....	67
	Appendix 4: Acronyms .....	71

## 1 INTRODUCTION

Much has been written recently about the role English will play in post-modernist times. English language is becoming a lingua franca for many countries, primarily due to the use of global communication tools such as the internet. Academics write that the role of English language is changing to meet the demands of ever changing global communication strategies. Some argue that if English is spoken fluently and well and can be used in the electronic arena, then there is a reduced need for a first language in many contexts.

This would suggest that the role of language testing is likely to become increasingly important. Bodies such as medical boards, nurse registration authorities, veterinary registration authorities, pharmacy boards and other occupational regulatory bodies have recognised that English language communication in an English language employment culture is critical to the public good. It therefore serves the public interest to ensure that individuals seeking to move from non-English speaking professional environments to English-speaking work contexts are able to communicate effectively in English. Further, there is recognition that in the global context, skills and employment mobility are now critical and will become more so in the future as currently disadvantaged peoples seek to move to the developed world to practise their profession.

The International English Language Testing System (IELTS) was originally intended and designed as a set of international benchmarks to assess an individual's proficiency for academic study in English-speaking teaching contexts. This test, in addition to other international tests designed for a similar purpose, has now been adopted by a number of professional associations throughout the English-speaking world as a reliable means of assessing the quality of communication in English language for the professional workplace. The growing trend for IELTS to be adopted by users outside of academia, including governments, professional organisations and employers, may constitute a risk for the test owners if the assessment system cannot be validated for the purposes for which it is being used.

A previous study by this author (2008) focused on the use of IELTS to assess occupational communication skills in Australia, New Zealand and the USA. This study seeks to build on the former research parameters and outcomes, focusing this time on Canada, the UK and Ireland.

In Canada, the number of bodies registered as accepting IELTS has grown from five to ten in the last four years. In the UK and Ireland the number of registered users in this category has grown from nine in 2004 to thirteen in 2009, most of these in the UK.

The majority of associations, as was found in Australia, New Zealand and the USA (Merrifield 2008), represent the health professions and so the accuracy of the assessment tool constitutes high stakes to both the sector and to IELTS. If the number of users continues to grow at the rate it is currently growing, development of a body of knowledge of what is happening in the professional world is an important aspect of risk management for the managing partners. The IELTS partners need to understand the non-expert's perception of how good the "fit" is, the means by which entry level band scores are established, how often minimum levels are revisited and what support might be needed from the IELTS administration. This report addresses these issues.

The second main area of risk to IELTS is the competitive environment in which it operates. Prior to the global expansion of IELTS, the best language assessment tools available and accessible were traditionally the Test of English as a Foreign Language (TOEFL) and the Cambridge ESOL (English for Speakers of Other Languages) suite of tests. The TOEFL is designed and operated by Educational

Testing Service (ETS) and has been available internationally for many years, particularly in Asia and the Americas. Recently a new international testing system has been launched, the Pearson Test of English (PTE), which has introduced high level technology to address some of the very complex and difficult issues inherent in international language testing, like identity verification of candidates. The impact of this testing system has not yet been felt.

This report examines the major competitors to IELTS in the Canadian and European contexts, and attempts to identify attitudes to IELTS and its competitors as effective benchmarking tools.

## **2 OBJECTIVES, SCOPE AND APPROACH**

### **2.1 Objectives**

Key issues to be considered in this study were:

- Value of IELTS as key indicator
- Value of IELTS in comparison with its main competitors
- Appropriateness of the test for the purpose of association membership or registration
- Perceptions of IELTS by key stakeholders
- Variations in awareness and understanding of IELTS between users
- Initiatives to enhance recognition, understanding and reach of IELTS.

In accordance with the key issues, the objectives of this study were to:

- explore the history and rationale for selection of IELTS as a criterion for membership or entry to a profession
- identify minimum IELTS band scores being used as criteria for membership of professional associations or for registration
- identify the main competitors to IELTS and make a broad assessment of the risk they constitute to IELTS' market share
- assess the overall degree of understanding of the test by professional association staff and identify similarities and differences between countries studied
- identify perceptions of users on how well IELTS fits the purpose
- develop an understanding on whether or not IELTS should actively seek to extend the reach of the test in non-academic sectors
- make recommendations if appropriate on the development of educational and marketing programs to enhance understanding of the test, its outcomes and its appropriate uses.

### **2.2 Approach**

This was a qualitative study; that is, an inquiry process to seek and report the views of individuals in their natural setting, examining personal experience with a variety of approaches which may be historical, textual or interactional (Creswell 1994; Denzin & Lincoln 1994). It should be noted that outcomes are reported in the form of case studies which may exemplify attitudes and approaches but which could not be used as valid bases for generalisation.

The research consisted of:

- Desk research of alternative examinations and assessment systems
- Desk research of target organisations that are registered with IELTS
- Development of a list of key contacts to be invited to contribute to the research
- Telephone and electronic contact to establish a relationship, to introduce the aims and purpose of the project and where possible, to set up interviews
- Conduct of semi-structured face-to-face, telephone and electronic interviews with key contacts
- Follow-up by telephone and email.

## **2.3 Scope**

The study focused on Canada, the United Kingdom and Ireland. Professional associations and registering or licensure bodies which formally accepted IELTS formed the basis for initial contact. This meant that the organisations had officially included IELTS as the language assessment benchmarking system, or one of two or more systems, as part of their criteria for assessing language skills, and had established a formal relationship with IELTS which allowed them to access information and secure verification links for checking candidate outcomes.

The initial list of contacts was as follows:

### **2.3.1 Canada**

- Association of Registered Nurses of Newfoundland and Labrador
- British Columbia College of Teachers
- College of LPNs of British Columbia, Burnaby
- College of Pharmacists of British Columbia
- College of Registered Nurses of Manitoba
- National Association of Pharmacy Regulatory Authorities
- College of Nurses of Ontario
- Ontario College of Pharmacists
- Ontario College of Teachers
- College of Registered Nurses of British Columbia

### **2.3.2 The United Kingdom**

- British Acupuncture Council
- Chartered Institute of Marketing
- Faculty of Public Health
- General Dental Council
- General Medical Council
- General Optical Council
- Health Professions Council
- Nursing and Midwifery Council
- Professional Linguistic Assessment Board

- Royal College of Veterinary Surgeons
- Royal Pharmaceutical Society of Great Britain

### 2.3.3 Ireland

- Irish Nursing Board/ An Bord Altranais
- Irish Medical Council
- Pre-Hospital Emergency Care Council

## 2.4 Responses

A total of twenty-four professional associations were approached to seek their participation and response to the study. Of these, fourteen agreed to actively participate and the remainder were investigated primarily through desk research. The PLAB, which is an examination taken by medical professionals in the UK, was considered together with the General Medical Council.

There were some key constraints, however, in conducting a research project in which professional people were asked to participate in a study which focused on an area which was outside the bounds of their qualifications and expertise. In the UK in particular, contact with several professional associations for interview could not be made, and some indicated that they were unable to fully participate.

Possible reasons for this were as follows:

- Firstly, the twenty-first century workplace is characterised by staff who are “time poor” and who are dealing with a broad range of demands on their time. Some individuals may therefore have been reluctant to participate in a research project which they may have seen as not value-adding to their job or workplace.
- Secondly, the typical workplace for the associations being targeted in the project used electronic means to mass-communicate, making use of websites, email contact and virtual switchboards. Security firewalls on websites and email accounts, and limited options on the virtual switchboard, tended to be a barrier to communication, even in the case where a key contact had been identified.
- Thirdly, some staff in these organisations indicated that they were reluctant to be interviewed, even in an informal manner, on a subject about which they had limited knowledge.
- Finally, some of the smaller organisations were largely staffed by Council members who were still practising their profession and lending their expertise on a voluntary or consultancy basis, and were unavailable for what they saw as an extraneous matter.

In addition, some areas to be explored in the study depended upon historical records. Many of the professional associations and regulatory bodies approached had been operating for many years. Staff turnover and loss or lack of historical records meant that early policy and aspects of the history of the organisation (for example, the initial decision-making process on use of IELTS for language assessment) could not be determined.

Those who participated fully contributed valuable information and were generally pleased to have the opportunity to discuss their needs and gain some feedback about management of English language assessment by other similar organisations.

### 3 NATIONAL AND INTERNATIONAL TESTS

In the three countries on which this study focused, English language testing and assessment has been in place for many years. Professional mobility is increasing, encouraged by access and equity legislation and political policies which encourage migration of the educated and the skilled. This in turn is increasing the need for internationally accessible English language testing systems.

The associations examined in this study varied quite widely in the choice of national and international tests which they considered met their needs, generally stipulating somewhere between one and four language assessment options. By far, however, the most comprehensively used assessment systems were IELTS and the various forms of the TOEFL.

The full range of assessment systems accepted was as follows:

- International English Language Testing System (IELTS)
- The Test of English as a Foreign Language (TOEFL) in its various modes:
  - Paper-based TOEFL
  - Computer-based TOEFL
  - Internet-based TOEFL (iB TOEFL)
  - The Test of Written English (TWE) (combined with TOEFL)
  - The Test of Spoken English (TSE) (combined with TOEFL)
- Cambridge ESOL Tests
- Trinity College London Integrated Skills in English (ISE)
- The Test of English for International Communication (TOEIC)
- The CanTEST
- The Canadian English Language Benchmark Assessment for Nurses (CELBAN)
- The Michigan English Language Assessment Battery (MELAB)

#### 3.1 International English Language Testing System (IELTS)

The International English Language Testing System was originally established in the late 1980s and according to the IELTS website, at the time of writing was available in 500 locations in 120 countries throughout the world. It is jointly managed by IDP:IELTS Australia, the University of Cambridge ESOL Examinations and the British Council, and in mid-2009 IELTS candidature reached the one million mark.

It was designed initially with an Academic Module with specific purpose content intended to test language proficiency for entry to studies in English, and a General Training Module which was aimed primarily at vocational language use. The specific purpose content was revised in subsequent versions of the test.

Recognition of IELTS globally has grown to well over 6,000 institutions and professional bodies as well as government immigration authorities. Recognition of the test in the USA, a relatively new and potentially very large market, had grown to over 2,000 institutions and professional bodies by 2009.

The test consists of an extended reading and listening test, a writing test in the form of an essay or report, and a speaking test between candidate and examiner in the context of a real-time and real-life discussion. Candidates are able to sit the test as often as they wish and are able to access their results in a short timeframe of thirteen days. Institutional and professional users are able to access candidate results directly through an online verification service.

Some of the perceived advantages of IELTS are:

- A very short waiting time to sit the test, broad accessibility and frequent test dates.
- Productive skills assessed through authentic writing tasks and a real-life person-to-person interview, which is a strong advantage of the test for professional associations
- Security and integrity of the test
- Research to underpin the quality assurance processes
- Brief turnaround between sitting the test and accessing the outcomes.

### **3.2 Test of English as a Foreign Language (TOEFL, iB TOEFL)/Test of Spoken English (TSE)**

The Test of English as a Foreign Language is owned and operated by Educational Testing Service, which is based in the USA. The test has been in place as a paper-based test for many years and includes Listening, Structure, Reading and Writing modules. A speaking assessment was not included in the original paper-based test, and had to be taken as a separate (largely optional) process (the Test of Spoken English) at a limited number of access centres.

There are now revised TOEFL versions which give candidates a choice between the paper-based version (plus a Test of Spoken English (TSE)), a computer-based version and an internet-based version, which makes it more accessible on a global basis.

The internet-based TOEFL (iB TOEFL) was introduced in 2005 in response to concerns about the accessibility of the paper-based TOEFL and TSE, and was extended to include a speaking component. The format was revised to include an integrated tasks model, with all skills tested in one sitting. Originally better-known and more popular in the USA and Canada, the TOEFL is now offered and recognised in 130 countries by over 6,000 institutions and organisations, according to the ETS TOEFL paper, *Comparing TOEFL iBT and IELTS* (Rev: March 2009).

The integrated tasks include a Reading-Listening-Writing task, and the productive skills of Writing and Speaking are conducted on a computer interface, encrypted for security and scored by between three and six human raters. Scores are generally available two weeks after the test.

As with IELTS and other language assessment instruments, the iB TOEFL is designed to assess language proficiency within academic contexts. Since its introduction in 2005, ETS has sponsored research into such focus areas as validity and reliability of scores, comparison between human interactive scoring and E-rater scoring of components of the test, validity of the iB TOEFL for academic entry and the use of E-rater for assessing written texts. The use of iB TOEFL scores for entry to professions does not at this stage appear to have been a focus for research.

Since the introduction of the iB TOEFL, use of the assessment system has grown, and in 2009 a verification service for receiving institutions to gain direct online certification of candidate outcomes was launched. This is a service which IELTS has had in place for some time and is highly valued by users concerned about the security and reliability of certification.

Despite recent growth in candidature numbers for the various forms of the TOEFL, its popularity for language assessment is traditionally strongest in the Americas and parts of Asia. The TOEFL in its various forms is accepted by all Canadian professional associations involved in this study, one UK body and two Irish associations.



The paper-based version of the test is to be phased out, as is the computer-based version, as the necessary internet technology becomes available in developing countries. This will eradicate the necessity for candidates to engage in a two step process to complete their TOEFL assessment, which is time-consuming and could not always be completed in one day.

However, the previous study (Merrifield 2008) revealed reservations about the semi-direct approach to the Speaking test, in which candidates are responding to spoken stimuli rather than engaging in a face-to-face co-constructed dialogue, as is the case with other tests such as IELTS and the Cambridge tests. This remains an issue for professional associations in the countries involved in this study.

### **3.3 Cambridge ESOL Tests**

The Cambridge English language assessment examinations and tests are administered by Cambridge ESOL, a Department of Cambridge Assessment and part of Cambridge University. The suite of Cambridge tests is varied in its focus and has grown over a 150 year history to be one of the most highly respected testing systems in the world, with particular focus on the European context. IELTS is one of the assessment systems managed by Cambridge ESOL, in joint partnership with IDP: IELTS Australia and the British Council. However, a broad range of other tests offered by Cambridge ESOL have been used for many years by employers, in particular the Cambridge Proficiency in English (CPE) and Cambridge Advanced English (CAE) examinations.

Cambridge ESOL boasts a candidature of over three million in 2,500 centres covering 130 countries (Annual Review 2008), including IELTS, and has recently undergone review processes of some of the major tests to keep them current. The tests cover a broad demographic in terms of age as well as country of origin, and are widely recognised by governments and universities, particularly in Europe.

Computer-based versions of four of the tests were introduced in 2008. In addition, online marking for writing scripts for the Certificate in Advanced English (CAE) test was put in place to reduce the time taken for candidates to access their results. There is now a three week turn-around for results to be provided to candidates. These strategies will allow Cambridge ESOL to expand its examination dates, extend its reach and streamline application and marking processes.

The tests are supported by funded research and feedback, and there are strong quality assurance processes in place. Cambridge ESOL also makes a point of the fact that all its tests include a face-to-face interview, a feature which tends to be highly rated by professional bodies.

Despite the strong reputation of these tests, only one professional body based in Britain listed the CAE and the CPE among the accepted language assessment systems.

### **3.4 Trinity College London Integrated Skills in English (ISE)**

The Trinity College suite of tests has been in place for approximately 50 years. These examinations assess all four macroskills in an integrated and interactive testing situation, and are recognised internationally by academic institutions mainly based in the UK, North and South America and parts of Europe.

As part of the assessment, candidates prepare a portfolio of a varied sample of classroom work in addition to tasks completed under examination conditions. There are five levels of the ISE examinations. The portfolio forms the basis for the oral interview, whose broad aim is to conduct an unscripted conversation to replicate as far as possible real-life exchanges. While assessment levels are

frequently moderated, this introduces an element of subjectivity in assessing language competency, but is consistent with what many users of international language tests see as authentic communication.

Quality assurance procedures include occasional visits to test centres, monitoring a small sample of interviews, both live and recorded, re-scoring of approximately ten per cent of written tests and portfolios, an annual moderation session and professional development event for examiners, and statistical monitoring.

Despite a good network of test centres within the UK and some international availability of the test, only one respondent included the ISE III/IV as an alternative to IELTS. It is unlikely that it would be taken up by professional bodies generally because the process is lengthy and difficult for overseas professionals to access.

### **3.5 Canadian English Language Benchmark Assessment for Nurses (CELBAN)**

In Canada, control of language testing, professional standards and licensure has traditionally been devolved to the provinces, unlike the UK and Ireland, where there is national regulation.

However, an overarching body in Canada funded jointly by four of the larger provinces and the Department of Citizenship and Immigration Canada established the Centre for Canadian Language Benchmarks (CCLB) in the early 1990s, and the Canadian Language Benchmarks (CLB) were introduced in 1996. The purpose of developing the benchmarks was not to produce a language assessment test, but to devise a scheme which would act as a national standard for the description and measurement of second language proficiency. It was primarily intended to focus on education and the workplace for migrants settling in Canada, and to provide a practical curriculum guide. The benchmarks range from 1 to 12, one being the lowest level of communication and 12 being the highest.

The Canadian English Language Benchmark Assessment for Nurses (CELBAN) was an assessment system created as an extension of this project to assess specific purpose English language skills of overseas trained nurses seeking licensure in Canada.

The CELBAN was developed as an alternative to more generic English language assessment systems like IELTS and TOEFL, and evaluates the four macroskills of speaking and listening, reading and writing. The CELBAN is now administered by the Canadian English Language Assessment Services (CELAS) Centre, originally established in Winnipeg in 2004. The main disadvantage of the use of this test is that the CELAS Centre is the only location in which the test can be taken.

The CELBAN is recognised by ten Canadian provinces. In order to sit the tests, candidates must be qualified offshore and be in the process of applying for registration in Canada. Many of the provinces have their own registration body for nurses, and so a new assessment system like the CELBAN requires recognition by each of these bodies, and this recognition is growing. A primary feature of this test is that it has nursing-specific content, while the more global assessment systems do not, and this is seen by professional nursing associations in Canada as a major advantage.

However, the fact that it is only available within Canada means that overseas applicants for registration, many of whom complete a language assessment prior to applying for registration or recognition from offshore, need to look for an alternative. IELTS and TOEFL, with their international accessibility, are the primary alternative choices.

### **3.6 Canadian Test of English for Scholars and Trainees (CanTEST)**

The Canadian Test of English for Scholars and Trainees was developed by the Language Testing Service of the University of Ottawa as a tool for assessment of English language skills in both the academic and the professional arena. It is matched against Canadian Language Benchmarks 6 to 11 and is one of the approved tests for CLB. Assessment tests are offered at approved test centres in nine cities throughout Canada.

According to the University of Ottawa website, the CanTEST was designed for two main purposes: to assess whether a candidate is able to meet fluency requirements for entry to post-secondary education and training courses, and to test language proficiency for professional registration.

It is constructed of a listening test with approximately 40 multiple choice or short answer questions, reading tests which include a skimming and scanning exercise and a multiple choice and short answer section, a writing test of 45 minutes in the form of an essay and a fifteen minute face-to-face interview with two assessors.

The outcomes are a set of scores, called Band Levels, for each macroskill on a range of one to five. The Band Levels are described as Novice (1), Very Basic User (2), Limited User (3), Competent User (4), Very Good User (5) and Fluent User (5+). Candidates receive an “Unofficial Score Report” after the test and the official report is sent directly to their institution or association of choice.

Candidates may register separately for either the Writing or Speaking components, or all three skills tests for Listening, Reading and Writing. There is a six week waiting period to retake the Speaking test, and candidates may not retake the Reading and Listening test within three months of sitting.

The CanTEST is accepted by the national body for pharmacy regulation in Canada, and the two Canadian pharmacy associations, but like the CELBAN, it has limited accessibility and is little known outside of Canada.

### **3.7 Michigan English Language Assessment Battery (MELAB)**

The Michigan English Language Assessment Battery (MELAB) is an assessment system designed to assess language competency for entry to education programs and entry to professional training or work. It is administered by the English Language Institute of the University of Michigan in the USA, and is available at thirty-two locations in twenty provinces in the USA, and fourteen approved locations in six provinces in Canada.

The test consists of a composition, a Listening test and a written test, with an additional optional Speaking test available at some testing centres. Scores on the first three parts are averaged to reach a final score, and scores are issued within eight to ten weeks. Candidates receive an unofficial score and official scores are sent directly to the institution or professional body listed by the candidate in their registration form.

The MELAB can be re-taken every two months, up to six times per year, and results can take up to two months to be issued. It is accepted by eight Canadian associations.

This test is appropriate for North American residents, but is not available internationally. Its limited accessibility and long waiting period for results mean that it could not be considered a serious competitor to IELTS.

### **3.8 Test of English for International Communication (TOEIC)**

The TOEIC test was originally designed as an assessment of language skills for the workplace. It is run by ETS, as is the TOEFL, and is internet-based. Originally a Listening and Reading test, it has now had Speaking and Writing skills added to make it more comprehensive.

It is best recognised in the USA and Canada, where it has recently formed an new agreement with an associate, AMIDEAST, aimed at making it more competitive and extending its reach in the USA. AMIDEAST has formerly managed educational services including English language testing primarily in the Middle East and North Africa.

Three Canadian professional associations and one in the UK named the TOEIC as one of its recognised language assessment systems.

## **4 PROFILES OF PROFESSIONAL ASSOCIATIONS: CANADA, THE UNITED KINGDOM AND IRELAND**

### **4.1 Profiles of professional associations: Canada**

Canada is unique in that it has a dual language policy, English and French, and also has an indigenous population whose several languages are recognised in Canada. In such a multilanguage culture and environment, awareness of the significance of language competency for professionals is therefore deeply entrenched in regulation and legislation.

Professional organisations responsible for registration in Canada require evidence of an acceptable language fluency level in at least one of the two official languages, either English or French. Historically, Canada's political process has devolved control and monitoring of professional skills, training, standards and registration to the provinces. This means that for many occupations there is no standard national policy on minimum English language skills for professional registration. Politically, however, Canadian government legislation is moving towards reducing barriers and increasing mobility throughout Canada by establishing common standards for professions such as health care and education, and it is to be expected that this will include language assessment.

In this context, the larger provinces are collaborating so that standards become more uniform nationally. There are now organisations established in Canada which have a national mandate for regulation and accreditation. The Canadian Council for Accreditation of Pharmacy Programs (CCAPP), for example, is the Canadian accrediting body for courses of study in pharmacy, and only accredited courses with the CCAPP imprimatur or those accredited by the American equivalent (Accreditation Council for Pharmacy Education) are acceptable to provincial regulatory bodies. The Pharmacy Examining Board of Canada (PEBC) has a dual role in that it is a national examining board for the pharmacy profession and it also assesses the validity of pharmacy qualifications.

In the field of nursing, there has been a move in Canada to form a national organisation which will work towards developing a national set of standards and criteria for practising nurses. To date this has resulted in the CELBAN, which is outlined in 3.4 above.

It has been pointed out above that one of the critical factors in deciding whether or not to use a language assessment tool is accessibility. In Canada, there is accessibility to IELTS testing in the southern provinces and in the main towns. However, the northern parts of Canada, which are less densely populated, tend to have far more limited access to English language testing and the IELTS test

in particular. That being said, the vast majority of people seeking registration are overseas applicants for whom evidence of acceptable English language skills is a prerequisite to applying for registration. This means that most will need to present evidence of an appropriate language assessment prior to leaving their own country.

As is the case in the USA, TOEFL has a very strong position in Canada as a trusted and reliable testing system. The introduction of the iB TOEFL strengthened the position of ETS because it has made the test more accessible globally, and it includes testing of all four skills. However, there is still a perceived advantage in the IELTS system of assessment because of the work that has been done on test quality and security of outcomes.

At the time of the study, ten Canadian organisations were registered with IELTS. Of these, five were colleges of nurses, three were pharmacy professional associations and two were colleges of teachers.

#### **4.1.1 Nursing organisations**

The five nursing associations registered with IELTS were all established as provincial self-regulatory bodies under Canadian Health Professions legislation with a mandate to protect the public interest. Each association is responsible for establishing nursing standards of practice for the province, assessing fitness to practice, approving education and training courses for the industry (including assessing overseas qualifications for equivalence) and supporting registered nurses in meeting their standards. They are also responsible for conducting hearings on complaints against registered nurses. Since their authority is limited to the province in which they operate, this does not necessarily mean that there is consistency either in the standards of practice required for registration, or in the minimum standard of language competency.

With the exception of Quebec, provincial nursing associations require a language competency assessment as a prerequisite to sitting for the Canadian Practical Nurse Registration Examination (CPNRE). As the name suggests, success in this examination allows an international nurse to practise in Canada. However, policies vary on language assessments for applicants from English-speaking countries, or those who have been educated or have practised in an English-speaking environment. Despite variations in policy and standards, registration is generally, but not automatically, recognised from province to province.

Professional nursing associations which have an inbuilt process of review of standards have conducted research or gained expert advice on language needs for the profession. For those engaged in mental health, high level oral communication skills are critical. Nurses working in clinical environments need to be able to use and understand technical language, in addition to communicating directly with a very broad spectrum of clients, their families and advocates, and other health professionals. Writing is critical for completing accurate charts and records, and proficiency in reading is needed for deciphering doctors' orders, understanding care plans, dealing with medications and understanding a range of important medical records.

Although the range of acceptable tests is similar for each of the nursing regulatory bodies, there is significant variation in the minimum standards set in all assessment systems apart from the CELBAN (see Table 7 below).

##### **4.1.1.1 College of Registered Nurses of Manitoba**

The College of Registered Nurses of Manitoba (CRNM) began operating as a regulatory body almost a century ago and became a College in the year 2000. It is controlled by a Board of Directors consisting of both registered nurses and public representatives.

Evidence of English language competency is required from potential registrants from non-English language backgrounds and those qualified offshore. Applicants professionally trained within Canada are generally exempt. Numbers of applicants for nursing registration have reportedly increased significantly in recent years from approximately a hundred and fifty per year to over four hundred, which is attributed to an open door policy by Canadian immigration, coupled with active recruitment of healthcare workers, particularly in Manitoba. Ninety percent of applicants for registration are required to provide evidence of an acceptable English language fluency assessment.

The college officially accepts a suite of tests including TOEFL, the MELAB, the TOEIC and the CELBAN. As an alternative, it also accepts satisfactory completion of the Academic English Program for College and University Entrance (AEPUCE), a course offered by some universities in Winnipeg. While other evidence of fluency may be put forward by applicants, a formal assessment result is preferred by the college, and an IELTS Test Report tends to be the most commonly presented document.

It is not known how long IELTS has been accepted by the college. Acceptable band scores are an overall score of 6.5, with a minimum of 7.0 in Speaking. Applicants for registration may provide this evidence by presenting an amalgam of assessment outcomes achieved in successive tests over a two-year period.

#### 4.1.1.2 College of Registered Nurses of British Columbia

The College of Registered Nurses of British Columbia (CRNBC) is a self-regulatory body with a mandate under Canadian Health Professions legislation to protect the public through regulation of registered nurses. It tends to deal with large numbers of new applicants for nurse registration, with approximately 1,200 international applicants in 2008.

It operates under bylaws which describe “fitness to practice”, and evidence of language fluency is a prerequisite to registration.

Evidence of English language competency is required by all applicants for registration, regardless of their country of origin. The minimum overall score accepted in IELTS is 6.5, with a minimum of 7.0 in Speaking and no less than 6.0 in the other three macroskills. A candidate may combine two test outcomes if they have satisfactory scores in Speaking in one test and the other three skills in another. The association representatives were unclear on the reasons for this, saying that the minimum scores had been decided upon many years ago.

Also accepted are the various modes of the TOEFL (plus the TSE where applicable), the Canadian English Language Benchmark Assessment for Nurses and the Michigan English Language Assessment Battery. There is a preference for the CELBAN because it was specifically designed for nurses and has occupation-specific content, a perceived short-coming in each of the other accepted tests.

As with some other associations, there was the opportunity in the past for applicants to plead their case rather than sitting for international assessments in English language. However, this is no longer practised because it was considered to be open to subjective judgments, whereas an international assessment score is not.

The majority of applicants present an IELTS Test Report.

#### 4.1.1.3 Association of Registered Nurses of Newfoundland and Labrador

This organisation's statement of entry-level nurse competencies does not make reference to language skills, but the requirements for registration for international nurses specify a range of acceptable language assessment systems and outcomes.

The IELTS requirement is an overall score of 6.5, with a minimum of 7.0 in Speaking. There are no minimum standards specified for the Listening, Reading or Writing skills. The association also does not specify whether potential registrants should complete the General Training Module or the Academic Module, or whether both are acceptable.

Other acceptable tests are all versions of the TOEFL (plus the TSE where applicable), TOEIC (plus the TSE), the CELBAN and the MELAB.

#### 4.1.1.4 College of LPNs of British Columbia, Burnaby

This association represents and regulates the practices of Licensed Practical Nurses (LPNs), who are generally one year trained and responsible to Registered Nurses or Physicians.

Evidence of English language competency for overseas trained LPNs is required as a prerequisite to sitting for the Canadian Practical Nurse Registration Examination.

The acceptable testing systems are the Academic Module of the IELTS, the TOEFL (plus the TSE where applicable) and the CELBAN. IELTS minimum band scores are

- an overall outcome of 6.0, with a minimum of 7.0 in Speaking and 6.0 in each of the other skills.

#### 4.1.1.5 College of Nurses of Ontario

The College of Nurses of Ontario (CNO) is the registering body for both Registered Nurses and Registered Practical Nurses in the province, dealing with over 150,000 nurses.

Evidence of English language competency is part of the entry-to-practice requirements. Applicants for registration must gain at least IELTS 7.0 in Speaking, with a minimum of 6.5 in each of the other skills and an overall score of 6.5. The CNO does not appear to specify the Module applicants should take. The TOEFL, TOEIC, CELBAN and MELAB are also acceptable.

### 4.1.2 Pharmacy organisations

Canada's National Association of Pharmacy Regulatory Authorities (NAPRA) was established in 1995 as a voluntary organisation with the overarching role of bringing together provincial authorities. Its main aim was to establish a national voice for the pharmacy profession in Canada, and to facilitate development of best regulatory practice. It has thirteen pharmacy regulatory authority members and is governed by a board of twenty-four Directors and administered by an Executive Committee.

In 2000, a mutual recognition agreement for the pharmacy profession in Canada was signed by nine provincial regulatory authorities in a move to adopt common licensing requirements for pharmacists new to Canada. One of the major aims in forming this agreement was to facilitate free movement of pharmacists within Canada, and to remove potentially discriminatory requirements between provinces.

With its principal mandate the protection of the public, NAPRA established an entry-level set of professional competencies for pharmacists, and a National Model Licensing Program which included standards for language competency was published in 2006. This stated that applicants

who were educated in Canada or the USA in pharmacy may not be required to provide evidence of English language competency, and where evidence of English language competency was required, documentation should reflect a language assessment conducted in the two years leading up to the application for registration.

As a guide to provincial regulatory authorities, NAPRA's policy is to accept iB TOEFL, paper-based and computer-based TOEFL, MELAB, IELTS, and the CanTEST. Guidelines on acceptable levels of achievement were established as set out in Table 1 below.

While it was not possible to interview a representative of NAPRA, it could be assumed that the equivalencies for each of the language assessment systems accepted for benchmarking by its members may either be historically based or perhaps the result of research.

In the case of IELTS, a minimum of 6.0 in each of the macroskills is required; if a band score lower than 7.0 is achieved in one or more macroskills then this would need to be compensated for by scores of 7.5 or above in one or more macroskills. An overall band score of 7.0 is deemed to be sufficiently high to indicate that a pharmacist is able to communicate on a variety of levels and with a broad range of interlocutors.

This national approach has guided the provincial authorities in the standards of English language required.

#### 4.1.2.1 Ontario College of Pharmacists

The Ontario College of Pharmacists (OCP) is a self-regulatory body funded by public membership fees, and its role is to register pharmacists as a prerequisite to practising. Members include pharmacists, pharmacy students and interns, and membership will shortly be extended to pharmacy technicians.

The role of the college is to maintain quality standards, and it operates as a gateway to pharmacy practice through setting and monitoring professional standards. It operates as a complaints review body. It also works in cooperation with education providers of CCAPP or ACPE accredited courses, and sponsors the conduct of accredited communication skills programs.

As in other states, the Ontario College of Pharmacists operates under the NAPRA framework. A satisfactory result in IELTS or another accepted assessment system is a pre-registration requirement.

The college considers language proficiency to be a vital skill for registered pharmacists, who are required to communicate with patients with a broad range of demographics as well as fellow professionals. It has been using IELTS for a period of five years, and took on the NAPRA standards in 2006 following training and advice from a leading academic at the University of Toronto.

It is interesting to note that the Ontario College of Pharmacists departs from the NAPRA standards in the respect that it allows applicants for registration who have an IELTS band score of 5.5 in one or more of the macroskills to take an option of mounting an argument for its acceptability. Most present a test result, however.

According to college representatives, feedback from candidates indicates that many have a preference for IELTS, as they perceive it to be easier than alternative assessment systems. While some candidates question the relevance of IELTS because it does not include professional content, the college supports a more general context, taking the view that professional content is addressed in the training programs, and it is more important in language assessment to assess fluency and acculturation.



<b>TEST</b>		<b>MINIMUM SCORE</b>
iB TOEFL Internet-based	Speaking	27
	Writing	25
	Reading	Unspecified
	Listening	Unspecified
	Total	97
TOEFL Computer-based	Writing	26
	Test of Spoken English (TSE)	50
	Total	237
TOEFL Paper-based	Test of Written English (TWE)	5
	Test of Spoken English (TSE)	50
	Total	580
IELTS Academic Module	Speaking	6.0
	Writing	6.0
	Reading	6.0
	Listening	6.0
	Overall band	7.0
MELAB	Written composition	82
	Listening comprehension	Unspecified
	Grammar, cloze, reading, vocabulary, comprehension	Unspecified
	Oral component	3+
	Total	85
CanTEST	Speaking	4.5
	Writing	4.5
	Reading	4.5
	Listening	4.5

**Table 1. Minimum standards of English language competency for the Canadian pharmacy profession as set down by the National Association of Pharmacy Regulatory Authorities (NAPRA)**

#### 4.1.2.2 College of Pharmacists of British Columbia

The College of Pharmacists of British Columbia (CPBC) has a mandate to protect the public of the province by setting standards of practice for pharmacists and for the operation of pharmacies. International applicants for registration are required to successfully complete a National Qualifying Examination and a Jurisprudence Examination, and provide evidence of language competency whether or not they have studied in an English speaking country.

Evidence of English language competency is consistent with the criteria set down by NAPRA. In order to develop a good understanding of band score levels and to arrive at levels which were appropriate for pharmacists, the organisation sought advice from an experienced academic who had a background in working with IELTS.

The college does not have a role in monitoring registrants after they commence practising. However, it does act as a grievance resolution body in cases where formal complaints are made. Some of these relate directly to communication skills, in particular, that communicative competency is not regarded as being sufficiently high to deal with the demands of the profession on a day-to-day basis.

The organisation did not consider that there was a need for direct communication with or support from IELTS administration.

### **4.1.3 Associations of Teachers**

#### **4.1.3.1 British Columbia College of Teachers**

The British Columbia College of Teachers (BCCT) is a self-regulatory body funded by member fees. The college is responsible for certification of individual teachers in public schools and some private schools prior to teaching in Canada. It also approves teacher education programs.

When assessing eligibility for teacher certification the college examines academic background, professional competency and evidence of professional practice within the previous ten years. While there is provincial autonomy in education, in the interests of labour mobility the provincial Colleges of teachers are seeking to move to a common standard. Language skills are considered to be crucial for maintaining professional standards in the classroom.

Evidence of English language competency must be provided with either an IELTS Test Report or the outcomes from any of the three versions of the TOEFL. Currently, the college accepts an IELTS standard of 6.5 in the Academic Module with no less than 6.0 in the Reading and Listening components and a minimum of 7.0 in Writing and Speaking. The macro skills of speaking and writing were considered more important than the receptive skills because of the need for teachers to speak professionally, to confer with parents, to communicate with children and to write reports. These scores must be gained within one sitting of the IELTS. Acceptable language assessment is required for all applicants for certification whose education was not conducted in English. At the time of the study, between a hundred and fifty and two hundred candidates annually were required to present a language assessment report.

The band scores required were reviewed upwards in 2007 to bring them into line with other similar organisations. IELTS test centres are available in British Columbia at the Simon Fraser University and the Victoria University. TOEFL, on the other hand, which is also accepted by the BC College of Teachers, is available at fifty centres throughout British Columbia. However, many applicants for registration gain evidence of language competency prior to moving to Canada, so that the availability of IELTS internationally works in its favour.

The BC College of Teachers has not requested or received specific support from IELTS administration apart from a package of information provided in 2003 when the original minimum standards were set. It was considered that occasional updates would be helpful.

There were no immediate plans to reassess the testing systems which were accepted. Any feedback about the testing system or any complaints against a particular teacher would be considered by the employer rather than the College of Teachers. The only perceived drawback of IELTS was the lack of subject specific content for those teaching in a particular discipline.

#### 4.1.3.2 Ontario College of Teachers

The Ontario College of Teachers (OCT) is responsible for regulation of teachers in publicly funded schools in the province, and it sets standards of practice and conduct, certifies teachers and accredits teacher education programs.

Evidence of English language competency must be provided prior to certification for all applicants educated in non-English language contexts. This College has set the IELTS levels for individual macro skills a little higher than the British Columbia College of Teachers. The overall minimum standard required in IELTS Academic Module is 7.0, with a minimum of 6.5 in Reading and Listening and a minimum of 7.0 in Speaking and Writing. The iB TOEFL is also acceptable.

## 4.2 Profiles of professional associations: The United Kingdom

Professional associations and registration bodies in the United Kingdom which require an IELTS score from examination candidates, potential members of the organisation or professionals seeking registration are generally within the health care professions. For these professions, the UK has a relatively complex and diverse network of regulators, and all have responsibility for setting standards and regulation of those standards. Membership or registration is individual, and some associations have both a representative role for members and an arbitration role in that they investigate complaints against practitioners, potentially resulting in sanctions which may include suspension or striking-off.

Some association representatives put the view that with its National Health Service, the UK is an attractive destination for migrants. Since the establishment of the European Union, increasing numbers of health professionals have been entering the United Kingdom from Europe and further afield in search of a better lifestyle and work in their chosen profession. High level communication skills are considered crucial if the UK is to protect its reputation as a quality health provider.

However, it appears to be a source of some frustration to regulators that under European Law, professionals from within the European Economic Area (EEA) entering Britain are exempt from having to provide evidence of English language competency on the grounds that it may be discriminatory. This was considered by many of the professional associations involved in this study to be a high risk policy for the UK, given the vital importance of effective communication when serving public health interests.

There was an indication from some organisations that pressure may be brought to bear on European governments to review this aspect of the legislation. Some suggested that if changes should be made, a possible outcome may be to require professional associations to accept a broader range of assessment systems rather than focusing on IELTS as the only acceptable system, as is the practice for many associations.

Under current law, however, the only solution to this issue, according to association representatives, was to place the responsibility for assessment of language skills on employers. Professional associations and registration bodies would then only deal with issues arising from a poor standard of communication in English in cases where a formal complaint was made to the association as the body responsible for grievance resolution.

### 4.2.1 The British Acupuncture Council

The British Acupuncture Council (BAcC) is a voluntary organisation which is self-regulating and acts as the guiding body for acupuncture practice in the UK. It is a representative body for members, but

more importantly, is perceived as having a role as a protector of the public. In the absence of statutory regulation of acupuncture in the UK the British Acupuncture Council has established standards for acupuncturists. When statutory regulation is introduced in 2012 it is expected that an English language level will be established as a minimum requirement for all international candidates from outside the European Economic Area.

The perceived language needs of acupuncturists include proficiency in communicative skills for the purpose of effective record-keeping, providing advice and explanations to patients and understanding a patient's health and medical history. While there is no current policy on whether the minimum standard of IELTS should be gained in one sitting of the test, the perception of the BAcC is that this would better reflect the way skills are used in acupuncture practice.

Evidence of English language competency in the form of an IELTS outcome is a requirement for entry to membership of the British Acupuncture Council. The current understanding is that acupuncturists should have an overall minimum of IELTS 6.5. The module was not specified and so it is assumed that this applies to either the General Training Module or the Academic Module. No other formal language assessment systems are currently accepted.

#### **4.2.2 Chartered Institute of Marketing**

The Chartered Institute of Marketing (CIM) is a global organisation for professional marketers with a role in training, development and representation of the profession of marketing. It also accredits professional training courses and sets standards in the marketing industry within the UK. Individuals draw benefits from membership of the Institute including career and professional development, job search and marketing resources.

The establishment of minimum language assessment standards for this organisation appears to relate to entry to courses of training; that is, the purpose of the use of IELTS in this case is to assess the individual's ability to deal with academic study rather than performance of a profession.

Evidence of English language competency for the CIM is required in one of two forms. An overall IELTS level of 6.5 is required for entry to courses, but the criteria do not appear to draw a distinction between the General Training Module and the Academic Module of IELTS. The only other language assessment system accepted is the Trinity ISE III/IV. Trinity College London has equated this to a range of 6.5 to 7.5 in IELTS (Papageorgiou 2007). The minimum standards of IELTS were established after some initial web-based research.

#### **4.2.3 Faculty of Public Health**

The Faculty of Public Health (FPH) is a charitable organisation and is responsible for setting standards for medical specialists in the public health arena. It is a joint faculty of the Royal Colleges of Physicians of London, Edinburgh and Glasgow, and has a membership of over 3000.

The criteria set for applicants wishing to take specialist training in public health include an appropriate degree and at least three years of experience, in addition to evidence of English language proficiency.

The perceived language skills required by Medical Specialists are included in detailed criteria for applicants wishing to train as specialists, indicating that a good deal of research and analysis has been put into determining the language skills required to be a competent medical specialist. Evidence of an understanding of the principles of scientific research and evidence-based practice, the ability to communicate effectively with a variety of individuals and groups, particularly in written and spoken

English, and the ability to operate functionally in such areas as building rapport, persuasion and negotiation are required.

The association views the productive skills of written and spoken English to be particularly important. The requirement is that they should achieve in a single sitting of IELTS an overall score of 7.0, with a minimum of 7.0 in each of the four macroskills. This does not apply, however, to applicants who have gained their undergraduate degree in an English-medium university.

As an alternative to an IELTS Test Report, there is provision for making representations to the Faculty if an applicant believes that he or she can provide evidence of English language fluency by alternative methods. Most applicants, however, present with an IELTS test outcome.

No other language assessment system is accepted.

#### **4.2.4 General Dental Council**

The General Dental Council (GDC) is responsible for the registration of all dental professionals who wish to work in the UK, including dentists, dental hygienists, dental therapists, dental nurses, dental technicians and orthodontic therapists. It sets standards for the dentistry profession and assesses individuals' fitness to practise.

The Council offers an Overseas Registration Examination which tests the clinical skills and knowledge of any dental professionals who are from outside the European Economic Area, and whose qualifications have been gained from a country where there is no existing recognition of equivalency.

All dentists from outside the EEA who wish to apply for the overseas registration examination are required to provide an IELTS Test Report Form as evidence that they have achieved the minimum standard of communicative competency as set down by the General Dental Council. There has been a steady increase in numbers applying for registration in recent years.

The perceived language skills required by dental workers include the proficiency to interact with the public at all levels, from children and families to those from multicultural backgrounds. Proficiency in clinical and technical language is also critical, although this is not tested by an assessment system such as IELTS.

The GDC has a quality assurance process in place for approval of pathways into dental professions. Minimum standards of English language were most recently established by a Registration Committee after consultation with IELTS test professionals. These standards were last reviewed in 2000 for dentists, and 2003 for other dental professions, and a further review is due to be completed in 2010.

The minimum standards required vary according to the specific professional area. For a dentist, the minimum requirement is an overall band score of 7.0 in the Academic Module, with no score lower than 6.5 in any macroskill. The assessment must have been taken in a single test within the previous two years.

Dental nurses, on the other hand, require an overall band score of 6.0 with a minimum of 6.0 in Speaking and Writing and 5.5 in Reading and Listening, again emphasising the greater importance of the productive skills. IELTS is the only English language assessment system accepted by the council, although this may be subject to change if council policies or European legislation prompt a broadening of current criteria.

#### **4.2.5 General Medical Council and the Professional Linguistic Assessment Board**

The General Medical Council (GMC) is the regulatory body responsible for setting standards for medical practice in the UK, and it controls the register of doctors and specialists approved to practise in the UK. The council also plays a role in resolving complaints by holding disciplinary hearings.

Although the Professional Linguistic Assessment Board (PLAB) is registered with IELTS as a professional body, it is in fact the body which oversees assessment of medical graduates. The PLAB examination is a means by which international medical graduates demonstrate that they are appropriately qualified and experienced to practise as doctors in the UK. Applicants for registration complete a PLAB examination, which includes a written test and a set of clinical scenarios, provided they have met the minimum English language requirements. Many will also have had twelve months postgraduate clinical experience.

Evidence of English language competency must be provided in the form of an IELTS Test Report, or by providing evidence that they have been educated or have practised in a country where English is the first language.

However, interestingly, the alternatives to IELTS are not valid if an applicant has sat for the IELTS test and failed to achieve the minimum band scores required by the Council. This effectively blocks applicants from taking alternative pathways to achieve registration, and would appear to be a strong vote of confidence in the IELTS testing system and the reliability of test outcomes.

The minimum requirement is an overall score of 7.0 with minimum of 7.0 in Speaking and 6.0 in the other three macroskills. These outcomes must be obtained in a single sitting of the test.

#### **4.2.6 General Optical Council**

The General Optical Council (GOC) is the registration body for optometrists, dispensing opticians, student opticians and optical businesses and manages a registration of over 23,000 professionals. Overseas applicants for registration from outside the EEA are referred to the College of Optometrists, which is responsible for assessing the minimum criteria for registration.

A satisfactory IELTS Test Report allows potential registrants to sit for an entrance examination which is the gateway to applying for registration to practise in the UK.

The College of Optometrists requires that overseas optometrists have a minimum band score of 7.0 overall, with a minimum of 7.0 in Speaking and scores of no less than 6.0 in the other macroskills. It does not indicate whether this refers to the General Training or the Academic Module. No other language assessment systems are accepted.

#### **4.2.7 Health Professions Council**

The Health Professions Council (HPC) describes itself as a regulator which was set up to protect the public by establishing a set of standards for training, professional skills and behaviour for health professionals other than doctors, nurses and pharmacists. It boasts a registration of over 200,000 professionals from fourteen professions, including for example Chiropodists, Dietitians, Paramedics and Practising Psychologists. It also has a role in hearing complaints and conducting public hearings to assess fitness to practise.

An English language proficiency assessment is required for all applicants for registration for whom English is not their first language.

The perceived language skills required by health professionals are encapsulated in documents setting out competency standards for each of the professions, produced when a review of standards took place in 2005. Analysis of communicative competency expected of a health professional was undertaken at that time and was included under published standards of practice.

Standards of proficiency for health professionals include the ability to demonstrate effective and appropriate skills for providing information and advice, for professional instruction and for provision of a professional opinion to a range of interlocutors including patients, their families and carers, and colleagues. Practitioner communication skills also require effective non-verbal communication, and the awareness that effective communication may be affected by a number of factors including age, culture, gender and socio-economic status.

Evidence of English language competency must be provided in the form of an IELTS Test Report or a satisfactory score in a number of alternative tests.

The minimum band scores for Speech and Language Therapists is an overall score of 8.0 in the Academic Module with no score below 7.5. This is particularly high in comparison with other regulatory bodies, and the reason is that language communication is a core competency for speech and language therapists. The equivalent in Cambridge ESOL examinations is the CPE, and a score of 118/120 overall in the iB TOEFL.

For other practitioners, the minimum band scores are an overall 7.0, with no score less than 6.5. This is perceived as equivalent to the Cambridge Advanced English test or 100 in iB TOEFL.

#### **4.2.8 Nursing and Midwifery Council (NMC)**

The Nursing and Midwifery Council has as its primary aim safeguarding the health and well-being of the public. It does this through registration, setting standards of education, training and conduct, and ensuring that practitioners in nursing and midwifery maintain their skills and knowledge. It also has a role in investigating complaints of misconduct, which may come from employers, from other nurses or midwives or from the public.

Nurse and midwife training programs within the UK must be approved by the Council. In cases where overseas nurses and midwives are insufficiently qualified, they may be given access to a pre-registration program of two years. Generally, the minimum training program is three years.

This organisation is dealing with very large numbers of nurses and midwives, having a database of approximately 630,000, and requiring practitioners to renew their registration every three years. However, the number of overseas nurses and midwives entering the UK and seeking work in the field has been dropping in the past five years, from 10,000 in 2003 to approximately five hundred in the first half of 2009. This was attributed at least in part to an initial influx at the beginning of the twenty-first century from EEA countries when border restrictions were relaxed.

All overseas nurses apart from those from the European Economic Area are required to do an English language assessment, including those whose first language is English (e.g. Australians, Canadians, etc). Decisions on acceptable levels and review of those levels are made by the NMC Council of fourteen members.

Research was conducted by the NMC into the language needs of nurses and midwives in order to establish minimum standards of English language communication. They need to be able to communicate effectively with patients of all ages and sociocultural backgrounds, as well as foreign nationals. Practitioners also need mastery of medical terms and jargon, and a good understanding of Latin-based technical language.

After taking into account the outcomes of research into language needs and the standards of practice, the NMC established a minimum standard for IELTS of 7.0 overall in the Academic Module, with a minimum of 7.0 in each of the four macroskills. These levels represent an increase from a minimum of 6.5 overall and a minimum in each of the skills of 5.5 in either the Academic Module or the General Training Module. The revision was a result of feedback from trainers and employers, and the new standards are considered to be much more appropriate.

#### **4.2.9 Royal College of Veterinary Surgeons**

The Royal College of Veterinary Surgeons sets and monitors standards for veterinary surgeons in the UK as a statutory regulator. It is also a Royal College for advisory services and scholarships, and a charitable trust with an educational charter.

In order to be registered, veterinary surgeons qualified outside the UK and EEA in non-approved courses are required to pass a Statutory Examination for Membership prior to gaining the right to practise in the UK.

The perceived language skills required by Veterinary Surgeons are included in the core standards for the profession. In relation to language and communication, a core standard, according to the college's website, states that a practice "must have an effective policy for communication with clients", and must demonstrate the basic communicative competency to enable them to communicate effectively with a range of people, including clients, the public, colleagues and regulatory authorities. While the policy does not go into detail of what this may involve, it does acknowledge the need for functional skills such as listening effectively and responding sympathetically to clients and others, and a need to be able to select appropriate language to suit the audience and the context in which they are communicating.

Prior to taking the Statutory Examination for Membership, applicants for registration are required to provide an IELTS assessment of an overall band score of at least 7.0 in the Academic Module. No other language assessment systems are currently accepted.

#### **4.2.10 Royal Pharmaceutical Society of Great Britain**

The Royal Pharmaceutical Society of Great Britain was originally established to lead and regulate the pharmacy profession, to maintain a register of practising pharmacists and to act as the representative body for individual pharmacists. Following a government review of the organisation and its aims in 2007, however, it has been working on an eventual division of the dual roles of regulation and representation.

All non-EEA international pharmacists must provide a satisfactory English language assessment in addition to an appropriate qualification.

The society estimates that approximately one hundred and eighty candidates per year are required to provide evidence of English language competency out of a total of four to five hundred (including those from the EEA), and the majority of IELTS test takers sit the test in their own country prior to applying to the society for registration.

There are currently no plans to extend the range of language assessment systems which are acceptable, although this may be necessary if European legislation or government policy is reviewed. IELTS is considered by the Society to suit their purposes very well.

The purpose of the organisation's use of IELTS is twofold: to assess an individual's English language level for registration, and to assess language skills as a prerequisite to sitting for an examination to assess capability to practise.



The Society has been using IELTS as its English language benchmark since 2002. The minimum acceptable assessment is 7.0 in the Academic Module in each of the four macroskills. These levels have been revised from an overall level of 7.0 because the Society considered proficiency in all four skills to be critical for a practising pharmacist.

No other English language tests are currently accepted.

### **4.3 Profiles of Professional Associations – Ireland**

Irish regulatory bodies making use of IELTS are exclusively related to the health professions. Like the UK, Ireland is a member of the European Union and is bound by the same legislation barring Irish associations from making an English language proficiency assessment mandatory, and so evidence of English language proficiency is only sought from professionals wishing to practise from non-EU countries.

This has meant that responsibility for language assessment for EEA nationals has been devolved to the employer.

#### **4.3.1 Irish Nursing Board (An Bord Altranais)**

The Irish Nursing Board was established under Irish legislation in 1950 to maintain a register of nurses, midwives and individuals in ancillary health occupations. It is also responsible for maintaining a Code of Practice and Ethics and to assess fitness to practise of individuals.

There has been a drop in the number of overseas applicants registered in the last five years, according to Register Statistics 2008 published on the Irish Nursing Board website, with new registrations in 2008 of non-EU applicants numbering just over five hundred out of a total of 2,199, approximately half the number registered in 2004. This is consistent with the trend reported in the UK.

In addition to IELTS, the Board also accepts the various modes of TOEFL (combined with the TWE and TSE where necessary), or the iB TOEFL (minimum of 88), as alternative assessment systems.

The Board has set the minimum IELTS standard required for nurse and midwife registration as an overall 7.0 in the Academic Module, with a minimum in Writing and Speaking of 7.0 and in Reading and Listening of 6.5, thus giving greater weight to the productive skills than the receptive skills for nurses.

#### **4.3.2 Irish Medical Council**

The Irish Medical Council (IMC) was established under Irish legislation in 1978 with its main purpose being the protection of the interests of the public in health matters. It is responsible for quality assuring graduate programs, for setting and maintaining standards of practice and assessing fitness to practise for both doctors and specialists through an examination.

Doctors from non-English speaking countries are required to provide an IELTS Test Report prior to applying for registration. According to statistics on the IMC website, as of January 2009 there were almost 700 temporary registrations (that is, doctors under supervision) for the previous year, the vast majority coming from the Sudan, Nigeria, Pakistan and India. Approximately three to four hundred of these were required to provide language competency certification.

The perceived language skills required by doctors and specialists involve high level communication on technical matters, and effective communication with patients with varying levels of understanding. They also include skills in advising patients of diagnoses and breaking bad news in language the patients are able to understand.

The Medical Council conducted a review of standards for registration of doctors in the lead-up to June 2009, and language skills formed part of that review with a Working Group researching best practice.

Prior to June 2009, the IELTS band scores required were an overall score of 7.0 in the Academic Module, with a minimum of 7.0 in Speaking and 6.0 in the other three macroskills.

In revised standards which were implemented in mid-2009, however, these minimum scores were increased to an overall score of 7.5 and a minimum of 7.0 in all four macroskills. These minimum standards are required in one sitting of the IELTS, and were revised upwards to bring them into line with international best practice.

### **4.3.3 Pre-Hospital Emergency Care Council**

The Pre-Hospital Emergency Care Council (PHECC) is responsible for emergency services in Ireland such as ambulance services and paramedics. It is an independent statutory agency which sets standards of practice for emergency care, and has a role as an accreditation body for education and training. It also assesses fitness to practise and conducts hearings on formal complaints about practitioners.

The Council includes in its Code of Conduct and Ethics the statement that in the interests of safe practice, a registered practitioner should be competent in communicating effectively with patients and families, and be particularly concise in seeking informed consent for his or her actions.

The PHECC recommends that employees should have a minimum of 6.5 IELTS; whether this applies to the General Training Module or the Academic Module, or to both, was not specified. No other language assessment systems are recommended by the Council, but the TOEFL test may be acceptable on application. Threshold levels for the TOEFL are not published, but are available on application.

Because of EU legislation, the need for language assessment by the PHECC has been reduced considerably, the responsibility having been delegated to employers. However, there is the possibility that a registered practitioner may be subject to a “fitness to practise” enquiry by PHECC if language skills are lacking. Responsibility for the communication skills of employees represents high stakes to employers. If a complaint were to be made about a practitioner’s fitness to practise which related to English language skills, this would be considered by the PHECC and the provider would be held responsible, possibly resulting in withdrawal of approval to practice, which would prevent the employer from operating. This has reportedly not occurred to date.

## **5 OUTCOMES OF THE STUDY**

### **5.1 Summarising comments - Canada**

Eight of the ten associations involved this study were public health-related. The Canadian associations were the most diverse of the three countries in terms of the language assessment systems that were accepted. Because of the provincial nature of regulation in Canada, there was also greater variation between associations in the minimum levels of IELTS which were accepted.

IELTS is the predominant testing system for professional associations in Canada, despite Canada’s long history of using Educational Testing Service tests. IELTS has built a market advantage by ensuring that it maintains a broad testing network, prompt reporting of results, tight test security and reliable outcomes.

	<i>General Training Module</i>					<i>Academic Module</i>				
	Overall	Reading	Writing	Speaking	Listening	Overall	Reading	Writing	Speaking	Listening
College of Registered Nurses of Manitoba	N/A	N/A	N/A	N/A	N/A	6.5	No min specified	No min specified	7.0	No min specified
College of Registered Nurses of British Columbia	N/A	N/A	N/A	N/A	N/A	6.5	6.0	6.0	7.0	6.0
Association of Registered Nurses of Newfoundland and Labrador	Module is not specified					Module is not specified 6.5	No min specified	No min specified	7.0	No min specified
College of LPNs of British Columbia, Burnaby	N/A	N/A	N/A	N/A	N/A	6.5	6.0	6.0	7.0	6.0
College of Nurses of Ontario	Module is not specified					Module is not specified 6.5	6.5	6.5	7.0	6.5
National Association of Pharmacy Regulatory Authorities	N/A	N/A	N/A	N/A	N/A	7.0	6.0	6.0	6.0	6.0
College of Pharmacists of British Columbia	N/A	N/A	N/A	N/A	N/A	7.0	6.0	6.0	6.0	6.0
Ontario College of Pharmacists	N/A	N/A	N/A	N/A	N/A	7.0	6.0 Can argue case for 5.5	6.0 Can argue case for 5.5	6.0 Can argue case for 5.5	6.0 Can argue case for 5.5
British Columbia College of Teachers	N/A	N/A	N/A	N/A	N/A	6.5	6.0	7.0	7.0	6.0
Ontario College of Teachers	N/A	N/A	N/A	N/A	N/A	7.0	6.5	7.0	7.0	6.5

**Table 2. Minimum levels of achievement in IELTS required by professional associations and registration bodies in Canada**

### 5.1.1 Minimum IELTS band scores

Of the ten associations involved in the study, eight were clear in their publicity about the fact that the Academic Module was more appropriate for their purposes than the General Training Module. However, two of the ten associations did not make a distinction between the Modules, and the general understanding of the differences between the two tests appeared to be relatively superficial.

Table 2 above sets out the minimum levels established by each of the associations.

Two associations did not specify a minimum band score for Reading, Writing and Listening. All five of the nursing associations, however, had established a higher minimum outcome for the Speaking test, highlighting their perception that oral communication skills were a priority.

In overall scores, six of the respondents specified a minimum overall score of 6.5, and the other four, three of these involving pharmacy and one representing teachers, required an overall band score of 7.0. One association indicated that applicants had the option of presenting a case to a specially convened Committee to argue that 5.5 should be accepted. This was reportedly a relatively rare occurrence.

At least three of the associations indicated that they would accept the best result of two or more tests, provided they had been completed in the previous two years. This may be due to the fact that the associations are accustomed to dealing with tests that have separate components which may be taken individually.

Three associations had reviewed the minimum levels, either raising them or putting in place minimum scores for each macroskill in addition to an overall score. This was seen as minimising the risk of a registrant presenting with one particularly low skill competency, thus potentially interfering with their fitness to practise.

The organisations which responded to the study were unanimous in their strong appreciation of the IELTS verification service, indicating that the security of this system gave them “peace of mind”.

Although there is reportedly general acceptance of nursing practitioners from province to province in Canada, it is interesting to note from Table 2 that the language assessment levels they require differ somewhat. In the case of pharmacy professionals, however, the fact that there is an overarching body which sets benchmarks for pharmacists on a national basis means that the required language assessment levels have been standardised.

### **5.1.2 Alternative language assessment systems accepted – Canada**

All Canadian associations accepted at least one other test in addition to IELTS, most commonly the iB TOEFL. There appeared to be far more consistency in the minimum scores required in these alternative tests, the only deviation being those required by the College of Nurses of Ontario, which took into account a Standard Error of Measurement when setting minimum standards.

The numbers tested annually could be anything from just over one hundred (for example, the College of Pharmacists of British Columbia) to in excess of a thousand (for example, the College of Registered Nurses of British Columbia), with numbers reportedly increasing from year to year because of the government’s open door policy for certain professions.

Many of the Canadian associations accepted up to four or five alternative testing systems.

Table 3 opposite sets out the alternative assessment systems and language levels accepted by Canadian professional bodies.

For Canadian associations, the perceived advantages of IELTS were as follows:

- Wide accessibility throughout the world
- Frequency of testing dates
- Online validation of band scores
- Reliability
- Live examiners using “real day-to-day” English in interpersonal interactions.

However, they also had a history of working with the ETS tests, and expressed confidence in their reliability.

Minimum standards of assessment (other than IELTS)	TOEFL paper-based test	TOEFL computer based test	IB TOEFL	TSE	TWE	TOEIC	CELBAN	MELAB	CanTest	Other
College of Registered Nurses of Manitoba	550 plus TSE	213	79 *S 26	50		730 plus TSE	*R 8 *W 7 *L 9 S 8	83 S 3		Approved 3-month training course
College of Registered Nurses of British Columbia	550 plus TSE	213	RWL 60 S 26	50			R 8 W 7 L 9 S 8	83 S 3		
Association of Registered Nurses of Newfoundland and Labrador	550 plus TSE	213 plus TSE	86 R 20 W 20 L 20 S 26	50		800	R 8 W 7 L 9 S 8	83		
College of LPNs of British Columbia, Burnaby	550 plus TSE	213 plus TSE	80 S 26	50			R 8 W 7 L 9 S 8	83 S 3		Completion of Grade 12 English
College of Nurses of Ontario	537 RWL each 53	203 LW RWL each 20	74 RWL 60 S 26	50		720 LW 350	R 8 W 7 L 9 S 8	75 S 3		
National Association of Pharmacy Regulatory Authorities	580	237 W 26	97 W 25 S 27	50	5			85 S 3	R 4.5 W 4.5 L 4.5 S 4.5	
College of Pharmacists of British Columbia	580	237	97 W 25 S 27	50	5			85	R 4.5 W 4.5 L 4.5 S 4.5	
Ontario College of Pharmacists	580	237	97 W 25 S 27	50	5			85	R 4.5 W 4.5 L 4.5 S 4.5	Objective evidence of language proficiency
British Columbia College of Teachers	580	237	98 R 23 W 25 L 23 S 27	50						Minimum three credit units of study
Ontario College of Teachers			103 R 24 W 28 L 23 S 28							

\* R=Reading, W=Writing, L=Listening, S=Speaking

**Table 3. Minimum levels of achievement in assessment systems other than IELTS required by professional associations and registration bodies in Canada**

## 5.2 Summarising comments - the United Kingdom

As was the case with Canada, eight of the ten UK registered professional bodies were involved in public health. Most had either conducted research or sought advice from language experts in order to establish minimum standards of language proficiency in IELTS. Many were also guided by equivalent organisations in-country or overseas, and some had attended a briefing session run by IELTS in an effort to develop a better understanding of what the various band scores represented.

There appeared to be a clear understanding by the UK associations about the importance of high quality communication skills, and many of the associations had conducted research into the language needs of their profession, although this was generally functionally based and driven by the need to set language proficiency standards.

### 5.2.1 Minimum IELTS band scores

Table 4 below sets out a comparison of IELTS test outcomes required by UK professional organisations.

Minimum IELTS levels	General Training Module				Academic Module					
	Overall	Reading	Writing	Speaking	Listening	Overall	Reading	Writing	Speaking	Listening
British Acupuncture Council	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5
Chartered Institute of Marketing	6.5					6.5				
Faculty of Public Health	N/A	N/A	N/A	N/A	N/A	7.0	7.0	7.0	7.0	7.0
General Dental Council - Dentists - Dental Nurses	N/A	N/A	N/A	N/A	N/A	7.0 6.0	6.5 5.5	6.5 6.0	6.5 6.0	6.5 5.5
General Medical Council Professional Linguistic Assessment Board (PLAB)	N/A	N/A	N/A	N/A	N/A	7.0	6.0	6.0	7.0	6.0
The College of Optometrists	7.0	6.0	6.0	7.0	6.0	7.0	6.0	6.0	7.0	6.0
Health Professions Council - Speech and Language Therapists - All other professions	N/A	N/A	N/A	N/A	N/A	8.0 7.0	7.5 6.5	7.5 6.5	7.5 6.5	7.5 6.5
Nursing and Midwifery Council	N/A	N/A	N/A	N/A	N/A	7.0	7.0	7.0	7.0	7.0
Royal College of Veterinary Surgeons	N/A	N/A	N/A	N/A	N/A	7.0				
Royal Pharmaceutical Society of GB	N/A	N/A	N/A	N/A	N/A	7.0	7.0	7.0	7.0	7.0

**Table 4. Minimum levels of achievement in IELTS required by professional associations and registration bodies in the UK**

Three of the ten UK associations examined did not specify which module of IELTS was appropriate, which suggests that either the General Training Module or the Academic Module would be acceptable. This may indicate a limited understanding of the purpose and levels of the two test modules.

Secondly, eight of the ten associations had set individual minimum scores for each of the four macroskills in addition to a minimum overall standard. Only two relied on a single overall score. Both of these were outside the public health sector, one involving veterinary practitioners and the other a marketing organisation.

Minimum band scores were generally in the range of 6.0 to 7.0. The only exception applied to Dental Nurses, for whom the minimum band scores were 5.5 for the receptive skills of Reading and Listening. This may be due to the fact that they were generally regarded as support staff whose direct interaction with patients was likely to be less intensive than Dentists, Medical Specialists, Nurses and the equivalent. All others involved in the health professions required 6.5 or 7.0, with a minimum of 8.0 for Speech and Language Therapists, which is a language intensive occupation.

Three associations required a higher band score in Speaking than the other macroskills, which tends to highlight the importance placed on oral skills in communicating with the public about health matters.

At least three organisations had reviewed the required band scores set when they had first used IELTS, by changing from one overall score to minimum scores in each of the macroskills, or by increasing minimum levels across the skills.

All of those interviewed required minimum scores to be achieved in one sitting of the test, with outcomes valid for two years from the time of taking the test.

### **5.2.2 Alternative language assessment systems accepted – United Kingdom**

It is interesting to note from Table 5 overleaf that of the ten associations involved in the study, only two accepted alternative language assessment systems. Since most organisations had been using IELTS for many years, this would appear to suggest that the testing system suited the needs of the organisations well and was considered to be fit for the purpose.

The associations interviewed reported that the numbers requiring testing fluctuated, but could be from a few hundred to over two thousand.

Most organisations expressed their satisfaction with the IELTS assessment system and had no direct plans to review their decision on acceptance of the test or on levels required apart from reconsideration during regular reviews. This places IELTS in a strong position in the UK.

None of the associations reported disciplinary hearings which were prompted by poor language skills. It would seem that if such issues are raised, they are dealt with at employer level.

Several associations demonstrated a good understanding of the functional language needed for their profession.

<b>Minimum standards of assessment (other than IELTS)</b>	<b>TOEFL paper-based test</b>	<b>TOEFL - computer-based</b>	<b>IB TOEFL</b>	<b>TSE</b>	<b>TWE</b>	<b>TOEIC</b>	<b>Trinity College Ldn</b>	<b>Cambridge Suite</b>	<b>Other</b>
British Acupuncture Council									No other tests accepted.  Evidence of UG or PG qual
Chartered Institute of Marketing							ISE III/IV		Recommendation by educationalist.  Eng lang qualification.
Faculty of Public Health									No other tests accepted. Evidence language skills OR UG medical training in English
General Dental Council									No other tests accepted.
General Medical Council , PLAB									No other tests accepted.
The College of Optometrists									No other tests accepted.
Health Professions Council									
- Speech and Language Therapists	670	290	118/120			990		CPE	Hong Kong E AA
- All other professions	600	250	100/120			810		CAE	
Nursing and Midwifery Council									No other tests accepted
Royal College of Veterinary Surgeons									No other tests accepted
Royal Pharmaceutical Society of GB									No other tests accepted

**Table 5. Minimum levels of achievement in assessment systems other than IELTS required by professional associations and registration bodies in the UK**

### 5.3 Summarising comments – Ireland

Ireland is not at this time a significant market for IELTS in terms of professional associations, with only three associations registered as IELTS test users. One of these, the regulatory body for emergency medical staff, attracted most of its international registrants from the European Economic Area, and as they were exempt from the requirement to provide a language assessment, the need for language testing was minimal.



### 5.3.1 Minimum IELTS band scores

The minimum standards for IELTS for each of the three associations registered in Ireland are shown in Table 6 below.

Minimum IELTS levels	General Training Module					Academic Module				
	Overall	Reading	Writing	Speaking	Listening	Overall	Reading	Writing	Speaking	Listening
Irish Nursing Board/ An Bord Altranais	N/A	N/A	N/A	N/A	N/A	7.0	6.5	7.0	7.0	6.5
Irish Medical Council (IMC)	N/A	N/A	N/A	N/A	N/A	7.5	7.0	7.0	7.0	7.0
Pre-Hospital Emergency Care Council (PHECC)	N/A	N/A	N/A	N/A	N/A	6.5				

**Table 6. Minimum levels of achievement in IELTS required by professional associations and registration bodies in Ireland**

In setting its minimum levels for language proficiency, the Irish Nursing Board made a distinction between the productive skills and the receptive skills, requiring a higher level in the minimum scores for the Writing and Speaking tests.

At least one of the three Irish organisations, the Irish Medical Council, had undergone a process of review of the minimum levels of IELTS to bring them into line with international best practice. The outcome was to increase minimum levels by between 0.5 and 1.0 band scores; that is, the overall minimum of 7.0 in the Academic Module was increased to 7.5, and the Listening, Reading and Writing band scores were increased from 6.0 to 7.0, in line with the previously established Speaking score. It is a higher overall level than the minimum standards in place for its UK counterpart, the General Medical Council of the United Kingdom, but the Council believes it is more appropriate for its practitioners.

All associations required the minimum levels of IELTS to be achieved in one sitting of the test.

### 5.3.2 Alternative language assessment systems accepted – Ireland

Table 7 below sets out the alternative testing systems accepted by Irish professional bodies.

<b>Minimum standards of assessment (other than IELTS)</b>	<b>TOEFL paper-based test</b>	<b>TOEFL, computer-based test</b>	<b>IB TOEFL</b>	<b>TSE</b>	<b>TWE</b>	<b>TOEIC</b>	<b>London College Trinity Suite</b>	<b>Cambridge</b>	<b>Other</b>
Irish Nursing Board/ An Bord Altranais	570 R 56 TWE Structure 56 L 56 TSE	230 R 22 TWE L 22 Structure 23 TSE	88 R 21 W 19 L 22 S 26	50	5.5				Test results within last two years OR evidence of practice in English language since taking test.
Irish Medical Council									No other tests accepted.
Pre-Hospital Emergency Care Council (PHECC)	May be accepted.  Levels advised at time of applying.	May be accepted.  Levels advised at time of applying.	May be accepted.  Levels advised at time of applying.						Other testing systems may be accepted.

\* R=Reading, W=Writing, L=Listening, S=Speaking

**Table 7. Minimum levels of achievement in assessment systems other than IELTS required by professional associations and registration bodies in Ireland**

The only alternative tests acceptable for these organisations were the TOEFL tests, which were perceived as being almost as widely available as IELTS but less compact and less easily interpreted. IELTS was seen as a good quality, reliable and internationally accessible English language assessment with a person-to-person interview perceived as an advantage.

The number of applicants requiring a language test was reportedly decreasing. Candidate feedback indicated that some considered it to be a difficult test in view of the requirement of most associations to achieve the minimum band scores in one sitting of the test.

It was considered by one association that a more focused medical context, rather than the general academic focus, would be preferable.

One association indicated that more contact with IELTS administration, particularly on distinguishing between the various band scores and locating accessible testing centres, would be appreciated.

The online verification service was considered to be a major advantage.

## 5.4 General outcomes

### 5.4.1 Main competitors to IELTS

The major competitor to IELTS in the countries included in this study was the TOEFL, and in particular the iB TOEFL. Canadian associations have been using TOEFL in its various forms for many years and are familiar with ETS, the organisation offering the tests. Past forms of the TOEFL have been limited by the fact that there was a limited network of testing centres and the test did not include an integrated speaking component, and so candidates were obliged to take a separate oral test, which was inconvenient and time-consuming.

With the introduction of the iB TOEFL, access to the test in countries with the required technology was improved, and all skills were covered in one test. However, from the point of view of professional associations, IELTS has maintained an advantage over iB TOEFL to date for the following reasons:

- IELTS has a very extensive network of available testing centres, and test dates are frequent so that it is easily accessible to candidates. The iB TOEFL has not yet reached the same level of accessibility.
- IELTS has a high level of security, and the online results verification service is seen as a major positive factor. However, this will also become available for users of the iB TOEFL following the launch of a similar service in late 2009.
- IELTS has a face-to-face interview which is seen by many professional associations as superior to communication on a computer interface, as is the case with the iB TOEFL.

It is the last of the above points which continues to constitute a significant marketing advantage for IELTS. Despite the spread of electronic communication, there remains a strong scepticism about claims that communicative competency can be assessed through an electronic interface with the same accuracy and authenticity as a person-to-person exchange. This will no doubt be further tested when the Pearson Test of English, which has no person-to-person communication, becomes more widely recognised. In the meantime, conducting interviews as real time conversations with an interlocutor is a valuable marketing point for IELTS in maintaining or increasing its market share.

It was interesting that in the United Kingdom, most organisations accepted the IELTS test exclusively. The Cambridge ESOL tests are also internationally available, and yet only one user listed these tests in their standards. There was a suggestion from some associations that if European legislation should be amended so that European professionals from non-English language backgrounds were required to demonstrate English language competency, this may prompt a broadening of the range of tests accepted, including the Cambridge tests.

Of the other tests discussed earlier in this report, the national ones such as the CanTEST and the CELBAN are clearly serving a need in Canada. The CELBAN for nurses has the advantage of being profession-specific and is currently accepted by ten provinces. However, the disadvantages are that:

- It is available in nine cities throughout Canada, but is currently unavailable as an offshore pre-departure test
- It can be taken a maximum of three times, with a three month waiting time between re-sittings.

The CanTEST is seen as an appropriate test for Canadian post-secondary course admission or for registration purposes. However:

- It is generic in language focus
- It is available in only five approved locations
- It is unavailable offshore, where most registration applicants are located
- There is a waiting period of six weeks to three months to retake the test, depending on the skills required.
- The MELAB is similarly limited.

In addition, it is difficult for local tests to provide the same support to candidates in terms of practice materials and preparation courses as the international testing systems. It is unlikely that they will gain market share outside the countries in which they are offered.

#### 5.4.2 Rationale for selection of IELTS

Some associations used IELTS for more than one purpose, depending on the role or roles of the organisation. These are shown in Table 8 below.

<i>Role of the language assessment</i>	<i>Number of associations</i>		
	Canada	UK	Ireland
Registration to practise a profession	7	6	3
Eligibility to sit for examination to assess capability to practise a profession	4	5	1
Eligibility to enter professional training program	1	4	0

**Table 8. The role of language assessment in professional associations**

Many of the individuals interviewed, particularly in Canada and Ireland, while very well-informed about the professional qualifications they were expecting of registrants, were aware of the importance of English language communication skills but had only a basic understanding of the English language assessment tools their organisation was accepting. International testing and assessment systems vary widely in format and outcomes, but there is consistency in the characteristics and values sought by users who are not language experts.

Essentially, organisational users are seeking an assessment tool that allows them to meet all purposes. They need to be able to set realistic standards in simple terms, to understand without requiring detailed linguistic knowledge the language competencies signified by the attainment levels, to use or recommend a quality product to clients and to be assured that they are dealing with overseas professionals in a fair and equitable way.

In terms of quality, the most desirable characteristics of a viable assessment system appear to be:

- Global reliability of outcomes, so that whether a candidate sat for a test in China, the Philippines or Canada, the assessment would not differ
- Global accessibility for the many professionals who apply for registration from their home country

- Brief turnaround between taking the assessment and reporting of candidate outcomes
- Security and quality of the testing instruments and outcomes statements
- Regular training of assessment staff and facilitators to ensure that equivalent conditions and standards are used for all candidates in all locations
- Security of candidate identification procedures
- Both internal and external quality assurance processes
- International content that is not culture-bound
- A strong research base to underpin the system
- Sufficient flexibility to adjust to the changing global community.

Furthermore, language assessment differs from more traditional educational content testing in that it does not provide a pass/fail outcome, but rather a score or set of scores (or band scales, in the case of IELTS). This means that for the non-linguist, there must be very clear plain English guidelines about what a particular score or band means in terms of language competency. IELTS is perceived as meeting all the above criteria.

In addition, because of its very broad international appeal in the professions, it allows users to compare their needs and criteria with other similar organisations on a global basis.

The only circumstances in which associations may consider a change away from IELTS would be:

- If there were major changes to the test so that it was no longer meeting the needs of the association and the profession
- If there were serious complaints from candidates about the test which were upheld after investigation
- If there were serious breaches in reliability and security of outcomes
- If there were changes in European Union legislation which affected language assessment or the choice of language assessment systems, or required a broadening of the range of acceptable systems, for European professionals entering the UK and Ireland.

### **5.4.3 Review of minimum standards of IELTS**

Decision-making on tests and review of appropriate levels was generally the responsibility of the governing body of the association, assisted by a registration department or registration working group. The original decision on levels, at the time of taking up IELTS as a language benchmark, was in many cases unknown because of staff turnover. Whether or not the standards had been reviewed was also unknown for some associations.

This raises an interesting point. Knowledge of language assessment systems generally rests with an administrative person or small team who work with registration matters. This means that if the person or team members leave the association, or change jobs, that knowledge is lost, and records may also be lost. According to the respondents none of the organisations were in direct contact with IELTS administration. Although they could access information packs on IELTS, consult the website or choose to attend group briefings from time to time, it would seem opportune to keep IELTS at the forefront of accepted testing systems by maintaining personal contact with responsible registration staff.

The main strategies used to establish and review minimum standards of IELTS are set out in Table 9 below.

<b>Decision-making process</b>	<b>Number of associations</b>		
	Canada	UK	Ireland
Consulted IELTS administration	0	0	0
Consulted an expert in the field	2	2	0
Conducted a formal inhouse research project	3	4	2
Board/Council aligned levels with other similar organisations	2	2	1

**Table 9. Decision-making strategies for establishing minimum levels in IELTS**

Four Canadian associations have reviewed IELTS standards since establishing them, including NAPRA, two have not conducted a review, and four do not know. Of the four which reviewed standards, three left them unchanged and one increased the band scores by 0.5 in the productive skills.

In the UK, three associations have reviewed IELTS levels required, three have not, and four do not know. Of those which reviewed standards, one made a decision to no longer accept the General Training Module, and increased the scores by 0.5 to 7.0 in all skills. One changed from a single overall outcome to minimum standards in all four macroskills, and one left them unchanged.

One Irish association has reviewed the levels, increasing them by a full band score, to 7.0 overall and 7.0 in each of the macroskills. One left the standards unchanged at 6.5, and one does not know.

Although none indicated that the association had made direct contact with IELTS administration, a number said that they had received an informative information package.

#### **5.4.4 Consideration of IELTS as fit for the purpose**

The fact that associations are reviewing standards and making decisions on increasing minimum levels for individual skills, as pointed out in 5.4.3 above, is an indication that knowledge of the test and the band scores is growing, and there is confidence in the testing system.

Respondents were in general very satisfied with the IELTS test and its fitness for the purpose. Although the health professions in particular would prefer some content relating to the profession, IELTS is clearly meeting the needs of the organisations involved in this study.

#### **5.4.5 General feedback**

None of the associations interviewed received a great deal of feedback about the English language assessment. Some individual candidate comments were as follows:

A minimum assessment of 6.5 or 7.0 in each of the skills is deemed to be too high by some applicants, especially in one sitting of the test.

A small number have expressed a preference for the TOEFL test, which some applicants perceive as being easier than IELTS, but this test is not currently acceptable to many of the associations interviewed.

Some candidates (for example, Chinese candidates entering the health professions) have claimed that the requirement to sit for an IELTS test is potentially racist. The counterargument to this is that health professionals in particular practising in an English-speaking country must be able to communicate effectively with patients, colleagues and statutory authorities.

Some members of organisations who were dealing with IELTS on a regular basis expressed a wish to experience the test for themselves so that they were more familiar with the tasks that candidates were expected to handle.

#### 5.4.6 Risk management and opportunities

Some comments and issues raised by association staff in relation to the IELTS test are set out in Table 10 below. The most significant of these is the perceived preference by many associations involved in health care for subject-specific content in language tests. Historically, the IELTS test used to have discipline-specific modules of the test (IELTS 2009, History of IELTS). It may be timely to revisit this in test review processes.

<b>Feedback on IELTS - Issues</b>	<b>Number of associations</b>		
	Canada	UK	Ireland
Possible substitution of candidate		1	1
Unreliability of assessment in some countries where IELTS is conducted (e.g. India, China)	1		1
Lack of subject specific content of test for professions (especially health-related)	2	3	1
Lack of availability of the test in some source areas for applicants (e.g. Northern Canada, West Africa, Central America)	2		1
Fraudulent results document received		1	
Does not test "fitness to practise"	1		
Some language in the test is not international (e.g. boot of car, petrol ...)	1		

**Table 10. Feedback on IELTS - Issues**

## 6 CONCLUSIONS

The UK and Ireland are traditional markets for IELTS in the academic arena, and use of IELTS as an entry criterion to professional workplaces is a relatively recent growth market. In the English-speaking Canadian context, the increase in acceptance of IELTS has also been significant.

All participants interviewed demonstrated a strong concern for maintaining the integrity of their standards of practice, and this included language competency. Many had either looked to other similar organisations when making a decision on the minimum levels of English language competency they would require, or had sought advice from experts or academics familiar with the IELTS test and descriptors of levels.

The United Kingdom is a particularly strong market for IELTS, and stakeholder satisfaction is evident in the fact that association staff are knowledgeable about the test and its outcomes, and sufficiently confident of its reliability to use it exclusively as their English language assessment benchmark. Associations which specified IELTS as the only acceptable language assessment system and those which processed high numbers of applicants tended to have a more detailed knowledge of the test and the band scores.

European legislation which exempts European nationals from language assessment has posed a problem for many of the UK and Irish associations. As language testing has now been largely devolved to employers, this is a potentially new market for IELTS in the UK and Ireland. In addition, if there should be review of European Union legislation on language testing, it would be in the interests of the IELTS partners to lobby policymakers to position IELTS as the major international English language testing system accepted.

Canada is a growing market for the IELTS, which has been challenging the more traditional TOEFL and TOEIC tests in the country. Associations are becoming more familiar with IELTS and what the band scores signify, and are developing trust in the system and the outcomes. Security is also a high priority, particularly for the health professions. Provincial regulatory bodies can be targeted with information packages and seminars to assist them in decision-making on standards and to further familiarise them with IELTS.

Providers were generally clear in their understanding of the general day-to-day communication patterns between professionals and those with whom they would come into contact. However, none of the respondents was able to reference any in-depth research into the language skills required to operate effectively in the profession which they represented. Given that this is a relatively new use of IELTS, it may be to the advantage of the IELTS partners to develop a body of research which could then feed into the writing of content and questions for future tests.

For some associations, there appears to be a lack of understanding of the difference between the General Training Module and the Academic Module. Education and information packages would go some way towards remedying this issue, which constitutes a risk for both associations and IELTS.

The most common attitude of participants in this study tended to be that IELTS was designed and run by professionals in the field of linguistics and language assessment, and as long as it was working for the associations, it was an effective tool and they would continue to work with it.



## REFERENCES

- Charge, N and Taylor, L, 1997, 'Recent developments in IELTS', *English Language Teaching Journal*, vol. 51, no 4, Oxford University Press, Oxford
- Creswell, J W, 1994, *Research Design: Qualitative and Quantitative Approaches*, Sage, London
- Denzin, N K and Lincoln, Y S, 1994, *Handbook of Qualitative Research*, Sage, Thousand Oaks, CA
- Douglas, S, 2008, 'The registration and accreditation of international medical graduates in Australia: a broken system or a work in progress?' in *People and Place*, vol. 16, no 2, eds B Birrell and K Betts, Centre for Population and Urban Research, Monash University, Australia, pp 36-49
- ETS TOEFL, 2009, *Comparing TOEFL iBT and IELTS*
- Graddol, D, 2006, *English Next*, British Council, London
- Green, A, 2007, *IELTS Washback in Context: Preparation for Academic Writing in Higher Education*, Cambridge University Press, Cambridge
- 'IELTS Candidate performance in 2008', *Cambridge ESOL Research Notes*, no 36, May 2009, UCLES, Cambridge, pp 30-32
- IELTS 2009, 'History of IELTS' [online] [www.ielts.org/researchers/history\\_of\\_ielts.aspx](http://www.ielts.org/researchers/history_of_ielts.aspx) [Accessed October 2009].
- Merrifield, G, 2008, 'An Impact Study into the Use of IELTS as an Entry Criterion for Professional Associations – Australia, New Zealand and the USA', in *IELTS Research Reports*, vol. 8, IELTS Australia
- O'Neill, T, Buckendahl, C, Plake, B and Taylor, L, 2007, 'Recommending a Nursing-Specific Passing Standard for the IELTS Examination', in *Language Assessment Quarterly*, vol. 4, no 4, pp 295–317
- Papageorgiou, S, 2007, *Relating the Trinity College London GESE and ISE examinations to the Common European Framework of Reference: Final Project Report*, Department of Linguistics and English Language, Lancaster University
- Stupan, L, Rao, D, March, G J and Elliot, E, 2008, 'English Language Communication Skills – What are the Issues for International Pharmacy Students' in *People and Place*, Volume 16, Issue 2, eds B Birrell and K Betts, Centre for Population and Urban Research, Monash University, Australia, pp 62-70
- Trinity College London, 2009, *Table of comparisons between Trinity College London's ISE and a GESE examinations and other ESOL examination benchmarks*. [www.trinitycollege.co.uk/site/](http://www.trinitycollege.co.uk/site/) [Accessed 25/9/09]
- University of Cambridge ESOL Examinations Annual Review 2008*, University of Cambridge, Cambridge

## **APPENDIX 1: PROFESSIONAL ASSOCIATIONS AND LANGUAGE ASSESSMENT WEBSITES**

Association of Registered Nurses of Newfoundland and Labrador  
**www.arannl.nf.ca** [Accessed August 2009]

British Acupuncture Council  
**www.acupuncture.org.uk** [Accessed July 2009]

British Columbia College of Teachers  
**www.bcct.ca** [Accessed August 2009]

Cambridge ESOL  
**www.cambridgeesol.org/exams** [Accessed July 2009]

Canadian English Language Benchmark Assessment for Nurses  
**www.celban.org/celban** [Accessed August 2009]

Centre for Canadian Language Benchmarks  
**www.language.ca** [Accessed August 2009]

Chartered Institute of Marketing (CIM)  
**www.cim.co.uk** [Accessed June 2009]

College of Licensed Practical Nurses of British Columbia  
**www.clpnb.org** [Accessed August 2009]

College of Nurses of Ontario  
**www.cno.org/international\_en/reqs/req4\_lang/testing.htm** [Accessed May 2009]

College of Pharmacists of British Columbia  
**www.bcpharmacists.org** [Accessed May 2009]

College of Registered Nurses of British Columbia  
**www.crnbc.ca** [Accessed August 2009]

College of Registered Nurses of Manitoba  
**www.crnmb.ca** [Accessed May 2009]

Faculty of Public Health UK  
**www.mmc.nhs.uk** [Accessed August 2009]

General Dental Council, UK  
**www.gdc-uk.org** [Accessed June 2009]

General Medical Council, UK  
**www.gmc-uk.org** [Accessed August 2009]

General Optical Council UK  
[www.optical.org](http://www.optical.org) [Accessed August 2009]

Health Professions Council, UK  
[www.hpc-uk.org](http://www.hpc-uk.org) [Accessed August 2009]

Irish Nursing Board  
[www.nursingboard.ie](http://www.nursingboard.ie) [Accessed May 2009]

Irish Medical Council  
[www.medicalcouncil.ie/registration/statistics\\_page06.asp](http://www.medicalcouncil.ie/registration/statistics_page06.asp) [Accessed May 2009]

National Association of Pharmacy Regulatory Authorities  
[www.napra.ca](http://www.napra.ca) [Accessed July 2009]

Nursing and Midwifery Council, UK  
[www.nmc-uk.org](http://www.nmc-uk.org) [Accessed August 2009]

Ontario College of Pharmacists  
[www.ocpinfo.com](http://www.ocpinfo.com) [Accessed August 2009]

Ontario College of Teachers  
[www.oct.ca](http://www.oct.ca) [Accessed July 2009]

Pre-Hospital Emergency Care Council, Ireland  
[www.phecit.ie](http://www.phecit.ie) [Accessed June 2009]

Professional Linguistic Assessment Board, UK  
[www.gmc-uk.org/doctors/plab.asp](http://www.gmc-uk.org/doctors/plab.asp) [Accessed June 2009]

Royal College of Veterinary Surgeons, UK  
[www.rcvs.org.uk](http://www.rcvs.org.uk) [Accessed August 2009]

Royal Pharmaceutical Society of Great Britain  
[www.rpsgb.org.uk](http://www.rpsgb.org.uk) [Accessed August 2009]

## APPENDIX 2: ACKNOWLEDGEMENTS

The researcher would like to convey special thanks to the following professionals for their time and generous assistance.

### UNITED KINGDOM

**John Wheeler**

Secretary and Chair, Admissions Committee  
British Acupuncture Council

**Patrick Kavanagh**

General Dental Council

**Oliver Campbell**

Interim Team Manager  
Nursing and Midwifery Council

**Linda Price**

Education Policy Advisor,  
Education Department  
Chartered Institute of Marketing (UK)

**Martha Pawluczyk**

Royal Pharmaceutical Society of Great Britain

### IRELAND

**Ruth Thompson**

Executive Officer, Education and Training  
Medical Council

**Barry O’Sullivan**

Deputy Director and Registrar  
Pre-Hospital Emergency Care Council

### CANADA

**Corey Sigurdson**

Consultant, Registration Services  
College of Registered Nurses of Manitoba

**Lyn Cairns**

Registration, Inquiry and Discipline  
College of Registered Nurses British Columbia

**Doreen Leong**

Director, Registration/Special Projects  
College of Pharmacists of British Columbia

**Christyna Schillemore**

Manager, Registration Programs

**Diana Spizzirri**

Registration Advisor

**Deborah Byer**

Registration Program Assistant  
Ontario College of Pharmacists

**Beverley Maxwell**

Director of Certification  
British Columbia College of Teachers

My thanks go to the Project Manager, IELTS **Jenny Osborne**, for her patience and assistance throughout this project.

## **APPENDIX 3: GUIDING QUESTIONS FOR INTERVIEWS CONDUCTED WITH PROFESSIONAL ASSOCIATIONS.**

Questions for interviews conducted with professional associations and professional registration bodies.

### **IELTS Research Project**

#### **Interviews**

Name of Organisation .....

Brief explanation of the aims and purpose of the organisation .....

.....

.....

### **1 Introduction**

Explanation of the objectives of this study

Reference to the significance of this study to IELTS Australia and Cambridge ESOL

Brief background to the development and scope of use of IELTS

### **2 Guiding questions - IELTS**

**2.1** What is the purpose of your organisation's use of IELTS? (please provide details)

Registration to practise a profession

Eligibility to sit for an examination to assess capability to practise a profession

Assessment for migration purposes

Membership of a professional organisation

Other .....

## 2.2 What levels of IELTS are required?

<b>IELTS Module</b>	<b>Overall Band Score and/or Macroskill minimum score</b>	<b>Band Score required</b>
General Training Module	Overall score	
	- Reading	
	- Writing	
	- Speaking	
	- Listening	
Academic Module	Overall score	
	- Reading	
	- Writing	
	- Speaking	
	- Listening	

- 2.3** IELTS assesses all skills at a single sitting, unlike some assessment systems in which candidates can present with the best results of each of the four skills (Reading, Writing, Speaking, Listening) of two or three or more sittings. Is this in line with the way your organisation uses IELTS band scores?
- 2.4** What other professional requirements are there?
- 2.5** How many clients (approximately) will have their English language proficiency assessed in any one year?
- 2.6** Is this number increasing or decreasing?
- 2.7** How was the decision made on the appropriate threshold levels?
- 2.8** Who is responsible for making decisions on English language assessment levels?
- 2.9** Was any advice or support requested or received from IELTS Australia/Cambridge ESOL, or any other organisation, in making the decision on appropriate levels?
- 2.10** Has your organisation's decision on acceptable IELTS levels been reviewed?
- 2.10.1** If so,
- Why was it reviewed?
  - What did the review process consist of?
  - What was the outcome?
- 2.10.2** If not, is it planned to review it? Why or why not?
- 2.11** Are you aware of any gaps or inadequacies in the IELTS testing system for the purposes of your profession?

### 3 Guiding questions – Alternative language assessment systems/strategies

3.1 What other language assessment systems/strategies are accepted? What are the levels accepted?

✓	TEST	LEVEL(S) REQUIRED
	Test of English as a Foreign Language (TOEFL) and Internet-based TOEFL (iB TOEFL)	
	Test of Spoken English (TSE)	
	Test of Written English (TWE)	
	Test of English for International Communication (TOEIC)	
	Cambridge Certificate of Advanced English (CAE)	
	Cambridge Certificate of Proficiency in English (CPE)	
	CELBAN	
	MELAB	
	Trinity College Tests	
	Other	

3.2 In your view, what are the advantages and disadvantages of each? Which assessment systems/strategies suit your organisation best and why?

3.3 Have candidates expressed a preference for a particular test? If so, which one and why?

3.4 Have you had any feedback about how the candidates perceive the IELTS test?

3.5 Is there a strategy in place to follow up newly registered people or to seek feedback from their employers once they are in the workplace, to ensure that they are coping in terms of language skills? If so, please describe it.

3.5.1 If not, would this be useful?

3.5.2 What focus area(s) would be most useful to get feedback on (e.g. using technical vocabulary and language, language for dealing with general enquiries, using non-technical language when dealing with the young/aged/people of non-English speaking backgrounds/professional to professional ... )

#### **4 Guiding questions – Ongoing support**

4.1 Did you receive any information or support from the IELTS administration when making the decision?

4.2 Have you had any follow-up briefing or support from IELTS administration?

If yes, what support was given?

If not, would it be useful to you? What form of support would be most useful?

#### **5 Alternative language assessment systems**

5.1 Will you continue to use the IELTS test as an English Language assessment instrument? Why or why not?

5.2 Will your profession be accepting other language testing or assessment systems as equivalent to IELTS (eg iB TOEFL, TOEIC, the new Pearson Test of English, profession-specific assessment, ...)?

#### **6 Research**

6.1 Do you know of any research, articles or discussion papers that have been published into English language competency or testing and assessment in your profession by professionals in your field, academics, government or other relevant bodies? If so, please provide a copy or access details (name, author, publisher, year).

#### **7 Do you have any further comments or issues you wish to raise?**

Thank you for your time.

**Glenys Merrifield**, B Ed (TESOL), Dip TESOL, M Ed Administration

**www.gbm.net.au**

**Tel:** 61 2 9438 2428

**Mob:** +61 (0)407 095 913

**Skype:** glenys.merrifield



## **APPENDIX 4: ACRONYMS**

<b>ACPE</b>	Accreditation Council for Pharmacy Education (USA)
<b>AEPUCE</b>	Academic English Program for College and University Entrance (Canada)
<b>ARNNL</b>	Association of Registered Nurses of Newfoundland and Labrador
<b>BCCT</b>	British Columbia College of Teachers
<b>CAE</b>	Certificate in Advanced English (Cambridge ESOL)
<b>CanTEST</b>	Canadian Test of English for Scholars and Trainees
<b>CCAPP</b>	Canadian Council for Accreditation of Pharmacy Programs
<b>CCLB</b>	Centre for Canadian Language Benchmarks
<b>CEFR</b>	Common European Framework of Reference
<b>CELBAN</b>	Canadian English Language Benchmarks Assessment for Nurses
<b>CLPNBC</b>	College of LPNs of British Columbia, Burnaby
<b>CPNRE</b>	Canadian Practical Nurse Registration Examination
<b>CNO</b>	College of Nurses of Ontario
<b>CPE</b>	Certificate of Proficiency in English (Cambridge ESOL)
<b>CRNBC</b>	College of Registered Nurses of British Columbia
<b>CRNM</b>	College of Registered Nurses of Manitoba
<b>EEA</b>	European Economic Area
<b>ESOL</b>	English for Speakers of Other Languages
<b>ETS</b>	Educational Testing Service
<b>IELTS</b>	International English Language Testing System
<b>FCE</b>	First Certificate in English (Cambridge ESOL)
<b>iB TOEFL</b>	Internet-based Test of English as a Foreign Language
<b>IMC</b>	Irish Medical Council (Ireland)
<b>ISE</b>	Integrated Skills in English Examinations (Trinity College London)
<b>MELAB</b>	Michigan English Language Assessment Board

<b>OCP</b>	Ontario College of Pharmacists
<b>OCT</b>	Ontario College of Teachers
<b>PEBC</b>	Pharmacy Examining Board of Canada
<b>PHECC</b>	Pre-Hospital Emergency Care Council (Ireland)
<b>PLABE</b>	Professional Linguistic Assessment Board Examination (UK)
<b>TOEFL</b>	Test of English as a Foreign Language
<b>TOEIC</b>	Test of English for International Communication
<b>TSE</b>	Test of Spoken English
<b>TWE</b>	Test of Written English

## **2 Learning to play the ‘classroom tennis’ well: IELTS and international students in teacher education**

### **Authors**

**Wayne Sawyer and Michael Singh**

Centre for Educational Research, University of Western Sydney

Grant awarded Round 13, 2007

This study addresses the question of an appropriate IELTS score for graduate entry teacher education courses by investigating the extent to which current IELTS scores into graduate entry teacher education courses are considered adequate: by the lecturers of such students, by prospective teacher registration authorities and by the students themselves.

### **ABSTRACT**

Teacher education students whose language background is not English (LBOTE) not only need to perform adequately in English for the purposes of their academic study, but also need to be fluent in the public situation of teaching their own classes on practicum and in their future careers. Thus, for these students, the ‘public performance’ of English adds a significant layer of issues to those applying to other cohorts of LBOTE students.

This research sets out to address the question of an appropriate IELTS score for graduate entry teacher education courses through investigating: the extent to which current IELTS scores into graduate entry teacher education courses are considered adequate by the lecturers of such students, by prospective teacher registration authorities and by the students themselves, and, following this, what an appropriate score might be for entry into such courses. Academics from four Faculties of Education and one student cohort were interviewed, along with representatives of one state teacher registration authority. A range of language skills for teachers were identified. A key issue for these students in such courses is the potential for language growth in the course itself with a corresponding need to focus on exit abilities as well as entry ability. It is argued therefore on the one hand, that in short graduate entry courses, universities ought to consider setting an entry level corresponding to the requirements of the teacher registration authority. Some interviewees argued, however, that the complex of issues faced by these students – such as familiarisation with Anglophone schooling cultures – makes particularly high IELTS entry scores a distraction from these larger issues. There is also a need for universities to value the experiences and funds of knowledge brought by LBOTE students. Ultimately, IELTS entry scores are a function of a Faculty’s ability to provide language support. The nature of such support and of a richer approach to testing is discussed.

## **AUTHORS' BIODATA**

### **WAYNE SAWYER**

Associate Professor Wayne Sawyer is Head of Research in the School of Education at the University of Western Sydney and a member of the Centre for Educational Research. His central research interests are in the areas of effective teaching, English curriculum and educational policy. He has been widely published in the area of English (L1) curriculum. He supervises a number of international research students and is interested in the development of scholarly argumentation among students who are working in an L2 milieu.

### **MICHAEL SINGH**

Professor Michael Singh is engaged in researching the challenging opportunities presented by changes in the diversity of learners and teachers for making intellectual connections. His research focuses on the presence of international, migrant and refugee students and/or teachers in Australia's education institutions and the significant opportunities this creates for developing programs and pedagogies to improve their education and work/life trajectories. He leads a team of researchers, including research students from China, in a study of research-oriented, school engaged teacher education which aims to stimulate increased interest among school students in learning Mandarin.

## CONTENTS

<b>1</b>	<b>Introduction .....</b>	<b>77</b>
<b>2</b>	<b>LBOTE students, English language proficiency, testing instruments and teacher education: A review of literature.....</b>	<b>79</b>
2.1	International students in English language universities .....	79
2.2	Perceptions of IELTS testing.....	81
2.3	The language sub-skills .....	82
2.4	English language proficiency and academic results .....	83
2.5	International students, English language proficiency and teacher education .....	86
2.6	International students, teacher education and general support.....	92
<b>3</b>	<b>Research methodology .....</b>	<b>93</b>
3.1	Site selection.....	93
3.2	Site studies .....	94
3.3	Data analysis .....	94
<b>4</b>	<b>Baseline data .....</b>	<b>95</b>
<b>5</b>	<b>LBOTE teacher education students' perspectives .....</b>	<b>99</b>
5.1	Having been in your teacher education course now for some time, what are your views on the demands of the course in terms of your English language proficiency?.....	99
5.2	What interventions have been put into place to support you in terms of English language proficiency? Have these been suitable/adequate? .....	100
5.3	Where is the need for support greatest (e.g. through your study program, through the practicum)? .....	101
5.4	Are other issues more important than just language proficiency in supporting your success (e.g. becoming familiar with the cultures of Australian schools)? ...	101
5.5	Have your practicum schools identified English language proficiency as an issue for you? What about other issues (such as becoming familiar with Australian school cultures)?.....	102
5.6	In the light of this, what do you think of the adequacy of your IELTS score for entry into teacher education and what do you think an adequate score would be?.....	102
<b>6</b>	<b>The perspectives of representatives of one state teacher registration authority.....</b>	<b>102</b>
<b>7</b>	<b>Australian teacher educators' perspectives on the IELTS entry scores.....</b>	<b>103</b>
7.1	Background.....	103
7.2	What are your views on the English language proficiency of your international students who have been subject to IELTS testing? .....	104
7.3	Are you doing any internal testing of the English language proficiency of such students?.....	107
7.4	What interventions have you put into place to support these students in terms of English language proficiency? How have these been funded? .....	108
7.5	Are other issues more important than just language in supporting these students (e.g. becoming familiar with Australian schools)? .....	111

7.6	Where is the need for support most evident (e.g. through their study program, through the practicum)?.....	113
7.7	What have schools had to say about the English language proficiency of student-teachers? What about other issues (e.g. acculturation into Australian schools)? .....	114
7.8	What revisions, if any, has your Faculty made to entry procedures in the light of this experience? .....	115
7.9	What is required by your local Teacher Registration body in terms of language proficiency? Is this more or less demanding than the equivalent university entrance requirement?.....	116
7.10	In the light of this, what do you believe is an adequate IELTS score for entry into teacher education?.....	116
<b>8</b>	<b>Discussion .....</b>	<b>118</b>
	References .....	122
	Appendix 1: Glossary of terms .....	128

## 1 INTRODUCTION

In 2007, 17.3% of the student population in Australian universities were international students, with the top five source countries each Asian (IDP, 2007). These students have a number of adjustment issues in the Anglophone university environment (Deumert et al, 2005). Such students enter a complex environment of not only a foreign language, but a foreign language in an academic register (Canagarajah, 2002). IELTS assesses the listening, speaking, reading and writing abilities in English of potential Language Background other than English (LBOTE) students. In this paper, we use the acronym LBOTE to describe the language background of these students, unless quoting literature which refers to them in some other way, such as non-English Speaking Background (NESB).

A useful brief history of the IELTS testing system and selected associated research is contained in Hyatt and Brooks (2009, p 21). IELTS provides tertiary institutions with data about potential (LBOTE) students' English proficiency at a point in time. Universities use cut-off scores as an indication of a threshold level below which students are deemed unlikely to cope with the language demands of university-level study. Thus, setting threshold scores is intended to minimise the extent to which English language ability inhibits performance (especially early) in a course. Bayliss and Ingram (2006, p 1) describe the 'meaning' of an IELTS score for tertiary study as follows:

... the score a student achieves in an IELTS test is meant to indicate whether he/she has a sufficient level of English proficiency to cope with the linguistic demands of tertiary studies, (but) it does not imply that they will succeed academically or that they will not struggle linguistically

A similar meaning is ascribed by O'Loughlin and Arkoudis (2009, p 100):

... it predicts the extent to which a candidate will be able to begin studying through the medium of English

It is important to remember that categories such as 'LBOTE students' or 'international students' describe heterogeneous populations, 'from diverse cultural, economic, social and linguistic backgrounds... (that) cannot unproblematically be characterised as (all) having (the same) qualities' (Ryan and Viète, 2009, p 304). Thus, terms such as 'South Asians' disguise an immense diversity between educational cultures, intellectual heritages and students' learning experiences.

Teacher education students whose language background is not English and, in the case of graduate entry teacher education students who completed undergraduate degrees in cultures where English is not the majority language, provide a distinct sub-cohort of this category. These students not only need to perform adequately in English for the purposes of their academic study, but also need to be fluent in the public situation of teaching their own classes on practicum and in their future careers (to 'perform' in a different sense), while learning about the cultures of Anglophone schooling. There are also written abilities required of teachers that differ from academic work, such as preparing comprehensible written materials in English, and marking school pupils' work in English. Thus, for teacher education students, the 'public performance' of English adds a significant layer of issues to those applying to other cohorts of students subject to IELTS testing. Han (2006) has shown that Australia also confronts such students with unfamiliar pedagogies that in their turn produce certain dominant, contradictory or competing elements that make the task of identity transformation a challenge for such students – in our terms as public 'performers' of English. In addition, in Australia, teacher accreditation authorities require particular levels of language achievement (see Table 4 below). In the state of New South Wales the

main employer - the NSW Department of Education and Training - has developed its own instrument aimed at testing language skills specific to teaching: the Professional English Assessment for Teachers (PEAT). It is worth noting that in Australia, certain areas of school education – such as secondary Mathematics and Science – are increasingly heavily populated by LBOTE teachers.

This research investigates the question of an appropriate score for graduate entry teacher education courses. Through interviewing relevant personnel in universities and teacher registration authorities, the research addresses:

- 1 the extent to which current IELTS scores into graduate entry teacher education courses are considered adequate by the lecturers of such students, by prospective teacher registration authorities and by the students themselves
- 2 what an appropriate score might be for entry into such courses.

The research recognises that operating in Anglophone schooling cultures is not just a question of language difficulties for LBOTE students. The notion of ‘communicative competence’ is complex for teacher education students –including not only the ability to comprehend and communicate academic course material and to function effectively in learning situations such as tutorials, but also to comprehend and ‘perform’ English as teachers themselves, operating in front of school pupils who themselves may represent a diversity of cultural and linguistic experience – and all of this within a schooling culture that may be quite removed from their own schooling experience. The outcomes of this research are recommendations about appropriate IELTS scores for entry into teacher education in the context of other factors impacting on the success of these students.

The particular cohorts selected for investigation are those in graduate-entry pre-service courses in secondary education, such as a Graduate Diploma or Masters degree. The phrase ‘LBOTE students’ as used here, then, refers to students who have completed an initial degree in a country in which the majority language is not English. Some of these are ‘international students’, though many are also Permanent Residents or are Australian citizens. ‘International students’ is a term usually used to refer to overseas students on student visas. Many of the students to whom we refer here are not on student visas, but are, rather, Permanent Residents or Australian citizens. They have all, however, undertaken earlier education in countries in which English was not the majority language, and hence have been subject to IELTS testing in order to be accepted into their graduate entry courses in teacher education. In general usage, ‘LBOTE’ may still include those who have been schooled in Anglophone countries like Australia. However, we are specifically using the acronym ‘LBOTE’ here to refer to students who were subject to IELTS or similar testing before undertaking their graduate entry course. This includes both those who have undertaken earlier education in countries in which English was not the majority language and, when relevant, international students. Some of those on student visas intend to teach in their home countries; most of the total group under consideration will teach in Australia. All are educated on the assumption that they will teach in Australian schools. In this report we use the terms ‘student’ or ‘student-teachers’ to describe this group of teacher education students. The word ‘pupil’ is used to refer to school children. We also use the term ‘Faculty’ to describe ‘Faculties’ *and* ‘Schools’ of Education in universities, in order to distinguish these from primary and secondary schools.



## **2 LBOTE STUDENTS, ENGLISH LANGUAGE PROFICIENCY, TESTING INSTRUMENTS AND TEACHER EDUCATION: A REVIEW OF LITERATURE**

At the time of writing, international education is Australia's third largest export after coal and iron ore (Bradley, 2008, p 88). Over 80% of international students come from Asia, including twenty-one per cent from mainland China (Bradley, 2008, p 92). The Bradley review of Australian higher education links global engagement with international education, global research networks and student mobility because Australia has the 'highest proportion of international students in higher education in the OECD' (Bradley 2008, p 12) .

Australian higher education is seen as facing a risk because 'international students are concentrated in a narrow range of subject fields' (Bradley, 2008, p 12). There is pressure to broaden the fields and levels of studies being undertaken by international students. Sixty-seven per cent of the Chinese student cohort of 58,588 students in 2007 were undertaking degrees in management and commerce disciplines, while 'only 3.6 per cent (were) undertaking a research higher degree' (Bradley, 2008, pp 92, 93). In terms of research higher degrees, compared with other OECD countries, a 'relatively low proportion of Australia's higher degree students are international students' (Bradley, 2008, p 12). Significantly, it is argued that Australian 'immigration policies that target international students and scholars can yield positive results and can be critical in building the necessary skilled workforce for the future' (Bradley, 2008, p 99). In order to fill labour shortages and skills mismatches, international students are expected 'to become work ready in the Australian context' (Bradley, 2008, p 103). In particular, Bradley (2008, p 12) argues that because of 'looming shortages of academic staff (in Australia) it is time to consider how increases in higher degree enrolments from high-performing international students might be encouraged.' From these combined perspectives it is, therefore, 'essential for Australian (higher education) institutions to maintain high standards and excellent student support services to maintain (international) student numbers' (Bradley, 2008, p 107). The following review delineates the key debates and concepts concerning English language testing and tertiary students in general and teacher education students in particular, with particular emphasis on the Australian context.

### **2.1 International students in English language universities**

Over a decade ago, Barrett-Lennard (1997) reported that international students in Australia were not receiving the support needed for successful university study. These students needed help in preparing for IELTS and also with learning at Australian universities. Academic courses that addressed both needs were recommended. Krause et al (2005) investigated the changes over a ten year period in the experiences of first year students in Australian universities. They found that international students were less satisfied that their expectations had been met than were their domestic counterparts. Likewise, the analysis of the first year experiences of students from equity groups revealed sufficient subgroup differences to warrant institutional strategies designed to meet their specific needs.

Ramsay et al's (2007) study of first year students in an Australian university included comparisons of young (17–21 year olds), mature-aged, local and international students. While there was no significant difference between young and mature-aged students, local students rated their level of adjustment significantly higher than international students. The results of Ramsay et al's (2007) study indicate that the design of first year programs for all first year university students, and some categories of first year students in particular, need much more careful consideration by universities. They recommend an 'intercultural training approach which focuses on general or culture specific critical incidents and

involves the rehearsal of responses to potentially novel and complex situations in order to decrease stress and enhance coping within the particular context. Such (learning) activities could include both local and international students to try to facilitate links between the groups' (Ramsay et al, 2007, p 261).

Zhou et al (2008) report that a learning skills model has been advocated since the mid-1980s in response to the increasing presence of international students in British higher education. Training is seen as necessary for international students to acquire education-culture-specific skills that are required to engage in new learning. They point to practical guidelines for educational interventions to prepare, orientate and skill international students, for instance in areas relating to knowledge of the education culture, communicative competence, pedagogical engagement with local students, and building knowledge networks. With the increasing recruitment of overseas students by British higher education institutions, there has been a growing need to understand the process of students' intercultural adaptation and the approaches that can be adopted by these institutions in order to facilitate and support these students' learning experiences. Gill (2007) suggests that a common assumption in British educational institutions and among academics is that overseas students are 'problematic' and 'demanding'. Gill (2007) undertook a year long in-depth qualitative investigation of the experiences of a small cohort of Chinese postgraduate students' (N=10) in a British university. The investigation explored a three-fold 'stress-adaptation-growth' intercultural learning process for these participants by focusing on the discussions of their lived experience in the UK. It focused partly on their capabilities for engaging in critical reflection and stimulating their access to both Chinese and Western intellectual resources. The outcomes of this transformative intercultural learning included the students' ready accommodation of 'otherness' and adoption of constructive, tolerant, flexible and critical attitudes. Significantly, Gill's (2007) study indicates that British higher education institutions would do well to formalise academic credit-bearing studies in transformative intercultural learning to better ensure that these students acquire the skills and employ strategies to enable them to be successful in their studies.

Weisz and Nicolettou (2004) reported on the experiences of seventy students from China who articulated into various Australian university Business degrees. It was found that their English language proficiency as measured by the IELTS score was insufficient to meet their study requirements, despite having gained university entry. In order to build the students' English language skills, and to support their entry into mainstream classes by their second semester, an intensive teaching program in English language followed by discipline studies helped these students meet the academic challenges of their study programs. The students studying intensively achieved higher average marks in five out of seven subjects compared with the general student cohort who had completed the same subjects over thirteen weeks. It was also found that English language proficiency was only weakly correlated to academic success in two subjects. Indications were that small class sizes and specially designed support programs assisted students to overcome English language limitations.

Skyrme (2007) analysed the experiences of two international students from China beginning studies in a New Zealand university. The focus was on their negotiation of a single course and its assessment requirements. Neither student passed the course. However, one student did develop his English language reading skills, deepen his understanding of the course and improve his competence. The other student's previous successful learning practices proved to be ineffective, and the advice he received unhelpful. Large first-year classes provided no small group interactions with teaching staff, which seems to have hindered the recognition and adoption of suitable learning strategies. Skyrme (2007) recommended: better preparation for the practices demanded within the university; entry requirements addressing more than just English language proficiency, and universities making provision in academic workloads for greater teacher guidance within first-year courses.

Ryan and Viete (2009, p 304), however, have a quite different perspective on the issue of international students in Australian higher education. They point to contradictions in the internationalisation of Australian higher education in which pedagogies emphasise the 'one-way flow of knowledge from teachers to students ... (Whereby) Western knowledge is legitimised as international in focus, yet there is no indication that the focus is developing through genuine intercultural dialogue'. Learning by international students 'is expected to conform to seemingly immutable and often implicit norms laid down by the (Western, English language) academy' (Ryan and Viete, 2009, p 304). They report that those operating within this pedagogical framework construct international students as 'deficient' and advocate 'remedial' approaches in the areas of academic literacy and English language skills. The debate thus focuses on blaming the problem on the English-language proficiency of international students. However, Ryan and Viete (2009, p 306) argue that this position misrecognises the problem, noting that although international students 'will have been screened by an International English Language Testing System (IELTS) exam and are assumed to have adequate language proficiency to participate in their new learning environments, due to the disjuncture between the test and the demands of disciplinary discourses within the university, these may in fact not be a good indicator of their ability to operate within the language of Anglophone academia. ... English-language study ... may not equip them well for the discipline-specific and often fast-paced language in lectures or tutorials, which is saturated with unfamiliar local knowledge, pronunciation and mores of dialogic exchange'. Though it is logical and necessary for English-speaking Western academies to establish entry level English language requirements at an appropriate threshold, Ryan and Viete (2009, p 306) explain that the language proficiency threshold does not protect students against the complexity of academic reading and writing tasks and the cognitive overload students experience in their new learning environment. According to Ryan and Viete (2009, p 309), local Anglophone students prefer to form study groups separate from international and immigrant students, and do not listen adequately to LBOTE students. Likewise, lecturers often do not invite LBOTE students to present knowledge to enrich understandings of topics under discussion. Ryan and Viete contend that lecturers have not learnt to create pedagogies that give international students a sense of security in mixing with locals and growing an expectation that their opinions and knowledge will be valued. International students present opportunities for engaging different educational culture, intellectual heritages and transnational knowledge networks. However, Ryan and Viete (2009, p 304) point to the absence of reciprocity or mutuality in learning across intellectual cultures and to the absence of value being given to international students' knowledge. International students find that 'their own knowledge, linguistically mediated as it is in another language, is seen as being of lesser value' (Ryan and Viete, 2009, p 307). They are concerned with how students are taught to deal with academic tasks which are new to them and which are often tacitly understood by academics in English-speaking Western academies. The issue is less a matter of LBOTE students' general language proficiency than of supporting them in learning the target academic discourse. For O'Loughlin and Arkoudis (2009), this is a matter of gaining legitimacy within their disciplinary community of practice, which they find as particularly related to speaking.

## **2.2 Perceptions of IELTS testing**

Coleman et al (2003) undertook a study of student and staff perceptions of IELTS in Australia, UK and Chinese institutions. Staff and students were surveyed with respect to: their knowledge, perceptions and attitudes; beliefs about the predictive value of IELTS with regard to university demands; the appropriateness of entry levels, and awareness of unprincipled activities. Overall, Coleman et al found that the respondents perceived the IELTS test to have high validity, with students generally satisfied with the entry scores used by their institution. However, they found that staff wished to increase their institution's minimum IELTS entry score because they were less satisfied with the English language abilities of international students.

Brown and Taylor's (2006) survey of examiners on the revised IELTS Speaking Test reported approval of its interview format and assessment criteria. The examiners' main worries were about the use of prompts, the wording of the ranking scales, assessment of pronunciation and rehearsed speech, as well as concerns about topic familiarity and the appropriateness and equivalence of topics.

O'Loughlin (2008) studied the place of IELTS in the selection process of an Australian university by exploring the perceptions that administrative and academic staff and students had about the test. The central research question was: 'To what extent are IELTS test scores used in valid and ethical ways for the purpose of university selection?' (O'Loughlin, 2008, p 150). Among O'Loughlin's key findings, two are relevant to this study. First, the evidence pointed to the prevalence of 'folkloric' beliefs among university staff about English language proficiency and the IELTS test. Some of these beliefs had a firmer basis in research evidence than others. Such beliefs included scepticism about the validity, reliability and 'trustworthiness' of IELTS scores in terms of their power to predict academic success (as O'Loughlin notes, an unrealistic expectation). Second, there was a lack of clearly established equivalence between the IELTS test and other acceptable evidence of English proficiency in university selection policy. O'Loughlin contrasted the selection process of this Australian university with that used at a British university:

"the selection of postgraduate international students at Lancaster University... is radically different to the one described here. There the selection of international students is a complex, holistic decision-making process primarily based on the 'subjective' recommendation of an informed academic staff to the University's senior postgraduate admissions officer. The range of criteria taken into account is extremely rich, including the applicant's academic background, intellectual capacity, evidence of English language proficiency (IELTS or other recognised measure), work experience, the applicant's own argued case for selection, reports from academic and work referees, personal characteristics (such as motivation, age and adaptability) and, in some instances, a follow-up telephone interview ... Other factors influencing their decisions include the offer-acceptance ratio, recommendations from other academic colleagues, the reports of agents and scholarship agencies" (p 182).

### **2.3 The language sub-skills**

Elder and O'Loughlin (2003) investigated the connection between intensive English language study and gains on the band score on IELTS. They studied the progress of 112 LBOTE students enrolled in intensive English language courses at four different language centres in Australia and New Zealand. They gauged students' progress in terms of score gains in the academic module of the IELTS, which was administered at the beginning and end of a ten-to-twelve-week period of intensive English language instruction. Pre- and post-study questionnaires were administered to all participating students and semi-structured interviews were conducted with eighteen students sampled according to their level of gain at the post-test session. Interviews were conducted with administrators and teachers at each of the participating institutions in order to elicit information about the learning environment and the factors they saw as significant in influencing the English language progress of the students. Students made variable progress in English during the three month period with an average gain of about half a band overall.

Rogers (2004) reported that research over the past three decades had shown that repeated exposure to different accents measurably improves comprehension of them. This suggests that materials used in preparing international students from Asia for university study abroad might introduce them to the range of English accents that they are likely to encounter overseas. By incorporating such materials into pre-tertiary teaching, students can be exposed to a range of accents and be better prepared. Rogers

argued that because New Zealand universities accept the IELTS test as a valid measure of students' ability to cope with university demands, that IELTS preparation materials might include both native and non-native accents in their listening exercises.

Carpenter (2005), in reviewing previous studies, particularly highlighted NESB students' difficulties in understanding the content and intent of their lectures, difficulties in understanding everyday language and problems with oral language comprehension and competence.

Moore and Morton (2005) analysed the type of writing required in the two domains of university study and the IELTS test. They compared the standard IELTS Task 2 rubric with a corpus of 155 assignment tasks collected at two Australian universities. They found that whilst IELTS writing bears some similarity with the genre of the university essay, there are also significant differences. Their findings suggest that the type of writing IELTS elicits seems to have more in common with certain public non-academic genres, rather than testing what is thought of as appropriate models for university writing.

Mayor's (2006) research indicates that there are recurrent features in the writing of candidates from Chinese language backgrounds under IELTS test conditions. These include a high level of interpersonal reference, combined with a heavily dialogic and hortatory style. Chinese candidates in Mayor's study used more interrogatives and imperatives than a similar sample of Greek candidates, along with grammatical devices which call for a response on the part of the reader or others. These features gave a polemical tone to the English-medium writing of these Chinese candidates. Mayor argues that it is important to recognise that some Chinese students who performed well in the Chinese education system may import into their English writing a range of practices valued in China, but which may have a negative affect on their scores. For Chinese students to succeed in English-medium universities they need to learn, and to be taught, the models of writing expected of students in those institutions.

Mahdavy (2008) argues that TOEFL and IELTS listening tests differ with respect to their history, theoretical basis, research support and form, leading to suggestions that IELTS is more content-based, task-oriented and authentic. Mahdavy undertook a comparative study of the cognitive demands of these two tests by giving 151 participants a TOEFL listening test and 117 of these same participants an equivalent IELTS test. The participants also completed the Multiple Intelligences Development Assessment Scales questionnaire. Despite the differences between these listening tests, Mahdavy showed that intelligence scores positively correlated with listening scores on both tests and that linguistic intelligence has a significant correlation with listening ability as calculated by these tests. Mahdavy suggests that English language teachers should provide further assistance to language learners who might not enjoy a high level of linguistic intelligence.

## **2.4 English language proficiency and academic results**

In 1993, Vinke and Jochems investigated the question of whether learning in a foreign language affected academic performance, arguing that the precise nature of the relationship between foreign language proficiency and academic success had not been established. Focusing on Indonesian engineering students at the IHE in Delft and using TOEFL scores as their baseline data, they found a cut-off point in the relationship of English proficiency and academic success. They then hypothesised that there was a range of TOEFL scores within which a better command of English increases the chance of academic success to a certain extent and within which a limited lack of proficiency can be offset by other factors (age, effort, mathematical ability). If this was correct, then it would not make a difference whether TOEFL scores were below the lower limit or above the upper limit - in

either case, improvement of proficiency would not contribute substantially to a better academic performance (Vinke and Jochems, 1993, p 282). These researchers suggested that additional research was needed to determine what variables account for international students' academic success or failure, so that 'criteria may be set accordingly and appropriate measures may be taken to raise the chance of international students being academically successful in an English-medium instructional setting' (Vinke and Jochems, 1993, p 284).

Cotton and Conrow (1998) reviewed a number of previous studies into the relationship between various English proficiency test results and academic outcomes, finding that the number of studies which found no statistical significance was roughly equal to the number which found significant correlations, while others yielded either inconclusive or mixed results. They concluded that the relationship between English proficiency and academic outcome was more ambiguous than one might initially suppose. Graham (cited in Cotton and Conrow 1998, p 75) explained the reasons for this thus:

First of all, there has been continued debate about the exact nature of language proficiency; secondly, the difficulties of testing language proficiency and the extent to which it could be measured with a high degree of reliability and validity; thirdly, the moderating variables which affect student performance in the testing situation and a number of intervening variables which affect students' academic performance, finally, the definition of academic success is ambiguous and is open to interpretation and hard to define.

One important study cited by Cotton and Conrow was the Validation Project for the (previous) ELTS conducted by Criper and Davies (1988) which sought to investigate not only predictive validity but also the construct, content and concurrent validity of ELTS. Criper and Davies concluded that the contribution of language proficiency to academic outcome is about 10%, a correlation of 0.3. This suggested that language does contribute to academic success but does not play a major part. Cotton and Conrow (1998) themselves tested the predictive validity of IELTS at the University of Tasmania. In correlating IELTS with students' Grade Point Averages (GPAs), only the reading subtest had a moderate positive correlation with academic results and in the case of the speaking subtest there was a negative correlation (Cotton and Conrow, 1998, p 92). Staff ratings of students' academic performance showed a weak positive correlation between the reading and writing subtests and academic performance. Staff ratings were then correlated with GPAs and these showed a reasonably strong correlation ( $r=0.73$ ). In the second semester of study by the students under investigation, there appeared a link between IELTS reading and writing subtest scores and students' self-ratings of academic performance. While no significant correlations were found for IELTS global scores, there appeared to be weak correlations between the reading and writing subtest scores with two of the three measures of academic outcomes. The reading subtest scores in particular were better able to predict subsequent academic performance. Cotton and Conrow (1998, p 109) concluded that language proficiency alone was no guarantee of success as other variables may have equal or more importance.

Dooley and Oliver (2002) studied the predictive validity of IELTS in Faculties of Business, Science and Engineering. Correlations between the students' semester-weighted averages (SWAs) and IELTS scores were calculated for the entire group and by discipline. Dooley and Oliver cited previous argument that students who scored higher on a standard English test would have a greater chance of future academic success, though some previous researchers had argued that it is mainly at low levels of proficiency that language makes a difference. Dooley and Oliver (2002) also found that English language proficiency was only one among many factors that affect academic success. They concluded that there was little evidence for the validity of IELTS as a predictor of academic success, although

they did find the reading module the better predictor of academic success as it was the only subtest of the four macro-skills to achieve a significant correlation. Dooley and Oliver went on to suggest that 'overseas students who do not fully meet admissions criteria in terms of their language may well have the potential to succeed academically' (2002, p 51).

Feast (2002) investigated the relationship between English language proficiency (as measured by IELTS) and performance at university (based on GPA). Feast also investigated the related issue of the trade-off between raising IELTS scores and the consequent loss of international students, i.e. should the current minimum entrance IELTS scores be increased so that the students who gain admission to university courses have a reasonable chance of success? Using multilevel regression analysis on the results of 101 international students, Feast found a positive, but weak, relationship between English language proficiency, as measured by IELTS scores, and performance, as measured by GPA. Feast recommended that the overall IELTS score be kept at 6.0 for undergraduate students and raised to 6.5 for postgraduate students (but that students be required to have a score of 6.0 minimum in the reading and writing modules in both cases). Feast projected that the implementation of this recommendation would involve a loss of just over 40% of prospective undergraduate international students and result in a GPA gain of 0.9 percent and a loss of 70% of postgraduate students with a GPA gain of 4%. Recognising that these figures may be 'unacceptably high', Feast recommended raising support levels for international students.

Bayliss and Ingram (2006) investigated the questions:

- 1 To what extent is the language behaviour implied by their IELTS scores reflected in the language behaviour (in all four macro skills) of university students during the first six months of their degree program?
- 2 To what extent is the language behaviour observed adequate for the study program being undertaken by the student?
- 3 Are there implications for raising or lowering common IELTS requirements for entry to undergraduate or graduate courses?

Twenty-eight international students were interviewed, given a questionnaire and observed in a variety of class types, and a rating scale was used against which researchers could measure the students' language performance and compare their language behaviour with that implied in their IELTS scores. On question (1), findings suggested that IELTS scores could quite accurately predict students' language behaviour in the first six months of their study program but that individual students might perceive their language proficiency levels quite differently. On questions (2) and (3), answers varied with the Faculty in which students were enrolled, however Bayliss and Ingram cautiously suggested that there may be implications for raising the IELTS entry levels for courses which require students to use spoken English in vocational training contexts in the early stages of their studies.

Phakiti (2008) reported on a study aiming to predict international postgraduate students' academic achievement using three variables, viz. English language proficiency, English reading proficiency and metacognitive knowledge of strategic reading. The participants were 125 Chinese international postgraduate students who were in their second semester in an Australian university. Their English language proficiency was measured by the IELTS tests, in particular their English reading proficiency, and their metacognitive knowledge of strategic reading was measured by a Likert-scale questionnaire. Through the analysis of the questionnaire, it was found that their English language proficiency, English reading proficiency and metacognitive knowledge of strategic reading accounted for 7%, 10% and 5% of their academic performance respectively.

A useful overview of the research on the correlation between English proficiency and subsequent academic performance is contained in Davies (2008) and a further useful discussion in Hyatt and Brooks (2009).

## 2.5 International students, English language proficiency and teacher education

Language use in the classroom has, of course, been the subject of detailed concern in education, especially since the early 1970s, with the growth of the language-across-the-curriculum movement. This concern with language and the huge number of studies devoted to it has tended to focus on language as an instrument of pupil learning. Teacher language use in this tradition tends to focus on language-as- pedagogical-strategy. This literature is not concerned with the teacher's language background though it is highly suggestive in terms of teacher language use. A selection of the classic works in this huge corpus includes Britton, 1970; Wilkinson, 1971, 1975; Creber, 1972; Cazden et al, 1972; Cashdan et al, 1972; Rosen and Rosen, 1973; Bullock, 1975; Marland, 1977. Some of this has focused specifically on the discourse structures of classrooms (Barnes et al, 1971; Barnes, 1976; Coulthard, 1977, pp 93ff; Sinclair and Brazil, 1982; Stubbs, 1976, 1983). Where the teacher's own proficiency with spoken language, in particular, has been an object of interest in teacher education it has tended to be dealt with often as a micro-skills issue, highlighting skills such as explaining, discussing, questioning etc. (e.g. Turney et al, 1983a, 1983b). If one consults generalist texts for the beginning teacher, the issue of the teacher's use of spoken language is touched on when such texts deal with specific language tasks, such as:

- questioning (Barry and King, 1988, pp 77ff, 363ff; McBurney-Fry, 2002, pp 87ff; Arends, 2004, pp 429ff; Zwozdiak-Myers and Capel, 2005, pp110ff)
- explaining (Barry and King, 1998, pp 63ff; McBurney-Fry, 2002, pp 84ff; Arends, 2004, pp 284ff; Zwozdiak-Myers and Capel, 2005, pp 109-10)
- discussing (Barry and King, 1988, pp 372ff; McBurney-Fry, 2002, pp 92-93;Arends, 2004, pp 430ff; Zwozdiak-Myers and Capel, 2005, pp 113-14)
- responding to pupils (Barry and King, 1988, p 83ff; McBurney-Fry, 2002, pp 90ff; Arends, 2004, pp 287ff, 312ff, 441ff; Zwozdiak-Myers and Capel, 2005, pp 114; Capel and Gervis, 2005, p 132-33).

In terms of specific language *qualities*, such texts tend to deal only with the very global notion of 'clarity of communication' (Barry and King, 1988, pp 61ff, 354ff; Arends, 2004, pp 283-84). However, based largely on research in the contexts of non-native-speaker teachers/teaching assistants and her own observations, Elder (1993a, p 237) has developed a more extensive list of 'desirable features of teacher communication', which she has applied to non-native-speaker teachers, viz:

- intelligibility
- fluency
- accuracy
- comprehension
- use of subject-specific language
- use of the language of classroom interaction
- overall communication effectiveness



Moreover, Elder has elaborated on a 'partially indicative' (Elder, 1994b, p 10) inventory derived from Ellis (in Elder, 1994b, p 6ff) of typical teacher tasks in terms of language use. These are derived from studies of L2 teachers, and therefore aimed specifically at the teaching of a language and address teacher use of the target language. *They include medium-oriented interactions, message-oriented interactions, activity-oriented interactions, interactions which create a framework within which teaching can take place and extra-classroom language use.* LBOTE teachers operating in English and teaching subject content in a number of curriculum areas are not focused on *medium-oriented interactions* to the same degree as L2 teachers, however other areas of the inventory are usefully suggestive for considering the tasks which LBOTE teachers have to address while teaching subject content in a language which is not their L1. These tasks include:

- *message-oriented interactions*: e.g. explaining, categorising, labelling, presenting information, narrating
- *activity-oriented interactions*: e.g. giving instructions
- *framework interactions*: e.g. directing, disciplining, explaining, questioning, responding, rephrasing
- *extra-classroom language use*: e.g. selecting and preparing material, simplifying texts, writing memos, talking to parents, reading professional development material, attending professional development seminars (Elder, 1994b, pp 6-9)

In Australia the history of providing teacher-specific language-proficiency scales has tended to be largely related to proficiency for prospective teachers of foreign languages rather than of the English language proficiency of international student-teachers across the full range of curriculum areas (Elder 1993c; Elder, 1994a, 199b; Elder et al, 1994; Wylie and Ingram, 1995a, 1995b; Iwashita and Elder, 1997; Consolo, 2006). However, there do exist in Australia two quite widely used scales of English language proficiency focusing specifically on the language of the classroom and the proficiencies assumed to be needed in that context. Both are for teachers fully or partly trained overseas who are seeking registration in Australia and both test the usual four macro-skills (reading, writing, speaking and listening). One of these tests is the version of the International Second Language Proficiency Rating (ISLPR) for teacher professional registration (cf ISLPR nd). In this test, speaking skills are judged throughout an interview on education-related topics, including the socialisation of young people. Some tests may include a simulation in which candidates take on the role of a teacher interacting with a student (the tester) about some aspect of the educational process (e.g. classroom behaviour). Listening is related to texts on education and the socialisation of young people, which may be news stories, news commentaries, interviews, talk-back, documentary material, community announcements or advertisements. The reading test is selected from curriculum materials (e.g. syllabuses or text-books), research reports, material from the media or professional journals (e.g. news stories, editorials, 'letters to the editor', columnists' opinions, feature stories), community information (e.g. brochures), advertisements or material related to conditions of employment (e.g. newsletters from an employer or union). Candidates are also asked to read aloud a short text, which is likely to be a school notice delivered to the classroom, or a self-contained section of a textbook or a 'big book' (depending on the candidate's teaching specialisation). The writing test in one section posits an audience of students, parents, colleagues or other members of the immediate school community, or officers in the education system. A second task is to write a memo, report, article or open letter (e.g. a letter to the editor of a newspaper) in which candidates will be expected to express opinions about education or the socialisation of young people (ISLPR nd).

The second teacher-specific scale available in Australia is the Professional English Assessment for Teachers (PEAT), developed for the New South Wales Department of Education and Training (NSWDET), but recognised in many jurisdictions. The PEAT instrument is designed to determine the level of competence in English of overseas trained teachers who wish to gain approval to teach in NSWDET schools. Like the teacher version of the ISLPR, PEAT is not an assessment of subject matter competence but an assessment of English language skills within the educational context. The result obtained in each language component of the PEAT is meant to be an indication of whether a candidate's proficiency is of a standard which would enable them to teach and interact effectively and confidently in a school setting in NSW. The reading test focuses on a NSW DET policy text, a workplace related text and a sample of handwritten pupil writing. Listening is to discussions and a monologue on education-related topics. In the writing test, candidates are asked to write an incident report, a pupil handout or a note home and to give written feedback on a pupil's work after identifying errors on a sample of pupil writing. Speaking is judged on an interview, a role play and a presentation (UNSW Institute of Languages, 2009). As can be seen from these brief descriptions, there is an attempt in each case to provide workplace authenticity, and certainly context-specificity. Relevant scores from both the ISLPR (teacher version) and the PEAT are included in Table 4 below as points of comparison to the IELTS test.

The question of 'authenticity' raises many questions about language testing instruments. Twenty five years ago Spolsky (1985) asked the question, 'What are the limits of authenticity in language testing?' In other words, to what extent can a testing task be made authentic 'so that it constitutes an example of normal and natural language behaviour on the part of both examiner and candidate (?) And if it cannot be made authentic, what difficulties does this present to the validity of the test and the correct interpretation of its results?' (Spolsky 1985, p 33). Elder, in a series of articles on rater reliability has raised interesting questions about the specific testing of language proficiency in the context of classroom performance. In Elder 1994a, she reported on an Italian oral proficiency test (for teachers of Italian) which required candidates to simulate the teacher role. Raters suggested that candidates who took the teacher role seriously and attempted to produce comprehensible input for an imaginary semi-proficient L2 audience placed themselves at a linguistic disadvantage by deliberately simplifying speech and slowing their rate of delivery. Elder concluded that there 'may be a fundamental incompatibility between the assessment of language proficiency, which assumes a developmental continuum involving an incremental increase in range and complexity of language use as proficiency progresses, and certain kinds of occupation-specific proficiency where certain features of pragmatic or strategic competence such as simplicity and clarity may be valued over elaborateness' (Elder, 1994a, p 56). Admittedly, students in the situation with which we are concerned here are not necessarily dealing with an imaginary audience who are less proficient in the language (in the case being studied here, the L1 language of instruction) than they themselves are – in fact, in schools where English is the L1 of most students, the reverse is likely to be true. However, in taking account of the need to use the language skills for teachers listed earlier – such as by simplifying explanations, for example - similar issues to those raised by Elder (1994a) apply. In fact, an earlier study by Elder addressed this issue. Elder (1993a) raised the question of linguistic experts such as trained language testers viewing second language performance differently from other 'linguistically naive' native speakers, in this case subject-specialist teachers. An implication from earlier research quoted by Elder is that if raters are concerned with gauging the impact of second language communication on the wider native speaker population, linguistic expertise may be a liability. In this article, she reported on a classroom-based assessment procedure to monitor the English proficiency of graduates from non-English-medium universities who were training to be teachers of Maths and Science in Australian secondary schools. The study arose

out of a concern that substantial numbers of LBOTE graduates entering teacher education courses were unable either to function effectively during their school-based teaching practice or ultimately to perform credibly as teachers. The assessment procedure offered a means for determining the extent to which their difficulties were related to language. While the findings are reported cautiously, Elder found trends that subject specialist raters of LBOTE teacher trainees emphasised interactive strategies above all else, and considered language proficiency in terms of real-world criteria (i.e., are teachers creating the necessary conditions for classroom learning to take place?) In behaving thus, they assessed language in terms of successful task completion. Language experts, on the other hand, tended to focus on the quality of the language sample elicited through teaching. Elder concluded that 'If we accept that there are instances where the formulation of an acceptable and intelligible message depends on discipline- or occupation-specific knowledge, the involvement of subject specialists as assessors (notwithstanding the strain that this may place on reliability) should be regarded as a condition of test validity' (1993a, p 249; see also Elder and Brown, 1997, p 77). The 'Hymesian' 'non-linguistic' factors (Elder, 1993a, Elder and Brown, 1997) in communicative competence which subject specialist teachers favoured in their rating of performance – sensitivity to audience, interactive skill and personal style – are indeed crucial attributes of teacher communicative competence, though they are not necessarily rated highly in language testing by language experts. In terms of using teaching simulations rather than on-the-job assessments, Elder and Brown (1997) built on Elder's earlier (1993a) conclusions by arguing that, 'Rather than being seduced by the appearance of authenticity into accepting that performance tests are necessarily more valid than traditional types of assessment, we need to find ways of ensuring that there is a reasonable degree of fit between behaviours elicited from candidates in the artificial environment of the test and actual performance in the target domain' (Elder and Brown, 1997, p 77).

In addition, Elder has suggested a number of principles for the testing of teachers of languages (L2) which can again be usefully applied in the context of LBOTE teachers operating in English. Because most possible uses of teacher language cannot be accommodated in a test, she suggests that the following principles drive testing:

- *scope*: all four macro-skills should be included, a broad range of language functions and each category of Ellis' inventory (referred to above).
- *frequency*: giving preference to tasks most frequently performed and greater weighting to speaking
- *importance* (Elder, 1994b, p 10).

Elder reported on a performance test (Elder, 1994b, pp 11ff) based on these principles, which consisted of:

- story reading
- story retelling
- assigning and modelling a role play
- cultural presentation
- pupil error correction

Again, most aspects of such a test would be appropriate in the situation of LBOTE teachers working in English for a range of curriculum areas.

Elder has also investigated the predictive validity of IELTS with specific respect to teacher education students at a range of tertiary institutions in Melbourne. Specifically (Elder 1993b) she has asked:

- 1 Is performance on the IELTS test a reliable predictor of success in postgraduate Diploma of Education courses? Elder found that IELTS can be regarded as a reasonably good indicator of short term performance in teacher education courses, though in the long term, ‘the predictive power of IELTS diminishes’ because of improvements in English language ability and the nature of the language variables likely to affect teaching performance (Elder, 1993b, pp 78-80).
- 2 How does IELTS compare with each institution’s screening procedures as far as the accuracy of its predictions is concerned? Elder found that no substantial claims could be made either for or against IELTS when compared with other locally-applied procedures (Elder, 1993b, p 82).
- 3 What is the optimum IELTS threshold for entry to teacher education? To this, Elder answered that above global bands 4.5 and listening band 5.5, too many other factors (e.g. subject knowledge, cultural adaptability, understanding of classroom role relationships) were likely to interact with language ability in determining progress and should be taken into account in making initial student selection.
- 4 Do scores on the reading, writing, listening and speaking components of the IELTS test predict the degree of difficulty experienced by candidates in performing coursework tasks? No significant correlations were found, but the patterns in the data did show that the scores may have some value as a means of diagnosing the difficulties that candidates may experience with the language demands of their study (Elder, 1993b, p 86).
- 5 Does second language instruction/exposure during the training year affect the relationship between predictions and outcomes? The poor return rate of questionnaires made it impossible to calculate the impact of this variable on the strength of IELTS predictions (Elder, 1993b, p 87).

The data confirmed evidence from previous studies that it is at low levels of proficiency that language makes a difference (Elder 1993b, p 72). Patterns in the data did show that the IELTS scores have some value as a means of diagnosing the difficulties that candidates are likely to experience due to the language demands of their teacher education studies- thus, the lower their writing score, the more likely it is that the candidates will perceive essay writing as problematic; listening ability predicts difficulties with lecture and tutorial comprehension better than do scores for other test components, and the reading test score is the best predictor of difficulties reported in reading academic texts. This study indicated that English language support may continue to be necessary even at the higher levels of language proficiency. Elder (1993b, p 88) recommended that entry level thresholds regarding English language proficiency should be set by universities in accordance with their capacity to provide such support. (More recently, O’Loughlin and Arkoudis [2009] found that the degree of English language support students sought within the university and the degree of contact with English they had outside the university strongly influenced their language improvement).

Viete (1998) pioneered a culturally sensitive assessment procedure, the Diploma of Education Oral Test of English (DEOTE) for gauging the oral communicative competence in English of international student-teachers. This test was developed because ‘none of the large-scale testing systems (TOEFL/TSE/TWE, IELTS) or rating scales (ASLPRs) adequately reflected the demands of teacher training on communication skills in English’ (Viete, 1998, p 173; Viete’s reference is to the former Australian

Second Language Proficiency Rating, now superseded by the International Second Language Proficiency Rating). Specifically, the IELTS speaking test 'did not provide opportunities for candidates to produce the range of oral communication expected of student teachers in tertiary tutorials and secondary classrooms and thus lacked construct and content validity' (Viète, 1998, p 173). Some of the skills which large-scale tests are unable to elicit, but which DEOTE does elicit include 'the requirements of effective listening in multichannel conversation, the need to manage speculative language and the skill of explaining a concept in more than one way' (Viète, 1998, p 174). The DEOTE has been used to rate international students on five levels of performance from 'unsatisfactory' to 'advanced', with two levels used to indicate whether the candidate requires extended or minor support. Viète (1998, p 172) argues that support programs need to be 'designed to extend students' communicative competence and experience with the local educational culture once they are on course.' An important feature of the DEOTE is that it is administered by teacher educators. 'Two trained raters are used, one an expert in the disciplinary discourse in which the candidate has qualifications (e.g. Science), and the other a practitioner in TESOL and teacher education' (Viète, 1998, pp 179-180). The ratings generated against the assessment criteria by the two teacher-educator raters are moderated through discussion that clarifies differing interpretations. This means that the teacher educators who are to teach these students not only meet the students, but learn more about the relationship between linguistic performance and content area discourse. Thus in terms of cost effectiveness there is the added 'beneficial washback from the test. Teacher [educators] involved in testing have become more aware of the communication skills their [teacher trainees] need to develop, and have attempted to take account of these in their course design' (Viète, 1998, p 180). The dual function of this test, in assessing student-teachers and in educating teacher educators about how to address their needs, must be taken into consideration in judging the cost effectiveness of the DEOTE.

In a study of the authenticity of the testing of Language for Specific Purposes (LSP) for teacher education students, Elder (2001) investigated:

- 1 whether the domain of teacher proficiency was distinguishable from other areas of professional competence or from 'general' language proficiency;
- 2 what constituted appropriate task design on a teacher-specific instrument;
- 3 the role of non-language factors in a candidate's response to a contextualised test task.

Though agreeing that teaching required highly specialised language skills, Elder argued for caution in the application of LSP testing because of the indeterminacy of performance-based tasks as a means of measurement.

Woodrow (2006) investigated the predictive validity of IELTS for Education postgraduate coursework students, along with academic staff attitudes to English proficiency. IELTS subtest scores were correlated against students' semester 1 grade point averages (GPA). Other personal experience variables obtained through a questionnaire were included in the analysis. Academic staff completed questionnaires concerning their views on the English proficiency of students. The results indicated that English language proficiency as measured by IELTS is moderately predictive of academic achievement in the first semester of study for the student sample. Weak but significant correlations were obtained between overall IELTS bands and GPA. There were significant correlations between writing, speaking and listening subtests and GPA. No significant correlational relationship between other variables such as professional experience and academic achievement, as measured by semester 1 GPA, was found. The analysis indicated that at a lower level of English, the relationship is stronger than at a higher level. Thus, for students scoring 6.5 or lower, proficiency may influence their achievement, whereas with students scoring 7 and above, English proficiency does not influence academic performance.

## 2.6 International students, teacher education and general support

Success in teacher education depends not only on the grades assigned for academic essays and tutorial presentations, but also on performance during the school-based teaching practicum... It has been argued convincingly that the language of the classroom differs in its structure from other forms of discourse... and that particular types of interaction are more conducive to student learning than others... Teachers need to be able to correctly model important information, to tailor their language to make it intelligible to students, to give clear instructions and to process and synthesize learner feedback which may be expressed in non-standard varieties of English. Without high levels of comprehension and considerable flexibility and fluency of expression it is unlikely that non-native speakers who are training to be teachers will perform effectively in this crucial area of their professional education. Effective classroom management will also depend heavily on choice of register and appropriate non-verbal behaviour, both of which assume an understanding of role relationships between teacher and student. The norms underpinning classroom role relationships vary significantly across cultures and this may cause communication difficulties for those educated elsewhere (Elder 1993b, p 73)

Moving beyond the specific issue of language testing, in investigating the optimum kind of support available to NESB students in teacher education, Cruickshank et al (2003) focused on the development of provision for a group of 110 overseas-trained teachers undertaking teacher education at the University of Sydney between 1999 and 2003. They concluded that the most effective model of support for international students was one which was centred on content-based units, that is support which linked language with specific course content. This included support courses, individual mentoring, tutoring and self-directed learning (Cruickshank et al 2003, p 245).

McCluskey's (2004) study also focused on the experiences of an international teacher education student in rural Queensland and investigated the two questions: 'What are the major cultural differences encountered in schools by pre-service international teachers in Queensland? To what extent do these differences impact on their professional roles?' The subject of the study did not perceive large cultural differences that could impact on his professional role, but McCluskey (2008) found that there were issues not conveyed to the student at a point when strategies could have been implemented to counteract them.

Carpenter (2005) reported on a study in one Australian university in which NESB teacher education students had experienced difficulty communicating with children because of accented English language, as well as difficulty managing the culture of schooling and the expectations placed on them during the practicum. Each student experienced a level of 'shock' when confronted with the realities of classroom life especially with regard to behaviour management (pp 5-6). These students' lack of success on practicum showed their need for greater support than was available from the supervising teachers. The School of Education put in place a one-to-one mentoring project with a small group of students which, at the time of Carpenter's reporting on it, appeared to be working successfully. She called for adequate university support for NESB students which was discipline-specific.

Campbell et al (2006) studied the perceived concerns and challenges of international students prior to, and during, teaching practicum. They argue that, 'Whilst NESB student teachers perceived similar anxieties about practicum to their peers, they also confront language, communication and cultural differences which may hinder their successful completion of field experiences in schools.' (Campbell et al 2006, p 2). They reported on a program designed to enhance these students' confidence and skills in undertaking their field experience placements. The main findings from interviews during the program were that: participants claimed that they were more familiar with teacher-centred approaches (p 7);

that they were embarrassed to have their language corrected by supervising teachers, especially in front of children and parents (p 7) and that behaviour management was a prominent area of stress due to an unfamiliar pupil-teacher dynamic (p 8).

Spooner-Lane et al (2007) examined the changing perceptions of international student teachers enrolled in a one-year teacher training program and explored the issues they faced as they prepared for practicum in order to determine whether international student teachers' beliefs and expectations of practicum were incompatible with the realities of schooling in Australia. International student teachers' perceptions of practicum before and following school experiences were examined. Before they engaged in the practicum, most of the international student teachers felt relatively confident about the upcoming practicum experience. The findings suggested that international student teachers' positive perceptions of practicum were related to a somewhat simplistic view of teaching. They seemed unaware of how their teaching beliefs, formed in their country of origin, were misaligned with teaching in an Australian context. The practicum experience itself instigated a deeper awareness of the complexities of teaching and a more sophisticated understanding of the teaching knowledge and skills they lacked, which actually reduced their confidence. International student teachers realised that they needed to learn more about Australian school culture and teaching practices. Spooner-Lane et al (2007) concluded that Australian universities can better support international student teachers enrolled in the one-year teacher education training program by offering these students a differentiated mode of delivery for practicum, including a gradual developmental approach and mentoring training for supervising teachers.

In sum, for LBOTE teacher education students, successful course completion means that the language issues they must address encompass not just listening and reading comprehension or success in academic writing, but their 'performance' of oral and written English in classrooms, as well as a number of issues such as the differences in school cultures from that of their own backgrounds.

### **3 RESEARCH METHODOLOGY**

#### **3.1 Site selection**

This research focuses on graduate entry secondary teacher education students whose first degree was completed in a non-Anglophone culture. It was initially anticipated that the relevant courses might provide significant numbers of appropriate students, possibly clustered in particular secondary Method areas, especially Mathematics, Science, Technology and Languages. It was also anticipated that the Melbourne and Sydney metropolitan areas would potentially contain the greatest numbers of such students. During the first stage of the research in late 2008, all University registrars in the Melbourne and Sydney metropolitan areas were contacted and asked about the proportion of graduate entry secondary teacher education students who had been subject to IELTS (or similar) testing in their 2008 cohort. Relevant course co-ordinators were also directly contacted and asked to provide details on the proportion of students who had been subject to IELTS testing in their courses. To ensure reasonable proportions of students in the Faculties to be consulted, we widened the consultation to include one university in Adelaide with a significant proportion of international students enrolled in teacher education. Eventually four teacher education Faculties with the highest proportion of relevant students in graduate entry secondary education were chosen as case studies for interviews.

### 3.2 Site studies

The relevant universities were in Sydney (1), Melbourne (2) and Adelaide (1). For each of these Faculties, semi-structured interviews were conducted with: academic staff teaching in the courses, course co-ordinators and one Head of Faculty. In addition, a focus group was conducted with one group of students in one university, and with one group of academics in another. Representatives of one state teacher registration authority were also interviewed. The interviews were aimed at determining views on:

- 1 the English language proficiencies of LBOTE students (specifically those who have been subject to IELTS testing)
- 2 the degree to which English language proficiency is an issue in the practicum success of these students (relative to other factors such as lack of familiarity with Australian schooling practices)
- 3 the adequacy of current IELTS admissions scores into the relevant courses
- 4 what an adequate IELTS admissions score into teacher education courses might be

### 3.3 Data analysis

The analysis of the interviews and relevant policy texts used established procedures for data reduction verification (Emerson et al 1995; Ezzy 2002; Hatch 2002). Data reduction involved the usual procedures:

- 1 open coding: assigning labels to meaningful chunks of information (data segments) to analyse the whole data set
- 2 axial categorisation was used to subdivide this mass of data segments according to their emergent patterns, relationships and themes
- 3 interpretive summaries were produced focusing on the tentative explanations and conclusions to be drawn from this analysis; these summaries tested the logical consistency of claims
- 4 data verification was by means of the research team members' cross-checking each other's analysis to ensure the robustness of conclusions based on supporting evidence.



## 4 BASELINE DATA

Baseline data on IELTS scores into graduate entry secondary education in all metropolitan universities in Sydney, Melbourne and Adelaide which offer teacher education are contained in the following Tables. Included in each Table are comparative data on general postgraduate entry and the requirements of local teacher registration authorities.

INSTITUTION	Graduate entry pre-service courses	IELTS scores for entry into Graduate entry pre-service courses	IELTS General university entry level scores (Postgraduate)	IELTS Requirements of NSW Institute of Teaching	Is course specifically targeted at international students?
Australian Catholic University	Master of Teaching/ Graduate Diploma in Education (MTeach/DipEd)	7.0 overall (min. 7.0 in each subset)*	Varies	7.5 overall (min.8.0 in Speaking/ Listening; Min. 7.0 in Reading/ Writing)	No
Macquarie University	Graduate Diploma in Education (DipEd)	7.0 overall (min. 7.0 in Reading/ Writing; min. 6.0 in Speaking/ Listening)	6.5 overall (min. 6.0 in each section)	As above	No
University of NSW	Diploma in Education (DipEd)	6.5 overall (min.6.0 in each section)	6.5 overall (min. 6.0 in each section)	As above	No
University of Sydney	Master of Teaching/Bachelor of Teaching (MTeach/BTeach)	7.5 overall (min.8.0 in Speaking/ Listening; min. 7.0 in Reading/ Writing)	6.5 overall (min.6.0 in each sub-test)	As above	No
University Technology of Sydney	Bachelor of Teaching in Secondary Education (BTeach)	7.0 overall (min. 7.0 in Writing)	6.5 overall (min. 6.0 in Writing)	As above	No
University of Western Sydney	Master of Teaching (Secondary) (MTeach)	7.0 overall (min. 6.0 in each subset)	6.5 overall (min.6.0 in each subset)	As above	No

\* The terms 'sub-test', 'subset', 'section', 'module' and 'band' are the terms used on the universities' own websites and refer to the specific sub-areas of Reading, Writing, Listening and Speaking on the IELTS test

**Table 1. IELTS entry score data (Sydney universities)**

INSTITUTION	Graduate entry pre-service courses	IELTS scores for entry into Graduate entry pre-service courses	IELTS General entry level scores (Postgraduate)	IELTS Requirements of Victorian Institute of Teaching	Is course specifically targeted at international students?
Australian Catholic University	Master of Teaching/ Graduate Diploma in Education (MTeach/DipEd)	7.0 overall (min. 7.0 in each subset)	Varies	7.0 in each band	No
Deakin University	Graduate Diploma of Education (Secondary) (DipEd)	7.0 overall	6.5 overall (min.6.0 in each subset)	As above	Yes
La Trobe University	Graduate Diploma in Education (Secondary) (DipEd)	7.0 overall (min. 6.5 in each band)	6.5 overall with (min.6.0 in each band)	As above	No
Melbourne University	Master of Teaching (MTeach)	7.0 overall (min. 7.0 Writing/min. 6.0 in others)	6.5 overall (min.6.0 in each subset)	As above	No
Monash University	Graduate Diploma of Education (Secondary) (DipEd)	7.0 overall (min. 6.5 on Reading/Listening; min. 7.0 on Writing/Speaking)	6.5 overall with (min.6.5 on Reading/Listening; min. 6.0 on Writing/Speaking)	As above	No
RMIT University	Graduate Diploma of Education (Secondary)	7.0 overall with 6.5 in each section	6.5 over all with (min.6.0 in each subset)	As above	No
Victoria University	Graduate Diploma in Secondary Education (DipEd)	7.0 overall (min. 6.0 on all)	6.5 overall (min.6.0 in each subset)	As above	No

**Table 2. IELTS entry score data (Melbourne universities)**

INSTITUTION	Graduate entry pre-service courses	IELTS scores for entry into Graduate entry pre-service courses	IELTS General entry level scores (Postgraduate)	Requirements by Teachers Registration Board of South Australia	Is course specifically targeted at international students?
Adelaide University	Graduate Diploma in Education (DipEd)	6.5 overall (min. 6.0 in all subtests)	6.0 overall (min. 6.0 in Writing/ Speaking; min. 5.5 in Reading/ Listening)	7.0 overall (min. 7 in each of the modules)	No
University of South Australia	Graduate Diploma in Education (DipEd) Master of Teaching (MTeach)	7.0 overall (min. 6.5 in Reading/ Writing) 6.5 overall (min. 6.0 in Reading/ Writing)	6.0 overall	As above	No Yes
Flinders University	Master of Teaching (MTeach)	7.0 overall (min.7.0 in all sub-bands)	6.0 (min. 6.0 in Speaking/ Writing)	As above	No

**Table 3. IELTS entry score data (Adelaide Universities)**

Table 4 represents the language proficiency requirements of the teacher registration authority in each state and territory of Australia (except the Australian Capital Territory in which at the time of writing, there is no mandated teacher registration authority). Other relevant scores (e.g. ISLPR, TOEFL, PEAT) are provided for comparison.

<b>Teacher registration authority</b>	<b>IELTS scores</b>	<b>Other recognised tests</b>	<b>Who has to meet these requirements?</b>
New South Wales Institute of Teachers	Minimum overall score of 7.5 including a minimum result of 8.0 in both the Speaking and Listening modules and 7.0 in Reading and Writing.	ISLPR: at least 4+ in each of the four areas: Speaking, Listening, Reading and Writing. PEAT: Band A in each of the four areas: Speaking, Listening, Reading and Writing.	An applicant whose first language is not English and who did not gain the majority of his or her qualification in English in a country where English is the main language (i.e. Australia, Canada, New Zealand, Republic of Ireland, United Kingdom Republic of South Africa and the United States of America) unless they can show evidence of having taught successfully in English in a school in a country where English is the main language for a substantial period of time
Victorian Institute of Teaching	Level 7 in each of the areas of Speaking, Listening, Reading and Writing.	ISLPR: Level 4 in each of the areas of Speaking, Listening, Reading and Writing. PEAT: Band A in each of the areas of Speaking, Listening, Reading and Writing.	Applicants who have not completed teacher education qualifications at an Australian institution or in English

<b>Teacher registration authority</b>	<b>IELTS scores</b>	<b>Other recognised tests</b>	<b>Who has to meet these requirements?</b>
South Australian Teachers Registration Board	Minimum 7 for each of the modules Listening, Reading, Writing and Speaking.	ISLPR: minimum score of 4 in each of the macroskills of Listening, Reading, Writing and Speaking PEAT: Band A in all four components of Listening, Reading, Writing and Speaking.	All applicants for teacher registration, with the exception of graduates from pre-service teacher education programs in: Australia, Canada, New Zealand, Republic of Ireland, South Africa, United States of America, United Kingdom
Queensland College of Teachers	Minimum 7 in each of the areas of Reading, Writing, Speaking and Listening	ISLPR: 4 for Speaking, Listening and Reading and 3+ for Writing	Persons who have not undertaken their pre-service teacher education program in English in an exempted country (i.e. Australia, Canada, New Zealand, Republic of Ireland, South Africa, United Kingdom and United States of America)
Western Australian College of Teachers	An overall band score of 8 and if one individual score is less than 8, it must be 7 or 7.5 in either Reading, Writing, Speaking or Listening.		Applicants who did not complete secondary education in Australia, Canada, Ireland, New Zealand, United Kingdom or United States of America or did not complete teacher education in these countries or are not registered teachers in Australia or New Zealand or did not complete both secondary education and teacher education in English if from South Africa
Teacher Registration Board Tasmania	Minimum 7 in each of Speaking, Listening, Reading and Writing.		Applicants who have not completed their pre-service program of teacher education in: Australia, Republic of Ireland, Canada, United Kingdom, New Zealand, United States of America (or graduates from the University of the South Pacific)
Teacher Registration Board of the Northern Territory	7 or higher in each of Listening, Speaking, Reading and Writing.	ISLPR: 4 or higher in each of the macro-skills of Listening, Speaking, Reading and Writing. PEAT: Band A in all four components of Listening, Speaking, Reading and Writing	Applicants who did not graduate from pre-service teacher education programs in Australia or English-speaking Canada, New Zealand, Republic of Ireland, and United States of America.

**Table 4. Language proficiency requirements of Australian teacher registration authorities**

## 5 LBOTE TEACHER EDUCATION STUDENTS' PERSPECTIVES

Students in one university's Faculty of Education were interviewed as a focus group on their perceptions of English language proficiency issues. These students were from a metropolitan university which had created a specific secondary teacher education course for students on international student visas. These students' first degrees had been undertaken in a country in which the majority language was not English. Ten students took part in the focus group. One student was Japanese, one Sri Lankan and the rest Chinese or Indian. They represented the teaching areas of English, Mathematics, English as a Second Language, Japanese, Chinese and Social Sciences. All had gained an overall IELTS score to enter the course of between 6.5 and 7.5. All had completed their full practicum requirements and most had completed the course. The focus group responses are reported below by question.

### 5.1 Having been in your teacher education course now for some time, what are your views on the demands of the course in terms of your English language proficiency?

Students listed their specific language-related problems as follows:

- becoming familiar with the colloquial language and Australian idiom - both with respect to their fellow (local) students and with respect to their practicum pupils (one student referred to his pupils' use of 'the 'like' word'). All students referred to their pupils as 'speaking very casually' and contrasted this with their own schooling (*both* in learning English *and also* in their L1). The fact that their pupils' use of a more colloquial language was simply evidence of being comfortable in their own L1 may have been recognised by the students, but their focus was on this language as a barrier to their communication with pupils.
- the academic English demands of their course, both in terms of academic writing conventions and of the language of instruction
- the issue of accent. This was an issue in their understanding of others, but especially in others' understanding of them, including lecturers, fellow students and their practicum pupils
- certain pronunciation conventions, which differed between countries. Speed of speaking was also, of course, a related issue
- semantic differences. This especially impacted in schools and especially for Mathematics Method students. Indian students used the example of the phrase '2 into 3' meaning '2 x 3' in their own school education – semantically the reverse of the case in Australian schooling, where the term *into* refers to division. These differences had caused some problems in communicating with their classes, necessitating intervention by supervising teachers. Thus, for some students, the language of the subject itself was sometimes slightly problematic

Some of the language issues which they themselves identified may seem to be largely those which any person trying to cope in a foreign language might encounter. However, what may be ordinarily perceived as minor issues (e.g. teenage colloquialism) become magnified in the situation of communicating with pupils on practicum. The much more public 'performance' of English in all of its modes - especially speaking, writing and listening - is a major issue for teacher education students on practicum. This in turn means that the linguistic factors identified here take on a greater urgency than is the case for other users of English-as-a-foreign-language. One stark example of this is the case of

these students responding to their pupils in their practicum classes - an example of listening, thinking and formulating one's answer 'on the spot' in English, and a contingency that cannot be planned for in the way that much speaking, reading and writing can be planned in advance by a teacher. A key issue for these students was being comfortable enough to respond quickly and appropriately to their pupils. One student supplied the title for this research when he referred to local student teachers being able to 'play the classroom tennis very well' (referring to the back- and-forth of classroom dialogue and interaction with pupils) compared to him. Thus even those linguistic issues which we might see as very day-to-day, and not even necessarily specific to higher education, take on a different meaning when English has to be 'performed' in this context. The academic English of their coursework was listed as an issue (see next question), though the degree of support given for academic English within the Faculty of Education for these students was substantial and acknowledged as such by the students

## **5.2 What interventions have been put into place to support you in terms of English language proficiency? Have these been suitable/adequate?**

The Faculty of Education in question has in place a number of support mechanisms for these students. These include:

- a series of 4 course units which are peculiar to this international cohort – specifically:
  - an English language unit aimed at increasingly sophisticated understandings of language use in a range of contexts, including the discipline of Education
  - an introduction to Australian education. This unit is composed of school visits which include lesson observations. Some of these visits are to rural schools. This unit (4 hours per week of school visits for 10-12 weeks) also has an *extra* 2 hour per week 'support workshop' run by the Learning and Teaching unit – it consists of 'survival' English and 'academic' English (especially analysis and critical reflection) and is especially aimed at assignments, referencing, group assignments, giving presentations
  - cultural investigations in school
  - teacher-as-enquirer
- the Faculty employment of a PhD student to assist students with aspects of assignments. This student 'supports the support workshops'
- use of a retired teacher to mentor students in groups
- use of alumni as further mentors, who are themselves (mostly) teachers. All 2009 students had mentors, though prior to this, it was mainly 'at risk' students. The total group holds fortnightly meetings around language and acculturation.
- micro-teaching sessions
- 18 students are undergoing an 'international initiative' in 4 schools, in which they become the responsibility of one person in the school who runs a workshop or two for them on acculturation as part of the practicum.

The students find the English language unit especially 'very useful', as it covers a range of academic English concerns such as: report and essay writing; giving presentations, and critical analysis. Students also stressed the benefits of being in classes with local students - in terms of assistance and acculturation. This they saw as building confidence for interaction with their own pupils. While they believed skills taught in one area did carry over into other areas, all felt that more work on academic writing would be useful.

### **5.3 Where is the need for support greatest (e.g. through your study program, through the practicum)?**

Most students felt that there was an equal need for support in both academic university work and preparing for practicum. In terms of the latter, key issues were:

- how to teach the content well
- whether one is a good presenter
- avoiding linguistic errors
- understanding pupils' language and culture
- understanding the language of other teachers
- having a quick response ready to student questions and answers

All students stressed the need for both they and their pupils to 'slow down' in speaking. They argued that, over time, they and their pupils became used to each other, though they were more willing to be patient about this than were their pupils.

### **5.4 Are other issues more important than just language proficiency in supporting your success (e.g. becoming familiar with the cultures of Australian schools)?**

- becoming familiar with Western academic culture. Issues such as confronting 'plagiarism' as a concept for the first time were mentioned here
- becoming familiar with the cultures of Australian schools. Many issues were dealt with here and the following list captures the key ideas:
  - issues around respect for the teacher. All students argued that Australian schools were very different from 'home', where pupils were said to be 'more formal' and well-behaved. This was constructed as teachers being automatically respected 'at home', but having to earn respect in Australia. They felt that education was valued differently by both pupils and parents 'at home' and seen as 'the key to success'
  - Australian schools being perceived as 'more student-centred' than in their home countries
  - the differentiation of curriculum in Australia for specific pupil/pupil groups.
  - a less examination-centred mentality. The pervasiveness of examinations in their home countries was a big issue for these students, which they saw as driving the curriculum. They did not necessarily see this as a negative or positive, perceiving it rather as simply what goes along with an ethos of education as the 'key to success'.
- behaviour management. While this definitely tied into issues of acculturation - being linked to the issue of respect for the teacher, for example (most commenting that in their home countries, badly behaved students were simply excluded from school) - students also recognised that it was both just as big an issue for local students, and of such great importance that it deserved to be listed separately. They saw it as an issue for which they
  - needed more strategies
  - needed more knowledge ('the more we know, the more we can prepare') and
  - depended heavily on language for success ('When you deal with this issue, you need to negotiate with them, so communication is very important').

Despite these problems, a number of students preferred what they perceived to be the greater focus on individuals of the Australian curriculum, recognising that the ‘teacher-centredness’ of their ‘home’ curriculum did not necessarily ensure that pupils were learning. One student said he ‘couldn’t teach at home...it’s not actual teaching...I’d be happy to teach here’. They felt that the demands made of them in Australia made the standard of teaching very high because of the perceived need to make lessons more engaging. Though perhaps based on a romanticised view of students in their home countries as more compliant, the students in this group were identifying issues that were broader than issues of language alone and also had in common with local students the concern with behaviour management - a characteristic concern for all beginning teachers regardless of background (Arends, 2004, pp 29-31; Leask and Moorhouse, 2005, pp 22ff).

### **5.5 Have your practicum schools identified English language proficiency as an issue for you? What about other issues (such as becoming familiar with Australian school cultures)?**

Some students identified language competence as an issue for their schools in response to this question, but just as many felt that their issues were identified by schools as no different from those of local students

### **5.6 In the light of this, what do you think of the adequacy of your IELTS score for entry into teacher education and what do you think an adequate score would be?**

Students had strong personal views on the inconsistency of IELTS, telling stories of getting good grades in their course, but low grades on writing in IELTS, though ‘we do a lot of academic writing’ and after two years in Australia, ‘we are proficient in English’. They strongly felt that IELTS was not a consistent measure of ability (or a good predictor of success, raising again an issue which cannot be a function of such a test, but which is clearly perceived as important by these students). The inconsistency in IELTS scores with relation to what they perceived to be their actual language-functioning abilities in English was a big issue for these students. A strongly expressed opinion was their belief that success in the Academic tests of speaking and writing depended on one’s prior knowledge of the topic one was given to write or speak about. ‘Topics’, they argued, ‘are crucial’ - which in their view, made IELTS effectively a test of knowledge, not just language functioning.

However, when asked to discuss the scores on or near those required for entry into their course, all agreed that a score of ‘6’ would be too low for entry into a teacher education course, even with two years study in the language following it. ‘6.5’, on the other hand, was regarded as ‘not too low’, hence by default an overall score of 6.5 was regarded by these students as a baseline for entry into a teacher education course.

## **6 THE PERSPECTIVES OF REPRESENTATIVES OF ONE STATE TEACHER REGISTRATION AUTHORITY**

Interviews were conducted with the representatives of one state teacher registration authority.

Questions asked were:

- 1** What is your feedback from schools on the English language proficiency of newly graduated teachers who have been subject to IELTS testing?
- 2** Are other issues more important to schools than just language proficiency in the success of such newly graduated teachers (e.g. becoming familiar with Australian school cultures)?



- 3 What support mechanisms are in place in the state system itself to support such newly graduated teachers?
- 4 In the light of this, what do you think of the adequacy of IELTS scores for entry into teacher education?
- 5 What do you think an adequate score would be?

In addition, these people were asked to make comparisons with the PEAT instrument used by the NSWDET. Though devised in NSW, all other states are aware of the details of PEAT. Answers are reported here in their entirety, rather than by question.

The key response was that issues about culture and values posed by teachers from non-Anglophone countries were ongoing challenges to schools. The main employing authority in the state in which this interview was conducted ran a pre-employment program for the relevant teachers as they joined the system. This introduced people to the socio-cultural mix of schools, to curriculum and to professional experience and assessment practices. The team running this pre-employment program present ongoing professional advice and support. These interviewees believed that PEAT had been accepted by the broader education community. This becomes relevant to this research in light of the belief that a required score on PEAT was at a higher level than generally accepted IELTS scores for entry into teacher education. While there is no way of making a strict comparison, the belief of this group was that PEAT was 'definitely' higher than a 7.5 in IELTS. The significance of this, if the group is correct in this estimate, is that IELTS university entrance scores into teacher education are generally lower than an instrument required by one major employer in Australia. These interviewees felt that universities were being less than fair to students if they gave too much 'leniency' on entrance in this area – either through setting scores too low or allowing RPL. This disadvantaged students in the long run, either because they could not cope with the academic program or because they may not meet employer requirements. These people were aware that for students subject to PEAT in the relevant state, only about 30% passed the test on their first attempt. In general, they believed that a benchmark of at least '7' would be better to gain admission to teacher education because students may then be able to reach '7.5' by the end of the course (if the course is 18 months or more). They felt that a student with IELTS '7' at the beginning of a one-year Dip Ed would rarely reach IELTS '7.5' by the end of the 12 months. They also felt that a student with a score of '6' or '6.5' at the beginning of a program would be unlikely to reach '7.5' even after 2 years. In general, they felt that the correct principle for IELTS was that the entry score should be at least as high as the mandated state requirement for registration. This would allow time for students to improve enough to reach a score above the minimum mandated requirement.

## **7 AUSTRALIAN TEACHER EDUCATORS' PERSPECTIVES ON THE IELTS ENTRY SCORES**

### **7.1 Background**

Faculty of Education 1 runs a large graduate entry Masters level course for pre-service secondary teachers. LBOTE students are a relatively high (compared to other universities) proportion of students and are all already Permanent Residents or citizens. Most LBOTE students are in Maths, Science and Computing Methods. Most LBOTE students are from the Middle East and the Indian sub-continent.

Faculty of Education 2 runs a small course for international students only. It is at the Masters level and requires two years of study. Students are mainly Indian and Chinese, with small numbers of Japanese

and others. Students are spread across subject Method areas with slightly larger numbers in Maths, Science, Languages other than English (LOTE) (Chinese) – the latter being now subject to quota because of the difficulty of placement in schools (There is no Hindi taught in local schools, hence LOTE Method does not encompass the L1 of many Indian students). The course is full-fee paying. All students are on student visas, though many want Permanent Residency. Local students, in contrast, study a Diploma in Education, which is a shorter course and their English needs to be better on entry. From 2010, the Masters course will expand to take in local students and the international students will simply be a cohort within the larger course.

Faculty of Education 3 runs a one-year Diploma of Education (Secondary), which can be taken over two years part-time by domestic students. LBOTE students are a relatively high (compared to other universities) proportion of students and are mostly already Permanent Residents or citizens.

Faculty of Education 4 runs a graduate entry Masters level course over 4 trimesters (1.5 years) full-time for pre-service secondary teachers, with an embedded 3-trimester Graduate Diploma of Teaching. Students are permitted to exit with the Graduate Diploma. LBOTE students are a relatively high (compared to other universities) proportion of students. This Faculty had previously run a Master of Teaching specifically for international students.

In all Faculties of Education in the research, there are relatively high numbers of LBOTE students. Academics from one Faculty offered their perceptions of why LBOTE student-teachers study secondary teacher education:

Often they come through systems where there is more status teaching secondary and they think teaching adolescents is going to be easier. It (also) means they are able to work within their specific discipline area, (one) that they have studied at university. Whereas (in) primary, except for Languages and Music, you have got to be generalist

From my conversations with international students and students who might be children of migrants who grew up here, they see that in teaching, secondary school disciplines are something that they know very well... They probably have a tertiary degree in Maths or Science so therefore the comfort zone is there. In primary teaching ... they need to teach the English language as a language as well and that is something that might stop international students from applying

In what follows, the academics' answers to the interview/focus group questions are reported by question.

## **7.2 What are your views on the English language proficiency of your international students who have been subject to IELTS testing?**

### **7.2.1 Fitness for purpose of the test**

By its very nature, language testing does not reflect a student's overall ability to be part of an academic community. Because of this, some academics felt the reliance on language testing alone for entry for these students was a problem ('English language proficiency is not the only issue'). Others felt that an appropriate language test needed to be occupation specific:

The IELTS speaking test isn't relevant to teacher education students because it doesn't demand very much flexibility of expression from students. It's a measure that we in teacher education have to use because it's a university measure that the institution requires, but it's not always helpful

One critique of language tests was that similar (IELTS) scores reflected inconsistent language skill levels among students and the academics who argued this felt that this was more than just an issue of different scores on different sub-tests - one lecturer, for example, referred to the range of proficiency also varying across the sub-skills of speaking and writing. Students with the same scores who were well above the minimum entry into the relevant courses were still perceived to have marked differences between them. There was an expectation, then, among these academics that similar scores are meant to reflect reasonably consistent abilities and that this was not necessarily the case. In terms of fitness for purpose, however, an opposing view was:

As far as I know, IELTS is the best English language proficiency test that has been developed... I am an examiner of speaking and writing; the written test is quite useful for knowing how well international student-teachers might cope with academic studies. Even the General English test helps to see how well they'll cope with writing and reading. The speaking test is as good as it could be for that kind of test that has to be done in that limited time. It's quite informative. The IELTS test in itself does tell us quite a bit.

### **7.2.2 Student competence**

There was disagreement among academics about where they felt LBOTE students' problems manifested. Academics in one Faculty felt that students' general oral ability was often better than their written ability and problems therefore did not manifest in conversation. The students, they felt, could function conversationally but needed to be able to read, write and understand nuance, which they could not necessarily do at a level appropriate to their (Master's) degree. Others, however, argued that:

The main issue for (LBOTE students) in terms of communication is speaking and to some extent listening. The reading and writing is pretty good especially at the pre-service level; they can do assignments. They can write a very nice essay [but there] is no guarantee that they are able to engage in verbal communication - conversations about the content

... writing a piece of academic work; that is the benchmark for having a place in a particular course. At the Master's level there is so much emphasis on the writing of the assignments rather than participating at the classroom level. What do we do with these students? Should we not enrol international students who are interested in one of our Master's degrees just because they are not able to speak well? I am also not very sure that we should be stopping (international) students from enrolling in our courses because they are not able to speak as well as they are able to read and write... Of course associated with that is how... we test the speaking given the number of applicants that we may have.

Even comment on specific skills – such as speaking – needs to be nuanced, with groups of academics distinguishing between students' skills in private conversation and their public speaking, such as in tutorial presentations. These academics were arguing that when it came to academic discourse, LBOTE students in their courses became very visible. One academic summed up the issues complicating the performance of spoken language in practicum classrooms in terms which again suggested the need for occupation-specific skills:

... international students need to have skills that are strong enough to be able to re-formulate ideas in more than one way, and to be grammatically accurate enough to get the message across very easily... (Pupils) ask questions in ungrammatical ways... Students really need to be able to speak well enough in posing questions and they need to understand when (pupils) are asking questions even if they're making statements. So the listening and speaking skills

have to encompass those abilities. They also need to be able to use speculative language because it is very important in education, whichever subject area you're working with. Reasonable control of conditionals is important as well... These characteristics need to be present, along with intelligible pronunciation. By intelligibility I don't mean 'mainstream', because there is no mainstream. So I do not mean necessarily imitating an Australian accent. I accept diversity but (require) intelligibility of speech.

The conditions driving some Australian universities to recruit international students was noted by teacher educators and questioned on the grounds of students' marginal English language proficiency. This was noted in some Faculties as the tension between academics wanting higher IELTS scores and 'people making the decisions higher up' wanting lower scores in order to maximise international student numbers – the tension identified by Feast (2002). However, it needs to be noted that in other Faculties being discussed here, many of the LBOTE students are already Permanent Residents and not part of a cohort on student visas.

Some of the students who are on student visas enter teacher education programs through avenues created by universities themselves, such as through an English language college. The English proficiency of international students entering through these means was generally seen as relatively weak ('They haven't really got up to speed and they're not given another IELTS test before they enter the university course') and again seen as part of the broader university drive to enrol international students into courses. One concern about using university Language Centres was the sometimes unrealistic expectations by university management about how quickly international students could increase their English language proficiency. Another concern was the considerable pressure on the university language and learning support staff.

Proficiency in the English language is a complex issue in itself, subsuming a web of inter-related issues:

- functional English
- the technical language of the subject these students teach
- academic English - the critical and analytical
- the spoken and written language demands of the profession in general: not just functioning 'on their feet' in the classroom, but producing material for students, communicating with parents etc.
- the everyday language of the pupils in their classrooms
- the ability to be able to transform subject knowledge into explanations that are understandable by their pupils

Academics saw the need for assistance across this whole spectrum of language. Even having *knowledge about* language (such as grammar, of which many students see themselves as having good knowledge) is not the same as producing and understanding language. Academics find that students can have big gaps in understanding and though students learn to use the language enough to 'get through' the test, the ability to teach *in* the language is at another level. Perhaps only occupation-specific testing can address this. One academic argued that in her experience, making a mistake or saying the wrong thing led students from some cultural groups to withdraw, and not engage, leading in turn to lecturers and others not seeing their problems. Another academic highlighted written English as the key problem, identifying grammar, vocabulary and some semantic issues as of most importance. There are also a number of other issues in language:

(things) that we don't work on enough (are) face-to-face interviews... interaction with colleagues in school, or... interactions with parents. We may deal with this in the very last tutorial of the year, 'By the way you will be expected to do interviews with parents. Here are a few hints ... or a few "do's and don'ts."' But those pragmatic and strategic aspects of communication need to be dealt with much more explicitly

It seems we cottonwool them throughout their teacher training year or years, but once they are out on their own in the school situation there seems to be another big step which is another enculturation shock which then often affects their language. So they might have exited from our course with good communication skills in listening, speaking and writing. But the initial shock when they are on their own in a school, particularly a country or a city school where they are the only non-native speaker, then they can go backwards for the first term or semester. We had graduates drop out within a term because they were physically sick; they couldn't cope with the actual situation. They feel they haven't been prepared for this during their pre-service training.

Communication-wise some Asian students are quite amazed at the way in which we 'think aloud' in conversations. In South East Asian cultures what is said is more or less the summary of what has been thought through. (Given the time this takes) they therefore come across as being hesitant and not participating in meetings and discussions. They have problems in interviews for jobs and interacting with staff in meetings and working on projects.

In terms of classroom language, international student-teachers need to at least recognise the complexity of the issue if they're going to be able to teach well in Australia. One academic in Science education argued:

One of the difficulties... becomes if they can't recognise this, then they just give information to (pupils) who don't have an opportunity to learn that scientific language either - certainly not in a meaningful way. International students need a flexible understanding of the language of Science, the concepts of Science, in order to communicate it well. For somebody whose first language is not English, that's quite a challenge. It's possible if the international students are open-minded and prepared to give it a go. Being able to practice is important and getting support in the school from the supervising teacher who isn't judging the international students just on their ability to use the English language but as much on their willingness to try.

Finally, one comment on the issue of student competence did encapsulate the feelings of a number of academics interviewed and takes up a quite different perspective on the issue:

It always worries me that we only talk about language proficiency at entry. We must also talk about responsibilities we have for students' learning, and taking advantage of the knowledge that international students and others bring. This means looking at linguistic diversity as a benefit rather than a deficit.

### **7.3 Are you doing any internal testing of the English language proficiency of such students?**

One Faculty of Education conducts a literacy test of all students in the course at orientation in order to check their academic literacy. Invariably, students who have had to take the IELTS test are made

‘visible’ by this test. The particular test picks up comprehension and interpretation problems and also reveals some students who have been out of the academic milieu for a while.

Another Faculty threw into focus some of the problems with internal testing:

Until two years ago, we were able to use an oral proficiency test that was designed as part of a research project. So it is a validated instrument and it produced reliable results. We started to get more and more international students, but (the) only ... person organising and conducting the test... started to get exhausted. We simply couldn’t find the time to do (the test). We were often doing these interviews by phone to places like Hong Kong and Japan, coming in all sorts of hours to assist the students who were there. The Dean decided that (it) was a workload issue, and we just had to stop.

The details of this internal test are worth noting:

- it contained material from students’ subject specialisms (‘What does the diagram show? What topic would you be teaching about using this text? What knowledge do you have about this topic?’)
- students were given 15 minutes for preparation, during which time they were also asked to think about an issue about which they wanted to ask the two assessors (e.g. classroom management)
- this issue was discussed and then students were asked to summarise the academics’ responses and to ask further clarifying questions
- students were assessed on this discussion and on the subject-specific discussion in terms of both speaking and listening
- the two raters moderated their results
- scoring system: doesn’t need any support; needs some support; needs moderate support; needs a significant amount of support - those in the last category would not survive in the course
- occasionally, a writing assessment task was also required if a student was regarded as ‘borderline’.

This test is an example of an interesting and well-targeted response to the issue of student testing if issues of academic workload can be addressed. One other Faculty had conducted interviews in the past with all prospective students, but has had to cut this back to now interviewing only international students and only those whose applications in some way suggest they might need extra support.

#### **7.4 What interventions have you put into place to support these students in terms of English language proficiency? How have these been funded?**

In one Faculty, students are tested during orientation, and this is followed by the provision of academic seminars for students identified as in need. Seminars are run by the Learning Assistance Unit and tend to focus on: reading skills, interpretation/analysis, writing, using evidence and a course focused on spoken language in teaching. Students identified on the screening test also have sixty hours of mentoring by fellow (local) students, who undertake this mentoring as part of a professional experience unit. ‘Spoken language and teaching’ is a module taught after their first (stand-alone) week on practicum. Classes discuss the language they saw on practicum and role-play scenarios. These sessions are compulsory for all mentors and mentees. The strongest parts of this program, the

co-ordinator believes, are the discussions around practicum. The short, intensive nature of their (Master of Teaching) degree is seen as a problem by these academics, with many students attempting to complete the course in 12 months because of financial imperatives. Acculturation into the course itself takes time and the lecturers feel that students need to be supported more before the course begins. Many LBOTE students arrive feeling confident, do badly on practicum and then find they have missed the workshop support program. Funding for these initiatives has been dependent from year-to-year on the availability of funds at the Faculty level.

In a second Faculty, the following interventions are in place:

- (1) The inclusion in the course of 4 units which are peculiar to this international cohort. These are
  - an English language unit aimed at increasingly sophisticated understandings of language use in a range of contexts including the discipline of Education
  - a unit introducing Australian education. This unit is composed of school visits which include lesson observations. Some of these visits are to rural schools. This unit (4 hours per week of school visits for 10-12 weeks) also has an extra 2 hour per week 'support workshop' run by the Learning and Teaching unit – it consists of 'survival' English and 'academic' English (especially analysis and critical reflection) and is especially aimed at assignments, referencing, group assignments, giving presentations
  - cultural investigations in school
  - teacher-as-enquirer

The Faculty in question has recently taken the decision to make the international cohort a strand of the new Master of Teaching degree available to all secondary students and in this altered degree only the English language unit and the introduction to Australian education will be the distinguishing support units. Other Faculties recognised too the benefits of running credit bearing units that allowed international students to access appropriate support, though some areas of university administration sometimes see these units as inappropriate at Masters level. However, running courses which do not carry credit can impose insupportable costs on Faculties of Education.

- (2) the Faculty employment of a PhD student to assist students with aspects of assignments. This student 'supports the support workshops'
- (3) use of a retired teacher to mentor students in groups
- (4) use of alumni as further mentors, who are themselves (mostly) teachers. All 2009 students had mentors, though prior to this, it was mainly 'at risk' students. The total group holds fortnightly meetings around language and acculturation.
- (5) micro-teaching sessions
- (6) 18 students are undergoing an 'international initiative' in four schools, in which they become the responsibility of one person in the school who runs a workshop or two for them on acculturation as part of the practicum.

The course co-ordinator would like to further supplement this support in the future with more use of classroom scenarios on video, more videoing and critiquing by students of themselves in micro-teaching contexts and perhaps supplementing these using Web technology such as Second Life for role-playing scenarios. The Head of this Faculty noted that the Faculty had funded some of these

initiatives, while others were ‘funded’ out of the generosity of staff and hence were not necessarily sustainable. Secondary schools used for practicum also contributed resources and made a huge in-kind commitment. Interviews with individual lecturers confirmed the Faculty’s commitment - with each of them talking about giving these students extra time and extra help and making themselves more ‘available’. As one lecturer noted, while he had no deliberate interventions, his ‘pedagogical behaviour’ was affected by the presence of LBOTE students and being approachable, open and available for assistance was itself an ‘intervention’.

Various forms of Faculty-based supports are provided for international students in the remaining two Faculties. There was strong support among academics for LBOTE students’ support being provided within the Faculties, rather than being ‘farmed out’ to an academic skills unit with very little time to give to students because of their own pressures. Such Faculty-based supports currently included:

- (1) reading study groups, in which LBOTE students discuss set readings in small groups and have one-to-one consultations with a funded staff member- however, recurrent funding of such a position is an ongoing issue. Ideally, the Faculty could provide ‘a person ongoing, throughout the year for at least one day a week to support them...a trained person who would know the content, the subject matter, with a Masters Degree in the same area. It could be an ex-student or a good student who would help them in small group reading sessions and deal with other issues, such as counselling’
- (2) extra group tutorials provided for students by Faculty staff - though such service provision depends on the amount of time staff have available
- (3) individual assignment assistance
- (4) a teacher education camp. At the beginning of the year, in the second week of semester, one teacher education program takes all its students on a two day camp:

For many international students that’s the first time that they’ve been on a camp of this nature. We have a day where they’re looking at different activities that they could do outside of a classroom to help engage students with their Science learning. We take them around the beach and the bush, running activities like an ‘egg drop’ by building a parachute from natural materials and they have to figure out how not to break an egg if it’s dropped from a height. On the sensory trail they identify things on the beach that they see (could aid) learning about tidal action - all linked to what they’d be teaching. This time gives us an opportunity to get to know the students, see what some of the strengths and challenges there might be for the group, and gives them an opportunity to work with (pupils) from a local high school. On the second day of the camp, we bus in Year Seven (pupils) from a local school and our student-teachers work in small groups to teach them outdoor Science. That’s an opportunity for all the student-teachers to talk to the Year Seven (pupils) about what’s it like being a school (pupil), the transition from primary school, what school is like. We give our student-teachers a supportive environment to learn about schooling, learners and Science so as to develop some familiarity with the interests of school (pupil)s. It also gives the international students an opportunity to work with their peers and learn about the settings in which they’ll be working... We make the experience as inclusive as possible so that everybody feels very welcome... and they feel connected to the group... (Nevertheless) some students find it an extremely challenging and overwhelming experience because it’s nothing like what they’ve experienced before. They’re not quite sure what’s expected of them. Even though they’re working in small groups, they might have quite fixed ideas



about how (pupils) should learn or what teachers can do. That can be difficult for them to negotiate, becoming more open and flexible.

The university does not provide any fund for the teacher education camp other than for the academic staff. Over the years this camp has been reduced from one week to two-days-one night. Typically, such strategies are funded by the Faculties. However, many universities have decided to centralise student support services previously provided by Faculties, often then tending to render these services invisible, thereby reducing their usage and thus any claims for having such support.

### **7.5 Are other issues more important than just language in supporting these students (e.g. becoming familiar with Australian schools)?**

Academics see courses in which students have to go into the field and 'perform' (e.g. teaching, nursing) as raising particular issues. All international-oriented programs are geared towards testing and addressing language, but academics argue that this is only one dimension of such 'performance' and is secondary to 'adjustment':

The local students have a familiarity with the context and the cultural settings; it can make them look as though they know more than they really do. It can make them look more proficient compared to international students who are unfamiliar with the settings in which they'll be working. ... the first four-to-five weeks of the year, just before the practicum are always full of concerns about interacting with school (pupils), cultural issues and language issues, 'How am I going to understand what the student says?' 'How am I going to know whether the student is understanding?' As student-teachers they are very often very concerned about their own image in the first teaching round. They're more concerned about their teaching and themselves as teachers often than they are concerned about the students and their learning ... they're often concerned about their own image and what students will think of them. This is a very typical across all teacher education courses.

Familiarisation with Australian schools is a huge issue. Most academics argued that the biggest issues arose in practicum when students had to face both the foreignness of general Australian culture and the culture of Australian schooling. Many LBOTE students are from very different classroom cultures with different classroom dynamics. Academics strongly identified an issue of the disjunction between how their students were taught at school and 'how we teach here'. Some said the students tended to move into/favour a lecture delivery in teaching. LBOTE students who were residents rather than on student visas and who had some previous teaching experience in their home country could experience even greater problems:

They may have very well been able to teach English as a Foreign Language; that's absolutely fine. But they come into the school (and) their biggest problem can be 'how to be' in an Australian educational institution; how to relate to peers, how to relate to your supervisor, what your responsibilities might be. They need special preparation and acculturation to the Australian education system before going into schools. ... It is those local people we're actually letting down because they haven't had to take an IELTS test because they're already citizens.

One academic talked of the problem of language *in* the practicum as the biggest issue. Part of this concerns spoken delivery, but the key problem is with interaction, especially interactive questioning. Students can ask a question of pupils, then ignore the response because they do not understand it, so

questioning becomes a problem in terms of not responding appropriately to pupil questions and answers:

their ability to actually catch the message, and then to respond in what is considered in Australia to be a timely manner. That means not a long pause time or using fillers which indicate you have received the message but will need a few more seconds or so to think about how you are going to respond.

A big issue, therefore, is listening comprehension - something that teachers have to do well, reflecting Elder's finding that 'good listening skills are crucial for effective performance in teacher education courses' (1993b, p 83). Writing and speaking and questions-to-be-asked can mostly be prepared in advance, but listening and responding cannot and this creates problems. Language, of course, in all its complexity is ultimately inseparable from the acculturation issues:

They're all inter-connected. The more confident international student-teachers feel, especially in their speaking skills, the more potential there is for them to feel comfortable enough to question what's going on in their heads and what's going on around them. It's a sense of relationship, connectedness or community that leads them to ask for more help; they have to feel they're in a place where can relax and just try.

Academics also identified the clash between the sorts of learning experiences that Australian school students are likely to have or expect to have - and the willingness of the student to see this as part of their learning to be a teacher:

There is the attitude of the person who's coming in - how open minded they are to a different system or whether they're coming to reproduce their own system. From time to time there are international students who are so deeply entrenched in their own cultural context, who have very strong values around transmissive teaching or students' listening and not asking questions. These then become extra challenges for international students who are also trying to ensure they're proficient in the English language as well... Open mindedness is an essential characteristic of a teacher... so they can see learning and teaching differently. International students come in with a particular view that they may or may not be aware of. They need to start to question whether that view works in Australia, and what it means to be in a classroom as a teacher in Australia. Allowing one to question one's own views and allowing others to question these means using their teacher education program to experiment with other ideas.

In this situation, students' problems can manifest themselves in an overwhelming concern with classroom management, since many LBOTE students perceive their own schooling as having been composed of mainly passive classrooms in which pupils are not active or outspoken. A perceived need to earn respect, rather than have it bestowed automatically, can lead to feelings of inadequacy and then they just 'get used to' Australian classrooms when practicum ends. On the other hand, some academics saw this as largely no different to the problems faced by local students, while others felt that pupils were generally tolerant and willing to work with a teacher who was trying to build rapport.

As many argued, the whole issue of LBOTE students engaging in field-based experiences does raise the responsibility of the university program itself to address concerns about clashes of - in this case - pedagogical culture:

... the basic principles on which we have developed our teacher education program makes explicit what we're doing as teacher educators - we make our practice a basis for conversation with students. That doesn't mean that they'll agree with it or that they think that we're doing the right thing because there's always that disconnect in student-teachers' minds between what happens at university and what happens at school. We provide opportunities for them to question their own practice through the experiences we set for them in the classroom.

Making diversity a focus of educational practice was seen as valuable:

We are talking about educational diversity rather than just one way of doing things. The truth is our teacher education courses are very culturally positioned. This is a sad thing because even though they have the intention to be internationalised they're not entirely internationalised ... They do mention international examples and sometimes draw on students to shed a little bit of light on another way of doing things but that doesn't always happen. Of course, international students don't necessarily know much about their own education systems; they only know about their own experiences but they don't necessarily want to talk about that.

Assessment in their academic program is a key source of anxiety for these students:

There is an onslaught of assessment in teacher education programs. I said 'onslaught' deliberately. It's just interminable, and it's really different. There is a lot of reflective writing. They're not used to that. There's a lot of writing plans, assessment tasks, rationale writing. Its different sorts of academic writing from what most people would be used to. Some international students are not used to any academic writing at all because they've been through Science programs, so people aren't as used to extensive writing, not used to this kind of writing. So assessment is a big focus throughout the year.

Finally, the students' need to have employment can compound many problems (it is this financial component that often works against their being able to come early into the course, for example). In fact, one academic referred to students' 'double struggle' - the academic languages they had to master and the need for paid employment while studying, which gave them less time to achieve that mastery. Students from cultures in which examinations were the only form of assessment at tertiary level were also seen by some as having problems with attendance, though it was recognised that attendance was equally affected by their need for paid employment.

### **7.6 Where is the need for support most evident (e.g. through their study program, through the practicum)?**

Academics felt that support was equally needed through the practicum and the study program. Some also felt that a greater support system was needed in schools, along with professional development for teachers. One particular kind of appropriate support suggested was ensuring that courses were long enough for LBOTE students to gain the necessary skills and familiarisation with Australian schooling systems – not allowing the students in one course, for example, to take the course in an accelerated mode. LBOTE students, they argued, needed the time. Most academics were definite about LBOTE students not being allowed to take accelerated versions of the courses, though in some universities, marketing and recruitment practices worked against this (an issue discussed in Hyatt and Brooks, 2009, pp. 34ff, 42-44, 51-52).

At the same time as the number of international students has increased in universities in Australia, over the last decade universities have suffered severe cutbacks to funding, resulting in widespread cutbacks to teaching hours:

In a one year course what can you do? We now only have eight weeks contact in first semester and eight weeks contact in second semester – we are with the students for two to three hours a week. (The course) has a one hour support session once a week during those eight weeks. There is not a lot we can do.

This comment came from a university in a state which nominally has a mentor scheme for beginning teachers and academic staff feel that ‘Those mentors need to pick up the responsibility for supporting the international students... because they are all experienced teachers, so they know what the situation is in the first term of when you are out on your own’.

Sound pre-practicum familiarisation was also seen as of utmost importance. In some Faculties, LBOTE student-teacher support included pre-course school-based induction:

This pre-course is only a three-to-four day course that allows students to have a placement in school. Usually I have about 30 students in this course, which is held the week before their academic course starts. It’s not focused on language as much as on trying to give them a grasp of the system itself, how it works, how it is interpreted in a particular school, the culture of that school, the ways in which students interact with teachers... There is consideration of the language that teachers use in class for different purposes – for a task – for management. The student-teachers are placed with teachers in pairs; they shadow either a teacher or a class.

### **7.7 What have schools had to say about the English language proficiency of student-teachers? What about other issues (e.g. acculturation into Australian schools)?**

Schools have legal and moral responsibilities to educate pupils and to attend to their welfare, both of which can be affected if a teacher is not easily intelligible or cannot understand what a pupil might be saying. Moreover, through lack of ‘school culture’ knowledge, some LBOTE students behave inappropriately at school – such as through having wrong expectations of the degree of support to be provided to them; not taking the kind of initiative which Australian teachers expect; criticising the supervising teacher’s methods; or not following disciplinary policies in the school correctly. For these reasons, schools sometimes see these students as not functioning at a level required by them. This can lead to a culture in the schools that is critical of LBOTE students. Principals are sometimes perceived as acting as gatekeepers and being reluctant to accept these students on practicum. Most academics felt that schools tried to be accommodating on the whole, but recognised that schools had concerns about the students’ preparation for teaching and about their language. Schools, these academics believed, needed to see that they had a joint role in the preparation of teachers:

There is a need for closer connections with schools, more conversations about how these student-teachers learn to teach, what good teaching is, how to support their learning and teaching. It’s not easy to address this issue. It needs collective effort over time.

One group of academics saw ‘language’ in the schools as always being conceptualised as ‘speech’ and ‘speech’ as ‘accent’, leading to LBOTE students being labelled ‘deficient’ and ‘unable to

communicate'. They thus felt that the school system tended to collapse all issues into an issue of 'language'. Schools, for example, tended to say that students were not *listening* to advice, but in fact the students were not *understanding* the advice because of lack of 'school culture' knowledge and would not ask questions. Language testing, of course, cannot account for these factors and thus some argued that continually increasing IELTS scores was not an answer to these issues. Others, however, felt that 'Since (the IELTS was raised to '7.0') three years ago we've had fewer complaints from schools about the English language proficiency of the international students we send out'.

Academics again partly see their universities to blame, with the number of international students growing faster than schools could be prepared for them. One Faculty of Education with a relatively large cohort of such students serves a metropolitan area which is very Anglo-monocultural and with an identical teaching workforce. Academics from this Faculty felt that the gap was becoming harder to bridge with some schools, however, those schools involved in the Faculty's targeted 'international initiative' have grappled with these issues and it seems to be paying off. These are partnership relationships and rest on a different set of assumptions about who is responsible for preparing teachers, as well as throwing into relief questions about how one prepares teachers with a different set of capabilities, needs, demands and social capital from local students.

A final issue that is worth mentioning in terms of language preparation and the diversity of language tasks in which a teacher can be engaged, especially for Asian students, is the process of job-hunting:

... recognising that you have to apply to each individual school, and you have to write different essays to address the 'desirables' and 'essentials' that each school is looking out for. This is a 'cultural shock' for Asian students; they find themselves in a position where they have to sell themselves in interviews and their written applications. That is something they are not comfortable doing. (They expect) qualifications alone will get you a job, not telling people how good you are at communicating with parents.

... In the second half of the year I actually spend more time on helping international students to see how to present themselves to employers and how to write an application for a job because they're horrible. They have to address all the selection criteria. We also have an assignment called the 'Portfolio' which is really helpful in the end but is a total mystery to the students in the beginning. In the end it helps them to find all the evidence that they need to promote themselves in an interview for a job.

### **7.8 What revisions, if any, has your Faculty made to entry procedures in the light of this experience?**

At the time of writing, one Faculty from which interviewees were drawn is considering raising IELTS scores to '7.5' to enter the course. Academics interviewed nevertheless felt that this causes a loss of potential of those lower than '7.5'. They argue that a score taken at a moment in time does not account for the ability of students to improve and that this is lessening the pool of people from whom to develop teachers. Capacity for growth, they argue, is not being accounted for.

A second Faculty will raise the IELTS score from 2010 to '7', with '6.5' across all sub-tests. Prior to this, the Master of Teaching entry (6.5) had been lower than the Diploma of Education entry (7) because the former course was for international students only and was two years duration compared to the shorter DipEd.

### **7.9 What is required by your local Teacher Registration body in terms of language proficiency? Is this more or less demanding than the equivalent university entrance requirement?**

It will be noted from Section 4 above that only one Faculty of Education in Sydney has the same IELTS requirement on entry to the course as its state teacher registration authority. One Melbourne Faculty meets the VIT standard on entry, though this is lower than the NSW standard. One Adelaide Faculty meets the standard of the South Australian TRB, though again this is lower than NSW. In each case, there is an implicit assumption that time in the course itself will enable students to meet the teacher registration authority requirement (cf next question).

### **7.10 In the light of this, what do you believe is an adequate IELTS score for entry into teacher education?**

Part of the answer to this question depends on the length of the degree in which students are enrolled, since it can be (is) assumed that time in the course itself improves students' English – hence the University of South Australia in the Tables in Section 4 above can be seen to have a lower IELTS score for an international (student visa) cohort than for a more general cohort on the basis that the course for the more general group is shorter. Hence arguments from academics such as the following:

(The entry level) should be tied to what the (teacher registration authority) requires. If they're going to do a 4- year teacher education course it mightn't have to be '7'; but on the other hand, they would be going out to the practicum, so I think it's sensible for it not to go any lower than '7', because those students will go out on teaching rounds early on in their course. They could fall into a hole almost immediately on their teaching rounds and be put off. It's not really very fair on the students. They have to be at a higher level where they're going to be required to go into a practicum early on.

... if there was enough support available it would be alright for starters to be at '6.5' at the beginning of the 4-year degree. However, if there's no language and acculturation support, then no. If the Faculty doesn't have this support it makes a major difference to students. This includes what kind of lecturers they have, whether the lecturers will give them extra time to support them adequately. International students often don't ask for help; they don't use the lecturer like a local student would. They don't make an appointment to see you about their assignments when they should. The support is pretty important. In today's climate of lecturers' heavier workload, of having to do more and more, including more and more administration as well, it is understandable that lecturers may not say, 'Please come and see me anytime.' A more structured intervention where they meet a support person, have reading groups, is needed.

In the case of graduate entry programs, on which we have focused here, course length is considerably shorter (often one year in the case of a DipEd) and most academics have doubts about the capacity of these to provide the necessary English language development required for teacher registration and employment if the entry IELTS score is lower than that required for such registration. In such cases, the support provided by employers for beginning teachers becomes even more important:

In terms of finding a long term solution and support for international students who are becoming classroom teachers, the pre-service course is going to be too rushed and too crowded. The level and the nature of the support that they have as beginning teachers in classrooms are going to be very important. The mentor is going to be very important. Pairing such teachers with a native speaking

teacher, or Anglo-Saxon teacher, is going to be very crucial. The selection of these teacher-mentors is going to be important. The skills and support that we give these mentors is equally crucial as is the support that we give for the beginning international teacher they are mentoring. There is a place for professional development of all the professionals involved; that goes all the way up to the education administrators.

In the case of graduate entry programs, then, some argue that:

They should probably start their course with an IELTS score of '7'. I'm not quite sure how to balance what the universities require about different bands. For teaching, their speaking and listening should be higher to actually be able to work out what's going on. For their academic work they need to have good reading and writing skills; speaking and listening is where the Asian students are weaker. Local immigrant students are more likely to be worse at reading and writing.

On the other hand, in answering this question, academics in one Faculty returned to the key question that students have issues with the language *and* their academic subject matter and the socio-cultural aspects of Australian education. The key issue, they argued, is that these are not problems that will be solved by higher and higher IELTS scores, which they regard as the university 'missing' the problem. This group did feel that if support is in place for the student, then the student's language will grow and they saw this as the key issue – not just the student's entry score, but the potential for the course to increase the student's exit score. Institutions, this group argued, needed to rely not just on what a language test could provide but to focus on a greater number of dimensions of what students and the institution and the profession itself needed, especially in those courses in which field placement is fundamental. Testing people for reading/writing/listening /speaking was seen by this group as a limited world-view and only assessing what can easily be tested. In terms of a baseline score below which a student would probably not cope in the course because of language, about '6.0'-'6.5', they felt, was probably adequate. '7.5', they believed was too high because it excluded too many students, when potential success in the course depended on so many other factors. This view received support from other Faculties:

It's all mixed up with the other issues... like their willingness to think about their practice and how it might develop differently from what they've experienced as a learner. The personal and language attributes are closely connected.

Recognising that the question of lower IELTS is a marketing tool for the university even in the face of requests from academics for higher IELTS, one Head of a Faculty with large numbers of students on student visas (and therefore subject to international marketing imperatives) saw the key question as being, 'What should we wrap around IELTS (such as an interview system)'?

One Faculty in the research will raise its entry scores in 2010 from '6.5' to '7'. This decision was taken to lift the standard of English in the course. One lecturer in that Faculty had 'severe reservations' about her students teaching in Australia based on their language proficiency on entering the course at their current score. This group would like to raise scores to at least '7.5', but the university is resistant. Accepting students on a first-come-first-served basis means that the Faculty of Education cannot impose a higher score. They also felt that it was difficult to pin down a specific score as 'adequate' because of what they saw as the inconsistency problem - an important problem 'because IELTS is a standard and if it isn't a standard it isn't anything'. If all '6.5' students were like their 'best "6.5"s', they argued, it would be a good standard, but this is not the case. If this remains so, then raising the

standard changes nothing. One academic felt that ‘7’ was *not* adequate because students already at that score were not being set up for success. Thus the question for this group also became, ‘What does the IELTS test mean?’ In particular, if students are coached to pass it, what does it represent in terms of language ability? It becomes a hurdle to jump, they argue, not a reliable picture of actual ability.

## 8 DISCUSSION

The literature on language testing and teacher education – particularly the work of Elder – suggests some important principles in relation to language testing in teacher education. Elder (1993a, p 237) has identified the following ‘desirable features of teacher communication’:

- intelligibility
- fluency
- accuracy
- comprehension
- use of subject-specific language
- use of the language of classroom interaction
- overall communication effectiveness

In addition, she has usefully elaborated on a ‘partially indicative’ (Elder, 1994b, p 10) inventory derived from Ellis (in Elder, 1994b, p 6ff) of typical teacher tasks in terms of language use.

These include:

- *message-oriented interactions*: e.g. explaining, categorising, labelling, presenting information, narrating
- activity-oriented interactions e.g. giving instructions
- *framework interactions*: e.g. directing, disciplining, explaining, questioning, responding, rephrasing
- *extra-classroom language use*: e.g. selecting and preparing material, simplifying texts, writing memos, talking to parents, reading professional development material, attending professional development seminars (Elder, 1994b, pp. 6-9)

Moreover, Elder (1994a) has indicated the importance of taking account of discipline-specific competence in testing language – in the case of teacher education, recognising that language competence involves creating the necessary conditions for classroom learning to take place, which may, for example, necessitate simplifying language in the real-world situation. The need for subject-specific language support was also highlighted by the participants in Hyatt and Brooks (2009, pp 46-47, 54-55).

To Elder’s lists should be added the aspects of language identified by our interviewees. The range of English language skills needed by LBOTE teachers was identified by our cohort of student interviewees as including:

- becoming familiar with colloquial language and the Australian idiom - both with respect to their fellow (local) students and with respect to their practicum pupils
- the academic English demands of their course, and
- differences of accent and pronunciation.



Our academic interviewees identified as issues:

- functional English
- the technical language of the discipline
- academic English
- the spoken and written language demands of the profession in general
- the everyday language of the pupils in school classrooms
- the ability to be able to transform subject knowledge into language that is understandable by their pupils.

In terms of coping with the academic demands of the course itself, Woodrow's (2006) work suggests a minimum IELTS score of 7 might be useful for courses in teacher education. This reflects studies which show that it is at lower levels of performance on the IELTS that English proficiency influences academic performance (e.g. Elder, 1993b; Woodrow, 2006). However, given that communicative competence is a core issue within the teacher education course itself, the issue of language testing for students is not just one of entry-level ('Do they have an IELTS score that suggests they can cope with the academic language demands of the course?'). The course which these students undergo itself focuses on language demands in teaching – providing clear explanations, for example, as part of pedagogy. The issue is, then, above all, a question of exit-level competence ('Do they have a level of ease with the language that allows them to meet the demands of teaching? Can they use a range of language strategies to enhance pedagogy?').

Hence, language growth during the course itself may be a more central issue here than it is in some other professional areas. This suggests that if LBOTE students are to be allowed access to shorter courses (such as one-year Diplomas), then their entry level IELTS score should be at the level required by the relevant teacher registration authority. Though IELTS – or any language test - is not meant to be a predictor of academic success, since so many other factors are important, the findings of Elder (1993b), Woodrow (2006) and Bayliss and Ingram (2006) which suggest IELTS scores can be an indicator of short term performance (Elder, 1993b, pp 78-80), give some weight to the notion that in shorter teacher education courses, IELTS scores could be higher than in longer courses. This is a question of the opportunity which students have on graduation to find employment and the role which universities and teacher education programs have in contributing to that opportunity when accepting students. In Australia, this suggests mandating a range of IELTS '7'-'8' on entry, depending on the relevant state and territory requirements for specific bands, despite Feast's (2002) findings that this may lead to high rates of exclusion of those international students on student visas in shorter graduate entry courses.

On the other hand, some academics believed that the teacher registration scores are higher than they needed to be, arguing that a threshold score such as IELTS '6.5' overall, which they felt suggested an ability to cope with the academic language demands of the course, was adequate, because the issues students faced within the course (such as on practicum) were problems such as a clash of educational cultures and were often hidden under a banner of 'language problems'. For these academics, higher IELTS entry scores would not address the familiarisation and acculturation issues that could only be addressed in the course itself.

One of the Faculties included here had a specific minimum course length requirement for LBOTE students (who in this case were on student visas), which was longer for these students than for local students or for students with higher IELTS scores. This allowed them to focus on language growth as

well as other issues for LBOTE students, such as familiarisation with Australian schooling cultures. Such an approach then depends, of course, on specific programs being in place for these students in order to support them towards such growth, which was certainly the case in this Faculty. Such specially designed support programs have been shown to assist students (Weisz and Nicolettou, 2004). Obviously then, universities need to be prepared to invest adequately in the course preparation and school familiarisation necessary for these students to succeed. Programs in this particular Faculty included both language and acculturation support. Previous literature has identified the need for such multifaceted support and in that aspect which is focused on language, it should ideally address both areas of Elder's 'desirable features of teacher communication' and of her and Ellis' 'inventory of typical teacher tasks'. The inventory of tasks suggest the contexts around which such support could be structured and the 'desirable features' suggest the qualities to be aimed for in pursuing the tasks. Extending communicative competence into important areas of pedagogical usage, such as clarity of explanations, should also be the aim of such programs (Viete, 1998, p 172; Ryan and Viete, 2009). Ultimately, as Elder argues (1993b, p 88), entry level thresholds regarding English language proficiency should be set by universities in accordance with their capacity to provide such support. Addressing the range of unfamiliar cultural contexts – above all, local schooling cultures - probably involves a gradual introduction to practicum (Spooner-Lane et al, 2007) and to local schooling cultures, mentor training in schools and a targeted practicum experience involving much structured reflection. The importance of mentor training in schools was stressed by our academic interviewees as part of their view that schools needed to recognise their joint role in the preparation of LBOTE student-teachers. It is preferable for all support programs – whether focusing specifically on language or on schooling cultures - to be embedded in academic-credit-bearing studies so as not to place an unfeasible burden on teacher education Faculties (Gill, 2007; Skyrme, 2007). It is also preferable for them to be discipline-focused (Carpenter, 2005; Ryan and Viete, 2009).

Language proficiency alone is no guarantee of either success or failure (Cotton and Conrow, 1998; Dooley and Oliver, 2002). Our interviewees emphasised strongly the complex of factors which contributed to success or failure and suggested that in teacher education there is clearly a case for a system of richer testing, such as that discussed by O'Loughlin at Lancaster University (2008) and that may include interviews. Our academic interviewees showed that such measures have foundered in the past in some Faculties over the issue of the burden of workload. This seems a clear case for teacher education Faculties and teacher registration authorities to work together more closely on the issues of entry and exit requirements and ways to share the burden of a richer entry testing regime which includes, but goes beyond, language testing such as through IELTS. In terms of language testing specifically, the DEOTE instrument pioneered by Viete (1998) would appear to be one such useful instrument for diagnostic entry testing. The form taken by the NSW PEAT instrument which attempts to account for some aspects of context specificity is a useful model for exit testing, though no widely used test as yet would appear to deal in a nuanced way with 'pragmatic or strategic competence such as simplicity and clarity' (Elder, 1994a, p 56). While 'rich' occupation-specific language testing would obviously not want such 'pragmatic or strategic competence' to totally replace the kinds of skills tested by PEAT or ISLPR, exit-level competence could include them, as they are obviously aspects of language use important to teachers and are given emphasis in teacher education programs. Moreover, this would go some way to ensuring 'a reasonable degree of fit between behaviours elicited from candidates in the artificial environment of the test and actual performance in the target domain' (Elder and Brown, 1997, p 77). Any future move which IELTS might consider into teacher-education-specific testing would most usefully include a diagnostic component on entry and 'pragmatic or strategic competence' for exit testing.

This research has shown a strong emphasis being placed by both students interviewed and academics on the importance of listening comprehension, to 'playing the classroom tennis' – that is, being comfortable enough in English to be able to respond reasonably quickly and appropriately in spoken classroom exchange. Writing and speaking and questions-to-be-asked can be prepared in advance, but listening and responding appropriately and 'in time' cannot. Elder has suggested that 'special attention be given to listening scores in selecting applicants for teacher education' (Elder, 1993b, p 83). We would concur and we would point out the tendency revealed in Tables 1, 2 and 3 above for listening to be relatively undervalued in IELTS entry scores (with the ACU, Sydney and Flinders Universities the exceptions). We would also add that listening should be also an area of particular concentration for student support within the course.

Courses do need to deal openly with the perspectives that students bring with them about appropriate teaching practices, where these are markedly different from local pedagogical practices. However, it is true of both LBOTE and native-English-speaking student teachers that their perceptions of teaching are related to how their beliefs about teaching/learning were formed in their schooling (Spooner-Lane et al, 2007), and these need to be analysed and critiqued in all cases to develop a deeper awareness of the complexities of teaching and learning. As one interviewee argued, this means making 'our practice a basis of conversations with our students'.

The reverse side of this coin is that courses also need to recognise and build on the funds of knowledge LBOTE students bring with them. Ryan and Viète, as shown earlier (2009, p 304), point to the absence of reciprocity or mutuality in learning across intellectual cultures and to an undervaluing of LBOTE students' knowledge. LBOTE students present opportunities for engaging different educational cultures, intellectual heritages and transnational knowledge networks. A little explored issue in the internationalisation of Australian higher education is the prospects for knowledge exchange using the intellectual resources LBOTE/international students have gained, or have access to, from their homeland. Further research in this area is needed to disrupt the construction of LBOTE students, especially from developing countries, as 'deficient' and requiring 'remedial' education. Such research is needed to contribute to the debate over international students' knowledge being ignored. This is not a question of idealistically or simplistically ignoring the reality that students are being asked to teach in schools which are not going to change overnight to accommodate them. It raises questions about how one prepares teachers with a different set of capabilities, needs, demands and social capital from local students. It is partly a question of seeking out the funds of knowledge which they bring with them and inquiring into the pedagogical cultures with which they are familiar in order to ask what those funds of knowledge and what those cultures have to say about the kinds of issues which they face in their teaching – looking, as one of our interviewees argued, 'at linguistic diversity as a benefit rather than a deficit'.

## REFERENCES

- Arends, RI, 2004, *Learning to Teach* McGraw-Hill, Boston
- Barnes, D, 1976, *From Communication to Curriculum*, Penguin, Harmondsworth
- Barnes, D, Britton, J and Rosen, H, 1971, *Language, the Learner and the School*, Penguin Harmondsworth
- Barrett-Lennard, S, 1997, 'Encouraging autonomy and preparing for IELTS: Mutually exclusive goals?', *Prospect*, vol. 12, no 3, pp 29–40
- Barry, K and King, L, 1988, *Beginning Teaching: A Developmental Text for Effective Teaching*, Social Science Press, Wentworth Falls
- Bayliss, A and Ingram, DE, 2006, 'IELTS as a predictor of academic language performance', Paper delivered to Australian International Education Conference, available at: [www.aiec.idp.com/pdf/BaylissIngram%20\(Paper\)%20Wed%201630%20MR5.pdf](http://www.aiec.idp.com/pdf/BaylissIngram%20(Paper)%20Wed%201630%20MR5.pdf)
- Bradley, D (Chair), 2008, *Review of Australian Higher Education*, Australian Government, Canberra
- Britton, J, 1970, *Language and Learning*, Penguin, Harmondsworth
- Brown, A and Taylor, L, 2006, 'A worldwide survey of examiners' views and experience of the revised IELTS Speaking Test', *Research Notes*, vol. 26, pp 14–18, available at: [http://www.cambridgeesol.org/rs\\_notes/rs\\_nts26.pdf](http://www.cambridgeesol.org/rs_notes/rs_nts26.pdf)
- Bullock, A, 1975, *A Language for Life*, HMSO, London
- Campbell, M, Tangen, D and Spooner-Lane, R, 2006, 'A program for preparing ethnic minority student teachers for practicum', *International Journal of Practical Experiences in Professional Education* vol. 9, no, pp 1–10, available at: <http://eprints.qut.edu.au/6044/>
- Canagarajah, S, 2002, 'Reconstructing local knowledge', *Journal of Language, Identity and Education*, vol 1, no 4, pp 243–259
- Capel, S and Gervis, M, 2005, 'Motivating Pupils', in S Capel, M Leask and T Turner (eds) *Learning to Teach in the Secondary School: A Companion to School Experience*, Routledge, Milton Park, pp 120–135
- Carpenter, L, 2005, 'Managing diversity and difference: developing support initiatives for NESB students in teacher education programs', paper delivered to Australian Teacher Education Association Conference, available: [www98.griffith.edu.au/dspace/handle/10072/2475](http://www98.griffith.edu.au/dspace/handle/10072/2475)
- Cashdan, A (ed), 1972, *Language in Education*, RKP and Open University Press, London
- Cazden, C, John, VP and Hymes, D (eds), 1972, *Functions of Language in the Classroom*, Teachers College Press, New York

- Chalhoub-Deville, M and Turner, C E, 2000, 'What to look for in ESL admission tests: Cambridge Certificate Exams, IELTS, and TOEFL', *System*, vol 28, no 4, pp 523–539
- Coleman, D, Starfield, S and Hagan, A, 2003, 'The attitudes of IELTS stakeholders: Students' and staff perceptions of IELTS in Australia, UK and Chinese institutions', *IELTS Research Reports 2003* vol 5, pp159-235, IELTS Australia, Canberra
- Consolo, DA, 2006, 'On a (re) definition of oral language proficiency for EFL teachers: Perspectives and contributions from current research', *Melbourne Papers in Language Testing*, vol 11, no 1, pp 1-28
- Cotton, F and Conrow, F, 1998, 'An investigation of the predictive validity of IELTS amongst a group of international students studying at the University of Tasmania', *IELTS Research Reports*, vol 1, IELTS Australia and British Council, Canberra, pp 72–115
- Coulthard, M, 1977, *An Introduction to Discourse Analysis*, Longman, London
- Creber, JWP, 1972, *Lost for Words: Language and Educational Failure*, Penguin, Harmondsworth
- Criper, C, and Davies, A, 1988, *IELTS Research Report 1 (i) ELTS Validation Project Report* British Council and UCLES, London
- Cruickshank, K, Newell, S and Cole, S, 2003, 'Meeting English language needs in teacher education: A flexible support model for non-English speaking background students', *Asia-Pacific Journal of Teacher Education*, vol 31, no 3, pp 239–247
- Davies, A, 2008, *Assessing Academic English Testing English Proficiency 1950–89: The IELTS Solution*, Cambridge University Press, Cambridge
- Deumert, A, Marginson, S, Nyland, C Ramia, G and Sawir, E, 2005, 'The social and economic security of international students in Australia: Study of 202 student cases - Summary report', Monash Institute for the Study of Global Movements, Monash
- Dooey, P, and Oliver, R, 2002, 'An investigation into the predictive validity of the IELTS test as an indicator of future academic success', *Prospect*, vol 17, no 1, pp 36–54
- Elder, C, 1993a, 'How do subject specialists construe classroom language proficiency?', *Language Testing*, vol 10, pp 235–254
- Elder, C, 1993b, 'Language proficiency as predictor of performance in teacher education', *Melbourne Papers in Language Testing* vol 2, no 1, pp 1–17
- Elder, C, 1993c, *The Proficiency Test for Language Teachers: Italian, Vol 1: Final Report on the Test Development Process*, Department of Employment, Education and Training, Canberra
- Elder, C, 1994a, 'Are raters' judgements of language teacher effectiveness wholly language based?', *Melbourne Papers in Language Testing*, vol 3, no 2, pp 41–61
- Elder, C, 1994b, 'Performance testing as a benchmark for LOTE teacher education', *Melbourne Papers in Language Testing*, vol 3, no 1, pp 1–25

- Elder, C, 2001, 'Assessing the language proficiency of teachers: Are there any border controls?', *Language Testing*, vol 18, no 2, pp 149–170
- Elder, C and Brown, A, 1997, 'Performance testing for the professions: Language proficiency or strategic competence?', *Melbourne Papers in Language Testing*, vol 6, no 1, pp 68–78
- Elder, C, Ishiwata, N and Brown, A, 1994, *The Proficiency Test for Language Teachers: Japanese*, Department of Employment, Education and Training, Canberra
- Elder, C, and O'Loughlin, K, 2003, 'Investigating the relationship between intensive English language study and bandscore gain on IELTS', *IELTS Research Reports, Volume 4*, IELTS Australia, Canberra, pp 207–254
- Emerson, R M, Fretz, R I, and Shaw, L L, 1995, *Writing Ethnographic Fieldnotes*, University of Chicago Press, Chicago
- Ezzy, D, 2002, *Qualitative Analysis: Practice and Innovation*, Allen and Unwin, Crows Nest
- Feast, V, 2002, 'The impact of IELTS scores on performance at university', *International Education Journal*, vol 3, no 4, pp 70–85
- Green, A, 2005, 'EAP study recommendations and score gains on the IELTS Academic Writing test', *Assessing Writing*, vol 10, no 1, pp 44–60
- Green, A, 2007, 'Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses', *Assessment in Education: Principles, Policy & Practice*, vol 14, no 1, pp 75–97
- Gill, S, 2007, 'Overseas students' intercultural adaptation as intercultural learning: A transformative framework', *Compare: A Journal of Comparative and International Education*, vol 37, no 2, pp 167–183
- Han, J, 2006, 'The retention of "World English Speaking" student-teachers: A cultural studies perspective on teacher shortages in the context of the mobility of trans-national knowledge workers', unpublished PhD Thesis, University of Western Sydney
- Hatch, J A, 2002, *Doing Qualitative Research in Education Settings*, State University of New York, Albany
- Hawkey, R, 2005, 'A study of the impacts of IELTS, especially on candidates and teachers', *IATEFL Research News*, vol 15, pp 71–75
- Hyatt, D and Brooks, G, 2009, 'Investigating stakeholders' perceptions of IELTS as an entry requirement for higher education in the UK', *IELTS Research Reports*, Volume 10, pp 17–68
- IDP Education Pty Ltd, 2007, 'International students in Australian education', available at: [http://www.idp.com/research/fast\\_facts/higher\\_education.aspx](http://www.idp.com/research/fast_facts/higher_education.aspx)
- International Second Language Proficiency Rating (ISLPR), nd, 'Professional registration: Teachers' Available at: <http://www.islpr.org/>

Iwashita, N and Elder, C, 1997, 'Expert feedback: Assessing the role of test-taker reactions to a proficiency test for teachers of Japanese', *Melbourne Papers in Language Testing*, vol 6, no 1, pp 53–67

Krause, K, Hartley, R, James, R and McInnis, C, 2005, *The First Year Experience in Australian Universities: Findings from a Decade of National Studies*, Department of Education, Science and Training, Canberra

Leask, M and Moorhouse, C, 2005, 'The student teacher's role and responsibilities', in S Capel, M Leask and T Turner (eds) *Learning to Teach in the Secondary School: A Companion to School Experience*, Routledge, Milton Park, pp 18–31

Mahdavy, B, 2008, 'The role of multiple intelligences (MI) in listening proficiency: A comparison of TOEFL and IELTS listening tests from an MI', *Asian EFL Journal*, vol 10, no 3, pp 109–126

Marland, M, 1977, *Language Across the Curriculum*, Heinemann, London

Mayor, B M, 2006, 'Dialogic and hortatory features in the writing of Chinese candidates for the IELTS test', *Language, Culture and Curriculum*, vol 19, no 1, pp 104–121

McBurney-Fry, G, 2002, *Improving Your Practicum: A Guide to Better Teaching Practice*, Social Science Press, Katoomba

McCluskey, K, 2004, 'Teaching in Queensland schools: The rural experience of an NESB pre-service teacher', *Educating: Weaving Research into Practice*, vol 3, pp 1–9

Moore, T, and Morton, J, 2005, 'Dimensions of difference: A comparison of university writing and IELTS writing', *Journal of English for Academic Purposes*, vol 4, no 1, pp 43–66

O'Loughlin, K, 2008, 'The use of IELTS for university selection in Australia: A case study', *IELTS Research Reports*, Volume 8, pp 145–187

O'Loughlin, K and Arkoudis, S, 2009, 'Investigating IELTS exit score gains in higher education', *IELTS Research Reports*, Volume 10, pp 95–180

O'Neill, T R, Buckendahl, C W, Plake, B S, and Taylor, L, 2007, 'Recommending a nursing-specific passing standard for the IELTS Examination', *Language Assessment Quarterly*, vol 4, no 4, pp 295–317

Phakiti, A P, 2008, 'Predicting NESB international postgraduate students' academic achievement: A structural equation modeling approach', *International Journal of Applied Educational Studies*, vol 3, no 1, pp 18–38

Ramsay, S, Jones, E and Barker, M, 2007, 'Relationship between adjustment and support types: Young and mature-aged local and international first year university students', *Higher Education*, vol 54, pp 247–265

Rogers, S, 2004, 'Sound preparation: Accounts in IELTS listening materials', *TESOLANZ Journal: The journal of the TESOL Association of Aotearoa New Zealand*, vol 12, pp 59–66, available at: [www.tesolanz.org.nz/Site/Publications/journal.aspx](http://www.tesolanz.org.nz/Site/Publications/journal.aspx)

Rosen, C and Rosen, H, 1973, *The Language of Primary School Children*, Penguin, Harmondsworth

Ryan, J and Viète, R, 2009, 'Respectful interactions: Learning with international students in the English-speaking academy', *Teaching in Higher Education*, vol 14, no 3, pp 303–314

Sinclair, JMcH and Brazil, D, 1982: *Teacher Talk* Oxford University Press, Oxford

Skyrme, G, 2007, 'Entering the university: The differentiated experience of two Chinese international students in a New Zealand university', *Studies in Higher Education*, vol 32, no 3, pp 357–372

Spolsky, B, 1985, 'The limits of authenticity in language testing', *Language Testing*, vol 2, pp 31–40

Spooner-Lane, R, Tangen, D and Campbell, M, 2007, 'When your first year is your final year: Changing perceptions of practicum through NESB pre-service teachers' eyes', in *Proceedings of the First Year Higher Education Conference*, Queensland University of Technology, Brisbane, available at: [http://www.fyhe.qut.edu.au/past\\_papers/papers07/final\\_papers/pdfs/5b.pdf](http://www.fyhe.qut.edu.au/past_papers/papers07/final_papers/pdfs/5b.pdf)

Stephen, D, Welman, J and Jordaan, W, 2004, 'English language proficiency as an indicator of academic performance at a tertiary institution', *SA Journal of Human Resource Management*, vol 2, no 3, pp 42–53, available at: <http://sajhrm.co.za/index.php/sajhrm/article/view/48/48>

Stubbs, M, 1976, *Language, Schools and Classrooms*, Methuen, London

Stubbs, M, 1983, *Discourse Analysis*, Basil Blackwell, Oxford

Turney, C and Teaching Skills Development Project, 1983a, *Sydney Micro-Skills, Re-Developed: Explaining, Introductory Procedures and Closure, Advanced Questioning*, Sydney University Press, Sydney

Turney, C and Teaching Skills Development Project, 1983b, *Sydney Micro-Skills, Re-Developed: Reinforcement, Basic Questioning, Variability*, Sydney University Press, Sydney

University of New South Wales (UNSW) Institute of Languages, 2009, 'PEAT - Professional English Assessment for Teachers', available at: [www.languages.unsw.edu.au/testing/PEAT.html](http://www.languages.unsw.edu.au/testing/PEAT.html)

Viète, R, 1998, 'Culturally sensitive and equitable assessment of oral English for overseas-qualified teacher trainees', *Journal of Intercultural Studies*, vol 19, no 2, pp 171–185

Vinke, A and Jochems, W, 1993, 'English proficiency and academic success in international postgraduate education', *Higher Education*, vol 26, pp 275–285

Weisz, M, and Nicolettou, A, 2004, 'Making the transition to university : Do IELTS scores matter?', in *Doing Thinking Activity Learning*, eds, J Searle, C McKavanagh and D Roebuck, vol 2, Centre for Learning Research, Griffith University, Brisbane, pp 238–243

Wilkinson, A, 1971, *The Foundations of Language: Talking and Reading in Young Children*, Oxford University Press, London

Wilkinson, A, 1975, *Language and Education*, Oxford University Press, London



Woodrow, L 'Academic success of international postgraduate education students and the role of English proficiency', *University of Sydney Papers in TESOL*, vol 1, pp 51–70

Wylie, E and Ingram D, 1995a, *Australian Second Language Proficiency Ratings (ASLPR): Version for Japanese*, Commonwealth of Australia, NLLIA, Griffith University, Queensland

Wylie, E and Ingram D, 1995b, *Australian Second Language Proficiency Ratings (ASLPR): Version for Second Language Teachers*, Commonwealth of Australia, NLLIA, Griffith University, Queensland

Zhou, Y, Jindal-Snape, D, Topping, K and Todman, J, 2008, 'Theoretical models of culture shock and adaptation in international students in higher education', *Studies in Higher Education*, vol 33, no 1, pp 63–75

Zwodziak-Myers, P and Capel, S, 2005, 'Communicating with pupils', in *Learning to Teach in the Secondary School: A Companion to School Experience*, eds S Capel, M Leask and T Turner, Routledge, Milton Park, pp 105–119

## **APPENDIX 1: GLOSSARY OF TERMS**

<b>ASLPR</b>	Australian Second Language Proficiency Rating
<b>DEOTE</b>	Diploma of Education Oral Test of English
<b>DipEd</b>	Diploma of Education
<b>ELTS</b>	English Language Testing Service
<b>GPA</b>	Grade point average
<b>IELTS</b>	International English Language Testing System
<b>ISLPR</b>	International Second Language Proficiency Rating
<b>L1</b>	First language
<b>L2</b>	Second, and usually subsequent, languages
<b>LBOTE</b>	Language background other than English
<b>LOTE</b>	Languages other than English
<b>LSP</b>	Language for Specific Purposes
<b>NESB</b>	Non-English speaking background
<b>NSW</b>	New South Wales
<b>NSWDET</b>	New South Wales Department of Education and Training
<b>PEAT</b>	Professional English assessment for teachers
<b>RPL</b>	Recognition of Prior Learning
<b>SOSE</b>	Studies of Society and the Environment
<b>SWA</b>	Semester weighted averages
<b>TESOL</b>	Teaching English to Speakers of Other Languages (also: Teachers of English to Speakers of Other Languages)
<b>TOEFL</b>	Test of English as a Foreign Language
<b>TRB</b>	Teachers Registration Board
<b>TSE</b>	Test of Spoken English
<b>TWE</b>	Test of Written English
<b>VIT</b>	Victorian Institute of Teaching

### **3 A multiple case study of the relationship between the indicators of students' English language competence on entry and students' academic progress at an international postgraduate university**

#### **Authors**

**Gaynor Lloyd-Jones**

**Charles Neame**

**Simon Medaney**

Cranfield University

Grant awarded Round 13, 2007

An investigation into the selection practices and decision making rationales of admissions personnel in an international, postgraduate UK setting and the consequences for borderline non-native English speaking students' academic progress.

#### **ABSTRACT**

There is concern in the UK about declining degree standards due to the impact of internationalisation initiatives upon the expanded taught Masters postgraduate sector. Despite interest in the policy and managerial aspects of internationalisation of higher education, few studies have researched selection procedures that might illuminate current practices.

A case study approach was employed to study student selection in various Masters programmes in a postgraduate UK higher education institution specialising in engineering and management. The research revealed various selection processes in operation, some dependent upon English test scores, others reliant upon expert linguist assessments. There were differences between Schools in entry requirements for NNES students and in selection practices. Whatever the process or requirements, there was substantial support for complex, holistic rationales underlying Course Directors' selection decisions. Course Directors took into consideration academic qualifications and interests, motivation, readiness to adapt to UK HE culture, educational background and work experience.

Course Directors were most concerned about students' writing abilities which were difficult to assess reliably on entry and sometimes this resulted in failure to reach the required standard for the thesis. This impacted upon the workloads of thesis supervisors and cast doubts upon the reliability of entry assessments to predict academic writing abilities in this context.

The academic progress of students with borderline English language skills was followed during the year using several measures. Over half of the group was instructed to revise and resubmit their theses. In general, these students performed in line with their initial borderline status until the end of the year. The initial identification of students as borderline appeared sound whichever method was used to assess their language proficiency.

The unusual aspects of the institutional context and the nature of the enquiry discourage generalisation but offer opportunities for further comparative case study research in contrasting settings.

## **AUTHOR BIODATA**

### **GAYNOR LLOYD-JONES**

Gaynor Lloyd-Jones has worked in academic development in higher education for several years, formerly in undergraduate medical education. She was a member of the team that introduced the problem-based learning undergraduate medical curriculum at the University of Liverpool in 1996. For her PhD she conducted an ethnographic study of fresher students on the problem-based learning course. She currently holds a research role in Cranfield University's Centre for Postgraduate Learning and Teaching. Her research interests lie in the postgraduate sector of higher education, situated learning and the relationship between work and academic learning.

### **CHARLES NEAME**

Charles Neame is head of Cranfield University's Centre for Postgraduate Learning and Teaching. He has many years experience as a lecturer and senior lecturer in HE, designing, teaching and managing courses at undergraduate and postgraduate level for a very international student body. He manages Cranfield's Postgraduate Certificate in Learning, Teaching and Assessment for new academic staff, and his current research is focused on the process of fostering innovative practice in learning and teaching within academic communities.

### **SIMON MEDANEY**

Following a PGCE / TEFL at the Institute of Education in London Simon Medaney worked in mainstream TEFL for 15 years before joining Cranfield as Head of the Language Centre at Silsoe College. He has been Course Director of Cranfield's pre-sessional EAP programme since 1994 as well as providing language support for MSc and PhD students and screening non-native English speaking applicants for the Registry. He has worked throughout the European Union and designed pre-departure EAP programmes for Cranfield in Indonesia and Iran. Since 2006 his main focus has been Cranfield's European Partnerships Programme, where he assesses applicants for MSc programmes. His main interests are Academic Discourse and Intercultural Communication.

## CONTENTS

<b>1</b>	<b>Introduction .....</b>	<b>133</b>
	1.1 Background.....	133
	1.2 UK university entry requirements for NNES applicants .....	134
	1.2.1 IELTS test .....	134
	1.2.2 UK university entrance requirements for English test scores .....	135
<b>2</b>	<b>Literature Review .....</b>	<b>136</b>
	2.1 Predictive validity studies of IELTS and academic progress.....	137
	2.2 The knowledge of admissions staff about English language tests.....	137
	2.3 Decision making processes of student selection .....	138
<b>3</b>	<b>Context for Study .....</b>	<b>139</b>
<b>4</b>	<b>Aims of The study .....</b>	<b>140</b>
<b>5</b>	<b>Research Design and Methodology .....</b>	<b>141</b>
<b>6</b>	<b>Pilot study .....</b>	<b>142</b>
	6.1 Data collection methods.....	142
	6.2 Findings .....	142
	6.2.1 Selection procedures .....	142
	6.2.2 Academic English provision .....	143
	6.2.3 Analysis of selected application forms .....	144
<b>7</b>	<b>Interview study with Course directors .....</b>	<b>147</b>
	7.1 Method .....	147
	7.1.1 Sampling .....	148
	7.1.2 Interview schedules .....	150
<b>8</b>	<b>Findings .....</b>	<b>151</b>
	8.1 Models of selection practices.....	151
	8.1.1 Demand for places.....	153
	8.1.2 Large student groups with a shared language other than English.....	154
	8.1.3 Class size.....	154
	8.2 Course Directors' use of English test scores in the selection process.....	154
	8.2.1 General selection criteria .....	154
	8.2.2 Borderline cases .....	155
	8.2.3 Sceptics .....	156
	8.2.4 Management of the selection process and learning how to select .....	156
	8.3 Course Directors' views of NNES students academic and linguistic process.....	157
	8.3.1 Relationship between test scores and academic progress.....	158
	8.3.2 Speaking skills .....	158
	8.3.3 Writing skills .....	159

<b>9</b>	<b>Summer Programme Students' Progress .....</b>	<b>160</b>
9.1	Data collection methods.....	161
9.1.1	Documentary sources.....	161
9.1.2	Examination scripts.....	161
9.1.3	Questionnaire for thesis supervisors.....	162
9.1.4	Academic progress .....	162
9.2	Findings .....	162
9.2.1	Summer Programme students - language assessment at entry.....	162
9.2.2	Summer Programme students - Pre-test IELTS .....	164
9.2.3	Summer Programme students - Reports .....	165
9.2.4	Exam scripts of Summer Programme students .....	167
9.2.5	Summer Programme students - Workload of thesis supervisors.....	165
9.2.6	Summer Programme students - thesis outcomes.....	167
<b>10</b>	<b>Discussion .....</b>	<b>172</b>
10.1	Course Directors' admissions practices and experiences .....	172
10.2	The relationship between Summer Programme students' entry assessments and subsequent linguistic and academic progress .....	175
10.3	The consequences of different admission criteria and practices upon postgraduate students' academic progress .....	176
<b>11</b>	<b>Conclusion.....</b>	<b>177</b>
	Acknowledgements .....	177
	References.....	178
	Appendix 1: Interview Schedule for Course Directors .....	180
	Appendix 2: Focus group guide .....	181
	Appendix 3: Questionnaire for supervisors of masters theses of Summer Programme students .....	182
	Appendix 4: Supervisors' Responses to Q6 of the questionnaire for supervisors of Masters theses .....	183

## 1 INTRODUCTION

### 1.1 Background

The past decade has seen rising numbers of students seeking degree level study outside their home countries. UNESCO estimates that figures for international students have risen from 1.7 million in 2000 to 2.5 million in 2006. The Institute for International Education (IEE) which currently tracks student numbers across national borders estimates the equivalent figure for 2006/7 to be 2.9 million (IEE website, 2008). Of the eight most popular destination countries for international study, four are English speaking; the US, UK, Australia and Canada. The UK is the second most popular destination for foreign study, taking a 13% share of all international students in 2006/7. Whilst the IEE figures demonstrate variation in international student enrolment between countries and over time, the rising trend is consistent for the UK. Between 2002/3 and 2006/7, the number of enrolled international students in the UK rose from 305,395 to 376,190. The most popular countries of origin for UK international study in 2006/7 were China, India, the US, Germany, France, Ireland, Greece, Malaysia and Nigeria.

These effects of the internationalisation of higher education have been particularly pronounced in the UK where governmental policies have directly, and indirectly, encouraged the expansion of international student numbers. Successive initiatives by the then Prime Minister, Tony Blair, in 1999 and 2006 (Prime Minister's Initiatives 1 and 2) specified target increases in the number of non-UK students studying in UK higher education (DIUS, 2008; British Council, 2008). The targets of PMI were exceeded ahead of schedule and PMI2 aims to increase international students by a further 70,000 in 2011. The influx of overseas students has occurred simultaneously with the growth and diversification of the UK higher educational sector and the transfer of funding responsibility from the state to the individual student. However, the accompanying statutory limit on tuition fees for home students has inadvertently introduced an economic incentive for higher educational institutions (HEIs) to seek alternative sources of income and it is probably not coincidental that the same period has seen HEIs developing and implementing internationalisation strategies. Consequently, government policy has indirectly encouraged HEIs to maintain their own financial stability through the pursuit of growing numbers of international students.

In the UK, statistics from the Higher Education Statistics Agency (HESA, 2008) confirm that much of the expansion in international student numbers has taken place in postgraduate programmes. At the start of PMI1 in 1999/2000, the total number of postgraduates studying in the UK was 408,620 of whom 23% were classified as non-UK. For the year 2006/7, postgraduate numbers had risen to 559,390 of whom 8.6% were from the non-UK EU and 24.4% from elsewhere. It is the latter group that have contributed most to the increase, as the percentage of non-UK EU students has remained steady since 2002/3 when the separate categories of origin were introduced. Thus, there has been both an absolute and proportional rise in non-UK students over the past nine years. HESA statistics do not, however, differentiate between research and taught postgraduate students but a Higher Education Policy Institute report published in 2004 (Sastry, 2004) demonstrates that the rise in student numbers is directly attributable to international enrolment on taught Masters programmes as postgraduate research student numbers have remained stable during the period covered by the report.

The value of a degree gained in the UK holds attractions for foreign students. As well as the reputation of the UK degree and the quality of higher education, the conferring of a postgraduate degree implies English language proficiency of a high standard; one sufficient to preclude any necessity for formal English language testing in the future. Additionally, the 1-year duration Masters course

compares favourably in terms of cost and time with 2-year Masters programmes on mainland Europe and elsewhere. As well as the opportunity to improve English language proficiency, the British Council website cites other advantages of UK higher education in the variety of courses available, the flexibility of study and the multicultural experience of UK postgraduate study where at many universities 'more than 30% students may be international' (British Council, 2008).

However, success in recruiting greater numbers of non-native English speaking (NNES) students into higher education may pose a threat to UK degree standards. This aspect of internationalisation has recently been publicly highlighted in a debate involving the Parliamentary Select Committee on Innovation, Universities, Science and Skills, the Quality Assurance Agency (QAA) and the BBC. In an inaugural lecture at the University of Buckingham, Professor Alderman claimed that standards of English literacy were low in UK universities and particularly so for international students whose fees contributed essential revenue. A number of ensuing BBC online articles quoting unnamed academics appeared to support Professor Alderman's view (Coghlan, 2008). However, the QAA, in responding to invitations from the Select Committee, pointed out the difficulties of pursuing and evaluating such claims because open disclosure is protected by confidentiality, consent and legal issues. The QAA added that they were undertaking research into the recruitment and English language abilities of international students (Select Committee, 2008).

Similar debates have arisen in other countries, notably in Australia in the 1990s where there was a comparable influx of NNES students into higher education. Coley (1999) cites contributions from the media and the literature to justify a survey of the English proficiency entry requirements of Australian HEIs. She found a wide variety of sources of evidence in use in Australia at the time and little standardisation between institutions. Claims of apparent discrepancy between selection processes and entry requirements, on the one hand, and reported language proficiency, on the other, call into question the procedures surrounding the selection of NNES students and the forms of evidence upon which selection decisions are based.

## **1.2 UK university entry requirements for NNES applicants**

### **1.2.1 IELTS test**

As part of their admission criteria, UK universities typically require NNES applicants to produce evidence of language skills in the form of formal English test scores. The International English Language Testing System (IELTS) is the test most frequently cited on university websites and, although alternative tests are accepted, the IELTS is the benchmark against which other test scores are compared. The British Council, IDP: IELTS Australia and the University of Cambridge ESOL Examinations jointly manage IELTS. IELTS has a worldwide reputation and several years experience of providing a reliable measure of English language ability. It operates through a network of 500 locations in 120 countries around the world and around 6000 organisations have used its services.

In the academic form of IELTS designed for university entry, scores are reported in whole and half numbers which carry qualitative descriptions of the associated language abilities at each level (see Table 1) (IELTS, 2007). IELTS seeks to grade performance, in preference to establishing a particular pass score. Consequently, IELTS urges institutional users to interpret test scores in the light of course demands, their experience of teaching overseas students and a consideration of sub-test scores (IELTS, 2007). In view of this, IELTS therefore leaves to academic stakeholders the responsibility for setting their own entrance requirements in terms of test scores.



However, IELTS has issued guidelines relating test scores to courses categorised by linguistic demand and academic load. Only band score 7.5 is graded as 'acceptable' for the most demanding programmes such as medicine and law, although 7 is 'probably acceptable'. Conversely, 'acceptable' and 'probably acceptable' levels of 6 and 5.5 are suggested as suitable for animal husbandry and catering, which are classified as 'linguistically less demanding training' courses. Courses classified as either more academically or linguistically demanding fall in between this range. From the UK HEI perspective, whether the course guidelines classification adequately reflects the current diverse landscape of UK higher education might be questioned following the growth in interdisciplinary courses, the variations in course type (research and taught) and level (under- and postgraduate). However, IELTS issues guidance for institutions or departments wishing to set their own standard of score level which is more appropriate to their own context.

<b>Band Score</b>	<b>Qualitative description of capability</b>
7 Good user	Has operational command of the language, though with occasional inaccuracies, inappropriacies and misunderstandings in some situations. Generally handles complex language well and understands detailed reasoning.
6 Competent user	Has generally effective command of the language despite some inaccuracies, inappropriacies and misunderstandings. Can use and understand fairly complex language, particularly in familiar situations.
5 Modest user	Has partial command of the language, coping with overall meaning in most situations, though is likely to make many mistakes. Should be able to handle basic communication in own field.
4 Limited user	Basic competence is limited to familiar situations. Has frequent problems in understanding and expression. Is not able to use complex language.

**Table 1. IELTS band scores and descriptors (Taken from the IELTS Handbook 2007)**

Whilst it would seem desirable to restrict university entrants to a single test for comparison and continuity, this is not practical, so most UK HEIs accept specified alternative tests as evidence of English language proficiency. Some of the more commonly used tests include the Test of English as a Foreign Language (TOEFL), the Cambridge ESOL series of Certificates and the Test of English for International Communication (TOEIC). Each test has a different structure which makes standardisation difficult but tables of broad equivalence between the scores of different tests are published (Gillett, 2008) and several university websites quote requirements for the various tests. The TOEFL, for instance, exists in two different forms: paper and internet based which reflect emphases on different language skills in each test format. Certain UK HEIs have produced their own English language examination, for example the Test of English for Educational Purposes (TEEP) at the University of Reading and the University of Warwick English Test (WELT).

### **1.2.2 UK university entrance requirements for English test scores**

As we have seen, IELTS recommends scores between 6 and 7.5 for entry to tertiary study and the majority of HEIs conform to these recommendations as searches for institutional requirements on the IELTS website confirm, although 7.5 is only rarely cited and then almost exclusively for medical and veterinary courses. The great majority lie between 6 and 6.5, implying that students possess 'generally effective command of the language'. However, there are a number of institutions which accept lower

levels, a few as low as 4.5 (Brown, 2008). The band score 4.5 falls between the Modest user (5) who: ‘has partial command of the language’ and the Limited user (4) for whom ‘basic competence is limited to familiar situations’ but who ‘is not able to use complex language’ (Table 1). It appears that UK HEIs currently accept a wide range of English language proficiency as judged on test score requirements for university entry.

The array of entry test scores may express the contextual sensitivity to disciplinary, programme and institutional diversity which IELTS encourages HEIs to employ when setting test score requirements. That these figures represent broad guidance only is evident when searching individual HEI websites which reveal variation in entry requirements within, as well as between, HEIs. Higher scores and/or specific levels on sub-scores may be demanded for research students, specific programmes and, occasionally, for postgraduate study. The diversity may also be a means of dealing with problems thrown up by dependence upon a fixed score entry requirement such that the degree of change in underlying measurement is out of proportion to the consequences. This situation is a familiar one in education since it is analogous to pass-fail assessment decisions

There is, however, another explanation for the range of test scores, particularly those at the lower end. Several UK HEIs have embraced internationalisation strategies in which the development of academic English departments features. These departments may participate in selection, be responsible for delivering pre-sessional or foundation courses and provide ongoing support during degree study. Pre-sessional courses aim to ensure that students with borderline entry scores attain the required proficiency at the start of the programme and go some way to explaining why UK HEIs accept lower entry test scores. It is worth noting that students who participate in pre-sessional course at their admitting institution are not always required to sit a formal English test at the end of the period of language study. Furthermore, it has been estimated that an improvement of one band score requires full time study of 200-300 hours (Gillett, 2008).

The purpose of this brief review of UK HEI entry requirements is to draw a broad brush picture of the current state of regulation and guidance in the area. It is not intended to provide comprehensive coverage of the entry requirements for UK tertiary education. The wide range of acceptable English test scores is, perhaps, surprising and prompts questions about selection processes more generally. What is clear, though, is that regulatory information cannot reflect the actual selection decision making process, the criteria employed or the rationales for the judgments made. A case can therefore be made for exploring selection rationales in greater detail in order to examine the relationship between degree standards and linguistic proficiency.

## **2 LITERATURE REVIEW**

The internationalisation of higher education literature is extensive with a bias to the student perspective, in line with much current UK higher education policy. Apart from the specialised series of IELTS Research Reports, relatively little research has explored the academic staff perspective, especially in the areas of academic practice not directly related to teaching and learning or of the pedagogic consequences of classes of mixed nationalities, cultures and English language proficiency.

The literature review is divided into three sections by topic: predictive validity studies of the IELTS test; studies of the use of formal test scores in entry to higher education and finally, a recent study of postgraduate student selection at a UK university.

## **2.1 Predictive validity studies of IELTS and academic progress**

Given the current diversity of scores required for tertiary study, it is legitimate to question the evidence upon which entry score requirements are based. The commonest form of research employed to address this question is the predictive validity study which examines correlations between entry test scores and subsequent academic outcomes of NNES students. An early and influential study of this type (Criper and Davies 1987) claimed a score of 6.5 best differentiated between academic success and failure. Yet, subsequent studies have not produced such strong correlations between academic outcomes and English entry test scores (Cotton and Conroy 1998; Hill et al 1999, Kerstjens and Neary 2000; Lee and Greene 2007). Interpretation of results is further clouded by methodological issues (Banerjee, 2003) that leave HEIs and their admissions staff poorly supported when selecting a clear cut off point below which students struggle to progress academically. Several authors now consider that linguistic ability is only one influence amongst many in achieving academic progress (Rea-Dickins et al 2007; O'Loughlin 2008) and that the research model upon which predictive validity studies are based is unable to reflect the reality of the complex, multifactorial, dynamic process of learning. Further support for this view is gained from a recent interview and observation study of NNES fresher students which found that students' language proficiency correlated with their pre-entry test scores suggesting that IELTS scores were reliable (Ingram and Bayliss 2007). Significant influences upon academic success include motivation, subject discipline, programme structure, socio-cultural context and adjustment and ongoing language support each of which are complex phenomena in their own right.

The results of outcomes research therefore suggests that greater attention to learning process and context may help to delineate more clearly the relationship between learning proficiency and academic progress, for instance, by exploring whether admissions personnel take into account considerations other than English language test scores when making selection decisions.

## **2.2 The knowledge of admissions staff about English language tests**

Another research track has investigated the attitudes and knowledge of admissions staff, both administrative and academic, towards the test instruments used at their institutions. Studies based in Australia, China and the UK, and at under- and post-graduate levels have consistently found that staff knowledge about the tests and the significance of scores has not always been as sound or extensive as the authors considered it might be (Coleman et al 2003; Rea-Dickins et al 2007; O'Loughlin 2008). However, in interpreting the results of these studies, it is worth remembering that the authors have considerable expertise in the topic of English language testing. In each case, recommendations have been made that institutions and test providers should strive to encourage greater awareness and knowledge of test structure and scores. Ethical considerations apart, the assumption here appears to be that more fully informed admissions decisions will translate into improved selection processes and outcomes, a proposition that might be, but does not yet appear to have been, tested by an intervention study. Based on the limited knowledge of staff in these studies and the favoured use of questionnaires as methodology, O'Loughlin (2007) suggests that there is scope for more detailed research exploring how staff use their knowledge to interpret and use test scores in selection contexts.

One example regularly cited as a topic of limited knowledge is that of sub-scores and their meanings. Rea-Dickins et al (2007) tracked the linguistic progress of postgraduate NNES students in the Departments of Education and Politics at a UK university through interviews and learning logs. These students considered their listening abilities were underestimated by formal tests but, conversely, that the IELTS reading test was insufficiently faithful to the academic context to provide a reliable measure of their reading ability when faced with the heavy load of postgraduate study. The task

within the test is important here as the IELTS reading test assesses the candidate's ability to answer questions immediately after reading a passage of text. Academic reading, on the other hand, requires the assimilation of several texts as a preface to the production of written work that incorporates the student's reading and drafted in their own words. This is a far more difficult task than the IELTS reading test and one for which the students in Rea-Dickins' study appeared unprepared. The authors suggest that listening and reading test sub-scores may be better indicators of success than speaking and writing. Whilst further research is necessary to uphold or refute the findings, the research suggests how test sub-scores might be useful in selection decisions.

### **2.3 Decision making processes of student selection**

A singular example of micro-level research on student selection is found in Banerjee's doctoral thesis (2003) in which she investigated, via semi-structured interviews, the selection rationales of two admissions tutors on the MBA and MA in Politics programmes at a UK university. Banerjee found that these admissions tutors did not deal with applicants in the somewhat algorithmic model of selection described at undergraduate level (O'Loughlin 2008). Their selection decisions represented balanced judgments achieved through consideration of a variety of criteria, which were sometimes competing, and always considered in concert with one another. This is due, in part, to the fact that applicants, especially borderline cases, do not necessarily demonstrate neatly categorized experiences, skills or qualifications; applicants successful on one criterion may be borderline on another. Tutors took into account previous academic experience and attainment, work experience, secondary education, referees and the completed admission form. Occasionally, admissions tutors interviewed candidates to probe, refute and validate information on the application form. Evaluating the comparative merits of applicants therefore required the operation of judgement on a variety of competences, factors and circumstances including English language proficiency. Only under unusual circumstances or where the demand for places is high is the admissions tutor unlikely to face decisions of the type described by Banerjee. The model of decision making described in this study is compatible with the IELTS recommendation to admissions staff to employ test scores with some flexibility according to the circumstances of each case.

Banerjee then followed eight NNEST students who had been selected by the admissions tutors in the study, classifying them according to the degree of 'academic risk'. Employing student interviews and critical incident diaries, she found that the admissions tutors' assessments of risk were sound, those students considered at greatest linguistic risk, reporting more difficulties and time costs in surmounting problems.

Two distinct models of selection emerge from these studies. In one, the language requirement is treated as independent of other selection criteria, a simple accept or refuse decision with little attention paid to borderline cases. In the other, decision making is richly complex, the ultimate choice contingent upon multiple, interacting criteria. The former portrays the tidy world of audit and regulation, apparently free of risk; the latter, the messier reality of everyday life. How these models operate and relate to each other in practice and in different contexts is a matter of speculation, given the lack of research evidence. It is possible that operations of scale may affect the way university selections are made. At undergraduate level, where there are larger numbers of applicants, the process is more likely to resemble the simpler decision making model. At postgraduate level, however, where classes are smaller and the student body more diverse in age, education, nationality and work experience, then the Banerjee model may be more applicable. In summary, these findings require further confirmation through comparative research in different settings. In particular, probing what factors and circumstances selection personnel consider when deciding whether or not to offer a candidate a place.

### 3 CONTEXT FOR STUDY

The institutional setting for the study is Cranfield University, a wholly postgraduate UK university, classified as a specialist institution by the Higher Education Funding Council of England (HEFCE) because of its strengths in engineering and aerospace. Cranfield's focus on applied knowledge marks it out as unusual amongst its peers. Instead of traditional academic disciplines, the Cranfield campus is organized into four Schools; the Schools of Management (SOM), Applied Sciences (SAS), Engineering (SOE) and Health (CH) and many of the Masters programmes on offer are multidisciplinary in line with the applied focus and strong existing links with industry and management. The rural location of the campus is rare for a UK HEI and renders students largely reliant upon themselves for social and extracurricular activities. In these, and other, respects it provides a contrast to the more typical UK university setting in Banerjee's study.

During the 2007/8 academic session there were 1646 students registered on taught postgraduate Masters programmes on the Cranfield campus. The distribution of students across the four Schools was as follows: Management 33%, Applied Sciences 29%, Engineering 31% and Health 7%. Whilst the numbers of overseas students has increased over recent years, the international character of the Cranfield student body has been established for some time such that in 2008 it was ranked second in the world for its international student community in the Times Higher Education (THE) World University Rankings. In 2007/8 there were over 110 nationalities represented on campus. The breakdown in terms of student nationality on the Cranfield campus was 36% UK, 31% non UK EU and 33% from elsewhere in the world. Compared to the HESA figures for the sector, the Cranfield student population has fewer UK students and proportionately more non UK EU students amongst the overseas group. These descriptive institutional statistics operate at the institutional level and so fail to convey the range of national diversity evident in classes between different Masters programmes.

Around 65 taught Masters programmes were offered in the 2007/8 academic session at Cranfield, although a number of programmes sub-divide into as many as four to six Options. Most programmes are headed by a single named Course Director, with one or two exceptions where directorship is shared and it is unusual for a single person to have responsibility for more than one Masters course. Directorship of Options varies; occasionally it remains the responsibility of the overall Course Director but, more usually, it is awarded to another member of teaching staff whose academic speciality it is. Occasionally but not routinely, Options Directors may also undertake selection duties. In 2007/8, the Schools of Health and Management offered around ten programmes each with the remainder equally divided between the Schools of Engineering and Applied Sciences. The disciplinary specialisms of Cranfield are associated with a bias towards men amongst staff and students (28% women).

Cranfield Course Directors undertake full responsibilities for student selection akin to Admissions Tutors in larger institutions. They play a crucial role in maintaining academic standards, acting as the link between classroom practice and institutional policy. With their responsibilities for overall assessment, Course Directors are in an almost unique position to make judgements about the relationship between selection decisions and academic outcomes on their programmes. They are also likely to be the final arbiter in borderline cases of selection.

The design of one year taught course Masters programmes at Cranfield is remarkably similar across subjects and Schools. Courses are modular in structure, with module length varying between one and four weeks. Typically, in the first term, teaching takes place largely through the medium of lectures and practicals; in the second, students undertake a group project and in the final term, they embark on an individual research project and thesis. Assessment takes place after each module and

examinations are scheduled in January and April. Students have only short leave breaks at Christmas and Easter and the customary summer holiday period is spent on individual research projects. High value is attached to the group project because of the opportunities it affords for work outside academia through team working and/or the relevance of the project to a particular company or industry. This is not merely academic since the Cranfield specialisms of management and aerospace are themselves highly multinational areas of practice. Apart from the MBA cohort which exceeds 100, class size varies between 10 and 80 students and this small class size is reflected in Cranfield being ranked first in the UK and eleventh in the world for the staff student ratio in the THE World University Rankings in 2007.

In a recent research study into the influence of student diversity on teaching and learning at Cranfield, English language emerged as the prime topic for both students and staff (Lloyd-Jones et al, 2007). Lecturers reported concern about the standard of students' English language competence, particularly in academic writing. Whilst listening and speaking skills generally improved over the first two to three months, students' writing did not make concomitant progress. There was also concern about the influence of group size and dynamics on teaching and learning. Because of a preference for socializing within the mother tongue group, cohorts where one mother tongue group is in the majority may hinder English language development and cultural integration. Nor are the consequences simply confined to the national group concerned. As non-native English speaking students were keen to improve their English and welcomed constructive feedback about their abilities, there were good reasons from an institutional perspective to investigate English language proficiency in greater depth.

In summary, Cranfield University contrasts with the settings of previous studies, by virtue of being exclusively postgraduate, in boasting a diverse international student body, in a commitment to applied knowledge and in its bias towards science and engineering disciplines. It therefore provides an opportunity to conduct comparative case study research which may validate and extend our existing knowledge of NNES student selection procedures and rationales (Ward Schofield, 2000).

## **4 AIMS OF THE STUDY**

The overall purpose of the study is to explore current admission practices in relation to English language testing and the consequences of selection decisions upon academic progress and the need for ongoing academic English support at an international, postgraduate UK university.

The study has the following aims:

- 1** To describe and explain Course Directors' admission practices and experience in relation to IELTS scores.
- 2** To examine the relationship between non-native English speaking students' pre-admission IELTS scores and their
  - a. academic progress and
  - b. ongoing English language support needs
- 3** To compare the consequences of different admission criteria and practices upon postgraduate students' academic progress in a variety of courses.

## 5 RESEARCH DESIGN AND METHODOLOGY

A case study approach was chosen as most appropriate to the examination of a contemporary phenomenon in context (Yin, 2003) and as a suitable basis for comparison with previous research particularly with Banerjee's study (2003) which had taken a similar research approach. An inductive rather than a deterministic approach towards research design was also preferred such that the research progressed stage by stage allowing the emerging findings to be incorporated into subsequent lines of enquiry and data collection (Maxwell, 2004). This interactive approach served the enquiry well, since early findings from the pilot study were unexpected and had a major impact upon the subsequent research. As a result, the enquiry became a series of inter-related research studies with subsidiary objectives and data collection methods. The report will follow this chronological outline with each successive research study presented in separate sections each with its own methods and findings sections. The intention is to lay bare the problems that emerged during the research and to identify how these were dealt with at the time and the consequences for research process and outcomes.

From the start, the research design was based upon two assumptions. The first was that selection practices might differ between Schools and, possibly, between individual Masters programmes. Not only did this seem plausible given the disciplinary differences but sufficient was known from the earlier diversity study (Lloyd-Jones et al, 2007) about institutional structure and School practices to justify the hypothesis and make it an important principle of sampling throughout the study. The second assumption was that all NNES students would enter a Masters programme with a formal English language test score. This did not turn out to be the case (see Section 6) and had major implications for the research (see Section 9) since it appeared to undermine the second and third aims of the original research proposal. The lack of a baseline measure of English language proficiency across the institution therefore thwarted attempts to compare NNES students' academic and linguistic progress and alternatives had to be sought elsewhere. Because of these difficulties it was decided to focus upon the small group of NNES students who had taken the pre-sessional Summer Programme in English in 2007 before starting Masters study in the autumn. Around half the Summer Programme students entered with a formal test score which could act as a benchmark. Although individual students might demonstrate variable proficiency between different language skills, it could be reasonably assumed that the level of English language proficiency amongst the group was borderline on entry. Furthermore, as borderline cases, their progress would be of particular interest in terms of the consequences of the rationales that had led to their admission in the first place.

A range of research methods was employed within the case study. The primary focus on the phenomena of selection rationales and decision making which are largely unseen favoured qualitative research methods which were also in keeping with the inductive research design. Qualitative methods were supplemented by documentary analysis in the pilot study and in the review of the Summer Programme students. Finally, the progress of Summer Programme students' progress was evaluated using a series of available measures, a review of their examination scripts and an electronic questionnaire survey of staff supervisors of their Masters theses.

The remainder of the report is divided into three sections, each describing the methods and findings for the three separate research studies which comprise the overall case study. These are:

- a pilot study reviewing selection practices within the institution
- a semi-structured interview study with Course Directors
- a review of Summer Programme students' progress using a range of measures

## **6 PILOT STUDY**

### **6.1 Data collection methods**

The purpose of the pilot study was to establish the institutional regulations and practices surrounding the admission of NNES students on taught Masters programmes at Cranfield. This included details of the pre-sessional Summer Programme English course and the nature and availability of ongoing academic English support on campus. Initially, data was gathered from Cranfield University webpages relating to the admission of NNES students and the Summer Programme but was later supplemented by informal interviews with Registry staff (4) concerned with admission and selection. The roles of these staff varied and included linguists, administrators and academic English teachers directly involved in the selection process and/or the pre-sessional Summer Programme. Notes were taken at the time of the interviews which were not audio-recorded.

Student statistics provided by Registry comprised a third data source with detailed breakdowns of students' nationalities, MSc programmes and English language conditions. Additional, and more detailed, figures were available for students entering through the European Partnership Programme (EPP) in 2007/8 and for students participating in the 2007 English language Summer Programme prior to admission in October 2007. These searches revealed that sub-scores for English language test scores were not stored electronically but existed in paper format which were filed with the students' admission forms. Consequently, a hand search of the admissions files of 177 selected students, equivalent to 10.5% of all taught Masters students was conducted. The records included all students where the offer of a place was conditional upon the evidence of English ability in ten MSc programmes across four Schools: one from Cranfield Health, five from SAS and two each from SOM and SOE. The choice of Masters programmes approximately matched the initial sampling rationale for Course Directors in the interview study because interest lay in discovering whether differences in admissions practices between Schools and programmes could be detected from the admissions documents (see Section 7.1.1 for details). The admissions files of the remaining ten Summer Programme students (three of whom were research students) were also included but here the focus was on the individual student and their progress. Therefore no attempts were made to extend the search to all NNES students on the five additional Masters programmes introduced into the study through the inclusion of the additional Summer Programme students.

### **6.2 Findings**

#### **6.2.1 Selection procedures**

The pilot study confirmed that Cranfield is in line with many other UK higher educational institutions (HEIs) in requiring the following minimum English test scores from applicants whose first language is not English: IELTS 6.5, TOEFL 580 (paper test) or 237 (computer test) or 92 (Internet test). Other tests, such as TOEIC (830) and the Cambridge ESOL are accepted, in lieu. These levels have been in force for nine years and represent minimum entry requirements which Schools and individual Course Directors may raise, but not lower, if they wish. The date of the test result should not pre-date the start of the course by more than two years. There is some concern about authenticity of paper certificates as fraudulent versions are known to circulate so scores and sub-scores are verified against electronic test databases.

The main finding of the pilot study, however, was the identification of a separate and distinctive entry route for European MSc students through the European Partnership Programme (EPP). This well established scheme, which is linked to the EU Erasmus Programme, permits EU undergraduate



students who are currently studying at their home institution to apply to take the final year of study at Cranfield in order to gain a double degree award from both institutions, the final year being an MSc from Cranfield. The route is restricted to continental institutions that have signed a partnership agreement with Cranfield. Currently, the Cranfield website lists 64 institutions in 15 countries but new institutions are being added, particularly from central Europe. The scheme has grown in recent years and the majority of non-UK EU Masters students now enter Cranfield by this route (330 out of 508 in 2007/8). The change has been reflected in the nationalities participating in the English language Summer Programme. Formerly dominated by the developing countries and the Pacific Rim, around 50% of students now come from Europe.

Admission and selection procedures for students entering through the EPP route who are non-native English speakers are distinct from other applicants. After first applying to their host institution and then to Cranfield, applicants are interviewed in their own country by a member of Cranfield staff. In some cases, Course Directors may also interview EPP applicants, although this practice is not universal. Assessment of English language proficiency is then made by the Academic English staff and Course Directors, based upon the interview assessments and written applications. Some students in this group may have sat formal English language tests but by no means all. For example, to gain a Diploma in the French Grandes Écoles system, students must achieve a TOEIC score of 750 (which is lower than the Cranfield entry requirement of 830). The relevance of this finding lies in the large number of French students entering Cranfield each year since there are partnership agreements with 29 French HEIs. There are advantages to the EPP procedure. Over time, Course Directors become increasingly familiar with the undergraduate courses from which students apply and so are able to assess their academic abilities in some detail from their transcripts. Another advantage is an immediate opinion on a borderline student's need for pre-sessional academic English tuition. In this case the student will be offered a place, conditional on the successful completion of the on-campus, pre-sessional Summer Programme. In 2007/8, of 29 students on the Summer Programme, 14 were EPP taught Masters students.

The circumstances of students whose first language is not English and who cannot participate in the EPP necessitate different arrangements for the assessment of English language proficiency. European students who are attending institutions outside the EPP are required to produce a satisfactory test score in line with the regulations. For students beyond Europe, however, there are numerous grey areas such as applicants with prior experience of English at work or in education for whom a formal test might be redundant. Where there are doubts, the Course Director is encouraged to interview these applicants by telephone, if necessary, to assess the applicant's language skills and, depending upon the outcome, either request a test or waive the requirement.

### **6.2.2 Academic English provision**

In recent years the increasing number of overseas students at Cranfield has raised awareness of the need for academic English support. Students who have not gained a 6.5 score may be admitted provided they attend an intensive pre-sessional English programme (Summer Programme) before the start of the year in October. The Summer Programme runs from July to September and aims to improve students' ability and confidence in the four language skills in the context of academic study at Cranfield. Broadly speaking, an IELTS score of 6 requires a minimum of one month's tuition, a 5.5 score requires two months and a student with a score of 5 must take the full three month course. Using a similar rule of thumb, students may improve by 0.5 of a band score for each month's tuition but this is very variable. The programme has the following capacity which varies with the students' ability: up to 6 students may attend the full three months, 12 for two months and a maximum of 30 for the

final month. Class teaching takes place each morning followed by a variety of individual and group activities in the afternoon. A range of writing tasks is set regularly. Assessment is largely formative and students meet with a tutor on a weekly basis to set goals and discuss progress. At the end of the course, a report is sent to their Course Director commenting upon the student's achievement and their ongoing need for English support. Each language skill is graded (range A to E) and accompanied by comments. The grade E indicates a student's language skills are inadequate for successful study at Masters level. Around 10% of Summer Programme students do not reach the required standard to enter a Masters programme.

The reports may also advise a student to continue with weekly English tuition for academic writing that is provided by the academic English staff. In 2008 the sessions were opened to other NNES students who responded with such enthusiasm that a screening test had to be used to assess eligibility. Occasional students may have one to one tuition where resources allow. Other support is provided at School level; the School of Management offers oral English classes but places are neither free nor unlimited. Nevertheless, students from other Schools are welcome to join these classes if considered appropriate by their Course Director.

### **6.2.3 Analysis of selected application forms**

Some of the complexity and subtleties involved in the admissions and selections process is apparent in the 177 hand searched records from students where evidence of English language proficiency was made a condition of entry. Were the model of algorithmic selection to prevail, then almost all students would be expected to present test scores and other forms of evidence would not be prominent (see Table 2). This is not the case as only two thirds (118) of the sampled students possessed a test score and 50 of these students were interviewed. It is not possible to identify from the records whether the test or the interview takes precedence in terms of either time or significance. In just over a third of applicants (68) the test score is the sole cited evidence of English language proficiency.

The Registry records suggest that the interview plays a frequent role in the assessment of English language which is sometimes conducted by telephone. 61 students were interviewed and the English test score requirement was waived for 11 of them. Again, the records are not always informative on the reasons for the waiver but experience of degree level education or working in England or in English are cited. In view of the diverse population found in postgraduate education (Sastry, 2004), these waivers do not necessarily imply that institutional policy is being flouted, as there are many reasons why applicants for postgraduate study may have good English language skills which can render a formal English test irrelevant. Because the reasons for waiver are not always supplied, it is likely that the number of recorded interviews in the figures is inaccurate, probably an underestimation. The records do, however, validate the Registry staff data by demonstrating that interviews are conducted by a Cranfield representative or, specifically, by Course Directors and, sometimes, together. They share the load almost equally between them but this is likely to vary with the number of EPP students in the sample. A larger number of non-EU students will favour the Course Directors.

<b>Category</b>	<b>Student numbers</b>
Total number of students	177
Cranfield Health	4
School of Applied Sciences	73
School of Engineering	53
School of Management	47
All students with an English test score	118
Students with an English test score alone	68
All students interviewed and with an English test score	50
All students interviewed	61
Students interviewed but without an English test score	11
Interview with linguist	33
Interview with Course Director	29
Interview with Course Director and linguist	1
Total number of recorded English test scores	141
Students with more than one English test score	20
Students with IELTS	50
Students with TOEFL	41
Students with TOEIC	17
Students with other tests – Cambridge ESOL etc	19
Students with GMAT	13
Students with overall IELTS scores of 6/6.5 with a sub-score of 5.5 or below	9

**Table 2. Methods used to assess students' English language proficiency for admission to Cranfield of a purposive sample of 177 students in 2007/8 (taken from Cranfield Registry data)**

Distinctions were apparent in selection processes between Schools and programmes. Students (18) applying for places in SOM were more likely to present more than one test score whether these were repeated tests of the same or different types. Amongst these is the General Management Aptitude Test, GMAT, which is not a test of English language ability per se, but assesses analytical writing, verbal ability and numerical reasoning and is widely used in academic business and management disciplines. All the students with GMAT were applying for places in SOM courses. There were also variations between Masters programmes in the extent to which interviews were cited by Course Directors as a method of gathering evidence. This was independent of interviews with other members of Cranfield staff. One Course Director in SOE interviewed almost half of the listed NNEs students (9 out of 20) whereas one course each in CH and SOM did not mention Course Director interviews and three others in SOE and SAS interviewed less than 10% of NNEs students within the cohort (see Table 3).

<b>School</b>	<b>Sampling code</b>	<b>Total number of students</b>	<b>Number of NNES students</b>	<b>Number of students interviewed by CD</b>
CH	CH	14	4	0
SAS	SAS1	49	20	1
SAS	SAS2	30	12	1
SAS	SAS3	32	24	4
SAS	SAS6	4	4	1
SAS	SAS7	19	11	2
SOE	SOE1	36	26	1
SOE	SOE2	42	20	9
SOM	SOM1	42	34	7
SOM	SOM2	30	13	0

**Table 3. Numbers of recorded interviews conducted by Course Directors in the selection of NNES students in 10 Masters programmes (taken from Registry data)**

Although the IELTS test is the most frequently used test of English proficiency amongst the sample (50) it is only slightly more popular than the TOEFL test (41). The search showed that an acceptable overall score could mask discrepancies in sub-scores. Of 24 students with overall scores of 6 or 6.5, nine students had at least one sub-score of 5.5 or less and six of these were in writing (see Table 4). In an unusual example, a student with an overall score of 6.5 had a sub-score in Writing of 4 and two other students with overall scores of 6 had 2 sub-scores below 6 in Writing and Speaking.

In summary, the findings from the pilot study confirmed initial assumptions that selection practices and procedures differed between Schools and Masters programmes. In relation to the first aim of the study, the findings pointed to differences between individual Course Directors, and hence Masters programmes, in their preference for the use of a selection interview. The findings were specifically incorporated into the next stage of the enquiry by including the topic in the interview schedule for Course Directors, as was the topic of sub-scores. Consequently, and as intended, the pilot study shaped the subsequent enquiry.

However, the diverse measures employed to assess English language proficiency for entry to the institution held implications for the second aim of the study. At the start of the research, it was assumed that the great majority of, if not all, NNES students would enter with an English test score but the pilot study showed this was not the case. The lack of a general baseline measure of NNES students' English language proficiency at entry meant comparisons on a scale initially envisaged for the research would be logically insecure. In addition, the small numbers of students with English test scores for whom a comparison might be attempted would probably render any statistical technique invalid. For that reason, we sought alternative means of fulfilling the second aim.

<b>School</b>	<b>Overall IELTS score</b>	<b>Listening score</b>	<b>Reading score</b>	<b>Writing score</b>	<b>Speaking score</b>
CRH	6.5	7	5	6	7
SAS	6	6	5.5	6	6
SAS	6.5	5.5	9	4	7
SAS	6.5	5.5	7	7	7
SAS	6	7	6.5	5.5	5.5
SAS	6.5	7	7.5	5	7
SOE	6	5.5	6.5	5	6
SOM	6	6	6.5	5	6
SOM	6	6	6	5	5

**Table 4. IELTS scores for NNES students scoring 6 or above overall but below 6 on one sub-score (taken from Registry data)**

At the same time it is important not to infer too much from the findings of what was functionally a pilot study. Firstly, sampling was purposive so precluding any generalisation to the institution at large. Secondly, it is safe to assume that the admissions data is incomplete, not in a regulatory sense but because the records do exactly that, they record. There is no requirement to explain or justify decisions, or to elaborate upon what takes place in an interview, however it may be conducted. To investigate this further would require a more detailed exploration of Course Directors' practices and experiences of selection which follows in the next section.

## **7 INTERVIEW STUDY WITH COURSE DIRECTORS**

### **7.1 Method**

The interview research study addresses the first aim in the proposal: to describe and explain Course Directors' admission practices and experience in relation to students' IELTS scores.

As much decision making appears to be implicit, one to one semi-structured interviews were chosen as the method of data collection best suited to reveal, elaborate and discuss this aspect of the Course Director's role. Nevertheless, in choosing the interview method, the limitations of the data were recognised as a source of inference for reported action when compared to observed behaviour (Hammersley, 2006). What respondents report may not parallel their actions for a variety of reasons but interviews were preferred as a method suitable for discussion and elaboration of their decisions and underlying rationales in the light of the Registry data and the research literature. In contrast to other studies, (Coleman et al, 2003; O'Loughlin, 2008), the interview study neither set out to investigate policy compliance nor to test academic staff's knowledge of English language testing.

### 7.1.1 Sampling

The sampling rationale and selection of Course Directors for inclusion in the interview study were decided by the research team, two of whom (CN and SM) had several years' experience of the institution. Sampling also took into account the views of admissions staff taking part in the pilot study into admission practices. The findings of the preliminary study favoured a purposive sampling approach designed to reveal the variety of viewpoints that reflected range and scope in preference to the construction of a typical perspective. Based upon contextual knowledge, experience of the setting and the wider socio-economic context of UK higher education, the research team identified the following criteria that might influence a Course Director's viewpoint on NNEST student selection:

- Schools/disciplines
- Demand for places on the programme
- Class size
- Courses with large groups sharing a first language other than English

The interpretation of the demand for places requires further elaboration. The institution is currently paying attention to the admission and selection process, in particular, to the numbers of applicants, the number of places offered and accepted. Whilst such information may be held at departmental or School level, it was not publicly available at the time of the study. Therefore, the judgement of the demand for places on individual courses was based upon the experience and knowledge of the institution of the relevant members of the research team (CN and SM). In an attempt to corroborate the sampling decisions, questions were included in the interview about student applicants, offers and conversion rates, although these data were variably forthcoming.

One Course Director was included in the sample because of an interest in English language testing although the Masters programme he directed also fitted the sampling criteria. As the study progressed, a further category emerged in terms of the length of experience of directorship, when it became apparent that one Course Director was in his first year in the role. The realisation prompted an additional train of enquiry about how staff involved in making selection decisions learn to do so and whether there might be a need for academic development and training in the area. The prevalent use of selection interviews, which are generally conducted in one to one situations, provided another justification for the inclusion of the category. For this particular MSc programme, two interviews were included, one with the existing Course Director and another with his predecessor.

Table 5 lists the participating Course Directors, their Schools, the variation within the criteria and the interviewee codes. All Course Directors were male save for the Course Director in Cranfield Health. Eight of the programmes in the pilot study were represented in the interview study.

Of 16 Course Directors approached to participate, one declined on the grounds that he was about to relinquish the role. He suggested that the incoming staff might contribute instead. As this programme had been designated a low demand course, it was important to pursue alternatives to avoid omitting a potentially significant perspective. Two solutions were found to the problem. In the first, a departmental staff member with four years experience of admission and selection for one of the programme options consented to an interview. In the second, it was decided to follow up the Course admission and selection responsibilities. A group interview was preferred due to the shared situation facing the group members (Morgan, 1997) and the limitations of experience in the role upon data gathering in a one-to-one situation. As well as exploring departmental staff attitudes, the focus group aligned with the emergent criterion of role expertise already discussed above and brought into the

study the views of academic staff who were themselves multilingual, non-native English speakers (German and Egyptian). Some of the focus group members had experience of teaching in other UK HEIs and abroad and so were able to offer comparative viewpoints. Another MSc programme was omitted (although it had been used in the pilot study) when it was realised that the Director was on maternity leave and her role was temporarily being undertaken by a colleague.

<b>Interviewee Code</b>	<b>School</b>	<b>Demand</b>	<b>Monolingual groups</b>	<b>Class size</b>	<b>Course Directorship experience</b>
CH	Cranfield Health	unspecified	None	14	4 years
SAS1A	School of Applied Sciences	high	14 French	49	In first year of post
SAS1B					6 years
SAS2	School of Applied Sciences	high	None	30	10 years +
SAS3	School of Applied Sciences	low	None	32	4 years
SAS4	School of Applied Sciences	low	None	10-15	6 years
SAS5	School of Applied Sciences	medium	None	10	3 years
SOE1	School of Engineering	high	None	36	10 years +
SOE2	School of Engineering	low	10 Polish	42	10 years +
SOE3	School of Engineering	high	19 French	52	4 years
SOE4	School of Engineering	high	None	50	10 years +
SOM1	School of Management	high	None	42	3 years
SOM2	School of Management	high	None	30	1 year
SOM3	School of Management	high	None	55	5 years

**Table 5. Sampling for Course Directors' Interview study**

The selected interviewees were approached by email and/or telephone and given a brief summary of the research, an explanation for their selection and an invitation to participate in an interview about admission and selection practices of NNES students. They were informed that the interview would be audio-recorded and transcribed but that his or her identity would be protected in future presentations and publications. For the purpose of reporting, School titles have been retained but, for reasons of confidentiality, the MSc Programme titles have been coded in order to protect the confidentiality of participants.

### 7.1.2 Interview schedules

The interview schedule (Appendix 1) was drafted to explore Course Directors' practices of admission and selection in relation to English language proficiency and, particularly, the ways in which they employ English language test scores in selection decisions. The schedule was divided into three sections which covered the following topics:

- factual information about the MSc programme and the interviewee's experience of Course Directorship,
- selection and admission practices of non-native English speaking students on the programme and
- the relationship between international classes, pedagogy and academic progress of NNES students

The reason for pursuing the third topic was to illuminate Course Directors' views on the NNES academic and linguistic progress and to explore the practical means at a teacher's disposal to encourage learning in the international postgraduate classroom. An example of the latter is the extensive use of group projects in Cranfield MSc programmes where problems arising in classes with multilingual students may become evident. How teachers manage, moderate and overcome the problems arising might help to illuminate some issues in the literature about cultural difference and second language acquisition.

At the time of the study, the institution was engaging in a controversial debate about the publication of all Masters theses on the web and its topicality led to the inclusion of a question inviting participants' views. Finally, the choice of semi-structured interview method favoured an open and discursive approach towards questioning that allowed the interviewer to pursue and probe novel and emergent ideas, examples and concepts as they arose.

Course Director SOE1 (in Table 5) was chosen as the pilot interviewee for the interview schedule because of his openly expressed and considered views on the topic. It was anticipated, correctly, that his engagement would lead to a productive interview and provide a good test of the listed questions. Existing questions were satisfactory but additional questions were added about the programme's history, status within the UK HEI sector and the programme structure and assessment format, especially in relation to group project work. The latter was a fertile topic for discussing language and cultural issues.

The focus group guide (Appendix 2) was modelled on the interview schedule but with fewer questions and prompts to accommodate more open discussion and the emergence of relevant topics of interest to the participants. The single use of focus group method did not allow piloting of the focus group guide but, as the guide was a modified extension of the interview schedule which had been employed several times already, this was not seen as a major handicap. In practice, discussion and interaction in the focus group flowed well.

The duration of interviews and the focus group varied from 45-60 minutes and all were conducted by GL-J. Apart from two interviews and the focus group, they took place in her office between March and December 2008. The interviews with SAS2 and SOM3 took place, at their request, in their own offices and SAS2 declined to be audio-recorded. Notes were taken throughout this interview but all the remaining interviews were audio-recorded for later transcription by GL-J.



The data was analysed using MAXQDA software using deductive and inductive approaches. The former followed the line of questioning in the schedules; more inductive approaches led to the emergence of novel and unexpected themes within the data.

## 8 FINDINGS

This section is divided into three parts based on themes that link the research literature to the data. The first describes the different models of selection practices at Cranfield and the influence of the sampling criteria upon them as appears in the data. The second concerns Course Directors' knowledge and use of English test scores in the selection process and how these relate to other considerations in selection decisions. The final section deals with Course Directors' views of NNES students' academic and linguistic progress and how these relate to programme structure, delivery and pedagogy. Subsidiary topics and concerns recur across sections and are cross referenced as appropriate.

### 8.1 Models of selection practices

Three models of selection practice, related to School affiliation, emerge from the data. Although there are examples of Course Directors exerting autonomy in selection practices, the majority conform to the model prevailing within their School. As there was only a single contributor from Cranfield Health, these data are not included in this section.

The most distinctive of the Schools in developing a collective approach to the admission of NNES students is found in SOM for it has developed guidelines which extend the institutional regulations. As well as the IELTS score of 6.5, SOM demands a minimum sub-score on the writing component of IELTS of 6 and reserves the right to request an overall score of 550 and a writing analysis score of 4.0 on GMAT. The requirements apply regardless of whether the programme content favours text or number. SOM differs from the other Schools in admitting fewer EPP students (3 out of 127 in the programmes sampled) who are subject to the same requirements as other students so the linguists' involvement in selection is less than elsewhere. The findings in the pilot study provide supporting evidence for the operation of the School's language requirements. Nevertheless, six students with IELTS scores of 6 or less were admitted to Masters programmes, conditional on satisfactory completion of the Summer Programme. Where a Course Director wishes to offer a place to an applicant who scores below the required test score it is customary for him to seek the approval of the sub-Dean for academic performance or equivalent.

The requirements express the SOM's belief that it is preferable to filter out less linguistically able students at entry in preference to devoting extra resources to the support of borderline students. In justification, Course Directors stress the intensive and demanding character of Masters study and question whether spare time exists to devote to additional language study, a point that will be revisited in Section 8.2. Extract 1 is typical of the SOM viewpoint:

#### Extract 1

“I think we have to screen them (borderline cases) out if we think they are not going to be up to scratch in terms of language capability instead of admitting people with doubtful capabilities and trying to provide catch-up. It's a distraction of time and resources. We shouldn't be doing that. We are not here to teach them English. They should have reached a certain competence in English. If they're marginal, we can provide some help but I don't think it should be institutionalised at all because that's not what this programme is about.”

SOM3 Course Director

Course Directors are encouraged to interview applicants but there are resource implications so it is a matter of personal choice to what extent they do so. One SOM Course Director had chosen to interview as a matter of routine and was convinced that the standard of entrants had risen as a result.

### Extract 2

“When I’m talking about interviewing it’s for the majority of overseas candidates. So this is partly about providing verification of information that has been provided, it’s partly about assessing their suitability or their experience, their motivation and so forth but it’s also an opportunity to get an idea about English language ability as well ..... I have found that with overseas students you do need to delve into some of the detail because sometimes the information on the application form is not always 100% complete or accurate so it’s worth having a discussion about some of the details. We use interviews selectively but we are using them increasingly because we’re getting higher quality candidates overall.”

SOM1 Course Director

Selection practices in SOE share some features with SOM, principally in adherence to institutional requirements for test scores. However, there are no guidelines for selection at School level and three of the four Course Directors did not interview students routinely. There are larger numbers of EPP students admitted in the School, 56 out of around 175 students on the sampled programmes in this part of the research. SOE Course Directors display more autonomy in managing the selection process since two of the four interviewed described distinctive and contrasting practices which merit further description.

As already mentioned, SOE1 was particularly concerned with NNES selection because of students’ poor writing ability, an opinion that was shared by other teachers on his programme. Consequently, two years ago he had raised the entry requirement for the programme from IELTS 6.5 to 7 but he remained unconvinced of any gains in improvements in writing. The admissions records show that 18 NNES applicants presented appropriate English test scores and two borderline students with IELTS scores of 6.5 were admitted conditional on completion of the Summer Programme. SOE2, on the other hand, interviewed the majority of the NNES applicants for his course; the admission records show he interviewed nine of 19 NNES students for entry in 2007/8, the highest proportion of interviews of any programme in the pilot study. Six of the remaining students presented English test scores and others were interviewed by Cranfield staff. One of the latter students was advised to take the Summer Programme. The Course Director regarded the interview as an opportunity, first and foremost, to gauge the technical ability and, after that, the linguistic proficiency of the applicant, as displayed in Extract 3.

### Extract 3

“I ask them about their technical background and through that they have to use a fair bit of English describing what they’ve done what their aspirations are, what they want to do after they’ve Cranfield. Why they want to come to Cranfield? Things like that. We use it as a technical interview just to check that they are ok with the material and that they are happy with content of the course. That’s the thing I’m most concerned with. I mean, most of these students are taught English at their own university. They’re pretty good. They don’t really have any problems with the taught material. So spoken English is fine. The only, the areas where, that sometimes causes problems is probably the thesis project and writing an extensive report.”

SOE2 Course Director

The diversity evident in the selection processes amongst SOE Course Directors differs from the collective approach found in SOM and suggests that authority for selection is more likely to rest at Course Director than at School level.

The selection practices in SAS differ again from the other two Schools. The proportion of EPP students in the School is the highest amongst the sampled programmes, 66 out of around 125 students, and selection procedures reflect this. Whilst some Course Directors interview students, they leave the responsibility for assessing linguistic ability to their colleagues. There is a close working relationship between the SAS directors and linguist staff which extends beyond selection to ongoing language support for students on SAS programmes. How this works in practice is described in the following extract.

#### Extract 4

“It’s useful to differentiate between the EPP students and the non-EPP students. The EPP students have all come to us with a recommendation from the linguist staff and I’ve relied on them to provide the English language judgement because they’ve met the individuals. For non-EPP students, we look for the IELTS and go quite strongly on IELTS or equivalent tests. A small minority we will interview by telephone and usually if I do that, I will get someone like the linguist staff to help me. I will be concentrating on the more technical aspects and they will be looking at the English language.”

SAS1B Course Director

The association predates recent campus relocation before which some SAS programmes and the academic English staff occupied a separate campus several miles away from the main campus. SAS Course Directors who have always been located on the main campus did not evince such close working relationships with linguist staff.

#### 8.1.1 Demand for places

The hypothesis that demand for places might influence selection practices is directly related to the contemporary UK debate about degree standards and internationalisation. No evidence was found for the lowering of entry requirements in response to lessening demand for a specific course. However, there was support in courses where demand was high, as Directors were able to consider raising language entry requirements and consciously reflected upon using English language entry requirements as a filter to determine the allocation of places. This could take two concrete forms. It could, as in the case of SOE1, which has few competitors in its field elsewhere in the UK, simply raise the English test score required for entry. Alternatively, a Course Director might face fewer difficult decisions on candidates whose English entry proficiency was borderline. In Extract 5 a Course Director in SOE describes how demand and English language entry requirements are related on his course.

#### Extract 5

**Course Director:** “We were able to more or less guarantee that because with more than 400 applicants we could just pick the ones. You can really cherry pick.”

**Interviewer:** “Could you raise the entry level requirement?”

**Course Director:** “I don’t know that, I haven’t thought about it but I don’t know whether we would need to.”

SOE4 Course Director

Similar examples occurred in SOM and SOE but only in courses designated as high demand.

### 8.1.2 Large student groups with a shared language other than English

Course Directors considered that social and cultural integration of NNES students were important factors for academic and linguistic progress. Almost all Directors cited the challenge that a large minority group of monolingual students could present because of the attendant risk of delayed integration. The tendency to form social cliques diminishes opportunities to improve English language and discourages enculturation.

#### Extract 6

“About a third of our students are French and most of our European partner universities have a limit of four or five maximum on the MSc and, within an option, they’re not happy if there’s more than two because experience occasionally has shown that they form a clique, just speak in their native language and their English really does not improve at all.”

SAS1A Course Director

The resulting deleterious effects on learning and academic progress can affect the experiences of native English speakers as well as NNES students and are particularly evident in group projects. All save one Course Director paid considerable attention to the balanced allocation of students to groups for group work with regard to nationality and English language proficiency as well as academic ability. The courses selected to highlight this issue, SAS1 and SOE3, demonstrated how careful matching of selection decisions to course design and delivery can help to diminish any barriers arising from the presence of large minority groups of this nature within a cohort. In the SOE3 programme, for instance, every student gives a weekly oral presentation about their individual work to the group for the first six months of the course. In SAS1, the sharing of group projects between students on different Options permits contact with a wider and varying group of students. The same considerations extended to the selection of students on the smaller Options and some programmes with small classes.

### 8.1.3 Class size

There was no evidence that class size affected selection decisions independent of demand or due to the presence of a majority NNES student group with a shared language other than English.

## 8.2 Course Directors’ use of English test scores in the selection process

### 8.2.1 General selection criteria

In selecting suitable students, Course Directors are primarily concerned with the assessment of academic ability and potential to undertake study at Masters level (see Extracts 2 and 3). Directors repeatedly emphasised the difference between undergraduate and postgraduate study, referring to the ‘intensity’ of Masters study such that study time was jealously protected as described in Extract 7:

#### Extract 7

“It’s a case of the cost of time and how we can schedule it into what is already a nine to five programme, five days a week. There are lots of things that people say well can you introduce this or that on the course and I say we can do but something else has got to give. We can’t keep on adding more things in. If we add something in we have to take something out.”

SOE3 Course Director

Course Directors were looking for motivated, confident students capable of expressing themselves well and conveying arguments within their field (see Extract 3). These are important capabilities given

the importance attached to group projects and the UK higher educational values of independence and critical thinking. English language proficiency is fundamental to the development and demonstration of these abilities and interviews were conducted with these principles in mind. One experienced SAS Course Director who favoured interviewing considered it important to evaluate the applicant's ability to adapt to the UK HE system (Extract 8). His views were unique in this respect.

#### **Extract 8**

“You look at, obviously there's the academic ability of the student which you can get from their course results and then you have to make a decision on whether you think this person is capable of coming into a completely new system. It's partly language, partly confidence, partly motivation. So I don't think I would reject someone on language alone, if you see what I mean.”

**SAS2 Course Director**

Where students presented an English test score, Course Directors followed the institutional and School regulations and many regarded this as a simple matter of rule following (see Section 8.3). Although most were aware that English tests incorporated separate assessments of different language skills there was little detailed knowledge of variations between tests. Only Directors in SOM and two in SAS (SAS1B and SAS3) looked at sub-scores on a regular basis though others were more likely to do so in borderline cases.

#### **8.2.2 Borderline cases**

Course Directors treated borderline cases and applicants without test scores rather differently, examining application forms for validation of claims and for discrepancies between qualifications and displayed skills (see Extract 2). A perfectly scripted personal statement, for instance, might invite suspicions that it was not the applicant's own work. Evaluations were often subtle and based on prior experience of previous cases such as students from a particular institution or country. The preference for interviewing was a personal one although all who favoured the method claimed it was a highly useful verification of an applicant's speaking skills (see Extracts 2, 3 and 11). The assessment of writing skills is a more difficult challenge because of concerns about authenticity which it is almost impossible to guarantee. Most Course Directors were aware of the issue and some took steps to examine an applicant's spontaneous writing skills through emails and instant messaging.

#### **Extract 9**

“I might get an email saying I'm interested in the course, can you tell me a bit more about it? The one thing that does worry me when I get emails from some students is when it's all in text speak and that always worries me. You look at it and say hmm, ok. Is your English language so great? Why are you using text speak because you should be emailing me in proper sentences? Ok, when you speak to the student a bit more, you start to get the proper sentences coming through.”

**CH Course Director**

Two Course Directors, one each in SOE and SAS, had invited applicants to write a piece on a topic related to the course at interview.

### 8.2.3 Sceptics

Amongst the sampled group were two Directors, in SOE and SOM, who had considered selection in some depth and had come to view English tests with a degree of scepticism. Whilst not majority views, their experience and subsequent reflections justify further exploration.

The Course Director in SOE had more than 10 years of experience in the role and during this period had relied almost exclusively upon test scores to assess an applicant's English language ability, without recourse to interviews. Two years previously, worried by the poor writing standards of his students, he had raised the entrance requirement for the course to 7 but the desired consequences had not been realised. He now doubted his previous reliance upon English testing (Extract 10) and was searching for alternative, additional means of assessment such as another test, a writing task or an interview.

#### Extract 10

“Because, even now, having upped the barrier, even now we are getting students in the past couple of years who are still struggling on the course and I'm getting to the point of thinking, do you just keep on putting the barrier up? Or do we realise that maybe the measures, or the test, perhaps, is not testing in the way that mirrors what they are expected to be able to do when they come to Cranfield.”

SOE1 Course Director

The SOM Director had three years experience in the role but had set himself the task of developing the admissions and selection process during his tenure. He also followed the institutional and School guidelines on entry test scores but had seen many cases where the entry score had not correlated with either linguistic ability or academic progress. As a result, he had introduced interviews for most applicants and believed that the standard of candidates had improved as a result. He evoked a pragmatic view on testing, as displayed in Extract 11.

#### Extract 11

“These tests are not completely reliable. You may have two people with IELTS 6.5 and one is outstanding ..... but one of the concerns I have about examinations in general is that a test like this is how someone performs in certain circumstances on a specific day according to certain conditions and if they took the same test a month later under different circumstances they may well get a different result. We know that from people doing examinations. So I think it is not a completely reliable test so I like to have a bit of a second test by doing interviews. An interview is not a test of English language ability but it does give you some quite good indicators about their ability to communicate which is quite important.”

SOM1 Course Director

Nearly all Course Directors had similar cautionary tales of students with test scores that did not correlate well with subsequent linguistic and academic progress, including a student who failed an MSc despite having gained an undergraduate degree in the UK. These tales, which are grounded in experience, give primacy to experience rather than to the statistical evidence underlying formal testing.

### 8.2.4 Management of the selection process and learning how to select

Whilst some Course Directors were firmly of the belief that interviews gave good guidance for selection decisions, there were those with contrary views who were content to rely on the test score as

the sole measure. Sometimes these attitudes viewed selection decisions as a matter of administration, informed by issues of efficiency and resource use, such as the view below in Extract 12 gained from experience at another UK university.

#### **Extract 12**

“We had a set up in the department where there were a number of admissions tutors that were recruiting for each of the separate degrees and we were all spending quite a lot of time doing that task and we realised that a lot of the task wasn't academic. It was mainly administrative. The place where I came from we eventually appointed an administrator and her role was to carry out all the admissions tasks in the department except from where it was borderline. So she would say, no, these are clearly reject, you're going to see it factually, there are set criteria and all these are clearly accept and then there was a middle mound where she used to pick up the pile and go around the corridor and then go to talk to the academics and say what do you think from an academic point of view? And I would have said that probably most of my task so far on admissions, I'm only making an academic decision about 20% of the time.”

**Focus group participant**

Although this participant is describing events outside the institution under study, it is congruent with other Course Directors' views, albeit a minority. The attitudes contrast with the view in the research literature that greater knowledge about the tests amongst those concerned with admissions and selection will lead to sounder selection decisions. It suggests that there may be resistance on the part of some academics towards further training in this area. In connection with this, no consensus emerged from the interviewees about the necessity for any additional training in selection.

The privacy within which selection decisions are made calls into question how Course Directors learn about admissions and selection. When new to the role, almost all had been mentored by their predecessor and had learnt through a combination of experience and advice from colleagues (Extract 13).

#### **Extract 13**

“Some of it you just gain from chatting to your colleagues. What do you think of this person when you first start getting applications through and then you just start to do it on a feeling from experience. Well I've had someone through with a similar application before so we'll give them a go sort of thing.”

**CH Course Director**

Mentoring works well when the previous Course Director remains in the institution but if he or she leaves then the incoming Director may be remarkably unsupported. Unsurprisingly, the one Course Director whose predecessor left the institution felt there was a place for formal training for new Course Directors.

### **8.3 Course Directors' views of NNES students academic and linguistic process**

This section contributes to the second aim of the study by presenting the opinions of Course Directors about the linguistic and academic progress of their NNES students; particularly those with English test scores. As no attempts were made to refer to specific students currently studying at Cranfield in the interviews, the views expressed are couched in general terms but with some individual examples taken from experience.

### 8.3.1 Relationship between test scores and academic progress

As well as the two Course Directors quoted in section 8.2.3, opinions about the extent to which English proficiency test scores could be relied upon to indicate that an individual student was capable of successful study, uncompromised by linguistic problems, were contentious. Whilst some Course Directors appeared satisfied with institutional requirements and current selection procedures, others were unconvinced. The many stories of dissonance between the English test scores and subsequent academic progress chime with the inconclusive results of the research literature seeking associations between English language proficiency and academic outcomes and reflects Course Directors' beliefs that academic success depends on a complex array of factors and circumstances.

Several Course Directors stressed the importance of immersion into the UK educational culture as an essential prerequisite for linguistic and academic advancement. The diversity of Cranfield's student body is a help here because English is the common language in an international student body. Despite this, some Masters programmes may have significant numbers of students who share a first language other than English. Directors expressed concern about NNES students who live with other compatriots whilst in the UK, since it can compromise the degree to which NNES students use and practise English and hence limit the potential for linguistic development. However, it is considered inappropriate to intervene in a student's extracurricular affairs and so it remains a difficult area over which neither the institution nor Course Directors can exert any control. Immersion in the English language will appear later in connection with the Summer Programme students (Section 9.2.3)

### 8.3.2 Speaking skills

Exploring Course Directors' views of the development of NNES students' language skills demonstrated that the main issue lies with writing skills. Although skills in speaking may be initially limited, the majority of students overcome these difficulties within the first two to three months such that they are able to converse satisfactorily for course purposes. Only rarely were there stories of students whose pronunciation was so poor that they were unable to complete the course. In Extract 14, a NNES lecturer comments on students' spoken English.

#### Extract 14

“Spoken English generally, in the beginning perhaps, may be a little more difficult but students adapt very quickly and therefore I think spoken English is not the biggest problem actually. I never saw any big problems with spoken English.”

**NNES Lecturer, SAS, focus group**

Speaking as a student in class, however, is overlaid with cultural references and this can sometimes cause problems for a lecturer trying to assess an individual student's linguistic progress. In Extract 15, a Course Director muses on the interaction between culture and language for Chinese students on his course.

#### Extract 15

**Interviewer:** “How long does it take before the Chinese students' spoken English is reasonably good?”

**Course Director:** “Well, they can be very quiet but equally you can get some people who stand out, who are prepared to put themselves forward. I think it's more of a cultural thing but maybe it's not, maybe it's that they're not comfortable in English. If they were more comfortable in their English, they'd stand out more. I'm not sure which one it is.”



**Interviewer:** “Do they make any further progress or remain like that throughout the course?”

**Course Director:** “No, I think they come round. I think they come round in the second semester. That’s why I think it’s not their English, it’s more a cultural aspect but that could be argued to be not the case. You could argue that three months here improves their English. I’m not sure at this point.”

SOM2 Course Director

In general, though there was agreement that oral skills are not a major problem which may be a consequence of the Cranfield context, not least its geographical isolation which renders students more reliant upon themselves, the campus and the local environs for extracurricular activities.

### 8.3.3 Writing skills

One topic on which there was broad agreement amongst Course Directors was concern about standards of written English (Extracts 3 16).

#### Extract 16

“Out of 40 students, there’s probably a handful or so where it could be a problem if it’s not managed by the supervisor towards the time when they are writing up the report. If ever that slips then occasionally we do get a poor quality written report. So it’s a matter of identifying the students that are weak. Normally, it ties in with, it’s strange but, it ties in with their academic performance. So by the time they start the thesis we know which ones are borderline and not particularly strong and it’s those that we really have to focus on to make sure that they are supervised fairly closely in terms of what they put in the report, the structure and what they write. It’s a relatively small numbers but yes, it’s there.”

SOE2 Course Director

The thesis was a recurring problem primarily because, as a publicly available document, it is a manifestation of academic standards that is open to external scrutiny. Furthermore, the debate about whether to publish all Masters theses on the web had the effect of focusing Course Directors’ concerns about standards. Another feature of poor thesis writing was the impact upon supervisors and their workloads. Thesis supervision is a relatively unseen aspect of academic work because the process is largely undocumented and the incentives to produce a quality document are strong. Consequently, there is little evidence to demonstrate differential workloads related to students’ writing skills. Additionally, the timing of the thesis allows little scope for substantive improvements in writing proficiency.

Whilst there was consensus about student writing, there was little agreement about the remedies; particularly as academic staff may not regard it is part of their remit to correct English language (See Extract 1). Nevertheless, several Course Directors reported pedagogic strategies designed to deal with students most at risk. Firstly, early writing assignments, sometimes formative, had been introduced to identify students in greatest need of ongoing English tuition. Secondly, many courses maximised opportunities for students to practise writing. In one example, the structure of a group project report was deliberately modelled on the thesis to facilitate learning. Thirdly, a long standing Course Director remarked that much more learning support was now provided for thesis writing than had ever been offered previously.

A number of confounding issues emerged in discussions about writing skills. The first, and unanticipated, finding was that poor writing was not confined to NNES students alone; UK students were reported as less able than formerly to produce good written work. A second point related to

the assessment of English language in written scripts. Each school has its own policy on assessment but in general, marks are not allocated specifically for language per se although it may be subsumed under criteria such as ‘presentation’ or ‘structure’ for which marks are allocated in marking schemes. It was therefore not surprising that Course Directors varied in the degree to which they corrected grammatical and spelling errors on course assignments and theses. SAS5 and SOM2 proof read their students’ work, making corrections, because they believed it to be good feedback practice and hoped it might encourage students to improve their written English. Extract 17 exemplifies these points.

### Extract 17

“I will mark course work, and I think my colleagues do much the same thing; they will highlight where there are English problems within a report and encourage students, in future reports, to try and improve their English. The only time where it would have a bearing on the mark is if it becomes unreadable and we are not able to understand what they are trying to convey in which case it will cause a lowering of the mark. But we don’t have an element of the mark that is for English. We will, on occasion, have elements of the mark for presentation, but that isn’t necessarily looking at the language that is used. It’s more to do with the structuring of the report or the presentation of the map work that forms an important part of some assignments.”

SAS5 Course Director

Many Course Directors proposed that there should be more willing and capable ‘proof readers’ on campus to reduce the burden on supervisors. These suggestions came from Course Directors in all Schools but SOE Course Directors were particularly likely to voice the suggestion. It is worth noting that the nature of ‘proof reading’ was not explored in any depth during the interviews. Whether the implications are to attend to surface features of grammar or to deeper problems of editing how ideas and arguments are expressed cannot therefore be determined. The data indicate that the thesis is the academic product most likely to reflect the tension between academic standards and the rise in NNES students but that solutions are neither easy nor immediate.

## 9 SUMMER PROGRAMME STUDENTS’ PROGRESS

This section describes how the second aim of the study, seeking a relationship between NNES students entry test scores and subsequent academic progress, was attained. In Section 5 we explained how the research enquiry was diverted to a study of the Summer Programme (Summer Programme) students in consequence of the limited number of students entering with English test scores. Recognised as an alternative strategy, a study of Summer Programme students would be justified since they represented a discrete group of students whose language abilities had been identified as borderline and whose progress might be tracked.

In considering the concept of progress as applied to Summer Programme students, it was decided to construct profiles for each student that might log the development of their English language proficiency during the year. Initial intentions were to compare these with profiles of academic progress including marks from assignments and projects through the year but this was not possible because it contravened institutional regulations concerning the publication of individual students’ marks, so alternatives had to be found.

## **9.1 Data collection methods**

### **9.1.1 Documentary sources**

For this part of the study data sources included documents and questionnaires. The 2007 Summer Programme students underwent a number of linguistic assessments in the first few months at university which were accessible through the academic English staff. These included scores for students who had taken English tests and, following completion of the Summer Programme, reports for each student including grades for separate language skills, comments and recommendations for future study. In addition, all Summer Programme students participated in a pre-test IELTS exercise in the penultimate week of the Summer Programme towards the end of September 2007 which covered academic writing, listening and reading skills but not speaking. A pre-test is primarily designed to test new items for inclusion in future IELTS papers so marks are not strictly comparative with a true IELTS test. In fact, listening and reading papers are marked as raw scores, not band scores, although the maximum score for the combined papers is provided. Despite distinctions, advice to candidates states that the Pre-test IELTS results will give an indication of ability so the results supplied an approximate measure for those students without entry scores and for whom no baseline comparison measure existed. It is possible too, that relative differences, or similarities, between individual student's pre-testing scores are robust, albeit on a single examination sitting.

### **9.1.2 Examination scripts**

Apart from a student's need for ongoing language tuition there were no specific measures of student's language abilities once the student had commenced the MSc course. The earlier diversity study (Lloyd-Jones 2007) and the Course Directors' interviews had established that written skills were those giving rise to concern and, on this basis, a decision was made to focus upon students' writing abilities. Course assessments were preferred as data sources because of their contextually appropriate nature. The alternative of additional testing was considered but rejected on the grounds that it would be unacceptable to the students concerned. Of the three written assessment formats, course assignments, examinations and theses, the latter two were chosen for investigation. Assignments were rejected because of questions of authenticity and feasibility. Organising, and retrieving the large number of scripts necessary was deemed beyond the scope of the present study and sampling on a lesser scale appeared daunting. On the other hand, exam scripts were clearly authentic and the practice of two sittings before Christmas and Easter provided standardisation in time across the different courses.

Access to examination scripts was approved by Registry provided the students and Course Directors consented. Course Directors of Summer Programme students were emailed for their consent and all were willing. Summer Programme students were contacted individually by email with requests for consent to access their examination scripts. The email included an explanation for the request and a brief rationale for the study. Two students, both in SOE, declined, leaving 22 students participating in the exam script study. Course Directors or Course Administrators provided the exam scripts which were copied and the originals returned to their departments. The scripts of three students and one paper each for two more students were unavailable as a result of external examining requirements. By agreement, the copies will be destroyed following the completion of the study.

The scripts were examined with two questions in mind. The first was to gain an impression of the amount of textual content in the scripts which might corroborate the interview data for certain programmes (see Extract 3). The second was to identify examiner's comments about language, either critical or positive. Two reviewers separately read the scripts, a lecturer from the Academic English

staff and a member of the research team. There was good agreement between the two reviewers; the only shared doubts concerned how to interpret a few comments about style or structure.

### **9.1.3 Questionnaire for thesis supervisors**

As access and evaluation to more than 20 theses from varied domains would clearly be impossible within the time scale and prevailing circumstances, a different approach was adopted towards the quality of students' thesis writing. The effect on supervisory workload identified in the interview study indicated that Summer Programme students' thesis supervisors could shed light on students' writing proficiency. A brief electronic questionnaire was constructed inviting the appropriate supervisor to comment upon whether the English language proficiency of his or her supervisee had affected the thesis mark and/or the supervisory workload (Appendix 3). Where respondents stated that the student's proficiency had had an adverse effect, they were asked to provide more details. An open question allowed respondents to make any further comments they felt to be relevant. The questionnaire was piloted within the CPLT with a previous Course Director and minor revisions made to the wording.

The names of supervisors of the Summer Programme students were obtained from the relevant Course Directors. The questionnaire was delivered electronically early in November 2008 after final thesis submission. Due to an oversight, the supervisors of the two Summer Programme students who declined to participate in the exam script study were not included so questionnaires were sent to 22 supervisors. The accompanying email named the Summer Programme student concerned and explained that the questionnaire referred to this particular supervision. Therefore each supervisor received an individual invitation although the questionnaire was identical. Non-responders received reminders one and two weeks later.

### **9.1.4 Academic progress**

The time of completion of the study, January 2009, coincided with the period during which students receive notification about the outcomes of their degree awards. The research team was granted access to these records for the Summer Programme students. Significantly for the present study and the focus on thesis writing, the records provided details about the outcomes of thesis marking. Cranfield University does not award distinctions for Masters programmes, so there is no differentiation in grading amongst those students who have passed. However, there are three possible categories of outcome. A straight pass, a pass subject to minor corrections and a revise and represent instruction where the student is required to make major changes to the thesis within a given time frame, around 3 months. Minor corrections are not a major impediment to the gaining of a degree but the necessity to revise and represent leaves a question mark over the outcome until the revised thesis is marked a second time. With the questionnaire data, the records provide further information about Summer Programme students' progress with tasks that most challenge their language skills.

## **9.2 Findings**

### **9.2.1 Summer Programme students – language assessment at entry**

29 students took the Summer Programme in 2007, 26 of whom were destined for taught course Masters programmes. One withdrew prior to the start of the Masters and another student transferred to a research programme early on in the year which left 24 students who completed both Summer Programme and Masters programmes and who contributed to the present research. Numbers were small and considered inappropriate for statistical analysis.

<b>Summer Programme student</b>	<b>School</b>	<b>Language assessment</b>	<b>Entry test score</b>	<b>Test sub-scores</b>
SAS1	SAS	Interview		
SAS2	SAS	Interview		
SAS3	SAS	Interview		
SAS4	SAS	Interview		
SAS5	SAS	Interview		
SAS6	SAS	Interview		
SAS7	SAS	Interview		
SAS8	SAS	Interview		
SAS9	SAS	Interview		
SAS10	SAS	Interview		
SAS11	SAS	Interview		
SOE1	SOE	English test	TOEFL CBT 220	L 20 St.W2 R 24 E 4.
SOE2	SOE	Interview		
SOE3	SOE	English test	TOEFL CBT 243	L 25 St .W 22 R 26 E 4
SOE4	SOE	English test	IELTS 6.5	L 7 R6.5 W 6 S 6
SOE5	SOE	Unknown		
SOE6	SOE	English test	TOEFL IBT 92	R 27 L 26 S 18 W 21
SOE7	SOE	English test	IELTS 6	<b>L 5.5 R 6.5</b> <b>W 5 S 6</b>
SOM1	SOM	English test	IELTS 6	L 6: R 6 <b>W 5 : S 5</b>
SOM2	SOM	English test	IELTS 6 TOEFL IBT 64	IELTS N/A TOEFL R 16: L 18 S 15: W15
SOM3	SOM	English test	IELTS 5.5	L 6 : <b>R 5</b> W 6 : <b>S 5</b>
SOM4*	SOM	English test	IELTS 6	L 6 : R 6.5 W 6 : S 6
SOM5	SOM	English test + Interview	IELTS 6	L 6 : R 6.5 <b>W 5 : S 6</b>
SOM6	SOM	English test	IELTS 5.5	L 5: R 6.5 <b>W 5 : S 5</b>

\*SOM4 IELTS score 3 months earlier 5.5

Bold type low sub scores

IBT Internet based test

CBT Computer based test

L Listening

St W Structured Writing

R Reading

E Essay

S Speaking

W Writing

**Table 6. Summer Programme students' language assessments at entry**

Table 6 shows the list of the 24 Summer Programme students who took part in the research, the method of language assessment, the entry test score and sub-scores, where available. The data confirm the earlier findings from the pilot and interview studies showing differences between Schools in admission practices for borderline NNES students. SAS students are solely assessed by interview, SOM by English test score and SOE use both methods of language assessment but are more likely to use formal English tests. Problems arose locating the records for Student SOE5 who transferred to another MSc programme within the same School.

Students SOE3, SOE4 and SOE6 comply with the university requirement for English entry requirements but SOE3 and SOE6 have not reached the standard for their MSc programme. SOE4 presented two IELTS scores almost a year apart; the initial score was 6, the later one, 6.5.

The score of student SOE1 lies between IELTS 6 and 6.5 but the remaining scores are below the minimum entry requirement. All of the latter group, save one student, have at least one sub-score of 5.5 or lower, which is somewhere between the modest and competent user in the IELTS band score descriptors. Student SOM4 presented the results of two IELTS tests, three months apart which showed an improvement of one band score in the later result which is reported in the table.

The majority of Summer Programme students attended the Summer Programme for four weeks but four students, SAS2, SAS10, SOM2 and SOM 6 attended for eight weeks.

### 9.2.2 Summer Programme students – Pre-test IELTS

All Summer Programme students sat the Pre-test IELTS in September a week before the conclusion of the Summer Programme and just before the start of the Masters course. For reasons already explained, the pre-test exam differs from IELTS so direct comparisons of the results with IELTS scores are not entirely reliable. There are other differences, notably in marking. Writing is the only skill given a band score similar to IELTS. Listening and Reading are reported as raw scores; the total marks available for Listening are 49 and for Reading, 45. The Pre-test IELTS results are shown in Table 7 with the students' IELTS entry scores, where available.

The overall results for the Summer Programme students are low when compared to pre-entry test scores and to the university entry requirements. Band scores for Writing average between 5 and 5.5. Only one student of the group with overall pre-entry IELTS scores of 6 or higher (or equivalent) scores 6 on Writing, the remainder range from 5.5 to 4. At the extremes, two students in SAS (SAS1 and SAS4) and one in SOE (SOE6) score 6 and at the other end of the range, two students in SAS (SAS2 and SAS11) and one in SOM (SOM3) score 4 or 4.5. The raw scores for Listening and Reading show approximately similar patterns of score distribution. The average score for the group on Listening is 27 (55%) and in Reading 26 (58%). The three students scoring 6 on Writing also score highest on Listening and Reading, when the two scores are combined. Of the three lowest scoring students on Writing, two also score low on Listening and Reading (36 and 38) but not the third (47). Three students with Writing scores of 5.5 have scores below 40 for combined Listening and Reading. Although the circumstances of the Pre-test urge caution in interpreting the scores, the findings indicate that these students' English language skills are borderline.

Summer Programme student	Language assessment	Entry test score	Pre-test IELTS scores (09.07)			
			Writing	Listening	Reading	Total
SAS1	Interview		6	33	42	75
SAS2	Interview		4	14	22	36
SAS3	Interview		5.5	8	24	32
SAS4	Interview		6	37	35	72
SAS5	Interview		5	19	22	41
SAS6	Interview		5.5	26	24	50
SAS7	Interview		5	14	15	29
SAS8	Interview		5.5	34	34	68
SAS9	Interview		5	21	19	40
SAS10	Interview		5	24	29	53
SAS11	Interview		4.5	20	18	38
SOE1	English test	TOEFL CBT 220	5.5	32	32	64
SOE2	Interview		5.5	29	25	54
SOE3	English test	TOEFL CBT 243	5.5	36	32	68
SOE4	English test	IELTS 6.5	5	34	23	57
SOE5	Unknown		5.5	25	35	60
SOE6	English test	TOEFL IBT 92	6	37	37	74
SOE7	English test	IELTS 6	5.5	24	23	47
SOM1	English test	IELTS 6	5.5	34	33	67
SOM2	English test	IELTS 6 TOEFL IBT 64	5	36	27	63
SOM3	English test	IELTS 5.5	4	29	18	47
SOM4	English test	IELTS 6	5.5	21	18	39
SOM5	English test	IELTS 6	5.5	27	25	52
SOM6	English test + Interview	IELTS 5.5	5.5	24	21	45

**Table 7. Summer Programme students' Pre-test IELTS and entry score**

### 9.2.3 Summer Programme students – Reports

On completion of the Summer Programme, academic English staff compile individual reports on Summer Programme students. Students are graded on an alphabetic system from A, A-, B+, B, C+, C on four course components: academic reading, academic speaking, academic writing and listening comprehension and note taking. The report includes specific feedback on the student's progress for

each component, the student’s attendance, advice for the direction of future work and the need for formal ongoing support in English language. Reports are sent to the relevant Course Director for guidance and information. Table 8 shows the students’ grades from the Summer Programme reports in comparison with entry scores, Pre-test IELTS scores and overall entry test scores, where available. It also includes specific recommendations which were given to students to attend weekly tutorials on academic writing. All students are advised to be aware of the dangers of spending too much time with other students sharing the same language.

Summer Programme Student	Entry Test Score	Pre-test IELTS scores				Summer Programme reports				
		W	L	R	Total	W	LN	R	S	Rec.
SAS1		6	33	42	75	A	A-	A-	B+	No
SAS2		4	14	22	36	C	C+	C+	B-	Yes
SAS3		5.5	8	24	32	B-	C+	B-	C+	Yes
SAS4		6	37	35	72	A-	B+	B+	B+	No
SAS5		5	19	22	41	C+	C	C	C	Yes
SAS6		5.5	26	24	50	B-	C	B-	C	No
SAS7		5	14	15	29	B-	C	B-	C	Yes
SAS8		5.5	34	34	68	B+	B+	B	B+	No
SAS9		5	21	19	40	B	B	B+	A-	No
SAS10		5	24	29	53	A-	B+	A-	B	No
SAS11		4.5	20	18	38	C	B-/C+	B-	B	Yes
SOE1	TOEFL 220	5.5	32	32	64	B+	B+	A-	A-	No
SOE2		5.5	29	25	54	B	B-	B	C	No
SOE3	TOEFL 243	5.5	36	32	68	B+/A-	A-	A-/B+	B+	No
SOE4	IELTS 6.5	5	34	23	57	B-	B	B	B	No
SOE5		5.5	25	35	60	B-	B	B	B-	No
SOE6	TOEFL 92	6	37	37	74	B	B+	A-	B+	No
SOE7	IELTS 6	5.5	24	23	47	B+	C	B	B+	No
SOM1	IELTS 6	5.5	34	33	67	B	B	B+	B	No
SOM2	IELTS 6	5	36	27	63	B+	A-	A-	B+	No
SOM3	IELTS 5.5	4	29	18	47	C-/C	C	C	C-	Yes
SOM4	IELTS 6	5.5	21	18	39	A-	B+	A-	B	No
SOM5	IELTS 6	5.5	27	25	52	B	B+	B+	B/B+	No
SOM6	IELTS 5.5	5.5	24	21	45	B	B	B	B	No

Summer Programme St Summer Programme students

ETS Entry test scores

\*Ongoing support required

Rec. Recommendation for continuing language support

W Writing L Listening R Reading LN Listening and note taking S Speaking

**Table 8. Summer Programme students’ entry and Pre-test IELTS scores and Report grades**



Generally, there is correspondence between the Pre-test IELTS scores and the Summer Programme Reports. However, the Pre-test IELTS report scores for Students SOM4, SAS 9 and SAS10 are less favourable than the Report grades, particularly so for SOM4. All students who score low either on the Writing score or on the combined Listening and Reading score were recommended to attend the weekly tutorials in English language. Student SAS2 was referred to the Disability Learning Support Officer for investigation because of writing difficulties. Students SAS4 and SAS7 who were entering courses where there were many students with whom they shared a first language were additionally advised to maximise opportunities for English language practice.

#### **9.2.4 Exam scripts of Summer Programme students**

The review of exam scripts demonstrated that nearly all examinations required students to write extended passages of text apart from one course in SOE where textual content was low. However, this was far from the case with the other SOE exam scripts where passages of text were common. Counter-intuitively, scripts from a quantitatively orientated course in SOM contained much less text than the SOE scripts.

Comments about the examinee's written language were not uncommon. The examination scripts of three students, SOM1, SOM2 and SAS3 all contained critical comments which specified language, such as 'generally superficial and with inaccurate use of English'; 'this comment is worrying but could just reflect a language barrier'; 'weak language' and 'not an answer, needs explanation, suspect language problems?'; 'appalling style, clearly has problems with written English but basics are here' (sic). The comments on the scripts of two other students (SOM6 and SAS9) referred to 'poor style', 'dumping not answering' and requests for more detail and explanations besides bulleted lists. Two students answered too many questions and one student received praise for presentation (SOM5). Statements that might be language related were found on other students' scripts but have not been reported because they cannot be securely attributed to language difficulties alone. Written comments were most frequently found on the scripts of SOM students.

It was not anticipated that exam scripts would provide a rich source of evidence of language difficulties because there is no requirement for examiners to mark English language specifically. It is possible too that examiners may take a more relaxed approach to the quality of English language in stressful exam conditions. The interview data had shown variation in the degree to which Course Directors paid attention to English language in course assignments so similar variation amongst examination markers was also to be expected. As described above, comments were strictly evaluated to exclude any ambiguous remarks. Consequently, those comments presented in the report are likely to underestimate the contribution of language issues in exam scripts and so can be regarded as robust evidence for the purposes of the research. The results are included in Table 9.

#### **9.2.5 Summer Programme students – Workload of thesis supervisors**

15 out of 22 supervisors responded to the questionnaire, a 68% response rate. Supervisors' experience of supervising Masters theses ranged from two to 20 years and five had been supervising for 15 years or more. The number of Masters theses each supervisor was responsible for in the academic session varied from one to sixteen but this range obscured the fact that 13 were supervising between four and seven theses.

The question relating the students' English language proficiency to the thesis mark was answered by only half the respondents and, in retrospect, lacked clarity. 20% of all supervisors considered the student's language proficiency had beneficially contributed to the thesis mark and 33.3% not at all.

Whilst none completed the ‘adversely’ box, seven supervisors supplied open comments which were critical either of non-native English speakers writing skills in general (2) or of their supervisee in particular (5).

80% of supervisors stated that their students’ English language proficiency had adversely affected their supervisory workload; the remaining 20% claimed it had had no effect. However, the supervisor of SOM2 reported that although his workload had been unaffected there had been a necessity to read several drafts and to focus on language rather than content. Twelve respondents wrote descriptive comments about the supervisory experience, detailing how their work had been affected. Most referred to the extra time involved in reading multiple drafts and the additional feedback that was necessary with each successive draft. The comments emphasised that it is not simply a matter of reading multiple drafts but a necessary qualitative change in the task of reading which involves interpreting poor English. As expressed in the words of one supervisor: ‘Several iterations of the same sections were sometimes necessary to establish clarity of thought and expression’. Two supervisors mentioned specific difficulties with data and their presentation. One supervisor had spent considerable time on early drafts and advised his supervisee to use a proof reader for help in preparing the final version. The replies to this question are shown for the students concerned in Table 9.

<b>Summer Programme student</b>	<b>Entry test score</b>	<b>Exam script</b>	<b>Effect upon supervisor workload</b>	<b>Thesis outcome</b>
SAS1		Unavailable	None	Revise and represent
SAS2		None Partial access	Adverse	PGDiploma subject to conditions
SAS3		Critical of language	Adverse	Revise and represent
SAS4		None Partial access	Adverse	Revise and represent
SAS5		None Partial access	Adverse	Pass
SAS6		Unavailable	Unknown	Revise and represent
SAS7		None	None	Pass
SAS8		None	Unknown	Revise and represent
SAS9		Critical of style	Adverse	Revise and represent
SAS10		None	Adverse	Pass
SAS11		None	Adverse	Pass
SOE1	TOEFL CBT 220	Unavailable	Unknown	Pass
SOE2		None	Unknown	Pass
SOE3	TOEFL CBT 243	None	None	Revise and represent
SOE4	IELTS 6.5	Unavailable	Unknown	Minor corrections

<b>Summer Programme student</b>	<b>Entry test score</b>	<b>Exam script</b>	<b>Effect upon supervisor workload</b>	<b>Thesis outcome</b>
SOE5		None	Adverse	Pass
SOE6	TOEFL IBT 92	None	Adverse	Revise and represent
SOE7	IELTS 6	Unavailable	Unknown	Revise and represent
SOM1	IELTS 6	Critical of language	Unknown	Revise and represent
SOM2	IELTS 6 TOEFL IBT 64	Critical of language	Equivocal	Revise and represent
SOM3	IELTS 5.5	None	Unknown	Revise and represent
SOM4	IELTS 6	Critical of language	Adverse	Revise and represent
SOM5	IELTS 6	Positive on style	Adverse	Minor corrections
SOM6	IELTS 5.5	Critical of style	Unknown	Revise and represent

**Table 9. Summer Programme students' entry scores compared with, exam script comments, supervisor's reported workloads and thesis outcomes**

Answers (12) to the final open question inviting additional relevant comments have been listed in Appendix 4. Whilst many reinforced the comments above by referring specifically to the individual supervisory relationship, others raised wider issues about the effects of internationalisation upon higher education, the role of the supervisor and the compromises that are sometimes necessary to produce a thesis that satisfies disciplinary and language standards and requirements. These findings establish that the English language proficiency of half of the Summer Programme students impacted negatively upon the workload of their supervisors.

### **9.2.6 Summer Programme students – thesis outcomes**

Table 9 shows the outcomes of the relevant Examination Boards for the Summer Programme students in terms of the degree award and the status of the presented thesis as at January 2009. Including theses requiring minor corrections, seven of the Summer Programme students have been awarded their Masters qualification and 14 students are required to revise and represent their theses for further assessment before receiving their Masters degree. One student, SAS2 was awarded a PGDiploma instead of a Masters degree, subject to passing a further examination. These figures confirm the findings in the pilot study that Summer Programme students are more at risk of failing academically.

Tables 10, 11 and 12 collate the results of the various studies employed to assess the progress of the Summer Programme students for individual Schools.

Four out of seven Summer Programme SOE students have gained their degrees without having to revise and resubmit their theses but this does not rule out the possibility that they were experiencing language problems as the example of SOE5 demonstrates (Table 10). The thesis of this student was passed but the supervisor reported an increased workload due to the student's language problems.

SP St	ETS	PI W	PI L	PI R	PI T	SP Rep W	SP Rep LN	SP Rep R	SP Rep S	Exam scr	S W	Thesis
SOE1	TOEFL CBT 220	5.5	32	32	64	B+	B+	A-	A-			Pass
SOE2		5.5	29	25	54	B	B-	B	C	None		Pass
SOE3	TOEFL CBT 243	5.5	36	32	68	B+/A-	A-	A-/B+	B+	None	None	Revise and represent
SOE4	IELTS 6.5	5	34	23	57	B-	B	B	B			Minor corrections
SOE5		5.5	25	35	60	B-	B	B	B-	None	Adverse	Pass
SOE6	TOEFL IBT 92	6	37	37	74	B	B+	A-	B+	None	Adverse	Revise and represent
SOE7	IELTS 6	5.5	24	23	47	B+	C	B	B+			Revise and represent

SP St Summer Programme student  
 ETS Entry test score  
 PI Pre-test IELTS W Writing L Listening R Reading T Total  
 SP Rep Summer Programme Reports LN Listening and note taking S Speaking  
 Exam Scr Exam Scripts None no comments about language  
 S W Supervisory workload

**Table 10. SOE Summer Programme students’ entry scores and measures of progress**

There are no responses from the supervisors of the other three students so language problems cannot be ruled out despite the successful academic outcomes. Of the students who are required to edit their theses for further submission, one is reported as having no English language problems so language deficiencies are not responsible for the student’s lack of success. However, the supervisory comments for SOE5 indicate that English language proficiency was contributing to the academic outcome for this student. Overall, the group has a high revise and represent rate (43%), 2 students with identified language problems, one student without and question marks over the remainder despite their academic success.

The SOM results have the highest proportion of revise and represent students of all the Schools (83%); only one student of six has been awarded a degree without going through major thesis revision. Three students, including the one who passed outright, are reported as having language difficulties during thesis supervision and critical comments about language or style were found in four students’ exam scripts. The results suggest that five students have experienced difficulties with written language. The supervisor of the fifth student, SOM3 has not responded to the questionnaire so no conclusions can be reached for this student, despite the low scores and report grades.

S sP St	E T S	PI W	PI L	PI R	PI T	S P Rep W	S P Rep LN	S P Rep R	S P Rep S	Exam scr	S W	Thesis
SOM1	IELTS 6	5.5	34	33	67	B	B	B+	B	Critical of language		Revise and represent
SOM2	IELTS 6 TOEFL IBT 64	5	36	27	63	B+	A-	A-	B+	Critical of language	Equivocal	Revise and represent
SOM3	IELTS 5.5	4	29	18	47	C-/C	C	C	C-	None		Revise and represent
SOM4	IELTS 6	5.5	21	18	39	A-	B+	A-	B	Critical of language	Adversely	Revise and represent
SOM5	IELTS 6	5.5	27	25	52	B	B+	B+	B/B+	Positive on style	Adversely	Minor corrections
SOM6	IELTS 5.5	5.5	24	21	45	B	B	B	B	Critical of style		Revise and represent

SP St Summer Programme student

E T S Entry test score

PI Pre-test IELTS W Writing

L Listening

R Reading

T Total

SP Rep Summer Programme Reports

LN Listening and note taking

S Speaking

Exam scr Exam Scripts None no comments about language

S W Supervisory workload

**Table 11. SOM Summer Programme students' entry scores and measures of progress**

Student SAS2 was awarded a PG Diploma, not a Masters degree, and on condition the student retook one exam. Six of the remaining students (54%) are required to revise and represent and of whom four have other evidence of language problems from their supervisors. However, the supervisor of SAS1 has reported that there are no language difficulties which is congruent with the student's Pre-test IELTS scores and Summer Programme report grades which suggests that academic reasons underlie the need for major thesis revision. Perhaps of more concern are the three students (SAS5, 10 and 11) who have passed the thesis but who are reported as having writing problems. This leaves seven students where language problems have been identified as contributing adversely to academic outcomes, two whose writing abilities are reported as satisfactory and two students who are revising their theses but whose supervisors have not responded to the questionnaire.

S sP St	E T S	PI W	PI L	PI R	PI T	S P Rep W	S P Rep LN	S P Rep R	S P Rep S	Exam scr	S W	Thesis
SOM1	IELTS 6	5.5	34	33	67	B	B	B+	B	Critical of language		Revise and represent
SOM2	IELTS 6 TOEFL IBT 64	5	36	27	63	B+	A-	A-	B+	Critical of language	Equivocal	Revise and represent
SOM3	IELTS 5.5	4	29	18	47	C-/C	C	C	C-	None		Revise and represent
SOM4	IELTS 6	5.5	21	18	39	A-	B+	A-	B	Critical of language	Adversely	Revise and represent
SOM5	IELTS 6	5.5	27	25	52	B	B+	B+	B/B+	Positive on style	Adversely	Minor corrections
SOM6	IELTS 5.5	5.5	24	21	45	B	B	B	B	Critical of style		Revise and represent

S P St Summer Programme student

E T S Entry test score

PI Pre-test IELTS W Writing

L Listening R Reading

T Total

S P Rep Summer Programme Reports

LN Listening and note taking

S Speaking

Exam scr Exam Scripts None no comments about language

S W Supervisory workload

**Table 12. SAS Summer Programme students' entry scores and measures of progress**

Overall, the results display some of the difficulties facing researchers who try to ring fence the contribution of English language skills to academic progress. Whilst the supervisors' perspective has provided helpful triangulation, the partial nature of the data and the difficulty in correlating different assessment measures hinder the aim of reaching secure conclusions about the contribution of linguistic competence to academic ability and outcomes for individual students. Nevertheless, it seems reasonable to conclude that these Summer Programme students, as a group, remain borderline in terms of English language proficiency throughout the year of study.

## 10 DISCUSSION

The discussion is framed by the three aims of the research study with references to the literature included where relevant.

### 10.1 Course Directors' admissions practices and experiences

The most striking finding of the study was the variety of selection practices within the institution. The variations were linked to School affiliation and, possibly, to former organisational arrangements when there were two campuses rather than one. The selection procedures varied in relation to the use of English test scores, ranging from SOM at one end which placed greatest reliance upon English testing to SAS, at the other, which largely delegated assessment of an applicant's English language

proficiency to linguist staff, often conducted by interview. The latter form of assessment operates within the EPP scheme which links HEIs around Europe prepared to exchange students under the EU Erasmus programme. The third School, SOE shared features of SOM and SAS selection procedures employing formal English testing and, where appropriate, assessment by linguist staff of EPP students. Although differences stemmed from School affiliation, there was evidence of autonomy of selection practice at the organisational level of Course Director, particularly in SOE where the greatest diversity of practices amongst Masters programmes was found. The strong collective culture in SOM is displayed in School regulations for the admission and selection of NNES student which supplement the institutional regulations. SOM Course Directors may modify selection procedures but only by employing additional measures that extend but do not replace the existing School and institutional requirements, for instance, through a preference for interviewing applicants. The consequences of these differences in practices between Schools and, to a lesser extent, programmes will be revisited under the discussion of the third aim of the research.

There was substantial evidence of complex decision making in selection rationales of the type described by Banerjee (1995) which are built upon multiple and sometimes competing, criteria. This was regardless of School affiliation or programme. Several extracts, especially those relating to applicant interviews, demonstrate how Course Directors view the integration of language and academic abilities. Interviews permit an interrogation of an applicant's disciplinary knowledge through the medium of English language, so allowing both to be evaluated. Correspondingly, there was little support for the alternative version of selection portrayed in the literature in which evidence of English language ability operates independently of other considerations. SOM, for instance, which is the School which places most reliance upon English entry test scores, yet prefers to use a variety of tests rather than rely upon a single one.

There may be several explanations for these findings. To some extent they may reflect the portrayal of different perspectives; academics being more likely to favour complexity of judgement, administrators, transparency and the requirements of quality assurance and earlier studies have incorporate both viewpoints in the same study. Secondly postgraduate applicants are not only diverse in nationality and language, but in age, educational background, academic qualifications and career experience. It is these aspects of an applicant's history, and others, that a Course Directors considers in deciding whether to make the applicant an offer of a place, yet they are not publicly recorded. In consequence of this diversity, it is common for applicants to satisfy one criterion but to be borderline on another; applicants do not display sets of skills and experience which fit neatly into selection criteria. This fact is neither sufficiently acknowledged in the literature nor in the spheres of policy, and audit. Postgraduate applicants differ substantially from the typical applicants for UK undergraduate courses who are often of comparable age and educational qualifications and where, as a result, selection procedures and practices may differ. These differences plus questions of scale imply differences between undergraduate and postgraduate selection.

A third reason relates to the demand for course places. Where demand for places is high, applicants are more likely to comply with entry requirements so reducing the need to consider borderline applicants who possess more varied profiles of ability and experience that entail more complex judgements. The data supported the influence of demand for places as Course Directors of programmes that were in high demand were more likely to consider raising the levels of English entry test scores and to be able to ensure that all their students complied with entry requirements. Selection decisions under these circumstances may be simpler where language requirements are concerned although choices amongst other criteria may be more challenging.

No clear consensus emerged amongst Course Directors about the predictive relationship between entry test scores and subsequent academic progress. A minority of Course Directors were outspokenly sceptical of inferring too much from a single test score and the majority could quote narratives of students where there was little correlation between English entry test scores and academic outcomes. This was true whether test scores were high or low. In describing selection rationales, Course Directors frequently referred to their own experience and sometimes of their colleagues, for instance, of students from a particular country or institution and this implicit knowledge appeared to play a considerable part in selection decisions. Selection decisions represented balanced judgments encompassing a variety of criteria which were considered in the round rather than singly or in isolation. In summary, the reported behaviour of Course Directors was congruent with that of Admissions Tutors in Banerjee's study (1995) and with the inconclusive results of studies of the relationship between English tests and academic outcomes.

Course Director's knowledge of English tests and structure varied amongst the interviewees but only one expressed interest in learning more about the available tests of English language proficiency. A minority of respondents were resistant to further development activities of this type, perceiving them as unnecessary. All Course Directors knew the institutional and School requirements for NNES applicants. Course Directors in SAS possessed less detailed knowledge of tests but such knowledge was redundant because linguist staff were responsible for making decisions on English language competence. These findings echo the findings in studies by Coleman et al (2003), Rea-Dickins et al (2007) and O'Loughlin (2008) but in the current study Course Directors regarded their knowledge as sufficient to the task in hand. It therefore contests the assumption, in the context of the present study, that greater awareness and knowledge of English will improve selection decisions and judgements. The findings in relation to Course Director's knowledge of English tests were congruent with the balanced judgement model of decision making and also reflected their view that a single test result contributed to, but did not solely determine, whether an offer should be made to an individual applicant.

The generally poor view of NNES students' writing abilities was evidence that test scores, even when in line with entry requirements, were no guarantee that a student could write satisfactorily in an academic genre, particularly for extended texts such as the thesis. Because of this issue, the assessment of writing skills was a matter of concern but one for which nobody had a clear solution. Whilst a selection interview provides useful information about oral skills, it does not contribute to assessments of writing skills. Student SAS2 from the Summer Programme whose English language was assessed by interview and was later suspected of having a learning disability exemplifies the attendant risks of some current practices. However, the assessment of writing skills is easier said than done when it is remembered that application forms, submitted papers or prepared texts are subject to questions of authenticity. The attempts of some Course Directors to engage students in electronic interaction or spontaneous writing exemplify their unease about assessments of writing. Writing assessment was another instance where Course Directors would resort to implicit knowledge and experience about an applicant's former educational background and HEI as evidence in selection decisions.

There was also ample evidence that Course Directors had responded to the present situation through the consideration and introduction of many modifications to their courses in order to ensure early identification of students most at risk and to facilitate students' writing skills early on in the programme. The outcomes of these initiatives were mixed. The introduction of a portfolio with an initial literature review was reported as having improved students' writing skills. On other courses, there had been less success. There was only one course where the opportunities for writing text, prior to the thesis, were limited and which might be contributory to the occasional poorly written thesis.



The overall concern about writing skills raises another question which is whether entry requirements should be higher for postgraduate courses in general, however it is assessed. Certainly, the majority view of these Course Directors is that the intensity of Masters study is such that there is little, if any, capacity for students to devote time and effort to anything other than academic activities. The results following the progress of Summer Programme students would support this view. Whilst some UK HEIs require higher English test scores for postgraduate than for undergraduate study this is not necessarily standard practice. The experience of Course Director SOE1 who had raised the entry requirement for his course is, however, not encouraging in this regard.

More positively, there was a reassuring consensus that students' oral skills were adequate for Masters study. To what extent this finding is particular to the context is a matter of speculation and further research will be required in contrasting settings to determine the answer. The relatively random nature of student diversity, the rural location of the campus and the prominence of group work, accompanying teamwork and interaction are all features that would encourage the development of oral skills. Where these are absent, or less prominent, the findings may differ. Another topic of agreement amongst all interviewees was the importance of immersion in the English language and culture of UK higher education, although this was not necessarily within a Course Director's control. Despite this, the degree to which a student might engage with host language and culture was yet another consideration for inclusion in selection decisions.

## **10.2 The relationship between Summer Programme students' entry assessments and subsequent linguistic and academic progress**

Whilst the varied selection practices within the institution produced rich data for analysis in the pilot and interview studies, it confounded intentions to compare students' entry test scores with subsequent academic trajectories and outcomes because of small student numbers. In addition, excluding NNES students without formal English entry test scores the variations would have resulted in the exclusion of several MSc programmes from the study so seriously limiting its scope.

The alternative option, which focused on the progress of the Summer Programme students, examined the progress of students identified as borderline in terms of English language skills. The results demonstrate that the group as a whole remained so throughout their degree studies, although the final degree results are unavailable at the time of writing. So far, nine out of 24 students have gained their Masters degrees, 14 have been instructed to revise and represent their theses and one student has failed to gain a MSc degree but may gain a PGDiploma, subject to passing an exam. Typically, around 10% of Summer Programme students do not gain MSc degrees so, provided all students pass their resubmitted theses, the pass rate is higher than anticipated.

These results show similarities with Banerjee's study (2005) of students who were considered to be 'at risk' by their Admissions Tutors on account of English language proficiency. In the Lancaster study, data were collected from eight 'at risk' students which showed that they suffered prolonged language difficulties and had less success in overcoming them than other NNES students. The data confirmed the judgements of the Admissions Tutors made at the time of selection, indicating that their diagnostic abilities to identify students who might struggle with Masters study were sound. The findings in the present study are directly analogous to the Lancaster study since the Summer Programme students constitute a comparative group to the 'at risk' students who also continue to experience language difficulties, especially in writing. The findings indicate that Cranfield staff involved in selection are equally able to identify students at risk from language problems.

Banerjee concluded that the ‘at risk’ group of students suffered in terms of extra study time and effort. The present study did not set out to reflect the student experience but did demonstrate that there were costs to teaching staff in terms of increased supervisory workload. 11 of the Summer Programme students were known, through supervisor’s reports, to have difficulties with writing which impacted adversely upon thesis supervision whilst three were reported as having no difficulties with written language. The present study therefore upholds and extends Banerjee’s findings by identifying costs from the teaching perspective.

Overall, the findings of the Summer Programme students’ progress indicate that they represent a borderline group throughout their study both academically and linguistically. However, it is not possible with any certainty to gauge to what extent linguistic problems contribute to academic failure for an individual student within the group.

### **10.3 The consequences of different admission criteria and practices upon postgraduate students’ academic progress**

The final aim of the study sought to compare the consequences of different admission criteria and practices upon students’ academic progress in a variety of courses. Because of small numbers it is not possible to determine differences at programme level but there are clear differences at School level. Distinctions between the three Schools in admission criteria and practices have already been discussed. Regarding academic progress, as measured by the need to redraft the thesis, students in SOM fared worse than students in the two other Schools. Of the six SOM students, five (83%) have been instructed to revise and represent their theses. For the 11 SAS students, one failed to gain a MSc award and six students (63%) are revising and representing their theses. For SOE, three out of seven students (42%) are revising their theses.

It appears that the more explicitly stringent are the admission criteria, the more likely are students with borderline English language skills to suffer academic setbacks. Taken at face value, the findings appear logical and predictable but there are caveats. One is the fact that the students with the lowest entry test scores (IELTS 5.5 and TOEFL IBT 64) are located in SOM. Conversely, the students in SOE have the highest entry test scores, three conforming to the institutional requirements (IELTS 6.5, TOEFL CBT 243 and TOEFL IBT 92). The differences could therefore be partially explained by variations in students’ English language proficiency at entry. Another possibility is the effect of disciplinary differences. In a discipline where written text is prominent, such as Management studies, not only are stricter English language requirements more likely but so are the assessment standards of written work. This explanation would be supported by the greater number of comments about language on the exam scripts of SOM students than from any other School. However, this explanation is not as straightforward as it may seem. Many courses in the other Schools at Cranfield are interdisciplinary and incorporate content from management, scientific and technological domains. Examples of these have been included in the SAS and SOE programmes in the current research study.

Unfortunately, the small numbers urge caution in reaching robust conclusions for the relationship between School admission criteria and academic outcomes. Further research, probably longitudinal, will be necessary to tease out the varying contributions to borderline NNEST students’ academic progress.

## 11 CONCLUSION

This study has explored how NNES students are selected for admission to taught postgraduate Masters courses in a single UK HEI and followed the progress of a group of NNES students whose English language proficiency was identified as borderline on entry. The study took place against a background of public concern about the effects of increasing internationalisation upon academic literacy and degree standards in the UK.

The accompanying literature is limited but the findings support an earlier UK study in demonstrating that academic admissions staff employ complex selection rationales in which English test scores contribute but do not ultimately determine ultimate choices. Academic staff took into account an array of factors and circumstances in reaching a judgement of an applicant's potential to succeed at Masters level study. Schools differed in selection practices and the methods and criteria employed to assess an applicant's English language ability. Some Schools preferred English test scores while others employed assessment by specialist linguist staff. No single method or set of practices appeared to be more reliable than any other at identifying students who posed the greatest academic risk due to borderline English language proficiency. However, this finding requires confirmation with a larger scale study. There was some evidence that high demand for places on a specific programme could result in higher test score entry requirements and limit the intake of NNES students with borderline English language abilities.

The main concern amongst academic staff centred on writing standards, particularly in connection with the Masters thesis that students complete at the end of the programmes. The trajectories of students with borderline English language indicated that the writing skills of this group continued to cause concern. Over half of the group were required to revise and represent their theses and the writing abilities of just less than half were reported as adversely affecting the workload of their thesis supervisors. Supervisors of students who passed their theses on first submission also reported adverse effects upon their workloads. These findings and the general concern about writing standards suggest that the problem may extend beyond the small borderline student group which was the focus of the current study. If so, it calls into question the reliability of methods for assessing the writing abilities of NNES students on entry.

A limitation of the current research is the lack of evidence from students whose entry test scores satisfy but do not exceed entry requirements. Future research should examine the writing abilities of all NNES students on taught Course Masters programmes to ascertain the scope of the problem. The case study approach employed in the study warns against generalising these findings to other HEIs. However, there are similarities with an earlier UK study. There are opportunities for comparative case study research in contrasting settings such as urban locations, the undergraduate level and where there are different compositions of national groups, to test and develop the findings further.

## ACKNOWLEDGEMENTS

We wish to thank the British Council, IELTS Australia and University of Cambridge for granting us the opportunity to carry out this research.

We are grateful to the Course Directors, lecturers and Registry staff of Cranfield University who gave up their time to contribute to the study and, in particular, to Chris Binch for his help reviewing the examination scripts.

## REFERENCES

- Banerjee, LV, 2003, *Interpreting and using proficiency test scores*, unpublished PhD thesis, Department of Linguistics and Modern English Language, Lancaster University.
- British Council, 2008, 'Why study a UK postgraduate degree?', accessed 16 December 2008 from <[http://www.educationuk.org/pls/hot\\_bc/page\\_pls\\_user\\_article?x=123348511603&y=0&a=0&d=4042](http://www.educationuk.org/pls/hot_bc/page_pls_user_article?x=123348511603&y=0&a=0&d=4042)>
- British Council, 2009, 'PM takes up the initiative on international education', accessed 14 January 2009 from <http://www.britishcouncil.org/eumd-news-pmi-ie-launch-takes-up.htm>
- Brown, L, 2008, 'The incidence of study-related stress in international students in the initial stage of the international sojourn' *Journal of Studies in International Education*, vol 12, no 5, pp 5-28.
- Coghlan, S, 2008, 'Whistleblower warning on degrees', accessed 18 November 2008 from <<http://news.bbc.co.uk/1/hi/education/7358528.stm>>
- Coley, M, 1999, 'The English language entry requirements of Australian universities for students of non-English speaking background.' *Higher Education Research and Development* vol 18, no 1 pp 7-17.
- Coleman, D, Starfield, S and Hagan, A, 2003, 'The attitudes of IELTS stakeholders: student and staff perceptions of IELTS in Australian, UK and Chinese tertiary institutions', in *IELTS Research Reports, Volume 5*, IELTS Australia Pty and British Council, Canberra, pp 160-207.
- Cotton, F and Conrow, F, 1998, 'An investigation into the predictive validity of IELTS amongst a group of international students studying at the University of Tasmania' in *IELTS Research Reports, Volume 1*, IELTS Australia Pty Ltd. Canberra, pp 72-115.
- Criper, C and Davies, A, 1988, *ELTS validation project report*, British Council and University of Cambridge Local Examinations Syndicate, London.
- Department for Innovation, Universities and Skills, 2008, 'Prime Minister's Initiative', accessed 14 December 2008 from <<http://www.dius.gov.uk/international/pmi/index.html>>
- Gillett, A, 2008, 'Using English for academic purposes. A guide for students in higher education', accessed 18 December 2008 from <<http://www.uefap.com/index.htm>>
- Hammersley, M, 2006. 'Ethnography: problems and prospects' *Ethnography in Education*, vol 1, no 1 pp 3-14.
- Higher Education Statistics Agency, 2008, 'Students and Qualifiers Data Table', accessed 14 December 2008 from [www.hesa.ac.uk/index.php?option=com\\_datatables&Itemid=121&task=show\\_category&catdex=3](http://www.hesa.ac.uk/index.php?option=com_datatables&Itemid=121&task=show_category&catdex=3)>
- IELTS, 2007, *IELTS Handbook*, co-produced by the University of Cambridge ESOL Examinations, the British Council and IELTS Australia Pty Ltd.

- Ingram, D and Bayliss, A, 2007, 'IELTS as a predictor of academic language performance, Part 1' in *IELTS Research Reports, Volume 7*, IELTS Australia Pty and British Council, Canberra, 2007, pp 137-99.
- Institute of International Education (2005a) 'Global destinations for international students', accessed 4 December 2008 from [www.atlas.iienetwork.org/?p=48027](http://www.atlas.iienetwork.org/?p=48027)
- Kerstjen, M and Neary, C, 2000, 'Predictive validity in the IELTS test' in *IELTS Research Reports, Volume 3*, IELTS Australia Pty and British Council, Canberra, pp 85-108.
- Lee, Y and Greene, J, 2007, 'The predictive validity of an ESL placement test' *Journal of Mixed Methods Research*, vol 1, no 4 pp 366-89.
- Lloyd-Jones, G, Odedra, H and Neame, C, 2007, 'An exploration of diversity in a UK postgraduate university.' in *Education in a Changing Environment*, Doherty, E, ed. Informing Science Press Santa Rosa, pp 79-98.
- Maxwell, J, 2004, *Qualitative Research Design: An Interactive Approach*, Sage Publications, Thousand Oaks.
- Morgan, DL, 1997, *Focus groups as qualitative research*, Sage Publications, Thousand Oaks.
- O'Loughlin, K, 2008 'The use of IELTS for university selection in Australia: a case study.' in *IELTS Research Reports, Volume 8*, IELTS Australia Pty and British Council, Canberra, pp 145-241.
- Rea-Dickins, P, Kiely, R and Yu, G, 2007, 'Student identity, learning and progression (SILP): the affective and academic impact of IELTS on 'successful' candidates' in *IELTS Research Reports, Volume 7*, IELTS Australia Pty and British Council, Canberra, pp 59-136.
- Select Committee on Innovation, Universities, Science and Skills, 2008, *Letter from Peter Williams, Chief Executive, Quality Assurance Agency to the Chairman of the Committee*, 30.10.08. accessed 16 December 2008 from <http://www.publications.parliament.uk/pa/cm200708/cmselect/cmdius/905/8071710.htm>
- Sastry, T, 2004, 'Postgraduate education in the United Kingdom', Higher Education Policy Institute, London.
- Ward Schofield, J, 2000, 'Increasing the generalizability of qualitative research' in *Case Study Method*, Gomm, R, Hammersley, M and Foster, P, eds. Sage Publications, London, pp 69-97.
- Yin, RK, 2003, *Case study research: design and methods*, Sage Publications, London

## APPENDIX 1: INTERVIEW SCHEDULE FOR COURSE DIRECTORS

Nature and reasons for research project

Topic areas for discussion

Confidentiality

Audio-recording

Any queries?

Can you tell me something about the course you direct?

How long has it been offered?

Are there similar courses around the UK?

How many students do you admit?

How many students apply?

What is the employment rate of your graduates?

What type of employment do they enter?

In which countries do they work?

What is your experience of a. teaching on the MSc programme and b. as a Course Director?

In relation to English language proficiency, how do you select students on your course?

What factors do you consider when assessing an applicant's English language ability?

What evidence do you use in assessing an applicant's English language ability?

What features of a student's CV and background would prompt you to recommend the student takes the Summer Programme?

What part does English language testing (IELTS/TOEFL etc) play in selection and admission to your course?

Do you consider different language skills separately e.g. listening, speaking, reading and writing?

Do you admit students under the European Partnership Programme?

What are your views on current admission practices in relation to English language proficiency?

How could current selection and admission practices be improved?

What do you think are the educational consequences of mixed English language ability classes for

a. non-native English speaking students?

b. native English speaking students?

c. teaching staff?

Can you describe the programme structure and its assessment?

Is English language assessed on the course? If so, how?

Have you observed problems in group projects related to mixed English language ability?

Sometimes student cohorts may include a large group of students who share a language other than English. Have you observed any educational consequences of this?

Native English speaking students may adopt specific roles in group projects such as proof reading and editing. What are your views on this?

Have you ever taught a non-native English speaking student who you thought might be dyslexic?

Have you recommended students for academic English support during the course?

What are your views on the web publication of Masters theses?

What are your views on the current ongoing provision of academic English support?

How could ongoing academic English provision be improved?

Would training for teaching staff in admissions and selection be helpful? If yes, what form should it take?

Thank you

## APPENDIX 2: FOCUS GROUP GUIDE

Nature and reasons for research project

Topic areas for discussion

Confidentiality

Audio-recording

Any queries?

Tell me about your experience of postgraduate teaching (Cranfield and elsewhere)?

(particularly, any experience of teaching non-native English speakers)

Tell me about the new roles you are undertaking within the department?

What are you looking for in selecting students for a Masters course?

What evidence will you look for?

How will you assess English language ability?

Does the international mix of students at Cranfield affect learning and teaching?

More specifically:

- a. NNES students
- b. NES students
- c. teaching staff
- d. management of group projects

What do you know about existing English language support for students?

What are your views on the form English support for students should take?

Is there a place for training staff in the selection of non-native English speaking students?

Thank you

## APPENDIX 3: QUESTIONNAIRE FOR SUPERVISORS OF MASTERS THESES OF SUMMER PROGRAMME STUDENTS

Thank you for taking part in this survey which is part of an IELTS/British Council funded research study exploring the effects of students' English language proficiency on progress in selected Masters courses at Cranfield.

The survey aims to discover whether deficiencies in students' English language proficiency are reflected in thesis marking and supervisory workload.

The survey should take about 5 minutes to complete. Please answer the questions with reference to the supervision of the student named in the accompanying email.

Once you press the submit button you will not be able to edit your responses. All data will be treated confidentially and anonymised in future publication.

- 1 I am a thesis supervisor on the  
*Title of the Masters programme*
- 2 Please enter the number of years you have been supervising MSc theses on this programme.
- 3 Please enter the number of MSc theses you supervised in 2007/8.
- 4 My student's English language proficiency has contributed to the thesis mark.  
*Adversely not at all beneficially*  
*If adversely, please specify:*
- 5 My student's English language proficiency affected my supervisory workload  
*Adversely not at all*  
*If adversely, please specify:*
- 6 Please provide any further information that you feel may be relevant to the topic

Thank you for responding to the survey



## APPENDIX 4: SUPERVISORS' RESPONSES TO QUESTION 6 OF THE QUESTIONNAIRE FOR SUPERVISORS OF MASTERS THESES

**SAS Supervisor:** Students are informed that their thesis needs to be grammatically correct - and therefore to discuss all issues re their thesis with their supervisor at an early stage. The onus is on the student to ensure that they use the services of the supervisor, and each student is different in this respect.

**SAS Supervisor:** English language proficiency varies considerably among my students - as noted above for some European students it has been a problem.

**SAS Supervisor:** As I understand the role of supervisor we do not correct English. So unless they are minor mistakes then I suggest that they get a proof reader.

**SOE Supervisor:** Technical writing skills require a level of English which is not being currently met by the majority of overseas students I supervise both from the EU and outside. This somehow needs to be addressed before the students join the course.

**SAS Supervisor:** Her English was generally good.

**SOM Supervisor:** I would appreciate any "great solutions" but I guess we just have to live with this only downside of our truly international, multicultural groups of students.

**SOM Supervisor:** I encourage my students to write the introductory chapter early on. I am very critical of the English not just the content in this first chapter. From this I can advise the student if they should use a professional proof reader or not. I also inform them that I am not the proof reader.

**SAS Supervisor:** Whilst every support was made available, this was at times ignored. Corrected and acceptable parts were sometimes changed for no apparent reason leading to either further necessary corrections, or in the case of the final submission, some sub-standard parts.

**SAS Supervisor:** It is primarily the written thesis document where greater time input is required to resolve the English problems. This may require more than one iteration particularly if new text has to be added which in turn then needs correcting.

**SAS Supervisor:** I think it crucial that non-native English speakers work hard to make their written work as fluent as possible because, as a supervisor and examiner, it can often be difficult to discern those who have a thorough understanding of the science, but who just can't express it in fluent English, and those who are not really understanding the science.

**SOM Supervisor:** Because I spent a lot of time draft-reading the work, it meant that I spent relatively less time on the evaluation of the actual content of the work.

**SOE Supervisor:** The standard of English from China / Taiwan has improved in recent years although individual students will have variable skills.



## 4 Construct validity in the IELTS academic reading test: A comparison of reading requirements in IELTS test items and in university study

### Authors

Tim Moore

Swinburne University

Janne Morton

University of Melbourne

Steve Price

Swinburne University

Grant awarded Round 13, 2007

This study investigates suitability of items on the IELTS Academic Reading test in relation to the reading and general literacy requirements of university study, through a survey of reading tasks in both domains, and interviews with academic staff from a range of disciplines.

### ABSTRACT

The study reported here was concerned with the issue of test development and validation as it relates to the IELTS academic reading test. Investigation was made of the suitability of items on the test in relation to the reading and general literacy requirements of university study. This was researched in two ways – through a survey of reading tasks in the two domains, and through interviews with academic staff from a range of disciplines.

Tasks in the two domains were analysed using a taxonomic framework, adapted from Weir and Urquhart (1998), with a focus on two dimensions of difference: *level of engagement*, referring to the level of text with which a reader needs to engage to respond to a task (local vs global); *type of engagement* referring to the way (or ways) a reader needs to engage with texts on the task (literal vs interpretative).

The analysis found evidence of both similarities and differences between the reading requirements in the two domains. The majority of the IELTS tasks were found to have a ‘local-literal’ configuration, requiring mainly a basic comprehension of relatively small textural units. In the academic corpus, a sizeable proportion of tasks had a similar local-literal orientation, but others involved distinctly different forms of engagement, including tasks that required a critical evaluation of material (i.e. more interpretative), or which stipulated reference to multiple sources (i.e. more global). The study also found a good deal of variation in the reading requirements across the disciplines.

The results of the study are used to suggest possible enhancements to the IELTS academic reading test. A useful principle to strengthen the test's validity, we argue, would be to push test tasks, where possible, in the direction of those more 'global-interpretative' reading modes characteristic of academic study.

## **AUTHOR BIODATA**

### **TIM MOORE**

Tim Moore works in the area of academic literacy at Swinburne University. His PhD was on the subject of critical thinking in the disciplines. Along with research into the IELTS reading module, he and co-researcher, Janne Morton, have also conducted IELTS-funded research into the academic writing module.

### **JANNE MORTON**

Janne Morton works in the School of Languages and Linguistics at the University of Melbourne as a lecturer in ESL. She is currently completing her PhD in the area of socialization into disciplinary discourse. Her research interests include academic literacies, spoken genres, and second language testing and assessment.

### **STEVE PRICE**

Steve Price works at Swinburne University and has provided language support to tertiary level students for many years with a particular interest in the development of disciplinary reading skills. He is currently researching how law students from non-English speaking backgrounds engage with common law discourses.

**TABLE OF CONTENTS**

<b>1</b>	<b>Introduction .....</b>	<b>188</b>
<b>2</b>	<b>Review of literature .....</b>	<b>189</b>
	2.1 The IELTS Academic Reading Test .....	189
	2.2 Construct validity.....	190
	2.3 Dimensions of reading .....	191
	2.4 Frameworks used in reading assessment studies .....	192
<b>3</b>	<b>Method .....</b>	<b>193</b>
	3.1 Towards an analytical framework.....	193
	3.2 Disciplines investigated.....	197
	3.3 Data and procedure .....	198
<b>4</b>	<b>Findings .....</b>	<b>200</b>
	4.1 IELTS reading tasks.....	200
	4.2 Academic reading tasks .....	221
	4.3 Findings from interviews - Comments on IELTS reading tasks .....	242
<b>5</b>	<b>Summary and discussion of findings .....</b>	<b>245</b>
	5.1 Main findings.....	245
	5.2 Specific findings.....	247
<b>6</b>	<b>Implications of findings for future development of the Reading Test.....</b>	<b>250</b>
	6.1 Should the IELTS Reading Test be modified? .....	251
	6.2 How could the IELTS Reading Test be modified? .....	251
	6.3 Further research .....	255
	Acknowledgements .....	256
	References .....	256
	Appendix 1: List of materials used in the IELTS Task Corpus .....	260
	Appendix 2: Schedule used in interviews with academic staff.....	261
	Appendix 3: Additional sample items showing more global and/or interpretative engagements .....	266

## 1 INTRODUCTION

Reading has always been a key element of university study. There was a time in fact when the preferred terminology for studying in a subject area at university was ‘reading the subject’. Nowadays, many recognise that it is the intelligent engagement with one’s sources that more than anything else defines the quality of being academically literate. Taylor (2009), for example, sees most student endeavours in the academy – whether the writing of essays, or engaging with the content of lectures, or the discussing of ideas in tutorials and seminars – as emerging from a “conversation” with one’s readings in a discipline (p 54). In the domain of language testing, the manifest importance of reading in university study is reflected in the prominence given to this skill area in the various language tests used by universities for the selection of students. Thus, in all the varieties of format found in the more widely-used language tests over the last 30 years (ELTS, IELTS, TOEFL), one single common element has been the use of a dedicated reading component.

Given the importance of reading within academic study, an issue of continuing interest for researchers and test developers is the validity of tests used to assess students’ academic reading abilities. A test is said to be valid if it ‘reflects the psychological reality of behaviour in the area being tested’ (Hamp-Lyons, 1990, p 71). In the case of a test of academic reading proficiency, this validity relates to a number of different areas, including:

- task stimulus i.e. the texts that candidates engage with on the test
- task demand i.e. the test items, which prescribe certain types of interaction between the reader and text
- task processes i.e. the reader-text interactions that actually take place in the completing of the test (McNamara, 1999).

Previous IELTS validation research has seen strong emphasis placed on the first of these areas – the task stimulus component of the reading test (see for example, Clapham 1996). Recently-commissioned research has also seen some attention given to task processes – in the work of Weir, Hawkey, Green and Devi (2009) into performance conditions on the test and how these might relate to the subsequent reading experiences of first year university students. To our knowledge, there has been limited validation work done in recent years (one needs to go back to Alderson’s (1990a; 1990b) major work on the testing of reading comprehension skills) on the second of these areas – that is, the task ‘demands’ of the current version of the reading test, and how much these might relate to the types of reading tasks and activities required of students on university programs.

The study described in this report investigated the suitability of test items in the Academic Reading Test in relation to the reading and general literacy requirements of university study. Specifically, the research sought answers to the following questions:

- 1 in what systematic ways can items on the IELTS academic reading module be analysed and classified?
- 2 what does a taxonomic analysis of test items reveal about the construct of reading underlying the IELTS academic reading module?
- 3 what is the degree of correspondence between the reading skills required on the IELTS test and those typically required on a range of undergraduate university programs?

Two methods were employed in the research: i) a comparative analysis of IELTS test items and assessment tasks from a range of undergraduate courses; and ii) semi-structured interviews with

academic staff involved in the teaching of courses covered in i). Findings from the research are used to make suggestions about how the IELTS Academic Reading Test could be adapted to make it more closely resemble the modes of reading required in formal academic settings.

## 2 REVIEW OF LITERATURE

The literature in the fields of reading research and reading assessment research is vast and complex. In the following section, we review briefly those areas thought to have particular relevance to the current study. These include the idea of construct validity; theoretical models of reading; and inventories of reading skills and strategies. We begin with a brief review of the IELTS Academic Reading Test, including an account some of the changes that have been made to the test over the 20 years of its use.

### 2.1 The IELTS Academic Reading Test

The IELTS system in its current form provides two different reading tests: a general training module and an academic module. The general training module is designed for a variety of cohorts and assesses “basic survival skills in a broad social and educational context”, while the academic module is said to “assess the English language skills required for academic study or professional recognition” (IELTS, 2007, p. iii). The present study is concerned only with the latter of these modules. According to test specifications, the skills tested in the IELTS Academic Reading include: following instructions, finding main ideas, identifying the underlying concept, identifying relationships between the main ideas, and drawing logical inferences (cited in Alderson, 2000 p 206; IELTS, 1996).

An IELTS Academic Reading Test is typically comprised of three sections (or testlets), each organised around a separate reading passage. These passages, which average about 750 words in length, are drawn from a range of sources including magazines, journals, books and newspapers, with topics designed to be of general interest, written for a non-specialist audience. Accompanying the reading passages are a range of tasks (40 in total) used to test students comprehension of material in the 60 minutes allocated. These tasks or techniques are characterised by IELTS (1999) as follows:

- multiple choice
- short answer questions
- sentence completion
- notes/summary /diagram/flow chart/table completion
- choosing from a heading bank for identified paragraphs/sections of text
- identification of writer’s view/attitudes/claims
- classification
- matching lists
- matching phrases.

Alderson (2000) notes that an “interesting” feature of the IELTS Reading Test is its use of multiple methods to test understanding of any one passage. This is a strength he suggests because in real life, readers typically respond to reading texts in many different ways (p 206). The *Official IELTS Practice Materials* (2007) include the following range of tasks used with each reading passage:

**Passage 1:** section-summary match; gapped summary; true/false/not given

**Passage 2:** true/false/not given; information-category match; multiple choice

**Passage 3:** section-summary match; sentence completion

The IELTS Academic Reading Test has been subject to several major changes since its introduction in 1989. The most important of these, the result of extensive monitoring and evaluation work in the early 1990s (eg Clapham 1996), saw the removal of subject-specific reading subtests, and the removal of the thematic link between Reading and Writing tests. The rationale for such changes has been extensively described in the IELTS literature (Charge & Taylor, 1997; Taylor, 2007). For example, the removal of the discipline specific component of the Reading Test was the outcome of findings that suggested that the range of subject-specific modules was not warranted, and that a single test did not discriminate for or against candidates from various disciplines (eg Taylor, 2007). The decision to separate the reading from the writing test was based on the observation that candidates varied considerably in the extent to which they exploited reading material in the Writing Test, with the implications this had for test fairness. It was thought further that having this connection also increased the potential for confusing the assessment of writing ability and reading ability (Charge & Taylor, 1997).

As mentioned, the focus of the current study is exclusively on the reading tasks and not on the reading passages that accompany them. It does need to be acknowledged however, that having a separation of these components limits the perspective somewhat. This is for the reason pointed out by Alderson (2000, p 203) that there may be a relationship between the text type and the sort of task or technique that can be used with it. This idea will be returned to briefly in the concluding section of the report.

## 2.2 Construct validity

The present study is concerned with investigating the construct validity of the IELTS Reading Test. In terms of reading tests, ‘construct validity’ is a measure of how closely a test reflects the model of reading underlying the test. In other words, the concept of ‘construct validity’ is related to those abilities it is thought readers need to possess in order to handle the demands of the target language domain. In the case of the IELTS Academic Reading Test, this domain is study at university level. Thus, if the ability to scan for specific information is considered an important part of university reading requirements, then the reading construct should include scanning and the test should diagnose the ability to quickly locate specific information (Alderson, 2000). Whilst construct validity is often associated with skills, another dimension is task structure. Bachman and Palmer (1996) suggest that a focus on the structure as well as the skills of target language use tasks might lead to the development of more ‘authentic’ test tasks (p.147).

The construct validity of a test is particularly important when the test is a large scale public test, and where there is a close connection between the operations of the test and the conduct of related educational programs. The construct validity of such tests thus has implications for curriculum and classroom practice through the so-called “test washback” (Alderson and Wall, 1993). As Messick (1996, p 252) points out:

[i]f important constructs or aspects of constructs are underrepresented on the test, teachers might come to overemphasise those constructs that are well-represented and downplay those that are not.

Washback is considered harmful then when there is a serious disjunct between a test’s construct of reading and the broader demands of real world or target language tasks.



The IELTS test is an example of a public test that is used to make crucial decisions about large numbers of people – whether they are eligible for English-speaking university entrance or not based on their English language abilities. An increase in the numbers of international students wanting to study at English-speaking universities and a concomitant increase in the number of universities requiring IELTS scores has led to a significant expansion of the IELTS test in recent years. This in turn has resulted in IELTS preparation programs being an important focus of many EAP courses taught in language centres throughout the world (Saville and Hawkey, 2003; Read and Hayes, 2003). The increased influence of IELTS and possible concerns about test washback suggest the need for, in this case, the reading construct underlying the test to be firmly based on a thorough understanding of the nature of reading demands in university study. It is this issue – the importance for the reading test to be as authentic as possible given practical and other constraints – that has motivated the present study.

### 2.3 Dimensions of reading

The current project is framed within broad theories of reading. Central to these are differing views about the nature of textual meanings and the relationships that exist between these meanings and the reader of a text. The more traditional view – the ‘transmission model’ – sees texts embodying relatively stable, objective meanings, ones that a proficient reader is able to locate and reproduce. Carroll (1964), for example, characterises reading as “the activity of reconstructing the messages that reside in printed text”. This conception of reading as the finding of pre-existent meanings is arguably the predominant construct in many reading comprehension tests, especially those that rely heavily on multiple choice formats (Hill & Parry, 1992; Alderson, 2000).

An alternative view, one that has gained increasing acceptance in many areas of the academy (particularly in education and in some branches of the humanities) is to see texts as having no single definitive meaning, but rather the potential for a range of meanings, ones that are created through the engagement of individual readers. As Widdowson (1979) states, “since conceptual worlds do not coincide, there can never be an exact congruence of coder’s and encoder’s meanings” (p 32). Despite the growing acceptance of ‘receptionist’ theories of meaning, there appears to be a reluctance – even on the part of more committed post-modernists – to accept fully the logical consequences of this position – namely, that any subjective account of the meaning of a text may ultimately be valid. It is the view of the researchers that both a strong receptionist and a strong transmissionist position represent rather idealised accounts of reading, and are best thought of as end points on a continuum of more reader-oriented and more text-oriented perspectives on meaning.

Related to these broad definitions of reading are differing ideas about what the processes of reading are thought to involve. Traditionally, accounts in this area have tended to aggregate around two broad approaches: bottom-up ‘information processing’ (with a focus on the processing of more micro-level constituents of texts – letter, words, phrases, sentences etc); and top-down ‘analysis-by-synthesis’ (with a focus more on macro-level constituents – genre, text structure, as well as the role of background schematic knowledge etc). Recently, there has been a move towards a more interactive, hermeneutic approach, one that assumes a degree of bi-directionality in these processes (Hudson, 1998). In the current project, research in the area of reading processes was useful as a way of identifying the type(s) of processing that test items appear to be principally concerned with, and also the levels of texts.

## 2.4 Frameworks used in reading assessment studies

Much of the research into the nature of reading in different domains has relied on taxonomies that seek to divide reading practices into a variety of skills and sub-skills. Particularly influential among these has been Munby's (1978) list of general language skills, used both for the purposes of syllabus and material design, as well as for the design of tests. In a list that he described at the time as "not exhaustive", Munby distinguished a total of 266 skills – sub-categorised into 54 groups, including such reading specifics as:

- understanding the communicative value (function) of sentences and utterances with explicit indicators
- understanding relations between parts of texts through grammatical cohesion devices of reference, comparison etc
- scanning to locate specifically required information: a single point/more than one point involving a simple search.

Amid the complexity of Munby's scheme, it is possible to detect a basic division between reading skills that are involved in the simple comprehension of texts (*eg understanding explicitly stated information* p 126), and those involving interpretation of some kind (*eg interpreting text by going outside it* p 128).

In recent years there have been efforts to pare such taxonomies down to a more manageable catalogue of skills (*eg Carver 1997; Grabe & Stoller, 2002*). Carver (1997), for example, recognises five basic elements: 'scanning', 'skimming', 'rauding', 'learning' and 'memorising'. Rauding is defined as a 'normal' or 'natural' reading, which occurs when adults are reading something that is relatively easy for them to comprehend (Carver, 1997, pp 5-6). For Grabe and Stoller (2002), the activity is best captured under seven headings:

- 1 Reading to search for simple information
- 2 Reading to skim quickly
- 3 Reading to learn from texts
- 4 Reading to integrate information
- 5 Reading to write (or search for information needed for writing)
- 6 Reading to critique texts
- 7 Reading for general comprehension

One notes that this latter list takes on a slightly simplified form in a recent study conducted for the TOEFL reading test (Enright *et al*, 2000):

- 1 Reading to find information (or search reading)
- 2 Reading for basic comprehension
- 3 Reading to learn
- 4 Reading to integrate information across multiple texts

Of the various taxonomies developed, the most useful for the present project was thought to be that proposed by Weir and Urquhart (1998), and used in another recent study into the IELTS academic reading test conducted by Weir *et al*. (2009). Rather than compile a list of discrete skills, Weir and Urquhart construct their taxonomy around two dimensions of difference: reading level and reading

type. For reading level, a distinction is made between reading processes focused on text at a more global level, and those operating at a more local level. For reading type, the distinction is between what is termed ‘careful’ reading and ‘expeditious’ reading, the former involving a close and detailed reading of texts, and the latter “quick and selective reading ... to extract important information in line with intended purposes” (Weir & Urquhart, 1998, p 101). The ‘componential matrix’ formed by Weir and Urquhart’s two dimensions has the advantage of being a more dynamic model, one that is capable of generating a range of reading modes.

In the literature on reading taxonomies, one notes a degree of slippage in what construct it is exactly that is being characterised. Most commonly, it is one of reading ‘skill’ (eg. Munby), but an assortment of other terms and concepts are typically used eg ‘processes’ (Carver, 1997), ‘purposes’ (Enright *et al*, 2000, Weir *et al*, 2009), ‘strategies’ (Purpura, 1998). Such terms, which are arguably somewhat inchoate in nature, all refer in some way to the putative abilities or behaviours of readers. In the present project, the construct we are dealing with is not related to any qualities of the readers as such. Rather the focus is on some entity that is external to the reader – the reading task. In this way, the preferred construct for the project is one of ‘activity’, or rather of ‘prescribed activity’.

### 3 METHOD

In this section, we outline the analytical framework used in the research, the disciplines investigated, and the nature of the data that was collected and analysed in the study.

#### 3.1 Towards an analytical framework

The approach adopted for the development of the analytical framework was a syncretic one, drawing initially on both IELTS tasks and academic tasks to establish broad dimensions of difference between reading tasks and then to refer to relevant theoretical frameworks later to refine the classification scheme. The method followed was similar to the one adopted in a similar validation study of the IELTS writing test conducted by several members of the research team (Moore & Morton, 2007). The framework that was used ultimately was derived in large part from the componential schema of Weir and Urquhart (1998), described in the previous section.

##### Dimension 1: Level of engagement

The first dimension used was what we term ‘level of engagement’ with text. For our study of IELTS and academic reading tasks, this dimension refers to how much of a text (or texts) a reader is *required* to engage with in the performing of a prescribed task. It was noted in our preliminary survey of reading tasks that some tasks were focused on quite circumscribed (or ‘local’) sections of a text (eg single sentences, or groups of sentences), whilst in others, there was a need to appraise larger textual units (eg a series of paragraphs, or a whole text). The most extensive ‘level of engagement’ related to those tasks that required engagement with a number of different texts.

For this dimension of reading tasks, the following two broad categories were used after Weir and Urquhart (1998), and Hill and Parry (1992).



As Weir *et al.* (2009) note, different types of reading activities are, of their nature, either more local or more global in their orientation. Thus, for example, the act of ‘scanning’ (i.e. locating specific information within a text) has a more local focus; on the other hand, the act of ‘skimming’ (i.e. obtaining an overview of a text) is necessarily a more ‘global’ form of reading.

### **Dimension 2: Type of engagement**

Our second dimension – ‘type of engagement’ – involved an adaptation of the Weir and Urquhart (1998) schema. Whereas their categories of ‘careful’ and ‘expeditious’ readings refer arguably to the reading ‘strategies’ (or ‘processes’) that students may adopt, our focus on academic tasks meant that the interest was more on what was needed to be done with texts, that is to say the prescribed outcomes of the reading. In our preliminary observations of tasks in the two domains (IELTS and academic study), it was clear that different tasks called for different types of readings. Sometimes, for example, the requirement was simply one of understanding the basic contents of a text; in other instances, readers needed to bring a more personal response to material.

In developing this dimension, the study drew initially on the distinction traditionally made in linguistics between semantic and pragmatic meaning. The semantic meaning of a text is typically characterised as the sum of the individual propositions contained within it; pragmatic meanings, on the other hand, refer to those meanings that emerge from the relationship between the text and the context of its production (Yule, 1996). As Yule (1996, p 4) explains it, whereas semantics is concerned with the literal meanings of sentences, pragmatics is concerned with probing less tangible qualities, such as “people’s intended meanings, their assumptions, their purposes or goals, and the kind of actions they are performing when they speak [or write].”

Related to acts of reading, a broad distinction can be made in this way between a focus on what a text says (semantic meaning), and what a text does, in saying what it says (pragmatic meaning). To illustrate this distinction, Taylor (2009, p 66) cites the following short text sample from a French History textbook:

The winter of 1788-9 was a very harsh one in France, inflicting untold misery on the peasants. The revolution broke out in July 1798.

These two sentences, as Taylor explains, can be read ‘literally’ i.e. as a sequence of propositions about events in late 18th century France (a semantic reading); or they can be read more ‘interpretatively’; in this case, as an attempt by the author to explain events i.e. to see the first event as a cause for the second (a pragmatic reading). Taylor (2009) suggests that while both types of reading are important in the context of academic study, it is the latter mode – the more interpretative readings – that is often missing in accounts of the types of reading students typically need to do in their studies.

This basic distinction in the way one might engage with a text (or be required to engage) provided the second category of our framework as follows (a similar basic distinction is often drawn in the broader area of learning theory, where engagement with materials is seen to divide between such binaries as surface vs deep learning (Marton & Saljo, 1976), higher and lower order skills (Bloom, 1956), reproductive vs. analytical (Ballard & Clanchy, 1991), critical and non-critical approaches to knowledge (Ennis, 1987)):



Whereas the 'literal' element of our binary refers to the unitary act of comprehending the propositional content of a text, there are arguably many different ways that one might engage with texts 'interpretatively'. These might include, for example, as Alderson (2000, p 320) suggests:

- identifying the function of a piece of writing
- recognising an authors presuppositions and assumptions
- distinguishing fact from opinion
- recognising an intended audience and point of view.

Catherine Wallace (1999, p 109), working within a more 'critical literacy' paradigm, provides a different list of skills, including:

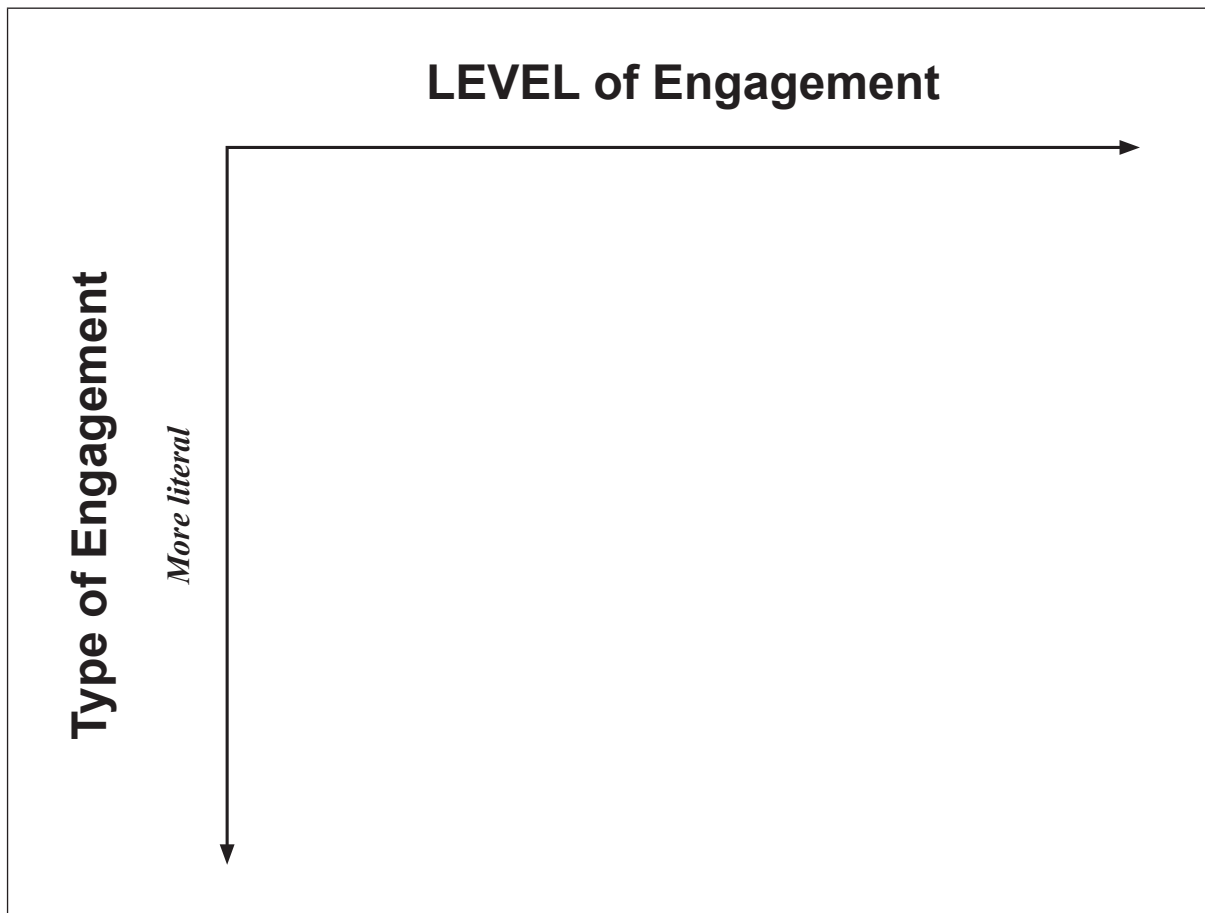
- understanding the hidden messages in texts
- identifying how texts persuade one to behave or think
- appreciating how texts are written for different audiences
- appreciating how texts might be read in different ways by different audiences

The present study resisted any effort to draw up a definitive, a priori list of these interpretative modes, and indeed to try to establish any hierarchical relationship between them. Instead, the approach employed was to rely on the broad brush distinction drawn between 'literal' and 'interpretative' forms of reading, and to assess whether reading tasks set for students (either on the IELTS reading test, or in academic study) seemed, on the face of it, to require more of one form of engagement than the other.

### Summary of analytical framework

The two dimensions of the analytical framework – *level of engagement* and *type of engagement* – are represented on the matrix shown in Figure 1 below. The *level of engagement* dimension, which describes a continuum from more 'local' to more 'global' engagement, refers to the level of text with which a reader needs to engage to respond to a task. At the extreme left of the axis (most local) would be tasks requiring engagement at the level of 'word'; at the extreme right of the axis (most global) would be tasks requiring engagement with multiple texts.

The *type of engagement* dimension, which describes a continuum from more 'literal' to more 'interpretative' engagement, refers to the way (or ways) a reader needs to engage with a text to respond to a task. At the top of this axis (most literal) would be tasks requiring a basic comprehension of textual material; at the bottom of the axis (most interpretative) would be tasks requiring a highly critical, and personal engagement with texts.

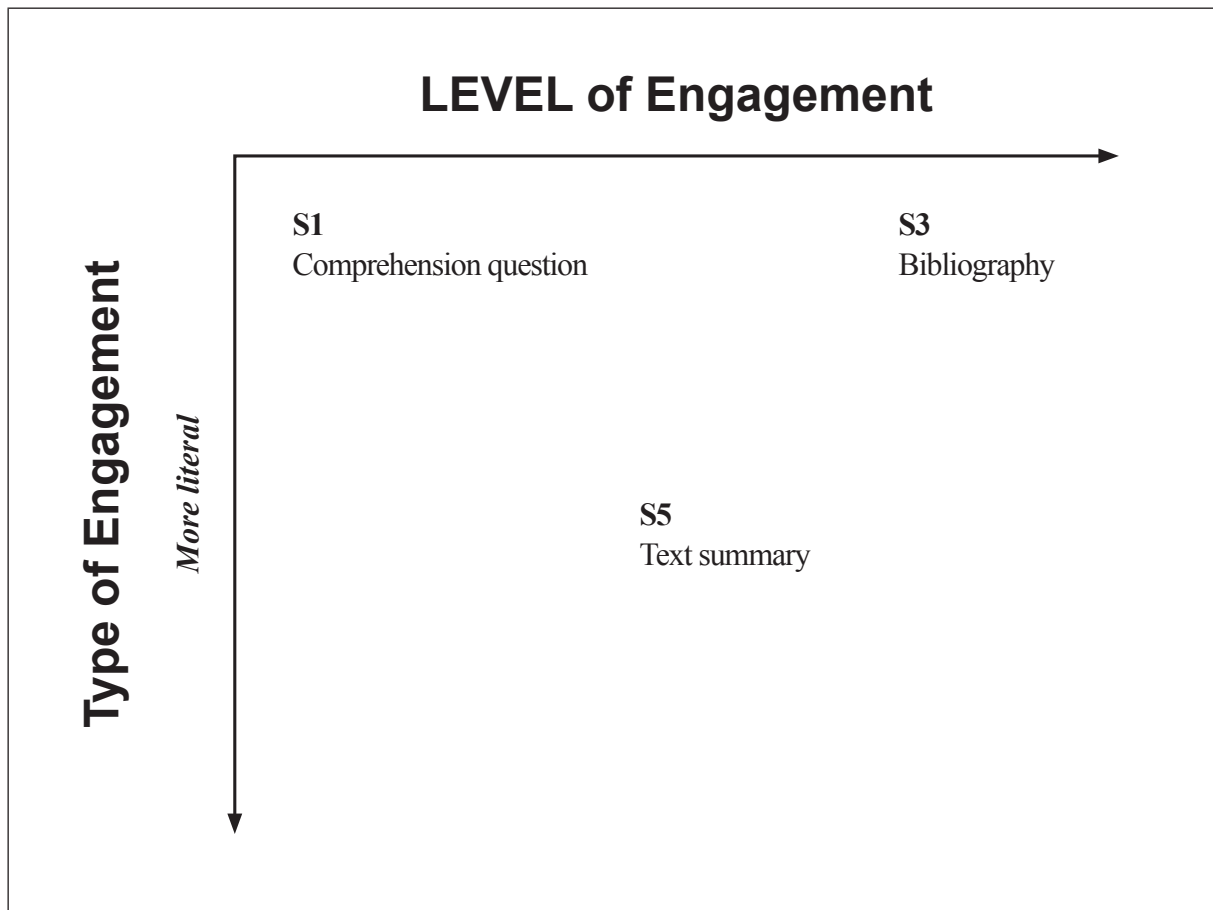


**Figure 1. Analytical framework used in the study**

To demonstrate the use of the analytical framework, a number of reading-related tasks are outlined in Table 1 below, with an analysis of each according to the two dimensions of the schema. In Figure 2 below we have shown how such tasks might then be plotted on the two continua of the matrix.

<b>SAMPLE READING-RELATED TASK</b>	<b>ANALYSIS</b>
<b>S1</b> answering a comprehension question relating to a single piece of information	HIGH LOCAL HIGH LITERAL
<b>S2</b> explaining the connotative meaning of a word in a text	HIGH LOCAL HIGH INTERPRETATIVE
<b>S3</b> compiling a bibliography of texts related to a specific subject	HIGH GLOBAL HIGH LITERAL
<b>S4</b> preparing a critical review of the literature on a specific subject	HIGH GLOBAL HIGH INTERPRETATIVE
<b>S5</b> summarising a single text	MID LOCAL/GLOBAL MID LITERAL/INTERPRETATIVE

**Table 1. Analyses of sample reading activities using analytical framework**



**Figure 2. Plotting of sample reading tasks on matrix**

Whilst the two dimensions used in the study were conceived of as separate features of reading tasks, it was noted in our preliminary survey of data that there was often an inter-relationship between the two. Thus, a general pattern was observed that if tasks were highly 'local' in their focus, it was often the case that a more 'literal' form of engagement was required. Similarly, for those tasks which took in larger more 'global' textual units, the tendency was for the engagement to be pushed more towards the 'interpretative' end of our continuum.

### **3.2 Disciplines investigated**

To obtain a picture of reading requirements across the academy, data were collected from two different universities, and from a variety of disciplines. One of the institutions was a long-established Australian university offering programs of a more traditional nature; the other was what is characterised as a 'new generation' university with a focus on more vocationally-oriented programs. Becher's (1989) matrix of hard-soft/pure-applied disciplines was used to ensure sampling from a cross-section of disciplines. Becher's typology groups academic disciplines on the basis of research methods and attitudes to knowledge. Whilst the disciplines selected in our study fit neatly within the four groupings (see Table 2), it is acknowledged that boundaries between groups may not be as clear-cut as a typology such as this suggests (see also Becher, 1989).

	<i>PURE</i>	<i>APPLIED</i>
HARD	Physics Biology Economics	Engineering Computer Science Architecture
SOFT	Media Studies Linguistics History	Business Studies Management Communications

**Table 2. Disciplines investigated in study: Becher (1989) taxonomy**

<i>DISCIPLINE</i>	<i>TITLE OF SUBJECT</i>
Physics	Life sciences and environment
Biology	Genetics and the Evolution of Life
Economics	Macroeconomics
Media Studies	The Media in Australia
Linguistics	Intercultural Communication
History	Contemporary History
Engineering	Engineering Systems Design
Computer Science	Informatics – Practical Computing
Architecture	Constructing Environments
Business Studies	Deriving Business Value
Management	Business in the Global Economy
Communications	Professional Writing

**Table 3. List of disciplines and subjects**

Within the twelve discipline areas covered, a single subject in each was selected for investigation (Table 3). All subjects were offered at first year undergraduate level, and were selected partly on the basis of their having relatively high enrolments of students from second language backgrounds. Whilst these subjects were chosen as representative of their discipline area, it is acknowledged that any single subject can only ever cover a portion of the intellectual content and educational practices of the parent discipline as a whole.

### **3.3 Data and procedure**

The study was mainly qualitative in nature involving the use of two research methods: an analysis of tasks (both IELTS and academic tasks) and interviews with academic staff. This combination of methods fits roughly with what Swales (1998) calls ‘textography’ – described as “something more than a disembodied textual or discursal analysis, but something less than a full ethnographic account” (p 1).



### IELTS task survey

A corpus of IELTS reading test samples was compiled for the study. These were from two sources: i) the official IELTS Practice Test (IELTS, 2007); and ii) practice test material published by Cambridge University Press (see Appendix 1 for list of corpus materials). It is understood that the CUP materials are made up partly of retired official materials, and so were thought to reflect better than many other commercial materials the actual nature of the official test. No live reading test materials were available to the study. A total of 13 complete tests were investigated, each made up of a variety of task types.

Reading tasks were analysed by the researchers according to the two dimensions of the study's analytical framework i.e. the 'level' and 'type' of engagement. Whilst a degree of interpretation invariably enters into any analysis of this kind, some objectivity was achieved on the study by having each researcher analyse tasks independently, and then for a consensual analysis to be arrived at through processes of moderation.

### Academic task analysis

To compile data for the university component of the study, lecturers from the twelve selected disciplines were contacted and invited to participate in the study. Participation involved initially the passing on of course reading and assessment materials, and then later being interviewed about these materials. A provisional analysis was made of the assessment tasks drawing on the same analytical framework used in the IELTS analysis. This analysis was also subject to processes of moderation.

### Academic staff survey

As a follow-up to the task analysis, interviews were conducted with the twelve participating staff. Prior to the interviews, a schedule of questions was sent to interviewees (see Appendix 2), along with a sample of IELTS reading test materials. The IELTS materials were selected so as to cover a representative sample of test tasks (see Appendix 2a).

The interviews were divided into three main phases, covering:

- general reading requirements on courses
- reading requirements on specific assessment tasks
- perceptions regarding the degree of correspondence between the academic reading requirements and those on the IELTS reading text.

The interviews were semi-structured and followed the procedure known as the 'discourse-based interview' (Odell, Goswami & Herrington, 1983). Such a procedure involves discussion with interviewees about specific text samples – in this case, the course materials provided by the lecturers and the sample IELTS reading test items. The interviews ran for an average of 1 hour. All interviews were audio-recorded, and transcribed. The main themes and ideas to emerge from our informants' commentaries are presented in Section 4.2.

The interview extracts presented throughout the report are in the main verbatim transcriptions of the interviews. In some instances, there has been some minor cleaning up of the text for the purpose of removing any extraneous features – false starts, hesitations, fillers and the like. As in Swales' (1998) study, the intention here was to make some small improvement to the readability of the spoken discourse of informants (p 26) while at the same time seeking to be faithful to the substance of their talk.

## 4 FINDINGS

The bulk of the research report is devoted to describing the findings of the study. In the first part of this section, findings from the IELTS task analysis are described. In the second part, we outline the findings from the academic task analysis and interviews.

### 4.1 IELTS reading tasks

The IELTS corpus compiled for the study consisted of a total of 13 tests, with each of these tests made up, on average, of three reading testlets (i.e. organised around three separate reading passages). In all, the total number of reading tasks across the corpus was 108, comprising 494 individual items.

A preliminary analysis found a variety of task types, with some featuring regularly in the corpus, and others less so. Table 4 lists the different task types identified, along with their relative frequencies. The figures in the left hand column show the total number of uses of each task type in the corpus, and those in the centre column, the total number of items under each of these types. Thus in the table, we can see for example, that the *True/False/Not given* format was used 23 times in the corpus, which included a total of 130 individual items (an average rate of 5.6 items per use of task type – see right hand column). Note that the order of frequency of task types in the table is based on the ‘total number of items’ – see centre column.

<b>Task type</b>	<b>No of occurrences of task type in corpus (% in bracket)</b>		<b>Total no of items under task type (% in brackets)</b>		<b>Average no of items per use of task</b>
1. True/False/Not given	23	(21)	130	(26)	5.6
2. Section-summary match	18	(17)	80	(16)	4.4
3. Gapped summary	14	(13)	78	(16)	5.6
4. Information-category match	12	(11)	61	(12)	5.1
5. Multiple choice	15	(14)	47	(10)	3.1
6. Short answer	8	(7)	33	(7)	4.1
7. Other (eg sentence completion, information transfer etc.)	18	(17)	65	(17)	3.6
Total	108 (100%)		494 (100%)		4.6

**Table 4. Task type by frequency**

In what follows, a description is provided for each of the task types identified, along with discussion of how each relates to the ‘level of engagement – type of engagement’ dimensions used for the analysis. Most space is devoted to describing and analysing the more frequently- occurring types. It is noted that in the corpus assembled for the study, the first three task types – *True/False/Not given*, *Section-summary match*, *Gapped summary* – accounted overall for more than half of the total items (57%). The category ‘Other’ shown at the bottom of the table included a range of additional task types, with each of these constituting less than 5% of items. No individual discussion is provided for these task-types.

**Type 1: True/False/Not given**

The most common task-type was *True/False/Not given*, accounting for about a quarter of all items (26% – see Table 4). In this format, test-takers typically needed to evaluate the truth status of summary information derived from the reading passage. In all cases in the corpus, this information was found to be in the form of a single sentence, and was normally related to a cognate sentence (or part of a sentence) from the reading passage. In those cases, where the true or false options applied, the sentence was typically constructed either as a synonymous (or near synonymous) paraphrase version of the related information from the passage, or was divergent in meaning in some way (eg. in a contradictory relationship). The exceptional case was the ‘Not given’ option, where the prompt was a proposition not included in the reading passage.

Sample 1:1 below is an example of the *True/False/Not given* task format, showing several sample items. Included in the sample are extracts from the associated reading passage showing relevant content for each item. Examples of both ‘true’ and ‘false’ formats are shown.

An alternative wording for this task-type noted in the data was to use *Yes/No/Not given* options rather than *True/False/Not given*. Thus, instead of writing true/false “if the statement agreed with / contradicted the information”, test-takers were asked to write yes/no. There would appear to be no substantive difference in these variable rubrics.

**True/False/Not given task**

Do the following statements agree with the information given in the reading passage?

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. It has been suggested that children hold mistaken views about the ‘pure’ science that they study at school.

**Relevant material from reading passage**

**Many studies have shown that children harbour misconceptions about pure curriculum science ...**

**Correct response:** TRUE

2. The plight of the rainforest has largely been ignored by the media.

**Relevant material from reading passage**

**Despite the extensive coverage in the popular media of the destruction of the rainforests, little formal information is available about children’s idea in this area.**

**Correct response:** False

**Sample 1.1: True/False/Not given task** (Sample 1:1 denotes that this is a Type 1 sample (i.e. True/False) and that this is the first sample of this type)

### Level of engagement

With respect to text ‘level’, it is noted that in the design of these tasks, the single sentence proposition contained in the prompt generally matches with a semantic unit of similar length in the passage, as seen in the first item above. This was not always the case however. In the second item above, for example, it is noted that whereas the prompt is a single sentence:

The plight of the rainforest has largely been ignored by the media.

the cognate information in the reading passage is realised in a smaller grammatical unit – a noun phrase:

Despite extensive coverage in the popular media of the destruction of the rainforests ...

The process was also found to work the other way, where the relevant information in the reading passage stretched over a larger grammatical unit than the prompt. In the following example (Sample 1:2), which shows ‘agreement’ between prompt statement and text, it can be seen that the relevant components of the prompt statement occur inter-sententially in the passage (shown in bold).

**Prompt statement:**

The approach to health during the 1970s included the introduction of health awareness programs.

**Relevant material from reading passage:**

The 1970s was a time of focusing on the prevention of disease and illness by emphasising the importance of lifestyle and behavior of the individual. Specific behaviours which were seen to increase risk of disease, such as smoking, lack of fitness, and unhealthy eating habits, were targeted. Creating health meant providing not only medical health care but health promotion programs and policies which would help people maintain healthy behaviours and lifestyles.

**Correct response:** TRUE

**Sample 1.2: Example of information occurring inter-sententially in True/False/Not given format**

Overall, however, it was found that most tasks of this type required engagement at or around sentence level. Accordingly in the analysis, such tasks were assigned to the more local end of the local-global continuum.

In performing such an analysis, one also needs to consider the additional component of the task – adjudicating on the ‘not-given’ option. This component suggests engagement at a different textual level. To establish whether certain information is or is not contained within a text requires some appraisal of the content of the whole text, and so for this component, the engagement is judged to be at a more global level.

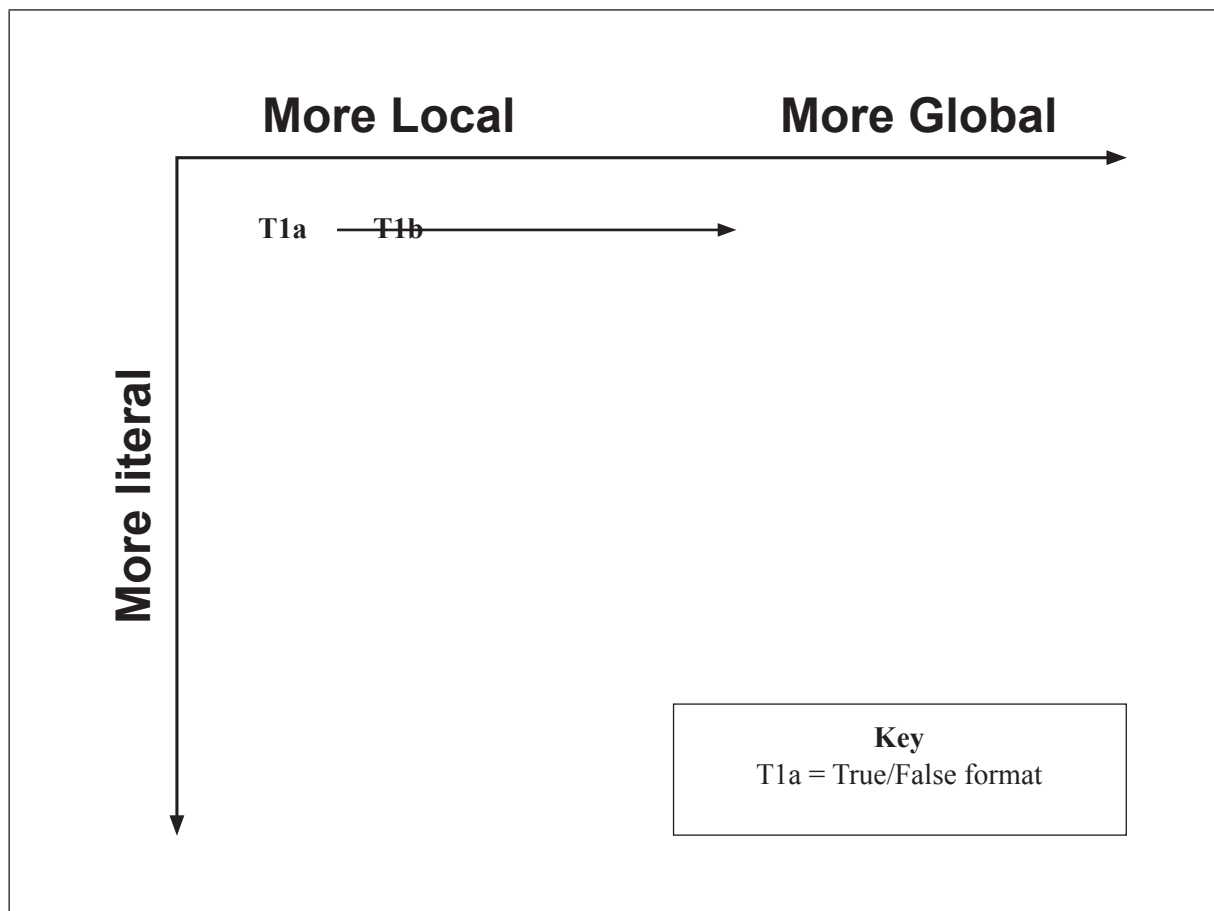
### Type of engagement

The type of engagement required for the completion of True/False/Not given tasks is one of establishing the semantic relationship between two discrete units of information (one in the prompt,

and a cognate one that needs to be located by the test-taker in the passage), and to decide whether the relationship is one of synonymy or non-synonymy (eg contradiction). The additional component of the task requires one to establish whether the propositional content of the prompt does in fact occur in some form in the reading passage – consideration of the ‘not-given’ option. Where this option applies, the task is thus one of detecting a lack rather than a presence.

The specific features of this task type – the need to establish the presence of certain propositional content in a text, and then to establish the relationship between this content and a variant version of it – suggest a strongly ‘literal’ engagement with reading material. Accordingly, this task type was assigned to the higher end of the ‘literal–interpretative’ continuum.

The preceding analysis gives the configuration shown in Figure 3 below (T1a refers to the ‘True/False’ component of the task, and T1b, the ‘Not Given’)



**Figure 3. Analysis of True/False/Not given task type**

### **Type 2: Section–summary match**

Section-summary match tasks were the second most common format, accounting for 16% of items in the corpus (Table 4). In this format, the task for test-takers was to match a section of the reading passage (usually a paragraph) with a statement that summarised the principal content of that section. An example of this format is shown below (Sample 2:1)

**Section – summary match**

Choose the correct heading for sections A-E from the list of headings below. Write the correct number i-x on your answer sheet.

**List of Headings**

- i) Contrary indications
- ii) Europe's Alpine glaciers
- iii) Growing consensus on sea level
- iv) Causes of rising sea levels
- v) Sea level monitoring difficulties
- vi) Group response to alarming predictions
- vii) The world 130,000 years ago  
etc

**Relevant section from reading passage:**

**RISING SEA LEVELS**

**SECTION A**

During the night of 1st February 1953, a deadly combination of winds and tide raised the level of the North Sea, broke through the dykes which protected the Netherlands and inundated farmland and villages as far as 64 km from the coast killing thousands. For people around the world who inhabit low-lying areas, variations in sea levels are of crucial importance and the scientific study of oceans has attracted increasing attention. Towards the end of the 1970s, some scientists began suggesting that global warming could cause the world's oceans to rise by several metres. The warming, they claimed, was an inevitable consequence of increasing carbon dioxide in the atmosphere, which acted like a greenhouse to trap heat in the air. The greenhouse warming was predicted to lead to rises in sea level in a variety of ways. Firstly heating the ocean water would cause it to expand. Such expansion might be sufficient to raise the sea level by 300mm in the next 100 years. Then there was the observation that in Europe's alpine valleys, glaciers had been shrinking for the past century. Meltwater from the mountain glaciers might have raised the oceans 50mm over the last 100 years and the rate is likely to increase in the future. A third threat is that global warming might cause a store of frozen water in Antarctica to melt, which would lead to a calamitous rise in sea level of up to five metres.

**Correct response: iv) Causes of rising sea levels**

**Sample 2.1: Section – summary match item**

It is noted that in this particular sample, the summary information is given as a ‘List of headings’ (i.e. Contrary indications; Europe’s alpine glaciers; Growing consensus on sea level etc), with the correct heading in this case being option iv) – Causes of rising sea levels.

A variation on this theme noted in the corpus was for the ‘headings’ category not to be used for the summary text, but instead for this material to be constructed in a more extended form. In these instances, prompts were designated ‘information’, as shown in Sample 2:2 below (emphasis added). Note that the relevant option for the reading passage extract is ii) Predictions regarding the availability of the synthetic silk.

### Section – summary match 2

Reading passage 1 has nine paragraphs, A - I

Which paragraph contains the following information?

- i) A comparison of the ways two materials are used to replace silk-producing glands
- ii) Predictions regarding the availability of the synthetic silk
- iii) Ongoing research into other synthetic materials
- iv) The research into the part of the spider that manufactures silk
- v) The possible application of the silk in civil engineering

Sample section from reading passage for option ii)

#### **SPIDER SILK CUTS WEIGHT OF BRIDGES**

#### **SECTION H**

At Du Pont’s laboratories, Dorsch is excited by the prospect of new super-strong biosilk materials, but he warns they are many years away. “We are at an early stage but theoretical estimates are that we will wind up with a very strong, tough material, with an ability to absorb shock, which is stronger and tougher than man made materials that are conventionally available to us”, he says.

### **Sample 2.2: Section – summary match item, using ‘information’ rubric**

The two samples provided above point to an additional variation in the Section – summary match format. This relates to the relative number of summary prompts and sections. Thus, for example, in Sample 2:1 above the number of summary prompts exceeds the number of sections, whilst in Sample 2:2, the ratios are reversed, with sections outnumbering prompts. This variation has implications for the process by which section and summary are matched up. In the former case (greater number of prompts), the process requires consideration of the text sections first, followed by identification of the appropriate summary prompt from the list given. In the latter case (greater number of sections), the sequence is reversed, with test-takers needing to begin with the summary prompt and then to match each of these up with the appropriate section of the text.

### Level of engagement

As the designated name of this task type indicates (i.e. *Section – summary match*), the level of engagement in this format is clearly at a supra-sentential level. In almost all cases in the corpus, the unit of text to be negotiated in the completion of tasks was the paragraph. Some variation was noted regarding the length of these paragraphs. In Sample 2.1 above, for example, the relevant paragraph is 10 sentences long (240 words); in sample 2.2 it is considerably shorter, running to only 2 sentences (67 words). In the whole corpus, the average paragraph length was 5 sentences. Overall for this task type, we can say that the level of engagement is on a more ‘global’ scale than for the *True/False* format analysed in the previous section (see Figure 4).

### Type of engagement

To complete *Section – summary match* tasks, test-takers need to be able to match up a putative summary of a section of text with the propositional content of this section. A feature of these summaries is their tendency to draw on a number of broad rhetorical categories eg cause and effect, comparison, prediction etc (Trimble, 1985). Thus, in Sample 2:1, we saw that the relevant rhetorical category for the section of text in question was ‘causality’ (*Causes of rising sea levels*); in Sample 2:2, this category was ‘prediction’ (*Predictions regarding the availability of the synthetic silk*).

The task for test-takers then, in many instances, is to be able to recognise the connection between the content of the designated section of text, and this broader rhetorical unit around which the summary prompt is structured. In the case of sample 2:1, this requires drawing a semantic link between the category of ‘causation’ in the prompt, and various ‘causal’ elements in the text – for example, i) certain key lexis (eg *cause, consequence, threat*) and ii) key structures (eg *Y would lead to a calamitous rise in sea level*). Similarly, in Sample 2:2, the task is to be able to recognise how key lexical items such as prospect, warning, as well as future time constructions – eg *we will wind up with a very strong, tough material* – equate to the rhetorical category of ‘prediction’. We note in passing the wide range of rhetorical functions used in the constructing of the summary prompts. The more prominent of these identified in the corpus are shown in Table 5, along with prompt samples for each category.

<b>Rhetorical category</b>	<b>Sample prompt</b>
Definition	<b>Definition</b> of health in medical terms
Role	The <b>role</b> of the state in health care The <b>role</b> of video violence
Importance/significance	Relative <b>significance</b> of trade and service industry The <b>importance</b> of taking notes on body language
Comparison	A <b>comparison</b> of the ways two materials are used to replace silk-producing glands
Causes/reasons	<b>Causes</b> of volcanic eruptions <b>Reasons</b> for the increased rate of bullying
Impacts/effects	The <b>impact</b> of the car on city development The <b>effects</b> of bullying on children
Changes	<b>Changes</b> to concepts of health in Western society
Problems / difficulties / failures	Sea level monitoring <b>difficulties</b> The failure of government policy
Merits/benefits	The relative <b>merits</b> of cars and public transport The <b>benefits</b> of an easier existence



<b>Rhetorical category</b>	<b>Sample prompt</b>
Reactions/responses	Group <b>response</b> to alarming predictions <b>Reaction</b> of Inuit communities to climate change
Methods/approaches	Holistic <b>approach</b> to health
Predictions	<b>Predictions</b> regarding the availability of the synthetic silk
Views/consensus	The <b>views</b> of the medical establishment Growing <b>consensus</b> on sea level
Suggestions / recommendations	A <b>suggestion</b> for improving trade in the future

**Table 5. Rhetorical categories used in summary prompts**

For this type of engagement, the moving between propositional content and summary, or what van Dijk and Kintsch (1983) call a mediating of ‘micro- and macro-processes’, is analysed as an ‘interpretative’ form of reading, or at least a more interpretative one than was seen in the *True/False/Not given* format discussed previously. The task for test takers in the *Section–summary match* format does not involve identifying a one-to-one correspondence between propositions as we saw in Task Type 1, but instead requires a ‘pragmatic’ understanding of material of the type identified by Taylor (2009 – see section 3.1). On the ‘literal-interpretative’ continuum on our grid, the generic Section–summary match task is therefore placed somewhat below the first task type (See Figure 4).

Regarding the degree of ‘interpretative-ness’, a variation in the design of Section – summary match tasks deserves brief comment here. Whereas most summary prompts were typically realised in a neutral, academic style; it was observed that in some instances a more idiomatic, ‘journalistic’ style of heading was used. Examples of this latter style are shown in Sample 2:3 below. (In this case the prompts relate to a reading passage describing the function of different types of security devices).

#### **List of Headings**

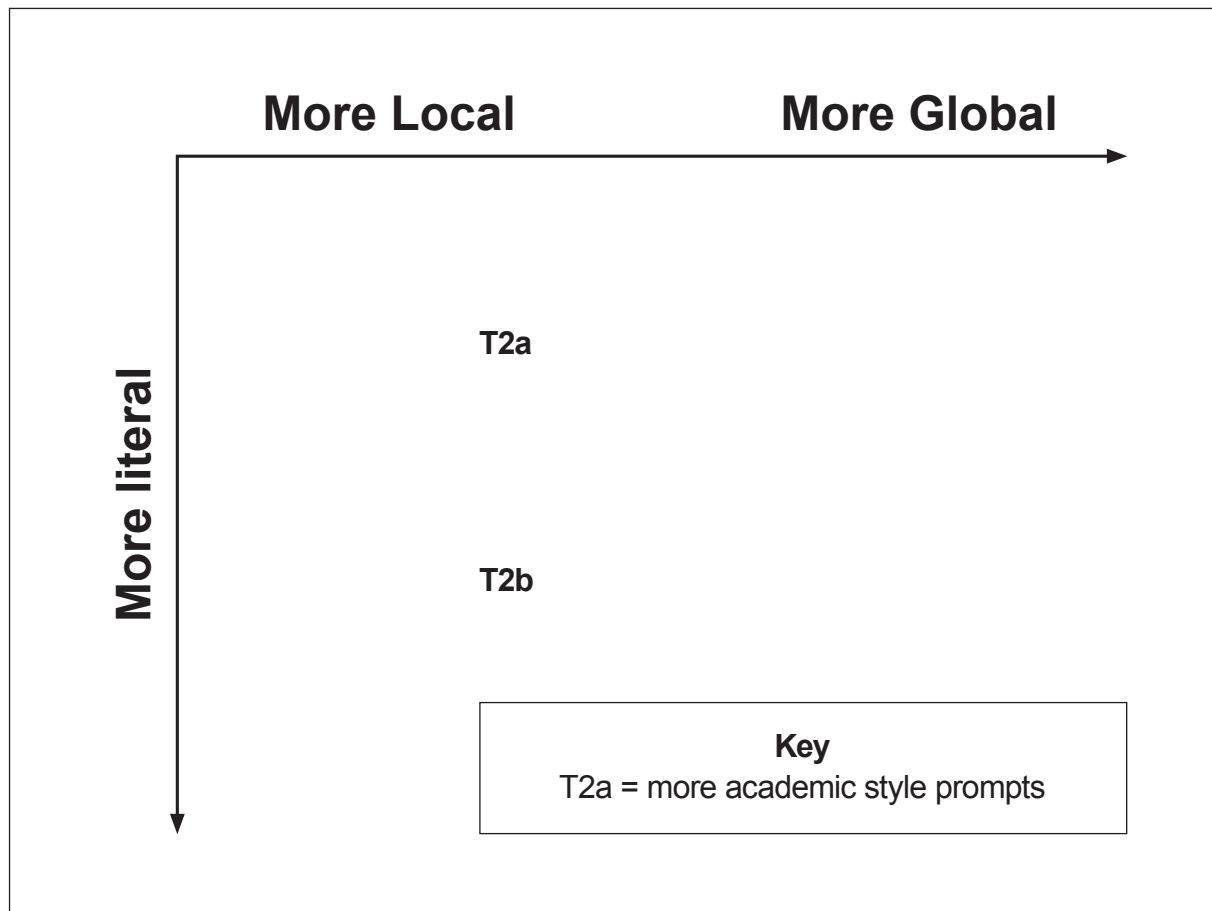
- i) Common objectives
- ii) Who’s planning what
- iii) This type sells best in the shops
- iv) The figures say it all
- v) Early trials
- vi) They can’t get in without these
- vii) How does it work?
- viii) Fighting fraud
- ix) Systems to avoid
- x) Accepting the inevitable

**Sample 2.3: Journalistic-style headings used in Section-summary match task**

These more journalistic-style headings are notable in the first instance for their lack of reference to the larger rhetorical units evident in many of the other prompt samples (eg cause, prediction etc). Other distinguishing linguistic features include the use of:

- a range of syntactic structures i.e. noun phrases (eg Fighting fraud, Common objectives); full sentences (eg This type sells best in the shops); question forms (eg How does it work?)
- more idiomatic phrasing or ‘prefabs’ (eg The figures say it all, Accepting the inevitable) and
- inexplicit pronominal reference (eg They can’t get in without these).

A number of writers have commented on the challenges generally involved in interpreting journalistic language (Nwogu 1991; Myers, 2003). It seems reasonable to suppose that dealing with less systematic categories of the type shown in Sample 2:3 is likely to require a greater interpretative stretch for the test-taker. In the grid shown in Figure 4, an attempt has been made to account for this task variety (see T2b).



**Figure 4. Analysis of Section – Summary match task type**

### Type 3: Gapped summary

The next most common format, by number of items in the corpus (16% of total items), was the Gapped summary. These tasks involved a different type of summary activity from that noted in the previous section. Here test-takers are presented with a continuous prose summary of a section of the reading passage from which key information/lexis has been removed. The task for test-takers is to draw on the reading passage to restore the omitted information.

We noted two alternative formats used for this task type: i) tasks where there was a bank of word/phrase options to choose from; and ii) where no options were provided. In the ‘no options’ format, test-takers are instructed to limit their responses to a maximum of two or three words from the passage. Examples of the two formats are shown in Sample 3.1 and 3.2. Relevant sections of the reading passage are provided for each sample.

#### Gapped summary 1

Complete the summary below.

Choose your answers from the box below the summary and write them in boxes 10-13 on your answer sheet.

There are more words than spaces, so you will not use them all.

The island will be partially protected from storms by ...(10)... and also by ...(11)... Further settlement caused by ...(12)... will be prevented by the use of ...(13)...

construction workers	coastline	dump-trucks
geotextile	Lantau Island	motorway
rainfall	rock and sand	rock voids
sea walls	typhoons	

#### Relevant section of reading passage:

##### AIRPORTS ON WATER

The airport, though, is here to stay. To protect it, the new coastline is being bolstered with a formidable twelve kilometers of sea defences. The brunt of the typhoon will be deflected by the neighbouring island of Lantau; the sea walls should guard against the rest. Gentler but more persistent bad weather – the downpours of the summer monsoon – is also being taken into account. A mat-like material called geotextile is being laid across the island to separate the rock and sand particles from being washed into the rock voids, and so causing further resettlement. This island is being built never to be sunk.

Correct responses:

10 = sea walls (either order possible)

11 = Lantau Island (either order possible)

12 = rainfall

13 = geotextile

#### Sample 3.1: Gapped summary sample, with options bank

### Gapped summary 2

Complete the summary of Paragraph G below. Choose NO MORE THAN THREE WORDS from the passage for each answer.

Write your answers in boxes 37-40 on your answer sheet.

A linguist can use a corpus to comment objectively on **37**..... . Some corpora include a variety of language while others are focused on a **38**..... . The length of time the process takes will affect the **39**..... of the corpus. No corpus can ever cover the whole language and so linguists often find themselves relying on the additional information that can be gained from the **40**.... of those who speak the language concerned.

#### Relevant section of reading passage:

##### OBTAINING LINGUISTIC DATA

A representative sample of language, compiled for the purpose of linguistic analysis, is known as a corpus. A corpus enables the linguist to make unbiased statements about the frequency of usage, and it provides accessible data for the use of different researchers. Its range and size are variable. Some corpora attempt to cover the language as a whole, taking extracts from many kinds of texts; others are extremely selective, providing a collection of material that deals only with a particular linguistic feature. The size of the corpus depends on practical factors, such as the time available to collect, process and store the data: it can take up to several hours to provide an accurate transcription of a few minutes of speech. Sometimes a small sample of data will be enough to decide a linguistic hypothesis; by contrast corpora in major research projects can total millions of words. An important principle is that all corpora, whatever their size, are inevitably limited in their coverage, and always need to be supplemented by data derived from the intuitions of native speakers of the language, through either introspection or experimentation.

#### Correct responses:

37 = frequency of usage

38= particular linguistic feature

39= size

40 = intuitions

### Sample 3.2: Gapped summary sample - without options bank

#### Level of engagement

Each item in the gapped summary tasks, it was noted, was focused on the locating of quite specific information. For example, in responding to items in Sample 3.1 above, candidates need to identify the various 'protective' measures that have been employed in the airport project discussed (sea walls, island, geotextile). On this basis, we would say that the level of engagement with the text is fairly local.

However, it was noted that in some Gapped summary tasks individual items could not be treated entirely in isolation, but instead needed to be considered in relation to the whole summary text, as well as to the relevant section of the reading passage. Thus, for example, in completing items 12 and 13 below (from Sample 3.1), one is not able to confirm the answer to 12 without looking further on in the reading passage to establish the likely response to 13.

Further settlement caused by ...(12 **rainfall**)... will be prevented by the use of ...(13 **geotextile**)...

Gentler but more persistent bad weather – the downpours of the summer monsoon – is also being taken into account. A mat-like material called geotextile is being laid across the island to separate the rock and sand particles from being washed into the rock voids, and so causing further resettlement.

We would say then that the ‘level of engagement’ for this task type relates to the span of text in the reading passage that is the subject of the summary. Some variation was noted in the length of these sections, ranging from summaries of a single paragraph from the original passage, to coverage of up to three or four paragraphs. This variation in engagement level is captured on the ‘local – global’ scale in Figure 5.

### Type of engagement

Whilst the level of engagement in the *Gapped summary* extends beyond the single proposition, the way in which test takers need to engage with material is arguably a fairly literal one. As was the case with the *Yes/No/Not given* format, the task for test takers involves, in essence, the matching of information from the reading passage with a paraphrased version of this information in the summary. Thus, the following items (taken from Sample 3.2) are completed by juxtaposing information in the item with corresponding information in the original passage.

#### Sample item 1

Some corpora include a variety of language while others are focused on a 38.

**Correct Response** = particular linguistic feature

#### Relevant section from reading passage

Some corpora attempt to cover the language as whole, taking extracts from many kinds of texts; others are extremely selective, providing a collection of material that deals only with a particular linguistic feature.

#### Sample item 2

The length of time the process takes will affect the 38 of the corpus.

**Correct Response** = size

#### Relevant section from reading passage

The size of the corpus depends on practical factors, such as the time available to collect, process and store the data: it can take up to several hours to provide an accurate transcription of a few minutes of speech.

The relatively ‘literal’ form of engagement suggested by *Gapped summary* tasks is indicated in our analytical matrix shown in Figure 5.

We note in passing that *Gapped summary* items can suffer from the problem of having items which it may be possible to complete (or partially complete) without referring to the original reading passage (Alderson, 2000). This is a characteristic however, only of the ‘options provided’ variant of this task type. In the sample items below, for example, we can see that certain items among the provided

options are semantically implausible within the information structure of the summary sentence, and so can be immediately discounted as possible answers (eg rainfall, typhoons).

The island will be partially protected from storms by ... (10) ... and also by ... (11) ...

construction workers	coastline	dump-trucks
geotextile	Lantau Island	motorway
rainfall	rock and sand	rock voids
sea wall	typhoons	

An additional dimension to this aspect were those cases where the provided options come in a variety of grammatical forms, and where some options could be automatically discounted on the grounds that they were syntactically anomalous in the summary sentence.

Alderson (2000) suggests that the problem with formats such as this is that they may be involved in testing constructs other than those that the instrument purports to test. Thus, with some of the Gapped summary tasks shown above, we might conclude that engagement with material is involved as much with grammatical competence or with principles of logical analysis, as with reading comprehension processes per se.

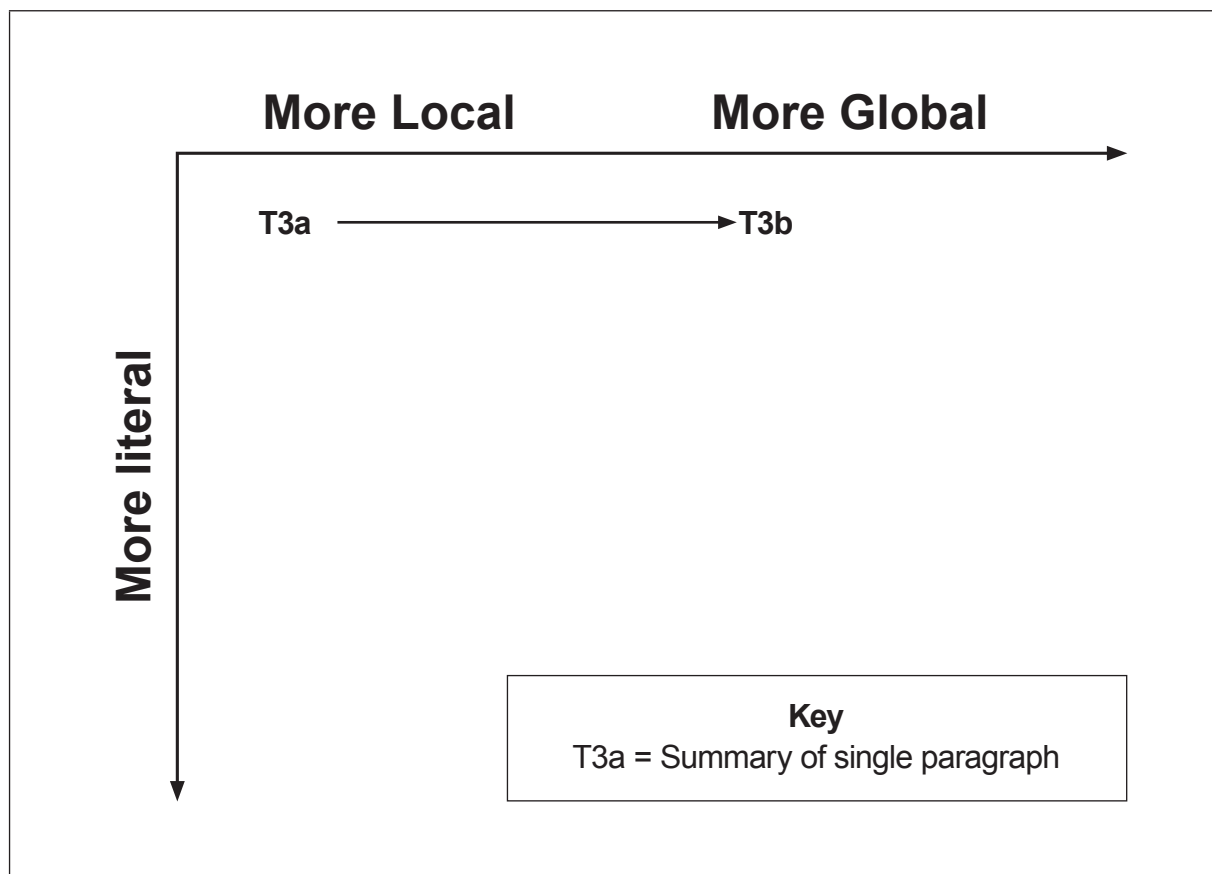


Figure 4. Analysis of Section – Summary match task type

**Type 4: Information-category match**

Information–category match tasks were one of the less frequently occurring tasks accounting for 12% of items (Table 4). Under this format, test-takers need to match information from the reading passage with a specific information category to be selected from a range of category options. The category-type used in the design of these tasks was found to be salient in some way in the reading passage, and which could be used as a basis for differentiating key information contained within it. Thus, in Sample 4.1 below, a task based on a reading comparing the design of different airports, the category of ‘airport location’ is used as the distinguishing element. Other category-types noted in the corpus of these tasks were places (eg cities); people (eg types of employees); time periods (eg decades).

**Question 1-5.**

Classify the following statements as applying to:

- A** Chek Lap Kok airport only
- B** Kansai airport only
- C** Both airports

Write the appropriate letter A-C in boxes 1-5 on the answer sheet.

**Sample statements:**

1. having an area of over 1,000 hectares
2. built in a river delta

**Sample 4.1: Information–category match item**

A specific type of *information–category match* task noted in the corpus was that which used individual scholars/writers as the category type. These were often used in tasks that accompanied reading passages consisting mainly of the attributed ideas or research findings of various individuals. The task for test-takers in this particular format then was to match a summary statement of a specific idea (or finding) described in the text with an individual scholar. Sample 4.2, based on a reading passage about endangered languages, is illustrative of this format.

**Question 5-9.**

Look at the following statements (Questions 5-9), and the list of people in the box below.

Match each statement with the correct person.

Write the appropriate letter A-E in boxes 5-9 on the answer sheet. NB You may use any letter more than once.

- A Michael Kraus
- B Salikoko Mufwene
- C Nicholas Ostler
- D Mark Pagel
- E Doug Whalen

**Sample statements:**

1. Endangered languages cannot be saved unless people learn to speak more than one language.
2. The way we think may be determined by our language.

**Sample 4.2: Information–category match – scholar as category**

**Level of engagement**

*Information–category match* items were generally found to be concerned with the locating of fairly specific information in the reading passage (eg size of airport in Sample 4.1). A feature of these tasks however, was that information often had to be retrieved from several different places in the text. Thus, for example, in the following item taken from the airport sample (Sample 4.1), test-takers need to identify whether the following statement concerning size of airport pertains to just one of the locations or both:

(Which airport) has an area of over 1,000 hectares

Completion of such an item thus necessitates engagement with several separate sections of the passage, as follows:

An island six kilometres long and with a total area of 1248 hectares is being created there. The new island of Chek Lap Kok, the site of Hong Kongs’s new airport, is 83% complete.

As Chek Lap Kok rises however, another new Asian island is sinking back into the sea. This is 520 hectare island built in Osaka Bay, Japan that serves as a platform for the new Kansai airport.

**Correct Response** = Chek Lap Kok airport only (**Option A**)

This particular characteristic of Information–category match tasks means that whilst engagement is generally at a local level, it is not as narrowly local as we have seen for other ‘specific information’ task types eg True/False/Not given (see Figure 6).



### Type of engagement

The airport example above suggests a highly literal engagement with reading material. In this case, the task for test takers is to identify specific information concerning the total area occupied by each airport site. A slightly less literal engagement is required arguably for the ‘scholar as category’ tasks (shown in Sample 4.2). In such tasks, the relevant ideas/findings of the scholar cited in the text are summarised in a relatively condensed form. The task for test-takers is to be able to link this condensed summary to the more extended version of the idea cited in the passage, as shown in the following example below.

#### Statement:

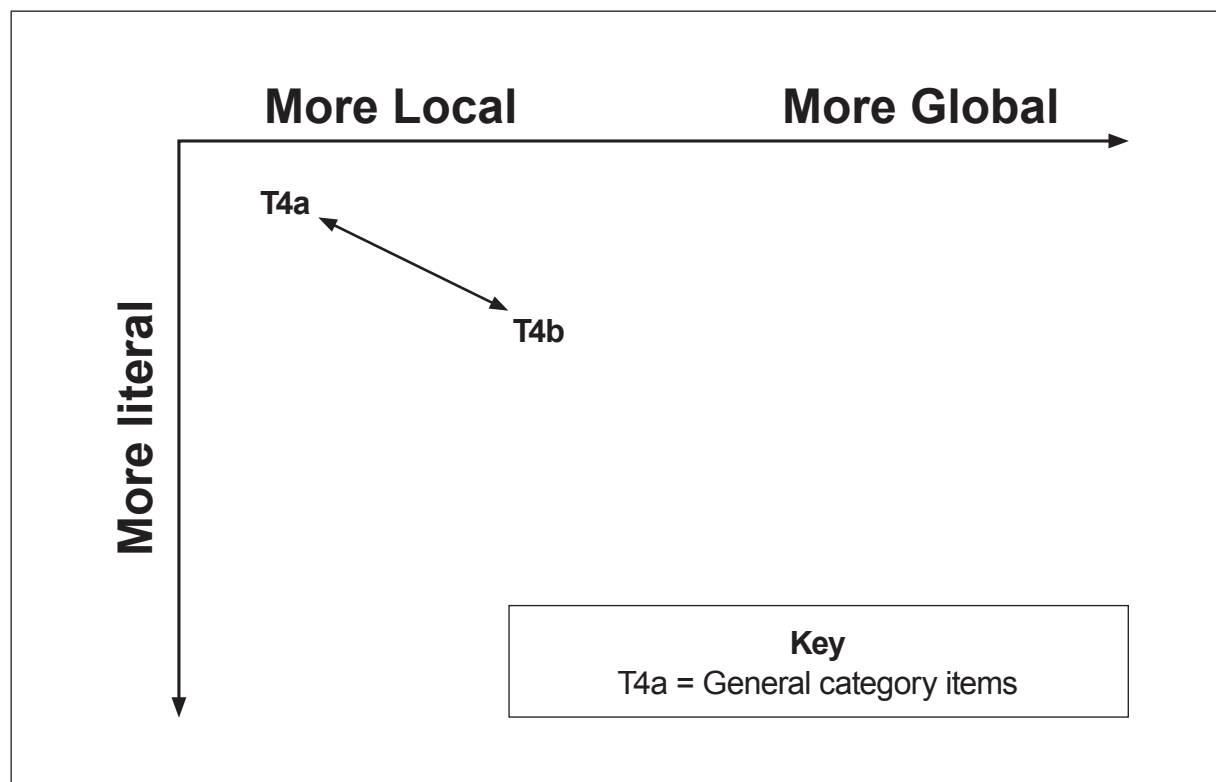
The way we think may be determined by our language.

Relevant section in reading passage:

There is mounting evidence that learning a language produces physiological changes in the brain. “Your brain and mine are different from the brain of someone who speaks French for instance”, Pagel says, and this could affect our thoughts and perceptions. “The patterns and connections we make among various conceptions may be structured by the linguistic habits of our communities”.

**Correct Response = Mark Pagel (Option D)**

Overall, the engagement with material in Information-category match tasks was concluded to be quite literal, but with some variation noted around the ‘scholar as category’ examples. An attempt has been made to capture this variation in Figure 6 below.



**Figure 6. Analysis of Information – category match task type**

### Type 5: Multiple choice

About 10% of items in the corpus used a standard multiple choice format, with almost all relying on a 4-option structure. For all items, test-takers were required to select a single ‘correct’ option. Sample 5.1 shows a range of multiple choice items related to a passage about the development of cinema.

#### Question 10-13

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

Write the correct letter in boxes 10-13 on your answer sheet.

- 10** When cinema first began, people thought that
- A it would always tell stories
  - B it should be used in fairgrounds
  - C its audiences were unappreciative
  - D its future was uncertain.
- 11** The writer refers to the film of the train in order to demonstrate
- A the simplicity of early films
  - B the impact of early films
  - C how short early films were
  - D how imaginative early films were
- 12** What is the best title for this passage?
- A The rise of the cinema star
  - B Cinema and novels compared
  - C The domination of Hollywood
  - D The power of the big screen

#### Sample 5.1: Multiple choice sample

#### Level of engagement

The *Multiple choice* task format in the IELTS corpus was found to be distinctive for implying no particular level of engagement with text. This is in contrast with the other task types considered so far. Thus, we saw for example that the *True/False/Not given* format was linked to engagement at a mainly sentential level; similarly the principal unit of analysis in *Section-summary match* was seen to be the paragraph. No such generalisation could be made however, for the multiple choice tasks in the corpus, with different items probing quite different levels of text. This is evident in the sample items above. In Item 10, for example, shown with the relevant section from the associated reading passage, the required engagement is at a more ‘local’, propositional level. (Correct response = D)

- 10** When cinema first began, people thought that
- A it would always tell stories
  - B it should be used in fairgrounds
  - C its audiences were unappreciative
  - D its future was uncertain.

**Relevant section from reading passage:**

When the Lumiere Brothers and other pioneers began showing off this new invention, it was by no means obvious how it would be used.

In contrast, Item 11 requires engagement with a more extended section of text – what in the passage is a full paragraph, as seen below (Correct response = B).

- 11** The writer refers to the film of the train in order to demonstrate
- A the simplicity of early films
  - B the impact of early films
  - C how short early films were
  - D how imaginative early films were.

**Relevant section from reading passage:**

One of the Lumiere Brothers' earliest films was a 30-second piece which showed a section of a railway platform flooded with sunshine. A train appears and heads straight for the camera. And that is all that happens. Yet the Russian film director Andrei Tarkovsky, one of the greatest of all film artists, described the film as a 'work of genius'. As the train approaches', wrote Tarkovsky, panic started in the theatre; people jumped and ran away. That was the moment when cinema was born. The frightened audience could not accept that they were watching a mere picture. Pictures are still, only reality moved; this must therefore be reality. In their confusion, they feared that a real train was about to catch them.'

Finally, the last question in this sample, Item 12, requires consideration of the whole reading passage – a text consisting of 10 paragraphs (Correct response = D).

- 12** What is the best title for this passage?
- A The rise of the cinema star
  - B Cinema and novels compared
  - C The domination of Hollywood
  - D The power of the big screen

Significantly, items of this latter kind – requiring test-takers to decide between different possible titles for a reading passage – were the only tasks found in the corpus that called for engagement at this whole text level. A total of five instances of this item type, all in a multiple choice format, were noted in the overall corpus, accounting for 1% of items.

From the examples above we can see that multiple choice items in the IELTS reading test probe a variety of textual units, ranging from the very local to the very global, as shown in Figure 7.

**Type of engagement**

As was the case with the level of engagement, IELTS multiple choice tasks in our corpus resisted any simple generalisation regarding the way test takers needed to engage with material. The sample items above suggest a variety of modes. Thus, Item 10, requiring identification of quite specific information (i.e. the perceived future of cinema), is clearly of a more literal type. In contrast, Item 12, which asks test-takers to consider how the contents of the whole text can be encapsulated in a single noun phrase title (i.e. 'The power of the big screen'), involves a more 'interpretative' engagement.

Between these two examples is the third sample item (Item 11), requiring test-takers to consider what point is made in the text through the citing of particular information (i.e. reference to the film of the train).

- 11 The writer refers to the film of the train in order to demonstrate
- A the simplicity of early films
  - B the impact of early films etc

Such an item, with its focus on the underlying rhetorical purpose of a span of text, was analysed as requiring a less literal form of engagement. The variety in the required form of engagement in these items is captured in Figure 7 below.

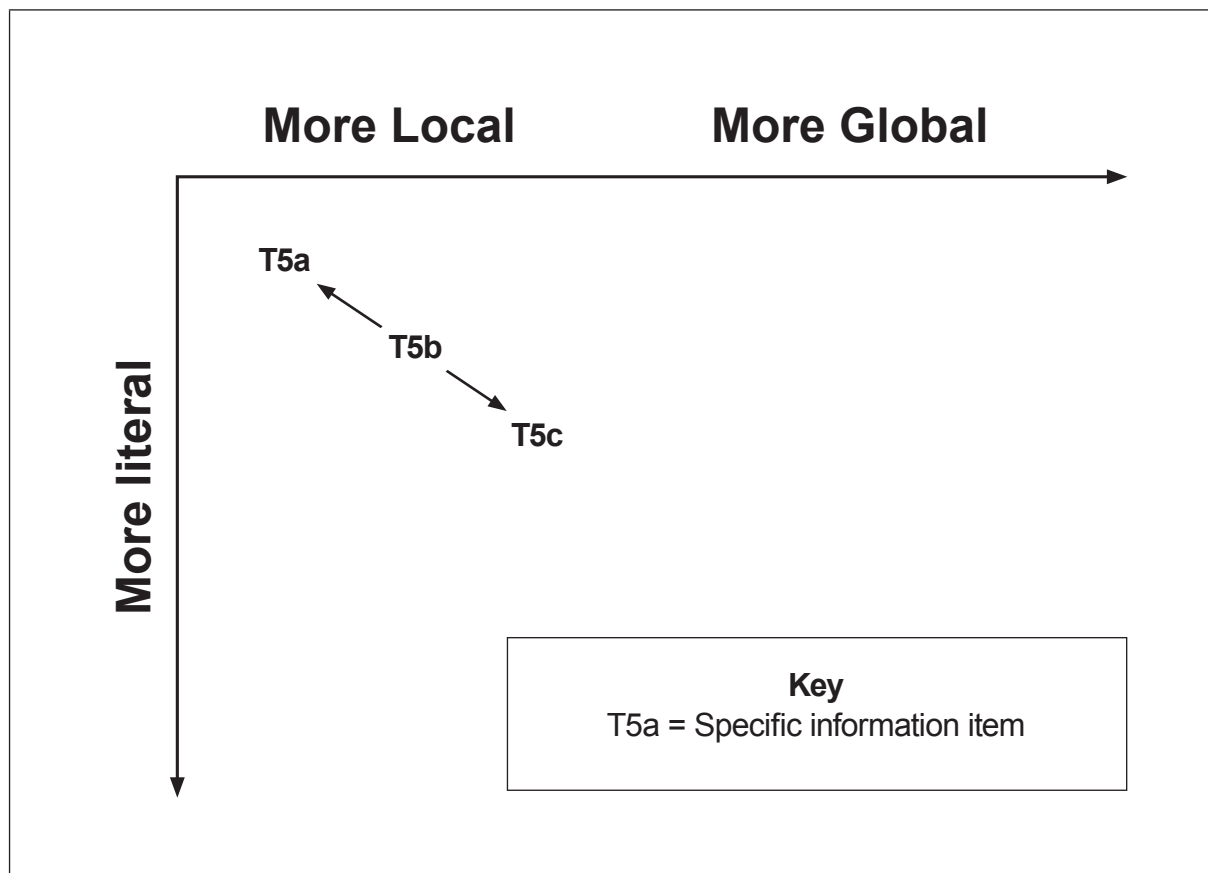


Figure 7. Analysis of Multiple Choice task type

### Type 6: Short answer

In Short answer tasks in the corpus (7% of total items), test-takers needed to locate quite specific information from the reading passage in response to basic wh-questions. A stipulation of responses in this format was that answers needed to be limited to no more than two or three words (or numbers), and that answers were composed only of lexis drawn from the reading passage. An example of this type, referred to by Bachman and Palmer (1996) as ‘limited production response’, is shown in Sample 6.1 below. The questions in this sample relate to a passage describing methods used to enhance the performance of athletes.

**Question 11 and 12**

Answer the questions below.

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

Write your answers in boxes 11 and 12 on your answer sheet.

- 11** What is produced to help an athlete plan their performance in an event?  
**12** By how much did some cyclists' performance improve at the 1996 Olympic Games?

**Sample 6.1: Short answer sample****Level of engagement**

Like a number of other 'specific information' task types we have seen previously (eg. *True/False/Not Given*; Gapped summary), engagement with the passage in Short answer tasks is at a local level, as shown in the examples below.

Question 11:

What is produced to help an athlete plan their performance in an event?

**Relevant section from reading passage:**

Well before a championship, sports scientists and coaches start to prepare the athlete by developing a 'competition model', based on what they expect will be the winning times.

**Correct Response** = 'competition model'

Question 12:

By how much did some cyclists' performance improve at the 1996 Olympic Games?

**Relevant section from reading passage:**

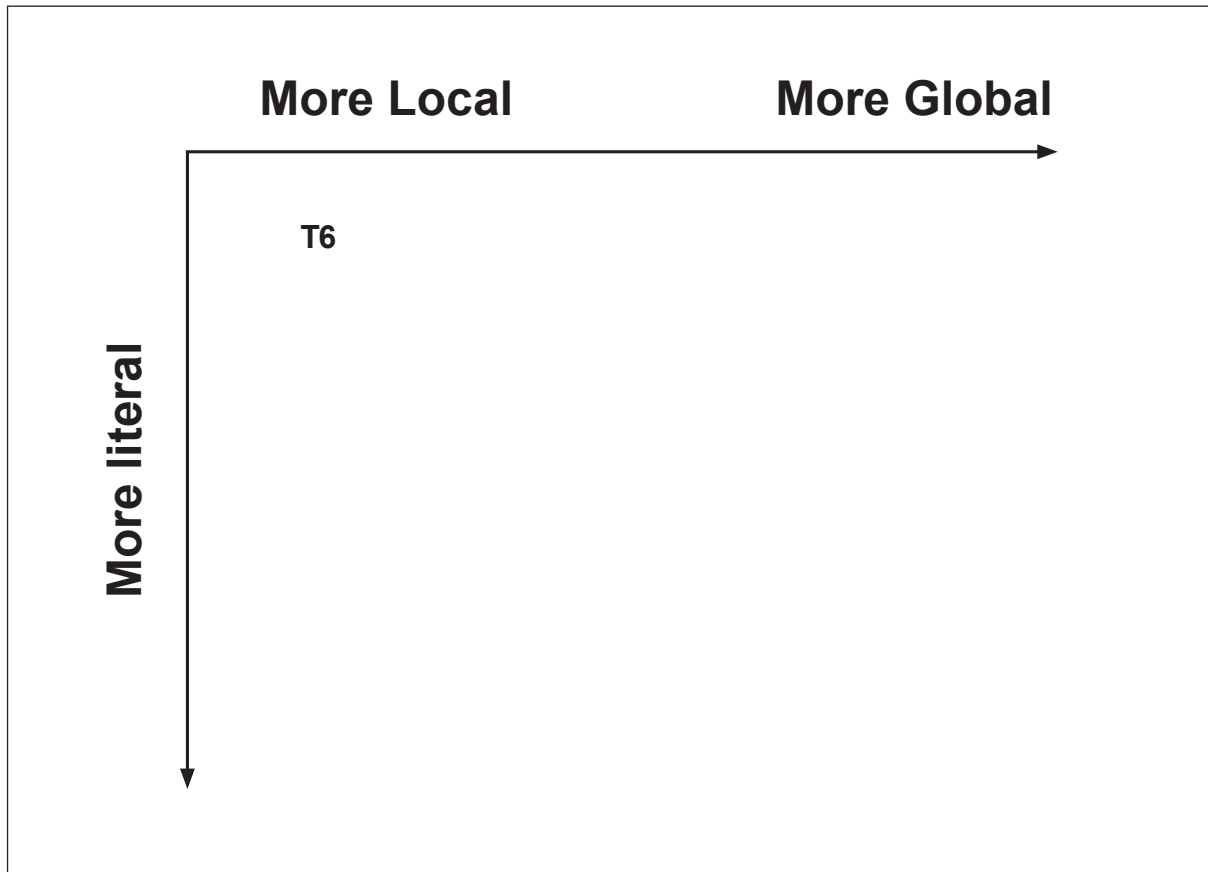
At the Atlanta Olympics Games in 1996, these [coolant jackets] sliced as much as two percent off cyclists' and rowers' times.

**Correct Response** = two percent

The requirement of these tasks – that is, to use a minimal number of words in relation to quite specific items of information – makes these tasks particularly 'local' in their orientation, as indicated in Figure 8.

**Type of engagement**

The *Short answer* format in IELTS reading, as we have seen, has a focus on quite specific items of information (eg the name of a specific performance-enhancement tool; the rate of improvement in a sports performance). We would say then that this involves a very basic form of text comprehension, and so this task type is located very much towards the literal end of our 'literal–interpretative' continuum. The allocated position of this task type on the matrix below suggests in fact that the short answer format constitutes the most 'literal' and most 'local' of all the task types considered so far.



**Figure 8. Analysis of Short Answer task type**

**Type 7: Other**

A number of other, infrequently-occurring task types were identified in the corpus, as follows:

- Sentence completion (4%)
- Information transfer (completion of table, diagram, flow chart etc) (4%)
- Information-paragraph match (2%)
- Information identification (0.8%)

All of these formats had a minimal presence in the corpus (<5%), and so were not analysed on an individual basis. It is noted that none appear to involve engagement of a distinctly different order from the patterns already identified in the previous task types.

**Summary of analysis of IELTS reading task types**

In summary, the analysis found that a majority of tasks in the IELTS corpus were of a distinctly ‘local-literal’ configuration, requiring mainly a basic comprehension of relatively small textual units (sentences, inter-sentences, paragraphs). It was noted that for a number of the more common task-types, the required engagement was in fact of a highly ‘local’ and ‘highly’ literal kind (i.e. *True/false/Not given; Short answer*). Other task types were found to be slightly less ‘local and literal’ in their orientation (i.e. *Section-summary match*), but were thought nevertheless to mainly inhabit the ‘local-literal’ region of our analytical matrix. The only items in our corpus that clearly traversed the ‘local-

literal' domain were certain multiple choice items that required an appraisal of the reading passage as a whole (eg items requiring the selection of a title for a reading passage). It was noted that the *Not given* option in *True/False* questions also required a more global engagement (i.e. in order to establish whether information is in fact present in a text). As was discussed however, items of this type arguably constitute a special case.

For the analysis overall, it does need to be acknowledged that the results are based on a limited sample of test tasks, and also one not deriving from actual live test materials. Assuming however, that the data used represent some approximation of current item-design practices at IELTS, we would argue, that the analysis provides at least a broad picture of the tests' overall construct– that is, a distinct orientation towards reading of a 'local' and 'literal' nature.

## 4.2 Academic reading tasks

In the second part of the study, investigation was made of the types of readings tasks required of students in undergraduate study in a range of disciplines. As discussed, a total of 12 discipline areas were investigated. This part of the study was informed by two different research methods: interviews with academic staff, and an independent analysis of reading-based tasks provided by these academics.

In what follows, we provide first of all a summary of the findings from the interviews, covering such issues as the quantity and type of reading prescribed on undergraduate courses, along with discussion of the way students are expected to engage with the material prescribed. This is followed by the analysis of the academic tasks. For this latter section, we draw on the framework used in the analysis of the IELTS tasks; that is to say, a consideration of the two key dimensions of the study – 'level' and 'type' of engagement with reading material.

### Findings from interviews

#### The importance of reading in university study

The first substantive finding from the interviews was that reading in some form was a requirement in all the subjects investigated. Some informants were at pains, in fact, to stress that it was the activity of reading more than anything else that underlay the educational processes in their subject area. Thus, the lecturer in Media Studies saw reading as providing the common thread for all the various activities on his course:

**MEDIA STUDIES:** The reading is just so essential, and it's just integral to the whole structure of my course. We set weekly readings which are of course what the lectures are based on, and then we use these readings as the basis for our discussions in tutes. ... And then later on hopefully [this material] will reappear in a different form when it comes to [students] writing their essays and assignments.

For the Linguistics lecturer, the development of key skills in reading was one of the more important objectives on her course.

**LINGUISTICS:** I am trying to encourage students to be critical thinkers and readers and reflect on the material that they have .... not just having an approach to learning where we transmit some knowledge and they absorb it. So being able to critically engage with texts is very much a primary generic skill on the course.

A number of other informants spoke of the effort they put in at the beginning of semester to persuade students to commit to doing the prescribed reading on their courses. For many, any adequate participation on academic programs (whether it be related to attending lectures, discussion in tutorials, participation in lab sessions) was contingent on students coming along to classes having read the relevant material. The lecturer from Communications, for example, drew attention in interview to the following ‘firm instruction’ to students contained in the course manual.

**COMMUNICATIONS:** I really hammer the following home to students at the beginning of the course (Referring to course outline document): “Please make sure you have done the reading before each lecture. The lectures and workshops will be based on the assumption that you have done this reading, and you will be expected to demonstrate some familiarity with the content”.

Whilst not emphasising reading to quite the same extent, some lecturers in the more technical (‘hard’) areas also affirmed the importance of this activity on their courses. The Architecture lecturer, for example, saw the development of good habits of reading as a key part of students’ professional training.

**ARCHITECTURE:** Even though we are a more technical subject, students need to appreciate that it is principally through reading that they will acquire key knowledge in the discipline. We’re aware of this not only for their university study, but for their ongoing development as professionals too ... I say to my students that good habits of reading will make them good professionals.

The overwhelming importance of reading in academic study was perhaps stated most emphatically by the informant from History:

**HISTORY:** What is very clear is that those students who do a lot of reading do better at whatever they are called upon to do than students who don’t do very much, and this seems to be the case with virtually all the academic work we set.

As we shall see later in the discussion of the interview findings, a number of staff reported a not-always-felicitous relationship between the expectations they had of reading, and the actual reading behaviours and attitudes that students brought to their studies.

### Quantity and type of reading prescribed on courses

Whilst there was general agreement among informants about the importance of reading, a fair amount of variation was noted regarding the amount and types of reading prescribed in specific subject areas. The differences observed here were mainly disciplinary ones, and perhaps not surprisingly, were found to divide mainly along the hard–soft distinction we have drawn between the disciplines investigated in the study.

Such differences were quite noticeable, for example, in the quantity of reading expected of students. In the ‘softer’ disciplines, informants spoke of the need for students to do ‘substantial’ reading on their courses, and generally to go beyond the set readings:

**MEDIA STUDIES:** There is a standard textbook. Every week there are particular pages of references they need to read, but then there are other important books, journal, magazine articles as well. To be properly informed in this subject, students also need to get into the habit of reading newspapers every day, and to be monitoring articles on media developments.



For the History lecturer, seeking out a range of sources and a variety of interpretations on a topic was an indispensable part of engaging with the discipline:

**HISTORY:** To properly engage with the subject is very much dependent on having a certain level of knowledge which ... is why we say to students you must do adequate reading otherwise you cannot respond to the questions [that we pose]. You might find a perfectly reasonable answer in a single book on this topic, but you're in no position to evaluate that unless you've read alternatives.

Other informants in the softer disciplines said they were quite precise to students about the quantity of materials that needed to be read each week, and the time that should be put in. The Linguistics lecturer, for example, said she advised students they should be reading the prescribed material from the textbook each week, and at least two relevant journal articles. The lecturer in Communications insisted to her students that they should devote at least 3-4 hours per week to reading in her subject.

In general, whereas the softer humanities disciplines required extensive reading, and from a range of different sources and genres, in the harder more technical areas reading was found to be less extensive, and mainly confined to the reading of the prescribed textbook in a subject:

**COMPUTER SCIENCE:** There is a textbook. Students are expected to read a chapter a week, but not every week, just for the first eight weeks or so. That's the first half of the textbook – which amounts to only about 150 pages for the course.

It was explained that in these harder disciplines the main purpose of weekly readings was to support the content of lectures.

**PHYSICS:** The textbook would be the main form of reading that students would do. We like students to be prepared for lectures and so we ask them to read the sections of the textbook that are relevant for a particular lecture.

Whilst in this case, the textbook material was intended to be read in advance of lectures, in other subjects, the purpose of textbook reading was mainly for review:

**ENGINEERING:** We have a textbook in the subject and after every lecture at the end of every lecture we put up the relevant pages of the textbook that they should read. So the idea is for them to read the PowerPoint slides, read the textbook and then write up their notes.

Several lecturers from other hard fields went on to explain that it was the nature of their discipline that the reading of texts was not always the only means of engaging with disciplinary knowledge.

**ARCHITECTURE:** Reading is important in this subject, though because of the nature of the discipline there are other literacies that come into play – visual literacy, kinesthetic literacy – to the extent that students are actually building things. Numeracy is also very important.

**COMPUTER SCIENCE:** We have a specialist type of reading in this subject which is related to the programming component. Students have to spend a bit of time reading other people's code, and this is a new type of reading for most.

### The nature of reading on courses

Along with variation in the quantity and type of reading material prescribed on courses, were perceived differences in the ways that students needed to engage with this material. Early piloting of the research suggested to us that it would not necessarily be a straightforward task for academic staff to expound at length on different types of required reading skills, nor indeed for them to be able to distinguish these skills in any substantive way. This was partly because the characterisation of such skills constitutes arguably an ‘insider’ educational discourse, one related to the study of academic literacy *per se*, and a discourse not necessarily readily accessible to academics working within their own disciplinary specialisations. As a way of facilitating discussion around this point in the interviews, it was decided to provide a list of possible reading skills (‘abilities’) drawn from the literature (Alderson 2000; Grabe, 1999), and to ask informants to comment on which of these they thought were relevant to study to their subject area (see below). This list seeks to capture some of the distinctions we have used in our analytical framework (i.e. type and level of engagement).

Ability to

- have a basic comprehension of key information in a text (LOCAL + /LITERAL +)
- summarise the main ideas in a text in one’s own words (GLOBAL + /LITERAL +)
- understand an idea for the purpose of applying it to a particular situation or context (LOCAL +/ INTERPRETATIVE +)
- understand the purpose for why a text may have been written (GLOBAL +/ INTERPRETATIVE +)
- critically evaluate the ideas in a text (GLOBAL +/ INTERPRETATIVE +)
- identify a range of texts relevant to a topic (GLOBAL +/ LITERAL +)
- draw on ideas from a range of texts to support one’s own argument (GLOBAL +/ INTERPRETATIVE +)

Given the relatively small sample size in interviews, the results are reported qualitatively (rather than quantitatively), with a focus on the key skill areas commented on by informants. Again basic differences were observed in the perceptions of academics across the disciplines. It was noted, for example, that those in the ‘harder’ disciplines thought skills towards the top of the list had the most obvious relevance to study in their subject area. The following are a sample of the responses from the more technical areas.

**ENGINEERING:** In Engineering I think we’re mainly concerned with basic comprehension (item 1) and summary skills (item 2). My sense of summary is students being able to convey the ideas back to us. So they need to understand the ideas and concepts, and report them back.

**PHYSICS:** I would be emphasising those skills more towards the top of the list. So we don’t really ask students to identify a range of texts relevant to a topic (item 6) nor draw on ideas from a range of texts to support one’s own argument (item 7). This is because students are not really making arguments at a first-year level. There are not so many things that are contestable at this level.

**BIOLOGY:** Well certainly basic comprehension and summarising of ideas (items 1 & 2), but understanding the purpose of why text is written is not important (item 4) Critically evaluate ideas (item 5), well only to a very limited extent – in most of first-year biology we

don't really challenge the ideas – we sort of present them as these are the ideas you need to know, and the last two are not important for us.

**ARCHITECTURE:** I think all have some importance, but apart from the first one (i.e. having a basic comprehension of key information in a text), they are not as important in this subject, as they might be in other subjects.

The main picture to emerge from these commentaries was that the important type of reading in these more technical disciplines was that related to basic comprehension of material. From these informants, it was generally felt that what was crucial in the first year of study in their disciplines was for students to come away with a good working knowledge of foundational ideas and concepts – and not to be spending too much time deciding whether such ideas were valid or not. A number pointed out that whilst more 'interpretative' forms of reading were clearly important in students' overall academic development, they had less obvious relevance in the early stages of training in the discipline. Among these more interpretative forms included in the list of skills, the least relevant, they thought, were those that involved 'critical evaluation' of material. As one informant explained it: "In the first year, we want students to accept certain things more-or-less as read, and to hold off with too much critical scrutiny of them at this stage".

It was explained however by several, that such a profile of reading skills was a specific feature of the first years of undergraduate programs, and that at more advanced levels in these disciplines, the other more interpretative types of reading had a much greater role to play.

**BIOLOGY:** As students move through the discipline I guess some of the things that were previously taken for granted start to become a bit more contentious – and we see these other skills like critically evaluating texts (item 5) coming more into play. Certainly students need to have a critical eye out when they read research articles, and the sorts of claims that are made in these.

The view expressed from the more humanities-oriented areas represented a clear contrast. For informants in these areas, all items on the list were thought to be important, and those at the lower end, particularly so. A common theme here was that in one's teaching, the more 'literal'-based skill areas were taken for granted to some extent, and that much of what students were expected to do simply assumed an understanding of basic concepts in the field.

**LINGUISTICS:** I think I make certain assumptions about those items further up the list, like being able to understand the ideas we present (item 1). That is, that students come to my course able to do these things.

**MANAGEMENT:** Having a basic comprehension (item 1), well obviously that's really important. If [the students] can't do that, the rest [of the skills] become a bit redundant.

For these academics, the focus was squarely on the more interpretative reading skills. Among those on the list, the idea of being critical of texts (item 5), and of being able to draw on multiple sources to support an argument (item 7) had particular resonance.

**LINGUISTICS:** The really important [skills] on the course are definitely critically evaluate (item 5) and drawing on ideas from range of texts to support argumentation (item 7). They are all important but those are the important formative skills for this subject. That's really the point that I try to get students to by the end of semester.

**MEDIA STUDIES:** All of the skills are important – having a basic comprehension, summarizing (item 1) is obviously important. On my course however, students being critical in their reading is absolutely essential (item 5). Students need to assess arguments, and part of this is identifying where arguments and ideas have been left out.

**MANAGEMENT:** the aim [on my course] is for [students] to develop an awareness of multiple types of sources, multiple viewpoints and to build confidence in their writing to draw on these different viewpoints in advancing their own view (item 7).

Among the more humanities-oriented areas, additional distinctions were observed at the individual discipline level. Our History informant, for example, attached special significance to students being able to “understand the purpose for why a text may have been written” (item 4). For him, such an ability related to a crucial part of the training students needed to undergo as novice historians – namely the ability to read and interpret primary source material.

**HISTORY:** working with primary source material is, I suppose, a specialist kind of reading in history, and we spend a lot of time on that. Students need to be able to see what’s surrounding a document, why it was created, what the author of the document is trying to achieve through it.

Additional variation was also found in the more applied disciplines. For informants in these areas, a key skill emphasised was the ability to draw on concepts in one’s reading for “the purpose of applying them to a particular situation or context” (item 3). Thus, the informant from the new applied Business discipline of *E-commerce* was keen to stress the essentially utilitarian nature of reading in the field:

**BUSINESS STUDIES:** The focus of E-commerce is very much about finding solutions to practical problems, and to develop electronic means to advance existing ways of doing things. Our sense of students learning is really about them grasping a concept, and then being able to apply it. Later on they might want to be critical of the concept, but in the first instance we just want them to focus on using it in some practical way.

In another of the applied disciplines, Communications, a similarly utilitarian conception of reading was emphasised. In this case, the focus was not so much on students being able to draw on conceptual resources for the purpose of solving real-world problems; but instead to draw on linguistic resources within texts for a different practical purpose – namely, the development of their writing. The lecturer in the subject explained this particular type of reading thus:

**COMMUNICATIONS:** Students need to write in a variety of genres, say for example the book review, and we get them to look at samples of these genres as a resource for their own writing.

**INTERVIEWER:** So how would you describe the nature of the reading that students have to do in this situation?

**COMMUNICATIONS:** Well, I tell them in the beginning that they are not reading so much as consumers anymore, but that they are reading it like a carpenter might look at a chair – not so much to sit in it, but to see how it is put together.

### Perceived changes in students' reading practices

A final area covered in this summary of interview findings is informants' perceptions of students' reading practices on their courses. Interestingly, this was an issue not directly probed in the interviews. As has been stressed, the primary construct that informed the research was the notion of 'task demand', and so it was not an imperative of the study to investigate issues of actual student behaviour and performance. We found however, that these were key issues for informants, and ones that many were keen to air in the course of our discussions. In short, concern was expressed by a number of informants – and indeed a degree of disdain by some of these – about the lack of commitment shown nowadays by students towards reading on their courses. The following are some representative comments on this issue:

**LINGUISTICS:** It is a constant struggle to get students to do the reading these days. So for example in the tutorial that I had earlier this week, I asked what I thought was a really self-evident question, and the answer was very clearly in the second reading from the week. Nobody got it. Literally nobody had even read the article.

**COMPUTER SCIENCE:** At the end of this semester we asked for a show of hands of how many of the students actually had a copy of the textbook and it was a slightly depressingly low proportion. So I think quite a lot of students [aren't] actually doing the reading.

**MEDIA STUDIES:** I've told you about what we expect, but one can't avoid mentioning what actually happens. So reading in fact has become a major problem. Students are just doing less reading than they've ever done before, and that would be local students as much as international ... Many complain that the standard of textbook is just too difficult ... We feel though that we have to resist dumbing these things down. It's a university textbook we prescribe; we can't go looking at something fit for secondary level.

Whilst the last informant, from *Media Studies*, thought vigilance was necessary to avoid any 'dumbing down' of requirements, others thought the pressures to scale things down – both the quantity and level of reading – difficult to resist at times. The *Management* lecturer, for example, described how the subject he taught had been forced in recent years to take a less rigorous, less 'literary' turn in response to changing student cohorts and preferences:

**MANAGEMENT:** I've taught the course for about five years. I took the course over from two of the older academics here who are probably well ... more literary in their take on the world. One was a trained historian; the other was an economic historian. But I've had to tone down the volume of reading and that's in response to the changing student mix and changing student behaviour. I have probably shifted it more to use of business press material, less academic material.

More ominously, another lecturer thought that on some programs, the reading load had to be reduced partly in response to certain pressures felt through formal processes of student review and feedback:

**ENGINEERING:** Students only have to read the textbook and the PowerPoint slides to be successful in this subject nowadays. And this is a lot to do with student expectations, because we have found that they are very reluctant to do extra reading. And again this has to do with quality of teaching. If you give them a lot of reading, you are going to get really knocked back on quality of teaching scores.

Mention was made in interviews of some of the possible factors underlying this apparent decline in reading, including a general sense of students disengaging from study, financial pressures, time spent in employment and so on. Another clear factor identified – one related to actual literacy practices – was students’ increasing use and reliance on digital resources, and the effect this seemed to be having on the way they engage with textual material. The view generally was that a lot of online working with material was not encouraging of good reading practices:

**MEDIA STUDIES:** There is a lot of material now that students access that they just typically browse. It’s a kind of trawling for information. They just don’t read this stuff in any serious and sustained way.

Concern was expressed too that access to the limitless resources available on the web has resulted in some students being less-than-judicious in their use of materials:

**COMMUNICATIONS:** Research is showing that the evaluation and management of material that’s coming out over the internet is the biggest issue. And some students do not have particularly well-developed evaluation skills.

Some thought however, that the considerable changes in the way that information is now accessed had major implications for teaching, and that there was need to address these issues positively with students. Several for example spoke of the importance of developing students’ ‘information literacy’ and ‘media literacy’ skills.

**HISTORY:** The web has been a boon to the study of history. But we have to help students to be a bit discerning about their use of websites. .... We actually have discussions in tutorials. How can we tell whether this is a reliable site or not? So its evaluation of who is producing this, in what context, and for what purpose.

**MEDIA STUDIES:** What I try to teach [students] is to get them to be selective in their reading of the media ... So, I want them to understand the context of what [they] are reading, and also the legitimacy of what they are reading.

For at least one informant, the lecturer from Linguistics, to resist such developments was really a futile pursuit, and that the onus was very much on the universities to adapt to emerging literacy practices. For her, the shift from a print-based academic culture, to a more digitally-based one, posed a much greater challenge for many academics than for students.

**LINGUISTICS:** So I think we in the university have to learn more about student’s reading habits and practices and to rethink our assumptions. And we are probably going to have to make big adjustments about what it is that students do these days when they study.

### Findings from task analysis

Along with the conducting of interviews with staff about reading requirements, the research also collected samples of the various academic tasks set by these staff on their courses. Each of the tasks compiled in the corpus involved a reading component of some kind, and included the following formats:

- 1 weekly exercises and questions, set principally for the purpose of lecture and tutorial preparation/review

- 2 questions and tasks set in examinations, tests etc
- 3 assignment tasks, set as part of the major assessment requirements on courses.

In the analysis that follows, we draw on the same framework used to analyse the IELTS reading task types, involving an allocation of each task to a category in the schema. This was done both through processes of moderation between members of the research group, and also through reference to informants' descriptions of these tasks in interview. The findings of this section are organised around the four analytical categories, viz:

- local – literal
- global – literal
- local – interpretative
- global – interpretative

### LOCAL + LITERAL

In our corpus, we could find many examples of tasks that required a mainly 'local – literal' form of engagement with reading material. It is significant to note also that such examples were present in *all* the discipline areas covered in the study. The following two items are presented as samples from the corpus fitting this 'local – literal' configuration.

Read Chapter 8 of *The Gecko's Foot – Bio-inspiration: Engineered from Nature*, and take notes around the following questions.

1. What is meant by the concept of 'ordered crumpling'? Why, according to the author, is this concept useful?

#### Sample A1. Weekly exercise task - Architecture

(A1 denotes that this is an Academic task, and this is the first task.)

2. Postmodernists basically believe that:

- |  |                          |
|--|--------------------------|
| a) All texts are closed                        | b) Most texts are closed |
| c) All texts are open                          | d) Most texts are open   |
| e) All of these depending on the circumstances |                          |

#### Sample A2. Exam question - Media Studies

The first example, Sample A1, is from a set of exercise questions, prescribed as part of the weekly readings in the *Architecture* subject. This task, as the rubric states, requires students to produce a short summary ("notes") of a specific concept from the readings ('ordered crumpling'), along with an account of its 'usefulness' – as described by the passage's author. This focus on explication of a single concept clearly fits with a more 'local and literal' engagement with material. Such interpretation was confirmed by the lecturer, who described the purpose of "basic comprehension-style" tasks like this as being "to help students come to grips with material and to get a grasp of key concepts in the subject". The lecturer went on to explain that in her subject such concepts were then typically drawn on in some practical way – in this case, it was to explore in tutorial discussion the issue of "how certain patterns in nature can be applied to design".

The second sample, Sample A2, is a multiple choice question set in an exam context. The lecturer in this subject (*Media Studies*) explained in interview, that this particular question, related to “quite specific information” contained in the textbook (and also covered in lectures), and would involve students, as he explained, recalling “basically factual information about one of the core beliefs of this intellectual movement” (Correct response: C). Other multiple-choice questions on the exam in this subject, the lecturer explained, followed a similar format, requiring the same literal recall of key content covered on the course. It was noted however, that the exam paper in *Media Studies* also included other question types (eg short essays), the specifications of which, as we shall see, fit with other configurations on our matrix.

The sample tasks we have described cover two of the task formats noted above, namely:

- 1 weekly exercises and questions, set principally for the purpose of lecture and tutorial preparation/review.
- 2 questions and tasks set in formal examinations

It is interesting to note that virtually all the ‘local – literal’ examples in our corpus relate to these two formats; that is to say, tasks set principally for the purposes of either inputting key concepts and knowledge during a course of study, or else for the testing of students’ acquisition of these concepts and knowledge at the end of a course (or course segment). We discuss each of these two formats briefly below.

### Weekly exercises and questions

A number of lecturers stressed the importance of weekly exercise tasks as a way for students to acquire (or to have reinforced) key content in their subject area.

**COMPUTER SCIENCE:** We set questions each week which are generic technical questions, and involve encyclopedia-style retrieval of the relevant information from the textbook and lecture.

As explained by some, such questions do not usually need to be answered formally, but often involve the taking of notes (or ‘scribbles’) from the set readings, that students would then bring to class to discuss:

**MANAGEMENT:** In terms of reading for the tutorials, there is definitely a set of study questions each week ... where the students can scribble things down and that will form part of the discussions of the tutorials. So those questions are guiding them through the reading, comprehension questions to make sure they have understood the reading.

As suggested in the discussion of Sample A1 above, a focus of these comprehension-style questions is often on key concepts in the discipline. This was a point taken up by a number of informants.

**BIOLOGY:** Students have a handbook of tutorial questions that they get at the start of semester. Their purpose very much is to filter out what is most important about a specific concept. So in their reading they have to be able to look for the concepts and fish out the most crucial points.

The lecturer in Physics explained that part of this ‘conceptual’ understanding involved students recognising how terms within a discipline often carry with them quite distinctive meanings, ones that differ from a term’s everyday usage:



**PHYSICS:** In physics [like other disciplines], there are particular conceptual challenges that students have, in part related to the fact that we do use some words which have a very precise meaning in physics that is different from how they are used in normal everyday usage.

This lecturer went on to explain that often the study questions she set were intended to draw out these distinctive meanings. Examples she gave were the terms ‘heat’ and ‘temperature’

**PHYSICS:** If we actually unpack a word like ‘heat’ and use it in the precise way it is intended then we are actually thinking about some energy being transferred from one object to another because their temperatures are different. That’s not normally how one would use the word, and there are lots of words like that. So words like ‘heat’, ‘temperature’ have really precise meanings and we try to bring this out in the questions we set for students.

Samples A3-6 show questions from a variety of disciplines, all which have this focus on facilitating students’ understanding of specific discipline-based concepts. It was noted, interestingly, that in the format of many such questions, the relevant concept with which students need to engage is often signaled through the use of inverted commas or italics.

Read Section X of the textbook on thermal energy:

Which has the greater amount of thermal energy, an iceberg or a cup of coffee? If the coffee and the iceberg were placed in contact, what would happen? Use this context to explain clearly the terms temperature, heating and thermal energy.

**Sample A3. Weekly exercise question - *Physics***

What is ‘currency risk’? Why does international business inevitably involve this type of risk? How can firms reduce these risks?

**Sample A4. Weekly exercise question - *Management***

What is the difference between the ‘Lossy’ and ‘Lossless’ types of media compression?

**Sample A5. Weekly exercise question - *Computer Science***

Explain what a ‘speech act’ is, and give several examples.

**Sample A6. Weekly exercise question - *Linguistics***

**Exam and test questions**

Exams and tests were set in almost all of the subjects investigated, the only exception being Communications. The reading requirements for exams, as explained by a number of informants, mainly involved reviewing the content of lecture and textbook material:

ENGINEERING: The exam is basically about testing students' understanding of key ideas and terms. As preparation for the exam [students] would need to look at the PowerPoint slides of the lectures and reread the relevant sections of the textbook.

Among the items set on the exam/test papers was a sizeable proportion requiring a mainly 'local – literal' engagement with material. These included 'Short answer' questions, very much in the vein of the 'study'-type questions discussed in the previous section – that is, questions requiring short summaries of key concepts and ideas. Samples of such question-types are shown below (A7-10).

What assumption about 'savings behaviour' is made by the Solow Swan model?

**Sample A7. Short answer exam question – Economics**

Briefly explain Fukuyama's End of History thesis.

**Sample A8. Short answer exam question – History**

What is meant by the concept of 'value configuration'?

**Sample A9. Short answer exam question – Business Studies**

What is the hypodermic model of media effects?

**Sample A10. Short answer exam question – Media Studies**

Also used in the exam/test samples collected were a range of more objective, 'closed' formats. The most common of these was *Multiple Choice*; *True/False*, and *Gap Fill* formats were also noted. Examples from the corpus of each of these formats are shown below:

An etic description of a cultural practice:

- A. reflects the culture member's own understanding
- B. focuses on sound differences
- C. takes an outsider's perspective
- D. requires a prolonged engagement and "going native"

**Sample A11. Multiple choice question - Linguistics**

The statements A-D are either correct or incorrect. Record whether the statement is Correct by entering 11 or Incorrect by entering 12.

- A The binomial name of a species is by convention printed in italics
- B. Phylogeny is the study of the life cycle of an organism
- C. Slime moulds get their name from the mucilage trails they leave behind
- D. Diatoms and dinoflagellates are important photosynthetic plankton contributing greatly to the productivity of the oceans.

**Sample A12. True/False question - Biology**

In a Keynesian short-run model of a two sector economy, suppose that savings are greater than planned investment. This will result in \_\_\_\_\_ in inventories.

- |                         |                         |
|-------------------------|-------------------------|
| a) unplanned, increases | b) unplanned, decreases |
| c) planned, increases   | d) planned, decreases   |

**Sample A13. Gap fill question - *Economics***

Such formats, on the face of it, bear a close resemblance to some of the IELTS reading task types we have seen. One needs to be mindful however, of an important difference in the processes involved in completing tasks in these two contexts. On the IELTS test, test-takers have access to the information required to complete tasks i.e. as information contained in the reading passage. This is not normally the case in subject-based examinations, where students are not usually permitted access to reading materials during the exam. Thus the two contexts rely arguably on different cognitive processes – in the IELTS test, these mainly involve the locating and comprehending of specific information to be found in the reading materials; in the examination format there is a distinctive “memory and recall” component to the required form of engagement.

Such differences relate very much to the very different purposes of reading in the two domains. In a reading test such as IELTS, the focus is more on assessing the extant skill level of tests takers; in university exam items, such as in the examples we have seen above, the focus is less on skill, and more on the extent to which students have acquired key knowledge in the discipline area. In short, in the university context, content is salient; in a language testing context, it is largely incidental. The implications of this difference for test design are discussed in more detail in Section 5 of the report.

**GLOBAL AND LITERAL**

According to the study’s analytical schema, tasks of a ‘global-literal’ configuration are those that require primarily basic comprehension of textual material (literal) in relation to larger textual units – i.e. whole texts as well as multiple texts (global). It was noted that almost all tasks in our corpus fitting these patterns were *assignment tasks* (i.e. out of class tasks, set as part of the major assessment requirements on courses). Most, but not all, came from the ‘softer’ humanities disciplines. Examples of such task types are presented and discussed below.

**Summary tasks - single text**

We saw in the previous section (‘local–literal’), a number of tasks requiring the summary of a single concept (eg ‘thermal energy’ in Physics; ‘speech acts’ in Linguistics). Tasks requiring the summary of a single *whole* text were relatively uncommon in the corpus. The following from the History subject, involving here the summary of a book chapter, was a rare example.

**Secondary source summary**

One of the most important skills in conducting research in history is the ability to comprehend a particular text, and then to summarise its major arguments and conclusions in your own words.

For this exercise, you need to read chapter X of *The path to genocide* by Christopher Browning, and then write a 500 word summary.

**Sample A14. Assignment task – *History***

In setting this task, the History lecturer explained that it was important for students not just to give “some simple blow-by-blow recapitulation of the text”.

**HISTORY:** What I stress to students is that they need to read chiefly with the purpose of identifying the author’s main argument. And the other part is then to identify the evidence the author presents to support this argument. All this needs to come out in their summaries.

### Summaries of arguments – multiple texts

A more common type of summary task was that requiring the summary of a number of texts, as in the following two samples from Management and Media Studies. As in the History example above, the main focus of these tasks was for students to give an account of arguments contained within texts. In both the sample tasks below, a key component is for these arguments to be seen as part of a larger debate – in the Management task (A15), it is one about how much globalisation has progressed as a phenomenon; in the Media Studies task (A16), it is a debate about different policy approaches to media ownership.

#### The globalisation debate

In no more than 800 words, address the following question:

What are the arguments for and against the idea that ‘the world has become flat’ in recent years?

Please base your discussion on readings for Week 3 and further research. You must meet the referencing requirements listed below.

Business in the global economy

#### Sample A15. Assignment task – *Management*

#### Media ownership

What are some of the basic for and against arguments in the debate about abolishing the cross media ownership AND foreign ownership laws in Australia? Refer to at least 4 primary sources in your response.

#### Sample A16. Assignment task – *Media Studies*

Both lecturers were keen to stress that such tasks were really focused on developing the skill of accurately representing the views of various writers on an issue. As the Management lecturer explained it:

**MANAGEMENT:** Students often struggle in just seeing what the main points of a piece of writing are, to have the confidence to say: “Yes, this is what this person is saying, this is what they’re arguing”.

This lecturer went on to explain that in such tasks, students were sometimes under the misapprehension that they should also be expressing their own view in the debate. For this lecturer, the ability to provide a clear summary of ‘the arguments of others’ in the field was a basic foundational skill, one which was then built on later in the course.

**MANAGEMENT:** One thing [students] struggle with is that it's actually a summary task. I'm after a set of arguments. I'm not after [their own] opinions which can throw them a little. We tell them that comes in later.

### Summaries of research findings

A different version of the summary task was that focused not on the identification of the arguments contained within expository texts, but rather on the purposes and findings contained within empirical research studies. In Sample A17 below, an essay set in the Biology subject, a major component of the work for students is to “summarise a range of recent studies” concerned with the search for a cure for malaria.

#### Malaria

Why do botanists study the malarial parasite (*Plasmodium*) and how do they hope to find a cure for this disease? In your response, you should summarise a range of recent studies, focusing on the kinds of drugs currently being tested, and why.

#### Sample A17. Essay task - *Biology*

Another example of a task requiring students to document a series of research findings is the following from the Linguistics subject (Sample A18). In this instance, students need to conduct their own research, but first of all to place their study in the context of previous work done in the area, involving “a summary of earlier studies in the subject”.

#### Speech act research

The purpose of this assignment is for you to collect and analyse speech act data. You will be expected to design a brief Discourse Completion Task (DCT) which elicits apologies or requests.

Write your paper with the following sections (including inter alia):

Introduction (about 400 words): Talk about the speech act you're investigating, and the role of politeness for realising it. Define your terms, and summarise some of the earlier studies on the subject (you may use your reader and lecture notes for this).

#### Sample A18. Research project task (extract) - *Linguistics*

### LOCAL + INTERPRETATIVE

Our category of ‘interpretation’ is a broad one and, as explained earlier, has been defined for the purposes of the study as “those forms of engagement with reading material that go beyond a literal comprehension of a text’s propositional content”. In this sense, as we pointed out, it is a more reader-focused than text-focused form of engagement.

Under the ‘local-literal’ category discussed earlier, we saw a range of tasks that were focused on students showing their understanding of key concepts in the discipline (eg ‘thermal energy’ in Physics; ‘speech acts’ in Linguistics; ‘value configuration’ in Business Studies). Tasks falling under this new category, ‘local-interpretative’, had a similar focus on key disciplinary concepts, but were distinguishable from these largely comprehension-based tasks in their requirement that students

engage in some reflective way with this material. Such a distinction is well-illustrated in the following quotation from one informant.

**ARCHITECTURE:** Some of the texts in the subject are difficult so we typically set some guide questions to help [students] pick out what we see as the key points in the reading. But we also want them to reflect on what they have read and always relate it somehow to their design work.

In the analysis of our corpus, we observed two main types of interpretative tasks around this more local material: i) tasks requiring students to show how a concept or idea in their reading could be utilised in their work in the discipline (application), and ii) tasks requiring some assessment of the validity, worth and so on of an idea, or concept (evaluation).

### Application tasks

The first of these task types, the ‘application type’, was the more common in the corpus, with instances identified in a range of discipline areas. In the following task, taken from the *Architecture* subject, we see exemplification of the principle enunciated above by the lecturer in this subject (Sample A19). As outlined in the task rubric, students here need first of all to consider certain concepts presented in their course reader (in this case ‘efficient structures found in nature’), and then for them to reflect on how these concepts might be applied in their ‘future design work’.

#### Structures in nature

The chapter Introduction to Building Structures gives a good overview of the structural systems you have been learning about. The author also looks at how efficient structures found in nature are good case studies in which to examine structural principles.

Make some notes from your reading on several of these structures, and suggest how you think the concepts discussed could be useful to you in your future design work.

#### Sample A19. Exercise task - *Architecture*

The following are additional tasks that have this focus on the *application* of key disciplinary concepts (Sample A20 and A21). In the Economics task (A20), students need to draw on a particular economic model (‘Solow-Swan model’) as a basis for analysing a particular economic state-of-affairs (or rather a state-of-affairs imputed by a particular economic commentator). A similar configuration is evident in the Physics task (A21), where students need to draw on a concept in the literature (‘gel electrophoresis’), as a basis for assessing the ‘accuracy’ of an example constructed by themselves.

Consider the following statement made by a leading Australian economic commentator:

Where once our economic growth was determined solely by the number of machines, today it is determined by our ability to generate new ideas and develop new ways of producing output.

Using the Solow-Swan model, assess this statement.

#### Sample A20. Exercise task - *Economics*

**Extended written answer**

- a) From a physics perspective, and using the simple model ( $F = CAv$ ), discuss how gel electrophoresis allows fragments of different mass and/or electric charge to be separated over time.
- b) Using an example constructed by you (i.e. you decide the mass, size, and charge of each fragment), demonstrate that two different fragments will separate over time.
- c) consult the literature on gel electrophoresis and briefly discuss one aspect of your initial analysis that is idealised or inaccurate.

**Sample A21. Exercise task - *Physics***

In their commentaries on these more interpretative tasks, informants emphasised the need for students to be operating beyond any ‘simple factual’ understanding of knowledge, where answers fall neatly into correct and incorrect responses. Interestingly, such a view was also enunciated by some from the hard technical disciplines, including the Physics lecturer who was keen to disavow students of the idea that studies in her subject involved a simple quest for the right answer:

**PHYSICS:** People think traditionally that Physics is really just about the mathematical solving of problems, and coming up with the right answer. In fact there’s a lot in it that’s just not that. A lot is about being able to understand concepts and working out how and when to apply them.

A similar view was expressed by the Architecture lecturer who also stressed the ‘open-ended’ nature of reading tasks in her discipline area. She pondered whether this in fact was a conception that was at odds somehow with those held by students from certain educational backgrounds.

**ARCHITECTURE:** In terms of tasks we set around reading, we have many open-ended tasks with no right or wrong answer. If students are coming from a different culture where there is an expectation that they need to get something right, then there are difficulties there I suppose.

**Evaluation tasks**

Less prominent among the tasks fitting a ‘local interpretative’ pattern were those that required explicit evaluation of material, involving the assessment of the value, worth, benefit etc. of some entity. Consistent with the finding from the interviews, it was noted that such tasks in the corpus were confined to the softer ‘humanities’ disciplines, as seen in the following examples. We note in passing that a range of different types of entities are presented here as the objects of evaluation; in Sample A22 it is a ‘policy’; in A23, a ‘thesis’; and in A24, a ‘concept’.

Explain what a ‘polycentric’ staffing policy is. What are the positives and negatives of a firm adopting such a policy?

**Sample A22. Exam question - *Management***

What is Samuel Huntington's 'Clash of civilizations'? How convincing is his thesis?

**Sample A23. Exercise task - History**

What is 'liquid modernity'? How useful do you find this concept? Discuss in relation to the phenomenon of reality television.

**Sample A24. Exercise task – Media Studies**

In such tasks, one sees clear expression of the 'critical' approach to knowledge advocated by a number of informants, as seen in the following remarks made by the lecturers from Media Studies and History:

**MEDIA STUDIES:** On my course ... students being critical in their reading is absolutely essential. Students need to assess arguments, and part of this is identifying where arguments and ideas have been left out.

**HISTORY:** I stress to students the need for a critical approach. The way I get at this is to say them: "Well just because this guy writes it in a book, it's not something you have to accept".

**GLOBAL + INTERPRETATIVE**

The category 'global-interpretative' refers to those tasks requiring students to bring a broadly interpretative approach to their reading in relation to whole texts or multiple texts. Most tasks in the corpus fitting this configuration were assignment tasks, taking in a range of genres: essays, reports and the like. The most prominent of these genres identified in the data are discussed below.

**Essays**

The assignment-type task most clearly requiring a 'global-interpretative' approach to reading material was found to be the expository essay. In the study, the essay genre was set for students in about half the subjects investigated – with the majority of these prescribed in the 'soft' disciplines. Below are two such samples, from *Management* (Sample A25) and *History* (Sample A26). In the rubric of these tasks we can see the need for students to engage with a variety of materials ('a range of views'; 'available evidence' etc.) and to bring a critical approach to these materials ('to 'examine, 'to assess', 'to come to your own judgment').

**Globalisation and Cultural risk**

"Globalisation is reducing cultural differences between countries and thus cultural risk. International firms can now pursue global strategies without fear of failure". Please assess the merits of this advice. Can firms ignore cultural risks?

In your essay you will need to consider a range of views on this issue before coming to your own final judgment.

**Sample A25. Essay task - Management**



**Kennedy's Vietnam policies**

In retrospect, JFK can be seen to have increased the American commitment in Vietnam. Many argue, however, that Kennedy would have resiled from extending the War. Examine the available evidence, including the views of Kennedy's contemporaries and the historians who have studied his presidency to assess the nature and impact of JFK's Vietnam policies.

**Sample A26. Essay task - History**

This 'global interpretative' approach was also outlined by informants in interview:

**HISTORY:** We require our students to read widely – both primary and secondary material. I stress to them that they need to do adequate reading otherwise they cannot respond to the questions that we pose. I say “You might find a perfectly reasonable answer in a single book on this topic, but you're in no position to evaluate that unless you've read alternatives”.

Accompanying such essay tasks in the subject manuals was a good deal of material covering issues of appropriate use of sources and plagiarism, including the following from the History manual:

Essay writing is an essential part of the learning process and a vital medium through which we can assess your understanding of the subject. The essay must therefore be your own work. This does not mean you should not make extensive use of the work of others. However when you quote or paraphrase the explanations of others, you must acknowledge your sources in full.

**Figure 9. Advice concerning plagiarism – History course manual**

In relation to essay writing, a number of informants spoke of the challenges of imparting to students how they should go about using reading material legitimately in their writing:

**MANAGEMENT:** Using sources appropriately is a tertiary skill, and in teaching that we try to inculcate some of the ideas of what plagiarism is ... but we do often face issues with students on that score.

**Reports**

Another assignment-type requiring a 'global-interpretative' engagement was the report. In the following example, a section of a linguistics research report task (cited earlier – Sample A18), students needed to interpret the results of their study against the findings of a range of studies described earlier in the report.

**Speech act research**

Discussion & Conclusion (400-500 words). Analyse and interpret your findings: Why did it turn out like this? What is the reason for the differences you found? How do these results stand with respect to some of the studies you reported in the introduction? End with a brief conclusion.

**Sample A27. Research project task (extract) – Linguistics**

In the following case study report from *Business Studies*, students needed to draw on certain discipline-related concepts ('value configuration', 'business structure') as a basis for analysing the effectiveness ('value') of a specific business organisation.

### Value proposition analysis

This assessment task requires you to analyse how the environment, value configuration and business structure affect the nature of a value proposition.

Task: Your tutor will assign you with a small to medium business (SME) example for you to focus your analysis. Drawing on key concepts from the course, you need to analyse various aspects of the business to explain and evaluate where and how an organisation delivers value to their customers.

### Sample A28. Report task – *Business Studies*

#### Text analysis

One final type of ‘global-interpretative’ reading task involved forms of text analysis. This type of task is arguably a more recent task-type set for students in the academy, and reflects the growing influence of notions of ‘genre’ and ‘academic literacy’ on teaching in university programs. In such tasks in our corpus, students were typically encouraged to see particular texts as ‘generic resources’ from which they could draw for their own writing, as seen in Sample A29 below. In this task, from the Communications subject, students need to investigate a range of Opinion pieces from a newspaper (Op-Ed articles) as a basis for writing their own pieces.

### Writing an Op Ed piece

For this task you need to research and write an opinion piece on a timely topic. You need to express an opinion and then to make an argument to support that opinion. This type of article is called in the industry an ‘op-ed’ piece. No ESSAYS please. Note that the op-ed is an entirely different genre from the academic essay.

To prepare for the writing of this piece, you should locate several examples of op-ed pieces written on a similar topic from a major newspaper (eg The Age). These examples of the genre can serve as a model for your own writing. In consulting the piece, you should consider what is said about the topic in question, but also – and very importantly – how the piece is put together (the language used, structure etc).

### Sample A29. Assignment task – *Communications*

This genre-based reading was elaborated on by the Communications lecturer, who saw such training as essential to the development of students’ writing abilities:

**COMMUNICATIONS:** Because they have to write in this subject, if they don’t read, then they will be completely ‘off genre’. They’ll just be writing stuff that they would have written at high school. So I get them to analyse texts. I actually get them to do things like count the words in the sentences, get the sentence range, what style of language it is. Is it elaborate or is it plain? And then they need to emulate that.

Whilst the setting of tasks such as this is quite understandable in the context of a course explicitly aimed at developing writing skills in students, we noted similar genre-based activities set on courses without the same focus on writing *per se*. Thus, in Sample A30 below from the Management subject, students are instructed to study a sample answer based on ‘The Globalisation Debate’ task discussed earlier (Sample A15), and to use this as an ‘indicative guide’ for completion of the ‘debate’ task set for students.

**Sample review**

The following text is a sample review in the manner of the debate review exercise (see sample X).

Study the text carefully. It should be used as an indicative guide to the sort of tone, analysis and structure expected of such a review. The references and quotations used are fictional and solely for illustrative purposes.

**Sample A30. Exercise – *Management***

A different type of text analysis task was one where students needed to adopt a ‘critical’ approach to language use. Examples of such tasks were confined to the Media Studies subject, such as the following ‘*Media Watch*’ task (Sample A31), requiring students to analyse different media representations of a particular story or issue.

**Media Watch**

Groups of 4–5 students will choose to look at one contemporary issue currently represented on media outlets – eg issues to do with politics, economics, religious affairs, sport music, celebrity or even the media itself. You should consult a variety of media outlets eg print media (including online sites), television news and current affairs.

The main purpose of this assignment is to analyse the similarities and differences in the coverage of the one story or issue that the different media organisations put out. Pay special attention to the language used and how this might involve distortion, bias, plagiarism or unethical reporting.

**Sample A31. Assignment task – *Media Studies***

The focus of such a task, as the rubric indicates, is very much on the way that language is used to construct particular representations of events. The lecturer in the subject described the approach thus:

**MEDIA STUDIES:** In the course we’re interested in students becoming deconstructive readers. The emphasis is not so much on what the meanings of the texts are, and whether I agree with them, but rather how meaning is being created. I want them to focus on how the words used in a text can carry particular nuances of meaning, or how images are used to create certain effects.

Such readings, which operate arguably at the most ‘interpretative’ end of our literal– interpretative continuum fit very much with recent developments in language analysis including critical discourse analysis (Fairclough, 1998) and critical literacy (Gee, 2008).

**Summary of academic task analysis**

The analysis of the reading tasks showed a wide variety of reading requirements across the disciplines investigated. As we saw, instances of tasks fitting all four configurations in our matrix were identified (i.e. local-literal; global-interpretative etc). Because of the largely qualitative nature of the study, it is not possible to make any definitive statements about which of these four reading modes was the most prominent overall. There are however, a number of broad generalisations that can be made:

- 1 most reading tasks in the corpus fitting a *local-literal* configuration tended to be in the form of short weekly exercise tasks or examination questions, and were set principally for the purpose of either inputting or testing a student's understanding of key foundational knowledge in the discipline. Such tasks were linked very much to readings from prescribed textbooks in subjects.
- 2 most reading tasks that fitted the other configurations from the matrix (*global-literal*, *local-interpretative*, *global-interpretative*) tended to be related to more extended written assignment tasks, and often involved readings from a variety of genres: monographs (or sections of monographs); research articles; newspapers and magazines; internet sites etc.
- 3 the variety of assessment requirements across the disciplines pointed to some discipline-based differences in reading modes, with *local-literal* patterns more prominent in the harder technical disciplines, and *global-interpretative* more so in the softer humanities disciplines.

### 4.3 Findings from interviews – Comments on IELTS reading tasks

In this final results section, we turn again to the responses from informants in the interviews. As described earlier, the interview was divided into three phases, covering discussion of: i) general reading requirements on courses; ii) reading requirements on specific assessment tasks; and, in the final phase iii) perceptions regarding the correspondence between reading requirements on courses and those on the IELTS reading test. To facilitate discussion in this final part of the interview, informants were provided with samples of IELTS reading tasks and asked to comment on perceived similarities and differences in reading requirements in the two domains (Appendix 2a). They were also asked to speculate on how useful they thought these sample IELTS tasks were likely to be as preparation for the reading demands of their courses. Responses to these questions were of three broad types:

- 1 an overall positive view of the relationship between reading in the two domains
- 2 a more qualified view of the relationship
- 3 a critical view of the relationship.

As has been the case in previous parts of the study, there was an identifiable disciplinary bias in the responses of informants, with those more positive about the relationship generally coming from the more technical areas, and those having a more critical view from the humanities disciplines.

Those who commented most favourably on the relationship between the test and study on their courses were the lecturers from Computer Science, Engineering, Biology, Business Studies, and Communications, comprising almost half the study's informants (5 out of 12). In general, these informants saw a clear connection between some of the IELTS task types and the types of tasks set on their courses, as seen in the following comments:

**BIOLOGY:** I think the skills required here [on the IELTS test] would be very closely aligned to what I would expect a student in first-year biology to come to terms with. There's a fair bit of reading there and a fair bit of basic comprehension, and that is certainly what our students need.

**COMPUTER SCIENCE:** Our exam questions are not dissimilar to some of the questions [on IELTS]. [This is] certainly true of the multiple-choice format, not so much true or false. One of the questions in our exam also involves the students rearranging lines of code in order to create a logical program, and that looks like at least one of the items in this material.

Several informants in this group expressed surprise at what one described as the test's 'unexpected complexity'. The view here was that the reading demands on the IELTS reading test appeared to them to be higher than those in their particular subject area – though it does need to be acknowledged that in such comments, a clear distinction was not always drawn between the demands of the test items and those of the reading passage on which the items were based:

**COMPUTER SCIENCE:** If anything, we're expecting less of students in terms of reading. The test is definitely relevant and having it at a higher level than what we're asking for in the course is a good thing. So it seems to be an appropriate sort of thing to be testing them on.

**COMMUNICATIONS:** I think [the IELTS reading test] would be good preparation actually. I found the science-based articles and items quite complicated actually. If I had to answer questions about the science, I'd have to go back and read it twice.

For informants in this 'more positively-disposed' group, the sense of correspondence between reading demands in the two domains, as well as the test's perceived difficulty led them to believe that IELTS would be an unequivocally useful form of preparation for tertiary study:

**ENGINEERING:** These sorts of skills [tested in IELTS] would definitely be useful in a generic sense ... and I can see that it would be good preparation for what we require on our course.

**BIOLOGY:** I think looking at [these samples], I would be happy if a student was coming to me with those skills.

**COMMUNICATIONS:** I think [the IELTS reading test] would be good preparation actually. ... I think if the students scored well on this then they would probably be OK.

Another group of informants had a generally positive view of the test – or at least of the sample materials provided in the interview – while at the same time, expressing some reservations about its overall usefulness. A similar number of informants fell into this group as the previous (5 out of 12), and consisted of the lecturers from Physics, Architecture, Economics, History, and Management. The main reservation expressed was a sense of a limited degree of correspondence between the test and reading requirements in their particular disciplinary domain, as suggested in the following remarks:

**ECONOMICS:** I think [the IELTS material] is fine. It's just comprehension really... I've got no problems with that whatsoever. Where economics is slightly different from this is that we use a combination of mathematical techniques, diagrammatic techniques and texts. ... It's a very abstract mathematical way of thinking about the real-world.

**HISTORY:** I'd see this as all useful. The test is very focused on reading comprehension ... that is a basic pre-requisite for our courses. It doesn't cover the quite discipline-specific methods of reading we're concerned with ... for example the way students need to be able to handle the reading of primary source material.

**ARCHITECTURE:** The topic area of the test - bridges - is spot on for our area. I think the type of questioning is also ideal for the level of language skill required in [our subject]. It's not clear though whether you just have to match words, or whether you have to read between the lines a bit – students certainly need to do some [of the latter] on the course.

In asserting these distinctions, a common theme among this group related to the importance of students reading to understand certain key concepts in the discipline, and to be able to show their understanding of these. This was felt by some to be a quite basic difference between the two formats:

**ARCHITECTURE:** I think a difference is that we want students to pull out key concepts from paragraphs. In IELTS it seems they are given the concepts and just need to match these up.

**PHYSICS:** In Physics, the majority of material in the text is trying to explain concepts and also explain problem-solving strategies, and this is what we want [students] to get from their reading. The IELTS tasks seem more arbitrary in what they pick out from the text ... and seem to be mainly about pattern recognition.

One other gap commented on was the lack of connection with processes of writing on the IELTS reading test. Several informants discussed the considerable challenges on their courses in getting students to understand and also adopt acceptable use of reading material in their written work. The view here was that this was perhaps an aspect of academic reading that could somehow be given explicit coverage in the test.

**MANAGEMENT:** [To use sources appropriately] students need to see concrete examples to know what is acceptable and what's not ... I can't see much evidence in the test of this aspect of academic study, and this would certainly be helpful.

Whilst identifying certain differences in the skills in the two domains, informants in this second group acknowledged that it would be most difficult to create a generic reading test that could accommodate in any systematic way the various discipline-bound forms of reading identified. One informant also thought it necessary to be realistic about the extant reading skills that students bring to their courses, and was sure that the responsibility for the teaching of any discipline-specific skills lay squarely with academics on their particular programs.

**HISTORY:** We just can't make too many assumptions nowadays about our students and their capacities. And this is irrespective of their background. ... the onus is clearly on us to develop these capacities within our courses.

A final group – a considerably smaller one than the previous two – had a more critical view of the test and its likely usefulness. This group was confined to just two informants – those from the humanities disciplines of *Media Studies* and *Linguistics*. The general view expressed by these two was that the construct of reading in the test was somehow at odds with that which operated in each of their discipline areas, and that, as a result, the test risked giving students a misleading impression of the nature of academic reading. Their takes on this disjuncture were slightly different ones. For the *Media Studies* lecturer the problem was at heart an epistemological one:

**MEDIA STUDIES:** In the tasks on the test, it seems to me that students are really just dealing with information. That's the way these texts are presented. And then it's mainly about regurgitating the information. This author is saying this. But it doesn't allow students options to engage with the material. Whether they think what is being said in the text is valid or not. I see it as pretty low level.

This lecturer went on to explain that from the outset on his course, he did not want students to see texts fundamentally as “repositories of information and facts”, but rather as the expression of particular ways of seeing and constructing the world:

**MEDIA STUDIES:** There’s a need for students to have an argumentative, conceptual, even ideological understanding of material. [I tell them that when] they come to university they need to learn how to critique ... well everything ... You question all that you read, and all that your lecturer gives you, and I can’t see much evidence of this in the test.

The concerns of the Linguistics lecturer related more to what she saw as the non-contextual nature of reading on the IELTS test. What was notable about reading at university, she thought, was that it always operates within a context, one which is shaped by the discipline itself and also by the particular task with which students are engaged. This, she thought, was a feature strongly lacking in the IELTS test:

**LINGUISTICS:** There is a broader context for interpreting the reading which university students have because they have a purpose for assignments, and the discipline serves to make it clear what is important and what is not. ... so [in the IELTS test], this is quite strange and difficult to relate to because the tasks are completely out of context. What is missing is the purpose for knowing this information.

This lecturer thought that a way to improve the test in this regard would be to construct tasks around particular contexts of study (or ‘scenarios’), which could serve to provide this sense of purpose:

**LINGUISTICS:** I think a good way to go would be if students had some background information like: “You are a student. You are studying blah blah blah, you need to know X,Y and Z in order to complete a certain assignment. This is the context for your reading. Now try and answer some specific questions. How would this information be useful to you and why?” Because that is the sort of expectations we have of students.

## 5 SUMMARY AND DISCUSSION OF FINDINGS

A range of findings have been presented so far, drawn from the two methodologies employed in the study; namely, the analysis of the IELTS and academic tasks, and the interviews with academic staff. In what follows we provide a summary of these findings focusing on:

- 1 main findings, which are those patterns to emerge from the data as a whole, and
- 2 more specific findings, which relate to particular areas of the data.

### 5.1 Main findings

#### IELTS reading vs academic reading

A feature of the study’s design was the development of an analytical framework that would allow systematic comparison to be made between reading requirements in the two domains – IELTS and academic study. As discussed, this framework took in two dimensions: i) the ‘level’ of engagement with text, which distinguished between a more ‘local’ and a more ‘global’ engagement with material; and ii) the ‘type’ of engagement, where the distinction was one between more ‘literal’ and more ‘interpretative’ readings of this material. Drawing on this analysis, we can say there is evidence in the

study of some correspondence between the reading requirements in the two domains, but also evidence of a fair degree of difference.

The main similarity is to be found in those forms of reading that required mainly a local and literal engagement with material. As was noted previously, this configuration was true for the vast majority of items in the IELTS corpus, with many tasks requiring mainly a basic comprehension of relatively small textual units (sentences, inter-sentences, paragraphs). In a similar vein, a sizeable proportion of tasks in the academic corpus were also found to have the same ‘local-literal’ orientation. Such tasks within the academic data, it was noted, tended to be set as weekly class exercises or on exams and tests, and had as their focus the need for students to understand certain discipline-based concepts.

But while this particular similarity was evident, the study also noted a good deal of divergence between the two domains. This was mainly found to arise from the considerable variety of reading tasks identified in the academic corpus, especially in those that related to more extended assignment tasks (eg essays, reports and so on). Thus, whereas the IELTS corpus saw virtually all task-types fall within the ‘local–literal’ area of our analytical matrix, the academic corpus was notable for incorporating tasks that covered all four areas. Amid this diversity were tasks which seemed, on the face of it, to be quite remote from the IELTS profile of tasks, including, for example, those which required a critical engagement with material, or which stipulated engagement with ‘a multiplicity of sources and viewpoints’.

These patterns – both of similarity and of difference – were largely confirmed in the interview commentaries of staff. Thus, some of our informants saw a basic congruence between the type of reading they expected their students to do on their courses, and what they perceived to be the demands of the IELTS test. Others, by contrast, were keen to point out what for them were clear differences.

### **Disciplinary variation in reading requirements**

The similarities and differences observed between the IELTS reading test and academic study can be accounted for in part by the variety in the types of reading required across the disciplines considered in the study. Much of this variety, as we have noted, related to the broad division in the disciplines investigated; that is between the ‘harder’ technical disciplines on the one hand, and ‘softer’ more humanities-oriented disciplines on the other. Thus, it was noted that in the more technical disciplines (eg Engineering, Architecture, Physics, Biology), less reading overall was required of students, and that much of this had the clear purpose of having students assimilate certain foundational concepts in the discipline. Such a view of reading was affirmed in the interviews, where it was suggested that the contents of reading materials on such courses were presented to students as essentially “the ideas they needed to know”.

In the more humanities disciplines, by contrast, reading was found to take on many different guises. While students on these courses (including Media Studies, Linguistics, History, Management) were also required to learn basic ‘concepts and viewpoints’ in their field, there were many additional ways they were expected to interact with material. In some contexts, for example, the task for students was one of comparing different ideas and viewpoints on an issue; in other contexts, it was to evaluate these ideas; in others again, students needed to synthesise a range of material as a basis for developing their own viewpoints. In contrast to the mainly ‘assimilationist’ approach to reading described by informants in the technical disciplines, the view from these latter fields was that students needed always to bring their own perspective to bear on material – an approach characterised by one informant as “the need to question everything”.



The findings from this part of the study suggest then, that in the first year of undergraduate study at least, the types of materials students need to read on their courses, and the ways they need to go about reading these material are subject to a good deal of variation. This feature of academic study points to the difficulties inherent in trying to conceive of some generalist construct of academic reading, one that has clear relevance to all disciplinary contexts. The implications of this situation are discussed in the final sections of the report.

## 5.2 Specific findings

Along with these general findings were a number of more specific findings that emerged from the data, ones that provide greater detail about some of the differences observed between the two domains.

### Epistemic entities

It was noticed in the two task corpora (IELTS and academic) that the types of entities that students/ test-takers needed to focus on in their reading were generally of a different order. In the IELTS test samples, for example, these entities were typically characterised as ‘information’, as exemplified in the following sample rubrics (our emphasis):

Do the following statements agree with the information given in the Reading Passage?  
(Sample 1.1)

Which paragraph contains the following information?  
(Sample 2.2)

In the academic tasks, by contrast, this knowledge tended to be characterised in a variety of ways. Firstly it was noticed that it was quite rare in fact for students to be asked to engage with ‘information’ per se; instead they needed to contend with a range of different entities. Most prominent among these was a characterisation of knowledge as ‘concept’ (or related entities – ‘model’, ‘definition’ and the like), as seen in a number of tasks in the academic corpus. Among the more humanities disciplines, we also saw an emphasis on entities associated with the ideas of particular scholars – including ‘arguments’, ‘viewpoints’, ‘theses’, ‘perspectives’ etc. Other entity-types were those related to the outcomes of empirical research eg ‘studies’ and ‘findings’.

This contrast in the epistemic entities in the two domains points to a more ‘constructivist view’ of knowledge in the case of the academic tasks, where knowledge is typically seen to arise from the thinking and researching of individual scholars in a field, or from the collective disciplinary community as a whole (Myers, 1992). The contrasting focus in IELTS on reading content as ‘information’ suggests instead a more positivist view of knowledge, where, as Hill and Parry (1992) suggest, “authorship is essentially anonymous” (p 439).

### Interpretative readings

These different ways of conceiving of academic knowledge were found to have implications for the way that this knowledge needed to be engaged with in the two domains. Thus, we saw that the essential task for students in many of the IELTS items was to demonstrate a basic comprehension of the propositional content of reading material. By contrast, the focus of many of the academic tasks was not only to arrive at a basic understanding of material, but also to ‘work’ with this material in order to proffer some interpretation of it. This basic requirement of academic study was well summarised by one informant thus:

we typically [want students] to pick out ... the key points in the reading. But we also want them to reflect on what they have read and always relate it to their ... work somehow.

In the academic corpus, it was noted that two types of interpretative reading tasks predominated – what we have termed application and evaluation. In application-related tasks, students were typically required to show how a concept or idea in their reading could be utilised in their work in the discipline; in evaluative tasks, the focus was more on making some explicit assessment of these concepts (eg with respect to their validity, worth etc). Of these two interpretative modes, the application-related tasks were found to be the more common.

We note in passing that interpretations such as these tend to be very much discipline-based (McPeck, 1992), evident not only in the specific concepts and ideas that students need to engage with, but also in the types of ‘interpretations’ they need to make of these concepts along the way. Indeed for some scholars, the process of being trained in a discipline is often characterised in these precise terms; that is, to learn the particular ways in which certain concepts are ‘applied’ and ‘evaluated’ within a field (Lave & Wenger, 1991). As Bourdieu (1990) points out, such practices are not only cognitive in nature, but are effective when assimilated into habituated dispositions. The strong discipline-base of these more interpretative forms of reading may provide some explanation for the apparent absence of these modes among the various IELTS tasks collected for the study. We can also recognise in this situation, the challenges that would be involved in incorporating such modes into any possible adapted version of the test.

### **Readings of multiple texts**

Another difference noted between the two domains was the quantity of reading required to complete some tasks. As we saw, all tasks in the IELTS corpus were focused on engagement with a single text (the relevant reading passage), and in the case of some task-types, a focus on relatively small components of the text. In contrast, a feature of some of the academic tasks, especially in the more humanities areas, was the need for students to engage with a range of texts. Examples of such tasks were: i) summary tasks which required students to give an account of a variety of sources in relation to a particular topic; and ii) essay tasks requiring the exploration of a range of views as a prelude to students presenting their own views on the topic.

Some of the academic tasks, as we saw, alluded again to a particular conception of knowledge, one that sees knowledge in a discipline being advanced through processes of debate and dialogue between scholars, as opposed to the furnishing of single, definitive answers to issues and problems. Several informants were sure that it was only through the engagement with multiple sources that students could develop a suitably critical frame in their studies. As one informant explained it, students might feel they have come across “a perfectly reasonable answer” to a question, but that they are in fact only in a position to presume this if they’ve had the opportunity to “measure this answer against alternatives”.

### **The contextual nature of reading**

Reading in the two domains was also seen to differ around the notion of context. One observation made about the IELTS samples provided to informants was the apparent lack of an underlying intellectual purpose for the particular questions posed in tasks; that is to say, that in many tasks, the particular items of information needing to be focused on appeared, on the face of it, to be rather arbitrary. In contrast, it was suggested that it is the nature of university study that there is usually a clear purpose and context for the type of reading that students need to do. As one informant explained

it, such a context – which is created at once by the broader knowledge base of a discipline, and also by the immediate demands of tasks and assignments set within courses – “serves to make it clear to students what [information] is important and what is not”.

This disparity between reading in the testing and academic domains has been commented on in the literature. Alderson (2000) notes after Barton (1994) that it is rarely the case in academic study, that reading as an activity is performed in vacuo; that is, without being related in some way to other academic activities. A concept invoked to capture this idea is ‘literacy event’, described by Barton and Hamilton (1998, p 9) as a series of observable activities mediated by text. As Alderson explains it:

Often literacy events – TLU reading tasks – are not undertaken in isolation. .. A coursework reading assignment leads to note-taking, which leads to further reading, to drafting a written paper, re-reading the draft critically (Alderson, 2000, p 148).

To accommodate this feature of academic study within the IELTS test is undoubtedly a challenge; as Weir et al (2009) suggest, full contextual authenticity “is generally unrealistic for language assessments” (p 12). The suggestion from one of the study’s informants was to construct tasks around specific study ‘scenarios’, ones that would seek to place the reading of test passages into some real-world context for students.

### **The reading – writing nexus**

Arguably one of the more significant literacy events in academic study is that which involves the integrating of one’s reading on a topic into some related writing activity (Horowitz, 1986; Moore & Morton, 2007). This was evident in many of the academic tasks analysed in the study, with virtually all of the assignment-style tasks in the corpus having a substantive reading component attached to them. A number of informants commented on the importance of this reading–writing nexus, seeing it as an area of particular challenge to students. Concern was expressed here about students’ abilities to use and document sources appropriately, along with perceptions about the growing incidence of plagiarism on courses. Several informants noted the absence of these reading–writing connections in the sample IELTS materials provided, and wondered whether this dimension of academic reading could be incorporated into the test somehow.

### **Information literacy**

Another area thought to have limited coverage in the test related to the skills involved in locating, selecting and evaluating information sources. In their discussions of the reading practices of students, a number of informants noted the opportunities, but also the considerable challenges created for students by the increasing online environment of academic study. As we saw, concern was expressed that students did not always bring a particularly ‘discerning’ attitude to the vast textual resources now available to them. The response of some of our informants to this situation was increasingly to adopt an ‘information literacy’ approach in their teaching, with students called upon to appraise texts in some broader social context, and to develop an awareness of such matters as the context of their production, their authorship, communicative purpose, and ultimately their ‘reliability’ as sources.

It was noted by some informants that the increasingly important skills related to the searching and selecting of sources appeared to have little or no coverage in the IELTS reading test. Indeed, the tendency of IELTS tasks to focus on quite specific items of information in reading passages, would seem to limit the possibilities of appraising texts in the broader social and contextual terms of an ‘information literacies’ approach (Shapiro & Hughes, 1996).

### Genre readings of texts

A final type of reading evident in the academic corpus is what we have called ‘genre readings of texts’. As noted, a number of reading tasks in the corpus required a focus not so much on the contents of texts, but rather on the ways in which ‘texts are put together’ (The focus of such tasks was on such textual features as rhetorical organisation, sentence structures, lexical choices and so on). In some of these tasks, it was noted, the main purpose was a more utilitarian one; that is, for students to ‘get a feel for the genre’, as one informant described it, so that they might emulate the particular written style in their own work. In other tasks, the purpose was more a critical or ‘deconstructive’ one, with students needing to identify how language operates in texts to create certain meanings – including ‘ideological meanings’.

As was mentioned, these types of ‘genre readings’, which take in both more ‘pragmatic’ approaches (Johns, 1997; Allison, 1996; Swales, 1990) and more critical approaches (Shor, 1999; Street, 2003), reflect the increasing role of textual analysis activities in academic study. It is fair to say that readings such as this were not really apparent in the IELTS corpus compiled for the study.

### An explanation for differences

The study has identified a number of differences between reading demands in the two domains, even if they are ones that can be readily accounted for. Arguably, the purpose of a test of reading is to assess students’ abilities to process written text. In this context, as we have seen, the actual contents of the reading tend to be somewhat incidental. In university study, by contrast, such content – which relates to study in a discipline – is of paramount importance. Thus, in university study, there is not the same interest in the skills of reading *per se*; instead acts of reading, as we have seen, are tied intimately to the acquisition, application, and ultimately to the advancement of disciplinary knowledge. This contrast in the role of knowledge in the two domains necessarily entails some quite basic differences in the nature of the texts students need to read, and what it is students need to do when they read them.

## 6 IMPLICATIONS OF FINDINGS FOR FUTURE DEVELOPMENT OF THE READING TEST

In considering the implications of the study, there are arguably two key questions that need to be addressed:

- 1 Is there a case for making some modification to the IELTS Academic Reading Test?
- 2 If so, how could the test be modified?

### 6.1 Should the IELTS Reading Test be modified?

In relation to the first question, the general push in language assessment to maximise a test’s ‘authenticity’ would suggest that some modification to the IELTS reading test is at least worth considering. Bachman and Palmer (1996) define ‘inauthenticity as that situation where the link between the TLU task and the test task is weak’. Whilst the findings of the task analysis do not suggest overall a ‘weak’ link between tasks in the two domains, they do suggest that it is one that could at least be strengthened. Such a view was also reflected in the responses of some of the academic informants in our study, where it was felt that the demands of the test might be brought more into line with the type of reading required on their courses.

The ever-expanding influence of IELTS – and especially its curriculum effects on programs of English for Academic Purposes – provide additional impetus for modification of some form.

Clearly however, there are important practical considerations in any push to institute changes to a well-established test such as IELTS. One can point to a number of caveats. The first of these relates to the broad issue of achieving the right balance between the validity of a test and its reliability (Wigglesworth & Elder, 1996). For the IELTS academic reading test, this would include, among other things, ensuring that any modified version of the test fit with the overall structure of the current IELTS battery eg for the reading test to remain as a separate test of reading without significant overlap with other modules such as writing (Charge & Taylor, 1997); and for it to be retained as a clerically-markable module within the battery. A second caveat relates to the difficulty of accommodating the many different versions of academic reading we have seen in the study all within the one test. Much of this variety, as was noted, arose from the quite different reading demands evident in different disciplines and programs. This suggests a need to be prudent in selecting the type of reading tasks on the test, so as to avoid having items which may be pertinent in one disciplinary area, but have little relevance to others.

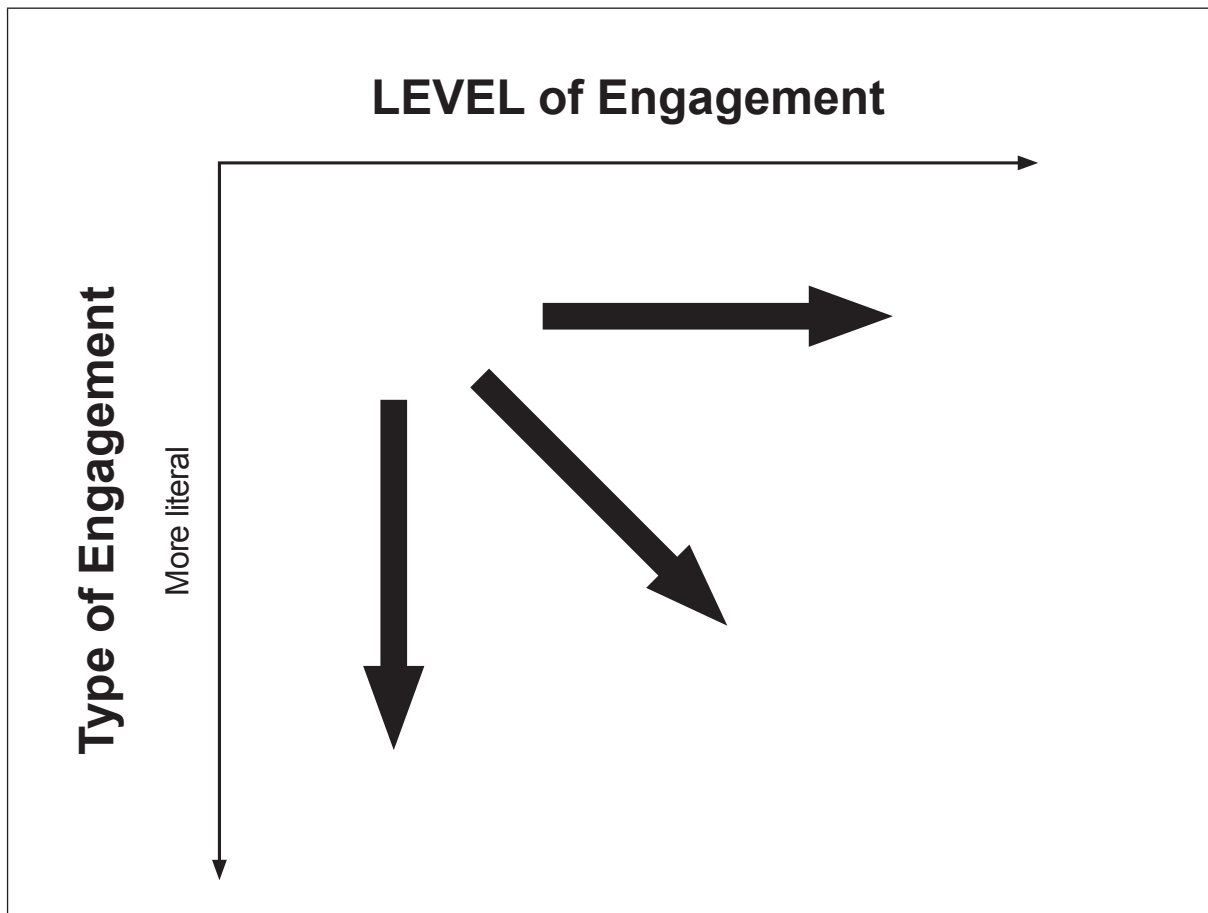
A final consideration is the matter of what one can reasonably expect an objective test of reading to cover. On this point, Taylor (2007) suggests we need to recognise the limits to which a test such as IELTS can simulate (and indeed should be expected to simulate) language use in the target situation. Thus, she notes that “IELTS is designed principally to test readiness to enter the world of university-level study in the English language”, and does not assume that test takers have already mastered the skills they are likely to need (original emphasis, p 482). Taylor goes on to explain that students will often “need to develop many of these skills during their course of study”, including those “skills ... specific to their academic domain”. Such an understanding was voiced, as we saw, by at least one of the study’s informants who suggested that the onus was clearly on academic staff to develop discipline-specific capacities “within courses”.

## 6.2 How could the IELTS Reading Test be modified?

If any modifications were to be made to the academic reading test, one useful principle to employ, we believe, would be to seek to push test tasks, or at least a proportion of them, in the direction of the more global/more interpretive regions of the analytical matrix used in the study, as shown in Figure 9.

In what follows we provide a number of sample tasks, where the intention is to indicate how we think some of these less-covered areas of the matrix could have some coverage in the test. The samples have been divided up into three areas of ‘extension’, each relating to the three under-represented quadrants of our matrix *viz*:

- 1 Extension 1: Local/Interpretative
- 2 Extension 2: Global/Literal
- 3 Extension 3: Global/Interpretative



**Figure 9. Suggested principle for modifications to reading test**

Several samples are provided for each extension; some additional samples are shown in Appendix 3. In the construction of these tasks, we have attempted to incorporate some of the specific differences noted between reading in the two domains (see section 5.2) with a focus on such dimensions as: authorial stance; specific academic entities (eg arguments); reading–writing connections; information literacy skills; genre readings of texts; text evaluation; and so on. In some of these tasks, there has also been an effort to structure tasks around the idea of relating tasks to specific study scenarios (see section 5.2).

It will be noted that all of the sample tasks provided follow a multiple choice format. This is for the reason noted earlier – namely that the multiple choice tasks of their nature, appear to have a greater versatility than some of the other task types currently used in the test, and on the face of it, seem better able to incorporate these more ‘global’ and ‘interpretative’ engagements with material. This is not to suggest however, that one would necessarily want to see a greater use of multiple choice items on the test. Following Alderson (2000, pp 211-214), we recognize that multiple choice tasks have a number of limitations, including the potential effect of candidates guessing the correct response. We would argue in fact that it is a major challenge for the test’s designers to develop certain conventionalised techniques that are able to test some of the more ‘interpretative’ and more ‘global’ dimensions of reading we have identified.

**EXTENSION 1 --> LOCAL/INTERPRETATIVE**

In Passage A, the writer states that ..... PROPOSITION (Line B)

The implication of this sentence is that the writer thinks:

- a X is a good thing and should be encouraged
- b X is a bad thing and should be discouraged
- c not enough is known about X, and it should be investigated further
- d sufficient research has been conducted into X

**Sample 1.1: Focus on inferential readings of sentences**

A student referred to information in Paragraph B of Passage A in an essay. Which sentence is a reasonable interpretation of the writer's view:

- a Smith (2000) argues that X is a Y
- b Smith (2000) argues that X is not a Y
- c Smith (2000) argues that X is a Z
- d Smith (2000) argues that X is not a Z

**Sample 1.2: Focus on reading-writing connections (scenario format)****EXTENSION 2 --> GLOBAL/LITERAL**

The author of Passage A claims that (Proposition Y). The main evidence presented in the text for this claim is:

- a Findings from a study she conducted
- b Findings from a study conducted by B
- c Findings from a study conducted by her, and by B
- d Findings from several different studies conducted by B and C

**Sample 2.1: Focus on macro-content of text (Epistemic entities= claim/evidence)**

Imagine you are writing an essay on the following topic (State topic X). Which paragraph from Reading Passage A do you think would be the most useful to draw information from.

- a Paragraph 1
- b Paragraph 2
- c Paragraph 3
- d Paragraph 4

**Sample 2.2: Focus on use of sources – information literacy (scenario format)**

### **EXTENSION 3 --> GLOBAL/INTERPRETATIVE**

Which of the following do you think best describes the main purpose of Reading Passage A:

- a to advise on the best ways to do X
- b to criticise the current ways of doing X
- c to provide background information on X
- d to predict what will happen to X

**Sample 3.1: Focus on overall rhetorical purpose of text**

The following are some possible criticisms that could be made of Passage A. Which particular criticism seems the most relevant to this text?

- a The writer states his support for X, but does not consider the other side
- b The writer claims that X is Y, but provides no evidence for this claim
- c The writer presents contradictory views about X
- d The writer gives practical information about X, but doesn't indicate how it can be used

**Sample 3.2: Focus on evaluation of text**

It will be clear from the samples above that the use of certain item techniques is very much dependent on having to hand reading passages which are relevant to the particular focus of the technique. For instance, an item that was focused on the relationship between claims and evidence in a reading passage would clearly only be able to be used in relation to text samples that were structured around these particular rhetorical characteristics. The study deliberately confined itself to a study only of reading tasks without consideration of the texts upon which they are based. It may be however, that any proposed shift in focus towards more global and/or interpretative modes on items would have major implications for reading passage design and selection on the test. The broad principle of the inseparability of reading technique and task has been commented on by Alderson (2000). Any modification to the test may indeed require substantial investigation into this aspect of reading assessment.



### 6.3 Further research

McNamara (1999), as noted earlier, has identified three areas of focus in appraising the validity of a reading proficiency test:

*task stimulus* i.e. the texts that candidates engage with on the test

*task processes* i.e. the reader-text interactions that actually take place in the completing of the test

*task demand* i.e. the test items, which prescribe certain types of interaction between the reader and text.

This list provides a useful framework for thinking about further study into the IELTS Academic Reading Test. In relation to ‘task stimulus’, the issue of text selection on tests has already been identified as an area of priority. Such an investigation would also be well complemented by additional research into the nature of texts typically used in studies in the disciplines in the contemporary university (Green, Unaldi & Weir, 2010). Whilst the present study observed the continuing importance of traditional texts such as textbooks and journal articles, the ever-increasing role played by various electronic media was also noted. Any efforts to enhance the validity of the text component of the test (‘task stimulus’) would need to be based on a thorough and up-to-date understanding of these developments, along with the dynamic effects they appear to be having on literacy practices in the academy.

Another area of interest is the way that students actually read and interact with reading materials when engaged with specific academic tasks (‘task processes’). Whilst the analysis used in the present study allowed us to make some estimate of what was required to complete certain tasks, it was not possible to know definitively from the data what the ‘psychological reality’ would be for students actually engaged in such tasks. Indeed research in the field of activity theory (Lantolf and Thorne, 2006) has shown that one must be wary about assuming any straightforward correspondence between the ‘task-assigned’ and the ‘task-performed’ (Coughlan & Duff, 1994). Weir *et al’s* (2009) study provides useful general information about student performance on the reading test and the TLU situation. Additional research could also be conducted to find out about how these processes compare between performance on specific test items and on larger ‘literacy events’ in academic study (Barton & Hamilton, 1998).

Finally, in the area of ‘task demand’, the present study was relatively small-scale in its design, investigating the assessment requirements in only a limited number of subject areas. The largely qualitative findings obtained could be complemented by larger-scale survey research which looked into reading requirements across a wider range of disciplines and institutions. To have a fuller picture of university reading would not only help in processes of test validation, but also assist us in a broader educational aim – to be able to prepare our students as best we can for the challenges and demands they will face in their studies.

## ACKNOWLEDGMENTS

The researchers wish to thank the following for their assistance with this project:

- staff from the two site universities who participated in the research
- Professor Tim McNamara, and Associate Professor Catherine Elder (University of Melbourne) who provided advice as members of the project's reference group

The researchers also wish to thank IELTS for their support of the project.

## REFERENCES

- Alderson, JC, 1990a, 'Testing reading comprehension skills (Part One)', *Journal of Reading in a Foreign Language*, vol 6, no 2, pp 425-38
- Alderson, JC, 1990b, 'Testing reading comprehension skills (Part Two)', *Journal of Reading in a Foreign Language*, vol 7, no 1, pp 465-504
- Alderson, JC, 2000, *Assessing reading*, Cambridge University Press, Cambridge
- Alderson, JC, and Wall, D, 1993, Does washback exist? *Applied Linguistics* vol 14, no 2, pp 115-29.
- Allison, D, 1996, 'Pragmatist discourse and English for academic purposes', *English for Specific Purposes* 15, no 2, pp 85-103
- Bachman, L and Palmer, A, 1996, *Language testing in practice*, Oxford University Press, Oxford
- Ballard, B and Clanchy, J, 1991, 'Assessment by misconception', in *Assessing second language writing in academic contexts*, ed L Hamp-Lyons, Ablex, Norwood, pp 19-36
- Barton, D, 1994, *Literacy: An introduction to the ecology of written language*, Basil Blackwell, Oxford
- Barton, D and Hamilton, M, 1998, *Local literacies: Reading and writing in one community*, Routledge, London
- Becher, T, 1989, *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines*, Open University Press, Buckingham
- Bourdieu, P, 1990, *The logic of practice*, Stanford University Press, Stanford
- Bloom, BS, 1956, *Taxonomy of Educational Objectives: The classification of educational goals*, Longman, New York
- Carroll, J, 1964, *Language and thought*, Prentice-Hall, Engelwood Cliffs NJ
- Carver, R, 1997, 'Reading for one second, one minute, or one year from the perspective of reading theory', *Scientific Studies of Reading*, vol 1, no 1, pp 3-43

- Charge, N and Taylor, L, 1997, 'Recent developments in IELTS', *ELT Journal*, vol 51, no 4, pp 374-380
- Clapham, C, 1996, 'The development of IELTS: A study of the effect of background Knowledge on Reading Comprehension', *Studies in Language Testing*, vol 4, Cambridge University Press: Cambridge
- Coughlan, P and Duff, P, 1994, 'Same task, different activities: Analysis of a second language acquisition task from an activity theory perspective', in *Vygotskian approaches to second language research*, eds JP Lantolf & G Appel, Ablex Pub Corp, Norwood, NJ, pp 173-191
- Ennis, R, 1987, 'A taxonomy of critical thinking abilities and dispositions', in *Teaching thinking skills*, eds J Baron & R Sternberg, WH Freeman, New York, pp 9-26
- Enright, M, Grabe, W, Koda, K, Mosenthal, P, Mulcany-Ernt, P and Schedl, M, 2000, '*TOEFL 2000 reading framework: A working paper*', TOEFL Monograph Series 17, ETS, Princeton
- Fairclough, N, 1998, *Critical language awareness*, Longman, Harlow, pp 1-29
- Gee, JP, 2008, *Social linguistics and literacies: Ideology in discourses*, Routledge, London
- Grabe, W, 1999, 'Developments in reading research and their implications for computer-adaptive reading assessment', in *Issues in computer-adaptive testing of reading proficiency*, ed M Chaloub-Deville, Cambridge University Press, Cambridge, pp 11-47
- Grabe, W and Stoller, FL, 2002, *Teaching and researching reading*, Longman, London
- Green, A, Unaldi, A, and Weir, C, 2010, 'Empiricism versus connoisseurship: Establishing the appropriacy of texts in tests of academic reading', *Language Testing*, vol 27, no 2, pp 191-211
- Hamp-Lyons, L, 1990, 'Second Language writing: assessment issues' in *Second Language Writing: Research Insights for the Classroom*, ed B Kroll, Cambridge University Press, Cambridge, pp 69-87
- Hill, C and Parry, K, 1992, 'The test at the gate: Models of literacy in reading assessment', *TESOL Quarterly*, vol 26, no 3, pp 433-61
- Horowitz, D, 1986, 'What professors actually require of students: Academic tasks for the ESL classroom,' *TESOL Quarterly* vol 20, pp 445-62
- Hudson, T, 1998, 'Theoretical perspectives on reading', *Annual Review of Applied Linguistics*, vol 18, pp 124-141
- IELTS, 1996, *The IELTS handbook*, UCLES/British Council, IDP Education Australia, Cambridge
- IELTS, 2007, *IELTS official practice materials*, Cambridge University Press, Cambridge
- Johns, A, 1997, *Text, role and context: Developing academic literacies*, Cambridge University Press, Cambridge
- Lantolf, JP, and Thorne, SL, 2006, *Sociocultural theory and the genesis of second language development*, Oxford University Press, Oxford

Lave, J and Wenger, E, 1991, *Situated learning: Legitimate peripheral participation*, Cambridge University Press, Cambridge

Marton, F and Saljo, R, 1976, 'On qualitative differences in learning I: Outcome and process', *British Journal of Educational Psychology*, vol 46, pp 4-11

McNamara, TF, 1999, 'Computer-adaptive testing: A view from outside', in *Issues in computer-adaptive testing of reading proficiency*, ed M Chaloub-Deville, Cambridge University Press, Cambridge, pp 136-149

McPeck, J, 1992, 'Thoughts on subject specificity', in *The generalizability of critical thinking: Multiple perspectives on an educational ideal*, ed S Norris, Teachers College Press, New York, pp 198-205

Messick, S, 1996, 'Validity and washback in language testing', *Language testing*, vol 13, no 3, pp 241-256

Munby, J, 1978, *Communicative Syllabus Design*, Cambridge University Press, Cambridge

Moore, T and Morton, J, 2007, 'Authenticity in the IELTS Academic Module Writing Test: A comparative study of Task 2 items and university assignments', in *IELTS collected papers: Research in speaking and writing assessment*, eds L Taylor and P Falvey, Cambridge University Press, Cambridge, pp 197-248

Myers, G, 2003, 'Discourse studies of scientific popularisation: Questioning the boundaries', *Discourse Studies*, vol 5, pp 265-279

Myers, G, 1992, 'Textbooks and the sociology of scientific knowledge', *English for Specific Purposes*, vol 11, pp 3-17

Nwogu, K, 1991, 'Structure of science popularisations: A genre-analysis approach to the schema of popularised medical texts', *English for Specific Purposes* vol 10, no 2, pp 111-123

Odell, L, Goswami, D and Herrington, A, 1983, 'The discourse-based interview: A procedure for exploring the tacit knowledge of writers in nonacademic settings', in *Research on writing: Principles and methods*, eds P Mosenthal, L Tamor and S Walmsley, Longman, New York, pp 221- 236

Purpura, JE, 1998, Investigating the effects of strategy use and second language test performance with high and low-ability test taker: A structural equation modeling approach, *Language Testing* vol 15, pp 333-379

Read, J and Hayes, B (2003), 'The impact of IELTS on preparation for academic study in New Zealand', *IELTS Research Reports* vol 4, Cambridge ESOL, Cambridge pp 153-206

Saville, N, and Hawkey, R, 2003, 'A study of the impact of the International English Language Testing System, with special reference to its washback on classroom materials', in *Washback in Language Testing: Research Contexts and Methods*, eds RL Cheng, Y Watanabe, and A Curtis, Lawrence Erlbaum and Associates, Mahwah, New Jersey

- Shapiro, J and Hughes, S, 1996, 'Information literacy as a liberal art', *Educom Review*, vol 31, no 2, pp 31-35
- Shor, I, 1999 'What is critical literacy?' in *Critical literacy in action*, eds, I Shor and C Pari, Boynton/Cook, Portsmouth NH
- Street, B, 2003, 'What's new in 'New Literacy Studies: Critical approaches to literacy in theory and practice', *Current issues in Comparative Education*, vol 5, no 2, pp 77-91
- Swales, J, 1998, *Other floors, other voices: a textography of a small university building*, Lawrence Erlbaum Associates, Mahway, N.J
- Swales, J, 1990, *Genre Analysis: English in academic and research settings*, Cambridge University Press, Cambridge
- Taylor, G, 2009, *A student's writing guide*, Cambridge University Press, Cambridge
- Taylor, L, 2007, 'The impact of the joint-funded research studies on the *IELTS writing test*', in *IELTS Collected papers: Research in speaking and writing assessment*, eds L Taylor and P Falvey, Cambridge University Press, Cambridge, pp 479-492
- Trimble, L, 1985, *English for science and technology: A discourse approach*, Cambridge University Press, Cambridge
- Van Dijk, TA and Kintsch, W, 1983, *Strategies of discourse comprehension*, Academic Press, New York
- Wallace, C, 1999, 'Critical language awareness: Key principles for a course of critical reading', *Language Awareness*, vol 8, no 2, pp 98-110.
- Weir, CJ, Hawkey, Green, RA, and Devi, S, 2009, 'The relationship between the Academic Reading construct as measured by IELTS and the reading experiences of students in the first year of their courses at a British University', *IELTS Research Reports 9*, British Council, pp 97-156
- Weir, CJ, and Urquhart, AH, 1998, *Reading in a second language: Process, product and practice*, Longman, New York
- Widdowson, H, 1979, *Explorations in Applied Linguistics*, Oxford University Press, Oxford
- Wigglesworth, J and Elder, C, 1996, 'Perspectives on the testing cycle: Setting the scene', *Australian Review of Applied Linguistics Series S*, No. 13, pp 13-32
- Yule, C, 1996, *Pragmatics*, Oxford University Press, Oxford

## **APPENDIX 1: LIST OF MATERIALS USED IN IELTS TASK CORPUS**

*Official IELTS practice materials*, University of Cambridge; British Council; IDP, IELTS Australia, 2007 (1 x academic reading test)

*Cambridge IELTS 2: Examination papers from University of Cambridge ESOL examinations*, Cambridge University of Press, Cambridge, 2000 (4 x academic reading tests)

*Cambridge IELTS 4: Examination papers from University of Cambridge ESOL examinations*, Cambridge University of Press, Cambridge, 2005 (4 x academic reading tests)

*Cambridge IELTS 6: Examination papers from University of Cambridge ESOL examinations*, Cambridge University of Press, Cambridge, 2007 (4 x academic reading tests)

## APPENDIX 2: SCHEDULE USED IN INTERVIEWS WITH ACADEMIC STAFF

### Interview schedule

The following questions will form the basis of the interview.

#### PART 1 Introduction (content, skills, general reading requirements)

- 5 How would you describe the main content of the course you teach on?
- 6 What do you see as the course's main objectives regarding the skills/attributes to be developed in students?
- 7 How would describe the general reading requirements for students on the course?
  - i How much reading do students need to do?
  - ii Are there weekly reading requirements?
  - iii What sorts of texts do students need to read?
  - iv Are there any activities they need to complete when doing the weekly readings?
  - v What purposes do you have for setting weekly readings for students?
  - vi Have the reading requirements on your course changed over the years?
  - vii What challenges generally do students face in handling reading requirements on the course? What about students from second language backgrounds?

#### PART 2 Reading and Assessment tasks

- 4 What are the main assessment tasks/activities you set for students on the subject?

Taking **each of these tasks** at a time:

- i What do students need to do to successfully complete the task?
- ii How much reading is required to complete the task? How many texts? What types of texts?
- iii How would you describe the nature of the reading they need to do to successfully complete the task? (eg basic comprehension of material? Some form of interpretation?)
- iv What type of material from the reading would students need to include in the written assignment?
- v What challenges do students face in drawing on reading material for this assignment? Are there particular difficulties for students from second language backgrounds?

- 5 The following is a list of specific reading skills required of students in their academic study. All are important in some way - which ones would you see as being **particularly important** on your course? Explain? Are there any other important skills not included on the list?

Be able to

- have a **basic comprehension** of key information in a text
- **summarise** the main ideas in a text in one's own words
- understand an idea for the purpose of **applying** it to a particular situation
- **understand** the purpose for why a text may have been written
- **critically evaluate** the ideas in a text
- **identify** a range of texts relevant to a topic
- draw on ideas from a range of texts to **support** one's own argument
- OTHER \_\_\_\_\_

### **PART 3 IELTS reading tasks**

Questions in this section concern comparisons between the assignment tasks you provided and the attached sample IELTS reading tasks.

- 6 What do you see as the main similarities and/or differences between the type of reading set on the IELTS test, and the type of reading you require of your students on the course?
- 7 On the evidence of these IELTS tasks, to what extent do you think training for the IELTS reading test would be useful preparation for the reading demands on your course? Explain.

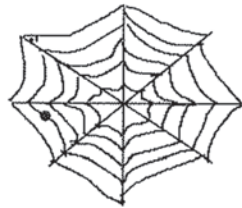


**Sample IELTS reading test material distributed to interviewees for comment:**

Official IELTS practice materials, University of Cambridge; British Council; IDP, IELTS Australia, 2007

**READING PASSAGE 1**

You should spend about 20 minutes on **Questions 1-14**, which are based on Reading Passage 1.

**Spider silk cuts weight of bridges**

*A strong, light bio-material made by genes from spiders could transform construction and industry*

- A** Scientists have succeeded in copying the silk-producing genes of the *Golden Orb Weaver* spider and are using them to create a synthetic material which they believe is the model for a new generation of advanced bio-materials. The new material, biosilk, which has been spun for the first time by researchers at DuPont, has an enormous range of potential uses in construction and manufacturing.
- B** The attraction of the silk spun by the spider is a combination of great strength and enormous elasticity, which man-made fibres have been unable to replicate. On an equal-weight basis, spider silk is far stronger than steel and it is estimated that if a single strand could be made about 10m in diameter, it would be strong enough to stop a jumbo jet in flight. A third important factor is that it is extremely light. Army scientists are already looking at the possibilities of using it for lightweight, bullet-proof vests and parachutes.
- C** For some time, biochemists have been trying to synthesise the drag-line silk of the *Golden Orb Weaver*. The drag-line silk, which forms the radial arms of the web, is stronger than the other parts of the web and some biochemists believe a synthetic version could prove to be as important a material as nylon, which has been around for 50 years, since the discoveries of Wallace Carothers and his team ushered in the age of polymers.
- D** To recreate the material, scientists, including Randolph Lewis at the University of Wyoming, first examined the silk-producing gland of the spider. 'We took out the glands that produce the silk and looked at the coding for the protein material they make, which is spun into a web. We then went looking for clones with the right DNA,' he says.
- E** At DuPont, researchers have used both yeast and bacteria as hosts to grow the raw material, which they have spun into fibres. Robert Dorsch, DuPont's director of biochemical development, says the globules of protein, comparable with marbles in

an egg, are harvested and processed. 'We break open the bacteria, separate out the globules of protein and use them as the raw starting material. With yeast, the gene system can be designed so that the material excretes the protein outside the yeast for better access,' he says.

- F** 'The bacteria and the yeast produce the same protein, equivalent to that which the spider uses in the drag lines of the web. The spider mixes the protein into a water-based solution and then spins it into a solid fibre in one go. Since we are not as clever as the spider and we are not using such sophisticated organisms, we substituted man-made approaches and dissolved the protein in chemical solvents, which are then spun to push the material through small holes to form the solid fibre.'
- G** Researchers at DuPont say they envisage many possible uses for a new biosilk material. They say that earthquake-resistant suspension bridges hung from cables of synthetic spider silk fibres may become a reality. Stronger ropes, safer seat belts, shoe soles that do not wear out so quickly and tough new clothing are among the other applications. Biochemists such as Lewis see the potential range of uses of biosilk as almost limitless. 'It is very strong and retains elasticity; there are no man-made materials that can mimic both these properties. It is also a biological material with all the advantages that has over petrochemicals,' he says.
- H** At DuPont's laboratories, Dorsch is excited by the prospect of new super-strong materials but he warns they are many years away. 'We are at an early stage but theoretical predictions are that we will wind up with a very strong, tough material, with an ability to absorb shock, which is stronger and tougher than the man-made materials that are conventionally available to us,' he says.
- I** The spider is not the only creature that has aroused the interest of material scientists. They have also become envious of the natural adhesive secreted by the sea mussel. It produces a protein adhesive to attach itself to rocks. It is tedious and expensive to extract the protein from the mussel, so researchers have already produced a synthetic gene for use in surrogate bacteria.

*Questions 1 – 5*

Reading Passage 1 has nine paragraphs, **A-I**.

Which paragraph contains the following information?

*Write the correct letter, A-I, in boxes 1-5 on your answer sheet.*

- 1 a comparison of the ways two materials are used to replace silk-producing glands
- 2 predictions regarding the availability of the synthetic silk
- 3 ongoing research into other synthetic materials
- 4 the research into the part of the spider that manufactures silk
- 5 the possible application of the silk in civil engineering

*Questions 6 – 11*

*Complete the flow-chart below.*

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

*Write your answers in boxes 6-11 on your answer sheet.*

Synthetic gene grown in **6** ..... or **7** .....



globules of **8** .....

*Questions 12 – 14*

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

*In boxes 12-14 on your answer sheet, write*

<b>TRUE</b>	<i>if the statement agrees with the information</i>
<b>FALSE</b>	<i>if the statement contradicts the information</i>
<b>NOT GIVEN</b>	<i>if there is no information on this</i>

- 12 Biosilk has already replaced nylon in parachute manufacture.
- 13 The spider produces silk of varying strengths.
- 14 Lewis and Dorsch co-operated in the synthetic production of silk.

## APPENDIX 3: ADDITIONAL SAMPLE ITEMS SHOWING MORE GLOBAL AND/OR INTERPRETATIVE ENGAGEMENTS

### 8 EXTENSION 1 --> LOCAL + INTERPRETATIVE

#### 1.1 Focus on connotative meanings of words

In Passage A, the author refers to X as a “Y” (Line B). This use of the term “Y” suggests that the writer sees X as:

- a** eg a positive development
- b** eg a negative development
- c** eg an expected development
- d** eg an unexpected development

#### 1.2 Focus on author purpose

The writer of Passage A refers to X in Paragraph B, in order to demonstrate:

- a** X is a good thing and should be encouraged
- b** X is a bad thing and should be discouraged
- c** not enough is known about X, and it should be investigated further
- d** sufficient research has been conducted into X

**2 EXTENSION 2 --> GLOBAL/LITERAL****2.1 Focus on macro-content of text (Epistemic entity = argument)**

Which of the following statements best summarises the author's main argument in Reading Passage A:

- a** that X is a good thing, and should be encouraged
- b** that X is not a good thing, and should be discouraged
- c** that X is neither a good thing nor a bad thing
- d** that X is a good thing for some, but not for others.

**2.2 Focus on macro-content of text (Epistemic entity = study)**

Reading Passage A describes a study conducted into X. Which of the following statements best summarises the study's main outcomes:

- a** that X is a Y
- b** that X is not a Y
- c** that X is neither an X or Y

### 2.3 Focus on macro-content of text (Scenario format)

Four different students wrote a one sentence summary of Passage A. Which one most accurately reflects the content of the passage?

- a** The writer discusses the main difficulties of X and describes some of the solutions that have been proposed
- b** The writer discusses the main difficulties of X, and recommends a range of solutions
- c** The writer discusses the main difficulties of X, and suggests that the problems are too difficult to solve
- d** The writer discusses the main difficulties of X, without recommending any solutions

### 2.4 Focus on multiple texts

Consider Reading Passage A and Reading Passage B. The main content difference between these two passages is best summarised as:

- a** Reading Passage A is about X and Reading Passage B is about Y
- b** Reading Passage A is about Y and Reading Passage B is about X
- c** etc

## **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)*. 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

<b>1</b>	<b>Aims .....</b>	<b>272</b>
<b>2</b>	<b>Background and related research .....</b>	<b>272</b>
	2.1 A socio-cognitive test validation framework .....	272
	2.2 Item writing.....	273
<b>3</b>	<b>Research methodology and design .....</b>	<b>277</b>
	3.1 Deduction and induction .....	277
	3.2 Design.....	277
<b>4</b>	<b>Analysis and findings from interviews and focus group discussions .....</b>	<b>279</b>
	4.1 Non-experienced IELTS item writer group .....	279
	4.1.1 IELTS text search, selection and characterisation .....	279
	4.1.2 Participant text search treatment and item development: flowcharts and discussions .....	280
	4.1.3 Participant focus group discussions.....	284
	4.2 Procedures with and findings from the experienced IELTS item writer group .....	287
	4.2.1 Participant text search treatment and item development: flowcharts and discussions .....	287
	4.2.2 Participant focus group discussions .....	297
<b>5</b>	<b>Analysis and findings on the texts .....</b>	<b>298</b>
	5.1 The non-experienced group.....	300
	5.2 The experienced group.....	303
<b>6</b>	<b>Analysis and findings on the editing process.....</b>	<b>308</b>
	6.1 The non-experienced group .....	308
	6.1.1 Choosing the text for the exam.....	310
	6.1.2 Change of view caused by the editing process? .....	311
	6.2 The experienced group.....	312
	6.2.1 Analysis and findings on the items.....	317
<b>7</b>	<b>Comparisons between groups.....</b>	<b>320</b>
	7.1 Item writing processes .....	321
	7.2 The texts .....	322
<b>8</b>	<b>Conclusions and Recommendations .....</b>	<b>324</b>
	References .....	327
	Appendix A: Commissioning letter .....	332
	Appendix B: Background questionnaires .....	333
	Appendix C: Item writer submissions .....	342

## 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 French, 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 i) the complexity of text and ii) the level of processing required to fulfil the reading purpose.

In Weir et al (2009a) IELTS texts were analysed against twelve 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 Haladyna 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 (e.g. 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 (e.g. 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 be 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](http://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 twelve and sixteen 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.

<i>Perceived IELTS text characteristics</i>	<i>Item Writer</i>		
	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.

<b>Text search</b>		
<b>Victoria</b>	<b>Mathilda</b>	<b>Mary</b>
<p>5-6 step flowchart (Victoria thinks now there are more steps than in her flowchart)</p> <ol style="list-style-type: none"> <li>1. task familiarisation</li> <li>2. topic selection (based on knowledge from past papers, website, course books)</li> <li>3. begin task to determine suitability</li> <li>4. research topic to test credibility and usefulness of text</li> <li>5. satisfied with text</li> <li>6. editing text for cohesion and text type</li> </ol> <p>Googled neuro-linguistic programming (NLP) and other potential topics &gt; decided on topic of content of dreams &gt; refining down topic &gt; sub-topics within dreams &gt; other articles &gt; also possible choices? &gt; so settled on the dreams text &gt; tried items out on her EL1 partner; 'apparently NS do really badly on IELTS reading'</p>	<p>5-main steps in flowchart</p> <ol style="list-style-type: none"> <li>1. looking at sample IELTS texts</li> <li>2. browsing for a suitable text</li> <li>3. selection of text from shortlist</li> <li>4. text adaptation</li> <li>5. selecting parts of text to target and writing questions / tasks based on the example of the sample tests</li> </ol> <p>Used practice IELTS tests (and her own experience as a candidate)</p> <p>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'</p>	<p>6-step flowchart:</p> <ol style="list-style-type: none"> <li>1. task assessment,</li> <li>2. background research</li> <li>3. text search and rejection</li> <li>4. text decision and editing</li> <li>5. text review</li> <li>6. item writing and text adjusting</li> </ol> <p>Used IELTS Express, Impact IELTS, past papers, old IELTS copies (Internet)</p> <p>searched under variety of topics, 'try to refine, refine, refine' e.g. science and nature, down to robots, 'using more and more refined words in order to be able to find an article that would be suitable'</p> <p>tested text and items on friend</p>
<b>Text editing</b>		
<b>Victoria</b>	<b>Mathilda</b>	<b>Mary</b>
<p>Believes in significant 'fixing up process' on text</p> <p>Did various things to make the text more academic: took out by-line, added more research-type 'rigour' (e.g. evidence-based), more hedging</p>	<p>Mathilda made hardly any changes: about 3 words</p>	<p>Text editing can mean: 'changing text structure, paragraphing, cohesion'</p> <p>Didn't want to change text too much but one person's 'changing a lot' not another's?</p> <p>Different texts need different amount of changing; editing is relative</p> <p>Is text editing for the sake of the tasks, changing text to fit a task type .... a validity issue?</p>

<i>Item writing</i>		
Victoria	Mathilda	Mary
<p>Knew the 10 task types, returned to IELTS Website hand out re format and stylistic aspects of task types</p> <p>her 'fixing up' of the text 'summons up the kind of task types there are'; so she could see e.g. 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'.</p> <p>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;</p> <p>MCQ items testing whether Cs 'have kept up with the order';</p> <p>linked items with reading purposes e.g. careful reading where you have to 'go back to text and work hard to understand it'</p> <p>MCQ distractors of similar lengths but not necessarily the same style?</p> <p>tried to keep the items in the order of the text as with IELTS</p> <p>wished there were only 3 alternatives; 4th just an 'add on', 'just rubbish', easy for test-taker to spot</p> <p>asks 'can you use words that you know, not in the text'; Must it be in the text? What's the rule?</p> <p>Victoria not much practice in SAQs; too many alternative responses; hard to generate all possible answers</p>	<p>Looked at task types (IELTS website says 10 different types) checked which would suit the text</p> <p>deciding which bits of info in text or which passages to summarise, making decisions on that in parallel; back and forth at same time</p> <p>decided to use matching paras with short summaries task as ...more suitable' for this type of text</p> <p>used true / false / not given task ...'put in a few correct ones, made up a few others' e.g. collapsing info 'that did not really go together ...' to reveal lack of understanding</p> <p>Tested vocab. e.g. 'if you don't know that adjacent means next then you don't know whether that info is correct or not...' i</p> <p>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'</p> <p>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'</p> <p>marked up text suitable for items i.e. 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'.</p> <p>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)</p> <p>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'.</p>	<p>Matching task (paras with researcher names) selected to test summary of main text topics</p> <p>summary completion task suited density of description of an experiment</p> <p>short paraphrasal text with candidates to use words from text in new context, to check their understanding</p> <p>didn't just want to test vocab. meaning; tried to elicit specific answers</p> <p>favoured the control offered by multiple choice (MCQ) but now felt she should have been more careful in designing distractors</p> <p>often had difficulty finding the 4th alternative</p> <p>should there be distractors not actually in the text but from test designer's mind?</p> <p>should we actually add to text to get distractors? Mary thinks no as it impairs authenticity</p> <p>never threw any questions away, but did dispense with 'a couple of distractors'</p> <p>IELTS items do not have to be in the order the item topic appears in the text?</p>

	<p>Threw away only one or two items but modified about half or her original items</p> <p>Thought the Website said all the items are in the order they are in in the text</p> <p>Short answer questions (SAQs) may be good for definitions, too</p>	
<b>Which of your sections you happiest with?</b>		
<b>Victoria</b>	<b>Mathilda</b>	<b>Mary</b>
<p>likes her T/F NG – it works</p> <p>Stylistically her MCQ wrong because the items are of uneven length, though the questions are ‘sort of OK’</p> <p>In her SAQs she is not convinced the answers are the only ones possible</p>	<p>MCQ strongest, not a NS so can ‘imagine what it’s like’ so easier to ‘make up the wrong ones’!</p> <p>Task type 7, summary info to match paras, too vague, so her worst</p>	<p>matching (sentences to researcher names) the best.</p> <p>summary completion task the easiest to write so perhaps the worst!</p> <p>MCQ task actually the worst because of her difficulty finding the final distractors</p> <p>summary completion the easiest – so the worst No her first section (the matchings)</p>

**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, i.e. 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)		STRAIGHT-FORWARD
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 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.

<b>Perceived IELTS text characteristics</b>	<b>Item Writer</b>			
	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.

<b>Overview of the Item Writing Process</b>			
<b>Jane</b>	<b>Anne</b>	<b>William</b>	<b>Elizabeth</b>
<p><b>6 step flowchart:</b></p> <ol style="list-style-type: none"> <li>1. Refer to commissioning letter to identify topics to avoid, sections needed (10 mins)</li> <li>2. Finding possible sources, read quickly to decide whether possible (1hr-2hrs)</li> <li>3. Collect likely sources and read again – topic suitability, suitable for task types, enough testable material (1hr)</li> <li>4. Start cutting to appropriate length, identifying information to test and which parts go with which item types (1hr-2hrs)</li> <li>5. Work on tasks, amending and cutting text as needed to fit tasks (1-2hrs per task type)</li> <li>6. First draft – check that tasks work, check for overlap between items, cut to word limit (1hr)</li> </ol>	<p><b>11 step flowchart:</b></p> <ol style="list-style-type: none"> <li>1. Text sourcing: check in files, investigate previously fruitful websites, Google a topic suggested in commission or that seems promising (30 mins-1 day)</li> <li>2. Careful reading (30 mins)</li> <li>3. Typing up with amendments (1 hr)</li> <li>4. Length adjustment (to target plus 100-200 words) (15 mins)</li> <li>5. Work on first (most obvious) task type (30 mins–2hrs [for MCQ])</li> <li>6. Mark up further areas of text for suitable items (30 mins)</li> <li>7. Work on further tasks – amending text as necessary (1hr-2hrs)</li> <li>8. Print off and attempt tasks (30 mins-1hr)</li> <li>9. Write answer key (10 mins)</li> <li>10. Check length and prune if necessary (10 mins-1hr)</li> <li>11. Review and proof read (10mins-30mins)</li> </ol> <p>Found text already in her file (keeps an eye on potential sources) – looking for a Section 1 (relatively easy) task</p>	<p><b>11 step flowchart:</b></p> <ol style="list-style-type: none"> <li>1. Think of subject – look at own books and articles for inspiration</li> <li>2. Google possible topics</li> <li>3. Locate a text and check suitability – how much needs glossing, any taboo subjects?</li> <li>4. Consider whether text will work with task types</li> <li>5. Scan or download text</li> <li>6. Edit text to roughly to required length (or slightly longer), modifying to keep coherence</li> <li>7. Choose and draft first task, modifying text to fit (abandon task if necessary)</li> <li>8. Prepare other tasks</li> <li>9. Revise text for coherence, length, to fit tasks, adapting tasks at the same time as needed</li> <li>10. Have a break</li> <li>11. Check and revise text and tasks</li> </ol> <p><b>Timings</b></p> <p>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</p>	<p><b>10 step flowchart:</b></p> <ol style="list-style-type: none"> <li>1. Keep eyes open for texts</li> <li>2. Choose from available texts</li> <li>3. Evaluate selected text</li> <li>4. Summarise main points and edit out redundant/inappropriate material</li> <li>5. Identify possible task types</li> <li>6. Write items</li> <li>7. Cut text to required length</li> <li>8. Tidy up text and items checking keys</li> <li>9. Leave for a day, print out and amend as needed</li> <li>10. Send off</li> </ol> <p>No timings given</p>

<b>Text search</b>			
<b>Jane</b>	<b>Anne</b>	<b>William</b>	<b>Elizabeth</b>
<p>'I don't normally have texts waiting'</p> <p>'I have certain sources that I go to regularly'</p> <p>There were quite a few texts and I made a decision</p> <p>'Texts are nearly always nearly three or four times the length we will need'</p> <p>If I can't understand a text, I wouldn't use it</p> <p>Opinion texts are more difficult to find</p> <p>'You can't assume that the candidates are specialists'</p>	<p>Sometimes has to reject texts at her typing stage 'one of the hardest decisions to take'</p> <p>'I think sometimes being interested in a text is a positive disadvantage'</p> <p>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'</p> <p>It may not be for the academic reader, it may be for the interested layperson... if the complexity is right</p> <p>The over complexity is when the research itself or the topic itself needs so much specialist language</p>	<p>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'</p> <p>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 and have an academic tone.</p> <p>'Adapt the text to fit the tasks', don't see the text as 'sacrosanct'</p> <p>'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.'</p> <p>Expressing a number of ideas and opinions, which would make it a Section 3</p> <p>If its fairly factual more Section 1</p> <p>Genuine academic texts are unsuitable because they assume too much knowledge and would require too much explanation</p> <p>I try and make sure that I understand it and can make it comprehensible</p>	<p>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]'</p> <p>Need to edit out references to 'amazing', 'surprising' or 'incredible' information in journalistic text</p>

<i>Item writing</i>			
Jane	Anne	William	Elizabeth
<p>I think I make a decision fairly early on about which task type I will use</p> <p>I decided this particular text was suitable for certain task types</p> <p>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</p> <p>With discrete information you can make it work with that</p> <p>Multiple choice questions fit best with an opinion text</p>	<p>Headings are difficult</p> <p>True-false is usually quite easy</p> <p>I don't like doing the diagram type ones or flowcharts</p> <p>Quite often in articles you get a little diagram, but it's too complex or guessable</p> <p>I read a lot of texts and cut them down before I decide which one to use.</p>	<p>My first main thing is how well the tasks fit that text. Chooses tasks that 'leap out from the text'</p> <p>Not something that could be answered by someone who knows the subject</p> <p>Considers which tasks pay more, which are worth the effort and so avoids MCQ if possible.</p> <p>Factual information you can test with true false not given</p> <p>We need to cover the whole text – every paragraph is tested</p> <p>A text ought to lend itself to having a topic in each paragraph that can be captured in a heading</p> <p>I think the paragraphs overlapped in this case</p> <p>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</p> <p>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</p> <p>I work on items at an early stage and will dump a text after ten minutes if I feel it will not work.</p>	<p>I think multiple choice can work across a range of texts including at a more basic factual</p> <p>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 people's ideas</p> <p>I often write notes on texts before deciding which one to use.</p>

<b>Which of your sections you happiest with?</b>			
<b>Jane</b>	<b>Anne</b>	<b>William</b>	<b>Elizabeth</b>
		<p>Unusually, I wrote all three tasks simultaneously</p> <p>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</p> <p>Paragraphs F and H each only have one item, which is not ideal</p> <p>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</p>	<p>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.</p> <p>This text does have one basic idea and really the whole text is saying that. I was testing the support for the idea.</p> <p>There is a stage when I think 'this is going to work and I'm not going to dump this.</p> <p>I thought there were enough discrete words that would make a key to support multiple choice.</p> <p>I am very conscious of how much of a text I am exploiting.</p>

**Table 2. Experienced Participants' descriptions of the item writing processwriting 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 (William)	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)		STRAIGHT-FORWARD
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.

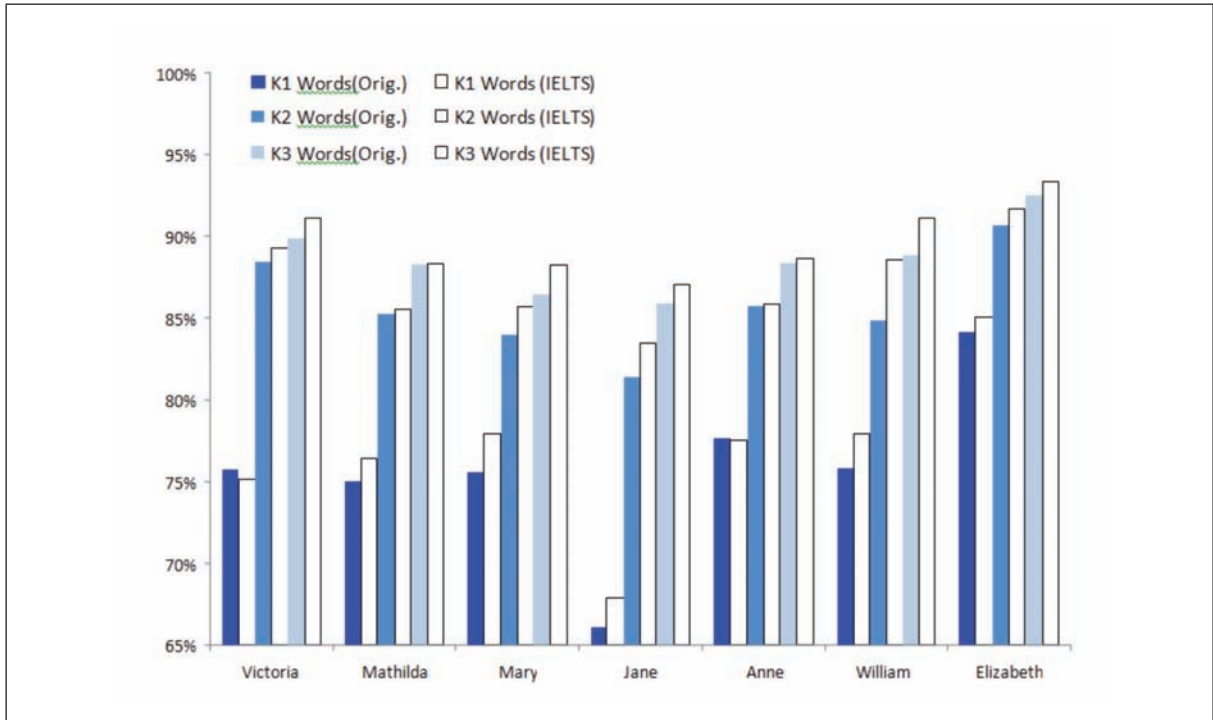


Figure 1. Results of word frequency analyses for original source texts and adapted IELTS text: percentage of very frequent words at the BNC 1,000, 2,000 and 3,000 word frequency levels

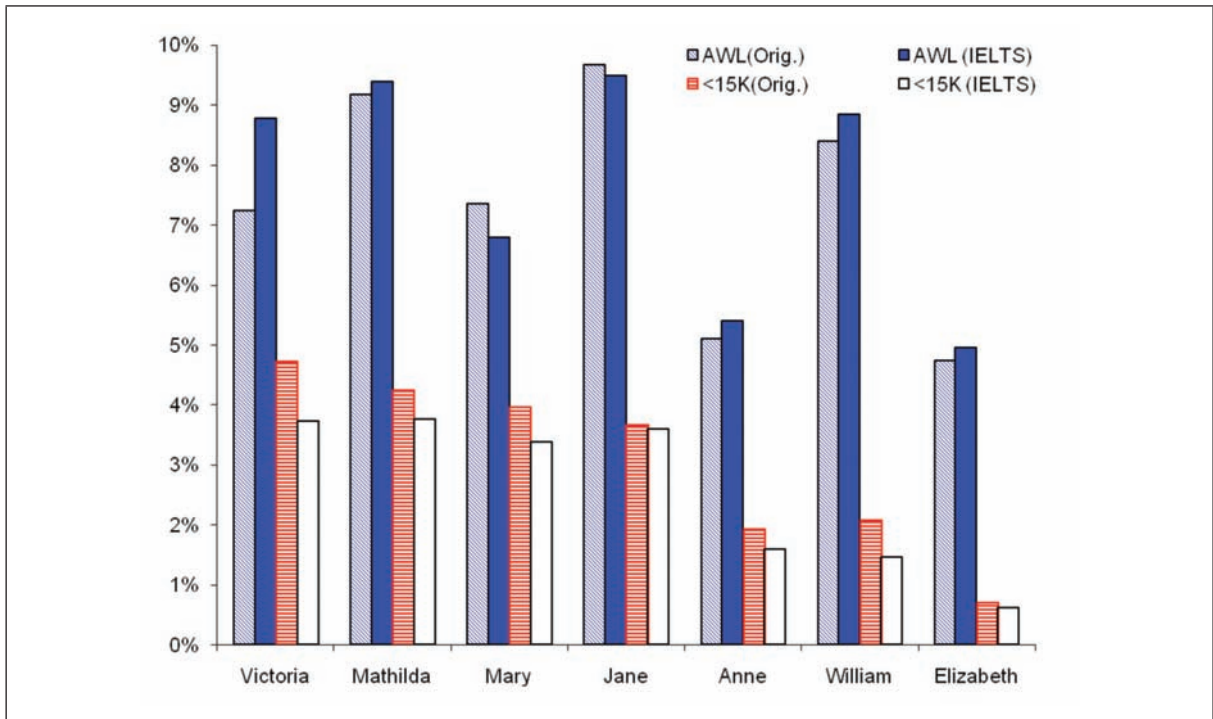
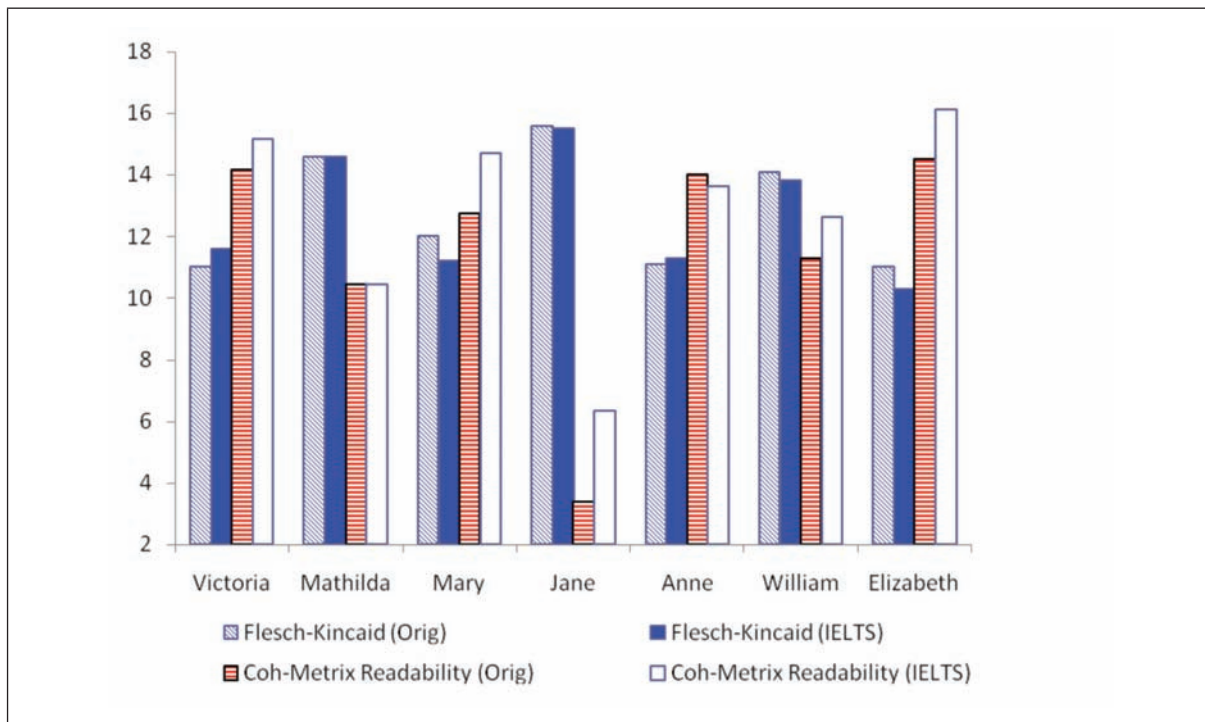


Figure 2. Results of word frequency analyses for original source texts and adapted IELTS text: percentage of sub-technical academic (AWL) and very infrequent words



**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: <http://www.msnbc.msn.com> published online Oct. 12, 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 3,000 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: <http://discovermagazine.com>, published online October 8, 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 05 April 2008**

As noted in Section 5 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 Section 6 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 1,592 to 2,518 words, the source texts used by the experienced writers were all very much longer than those of the non-experienced group (748 to 1,094 words). At 1,870 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 8,000 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 ('10m to 40mm' is retained to describe a camera lens focal range, and sound is 'sampled at 20kHz for up to 4h 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 (1,606 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 'sensitivity' 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 1,592 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 e.g. respect and love, were discussed.
- the discussion here raised a multiple choice task issue of whether all alternatives should be similar in function, e.g. 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 to 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 it 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 six, 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		√			√	√	√

<b>Task Types</b>		<b>Victoria</b>	<b>Mathilda</b>	<b>Mary</b>	<b>Jane</b>	<b>Anne</b>	<b>William</b>	<b>Elizabeth</b>
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](http://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 text from the experienced group; Mary's robot's 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](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 <http://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 <http://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 <http://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 - e.g. 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](http://www.cambridgeesol.org/teach/ielts/academic_reading/index.htm)

## APPENDIX B: BACKGROUND QUESTIONNAIRES

	<i>Victoria</i>	<i>Mathilda</i>	<i>Mary</i>
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	<p>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.</p> <p>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.</p> <p>Interestingly, native speakers are reputed to not score well on IELTS reading</p>	<p>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.</p>	<p>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.</p> <p>I find the reading topics to be questionable sometimes.</p>

**Table 10. Non-Experienced Item Writers**

	<i>Jane</i>	<i>Anne</i>	<i>William</i>
<b>Qualifications</b>	BA (Hons)	MA Modern Languages RSA Dip. TEFLA	BA (Hons) in English Language and Literature MA in Modern English Language RSA Dip TEFLA
<b>Experience in EL</b>	RSA Diploma in Teaching English as a Foreign Language to Adults (Dip TEFLA) 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
<b>Experience of IELTS</b>	7 commissions	Can't remember - 4 or 5 years	15+ commissions Published IELTS preparation book
<b>Training as item writer</b>	Editing meetings Formal training session at Cambridge ESOL		Training sessions plus informal training through participation in editing meetings
<b>Background and training you feel have most helped you in your item writing for the academic reading test</b> <b>(Please explain how this training has helped you)</b>	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 e.g. 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. <b>My academic background has been helpful for selecting and understanding texts.</b>



<b>Elizabeth</b>	<b>UK1</b>	<b>AUS1</b>	<b>NZ1</b>
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; e.g. 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.

	<i>Jane</i>	<i>Anne</i>	<i>William</i>
<b>Do you feel that your item writing for the academic reading test has improved with experience?</b>	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.
<b>Has the proportion of successful reading commissions increased in that time?</b>	This has remained fairly constant.	Yes	Most have been accepted
<b>IELTS support materials (e.g item writer guidelines) you consult before or while you select texts and write test items:</b>	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
<b>What additional support materials might assist you with text selection, editing and item writing?</b>	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 i.e. texts and tasks that have worked well	None
<b>Do you see your IELTS academic reading material in the meetings below? If so, how much of your material do you see at each? (e.g. If you see all your material edited at editing meetings, you should put 100%)</b>			
<b>pre-editing meetings:</b>	Writers don't always attend these		No
<b>editing meetings:</b>	I have attended one pre-editing meeting since I started	100%	No
<b>pretest review meetings:</b>	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%
<b>How would you describe the experience of attending these IELTS meetings?</b>	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.

<b>Elizabeth</b>	<b>UK1</b>	<b>AUS1</b>	<b>NZ1</b>
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 e.g. 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 e.g. 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.

	<i>Jane</i>	<i>Anne</i>	<i>William</i>
<b>What do you think are the key characteristics of more/less successful IELTS academic reading items?</b>	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 (e.g. 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
<b>What do you think are the characteristics of more/less successful IELTS academic reading item writers?</b>	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
<b>What do you enjoy/ dislike about IELTS academic reading item writing work?</b>	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
<b>Which aspects do you find easiest/ most challenging?</b>	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, e.g. flow-chart, 5-option multiple choice that isn't too easy

<b>Elizabeth</b>	<b>UK1</b>	<b>AUS1</b>	<b>NZ1</b>
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.

	<i>Jane</i>	<i>Anne</i>	<i>William</i>
<b>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?</b>	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
<b>What changes, if any, would you like to see in the IELTS academic reading test and why?</b>	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
<b>Other comments on IELTS academic reading module item writing.</b>		It's certainly a challenge at times but enjoyable on the whole.	None

**Table 11. Experienced Item Writers**

Elizabeth	UK1	AUS1	NZ1
<p>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 (e.g. use of contents/index or reading at chapter/whole book level) or text evaluation skills.</p> <p>Appropriate humanities texts (e.g. 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.</p>	<p>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; e.g. 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.)</p> <p>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.</p> <p>(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.)</p>	<p>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.</p>	<p>I think it's a good reflection.</p>
<p>It would be interesting to look at the possibility of varying text length - e.g. 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.</p>	<p>Perhaps a greater variety of text types and fewer potential short cuts for candidates. (The latter is really an editing issue).</p>		
	<p>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.</p>		-

## 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**

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

- 1 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 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 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 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 The CHP plant will generate electricity and hot water by using
- A wind turbines.
  - B solar technology.
  - C biomass.
  - D fossil fuel.
- 6 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.

**Answer Key**

- 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

**Answer Key**

- 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.

**Answer Key**

- 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 *robot*, 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

### Matching

### Questions 1–5

*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

### Summary Completion

#### Questions 6–13

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. \_\_\_\_\_.

### Multiple Choice

#### Questions 14–17:

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

### Matching

#### Questions 1-5

- 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

### Summary Completion

#### Questions 6-13

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.

**Multiple Choice**

***Questions 14-17***

*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

**17** 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 4?h 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 1?*

*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**

**Background to the study**

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

**2005 -2006 season**

- 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 ?



## Answer 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.

### **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.

**Answer 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

**Answer Key**

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

## 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

2 D

3 B

4 B

5 A

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

## **6 The cognitive processes of taking IELTS Academic Writing Task One**

**Authors**

**Guoxing Yu**

**Pauline Rea-Dickins**

**Richard Kiely**

University of Bristol

Grant awarded Round 13, 2007

An investigation of the cognitive processes of candidates taking IELTS Academic Writing Task One (AWT1) with different graphic prompts at two time points – before short training on how to do AWT1 tasks, and after. It explores the extent to which candidates' cognitive processes are affected by the use of different graphs, their graphic skills and English writing abilities, and by the training.

### **ABSTRACT**

This research investigated the cognitive processes of candidates taking IELTS Academic Writing Task One (AWT1) with different graphic prompts at two different time points – before short training on how to do AWT1 tasks, and post-training. It explored the extent to which candidates' cognitive processes are affected by the use of different graphs, their graphic skills and English writing abilities, and the short training. A grounded and multi-layered case study approach was employed to collect data on candidates' cognitive processes. 24 intending IELTS candidates from a large Chinese university completed eight AWT1 tasks while thinking aloud their processes of doing the tasks (four before training and four after training) under examination conditions. Samples of their English writing abilities and graphicacy were also collected, as well as post-task interviews with all participants.

The think-aloud protocols were analysed to identify the common patterns of cognitive processes. A model of cognitive processes was developed, consisting of three interrelated stages – comprehending non-graphically presented task instructions, comprehending graphic information and re-producing graph comprehension in written discourse in English as a foreign language. This model guided our analyses to address the four research questions: (1) How the participants processed the graphic information and how they followed the graphic conventions to re-produce their graph comprehension in written discourse in English were affected by the types of graphs they read. Such effects of different graphic prompts on the cognitive processes were clearly evidenced in the mean scores of the writings, in the use of vocabulary, and in whether and how they would make comparisons or trend assessments, following the graphic conventions in presentation, interpretation and re-production. (2) Although graph familiarity did not seem to affect task performance in terms of the marks of the writings, the participants clearly expressed some potential psychological impact of graph familiarity on their task performance. (3) There is a strong correlation between AWT1 writing performance and writing ability as measured via topic-based argumentative essays. (4) The influence of the special training was strong,

in particular, in terms of whether or not the participants tried to make interpretations, predictions and comments by linking the graphic information with their domain knowledge about the graphs.

The implications of these findings are discussed with reference to AWT1 task design, as well as other language test tasks that use graphs as prompts, particularly for listening, speaking and writing assessments.

## AUTHORS BIODATA

### Guoxing Yu

Dr Yu is Lecturer at the Graduate School of Education, University of Bristol. He earned his PhD in 2005 from Bristol; his dissertation was awarded the Jacqueline A. Ross TOEFL Dissertation Award by Educational Testing Service USA (2008). He was the Principal Investigator of this research project. He publishes in academic journals including *Applied Linguistics*, *Assessing Writing*, *Assessment in Education*, *Educational Research*, *Language Assessment Quarterly*, and *Language Testing*.

**Address:** Graduate School of Education, University of Bristol, 35 Berkeley Square, Bristol, BS8 1JA, UK.

**Email:** Guoxing.Yu@bristol.ac.uk

### Pauline Rea-Dickins

Dr Rea-Dickins is Professor of Applied Linguistics at the Graduate School of Education, University of Bristol. She has published widely in areas of language testing and assessment, particularly classroom-based assessment and language programme evaluation. She is Director of a major ESRC/DfID research project (2007-2010, RES-167-25-0263): *Student Performance in National Examination: the dynamics of language* (SPINE; [www/bristol.ac.uk/spine](http://www/bristol.ac.uk/spine)), which is a collaboration between Bristol and the State University of Zanzibar.

**Address:** Graduate School of Education, University of Bristol, 35 Berkeley Square, Bristol, BS8 1JA, UK.

**Email:** P.Rea-Dickins@bristol.ac.uk

### Richard Kiely

Dr Kiely is Reader of Applied Linguistics at the Graduate School of Education, University of Bristol. His research interests include language programme evaluation, language teaching and teacher development, and language learning explored from language socialisation and identity perspectives.

**Address:** Graduate School of Education, University of Bristol, 35 Berkeley Square, Bristol, BS8 1JA, UK.

**Email:** R.Kiely@bristol.ac.uk



## CONTENTS

<b>1</b>	<b>Background and rationale.....</b>	<b>376</b>
1.1	Introduction .....	376
1.2	Dearth of research into test takers' cognitive processes of completing AWT1 .....	376
1.2.1	Theories of graph comprehension in cognitive psychology and their implications for research into integrated writing tasks using graph prompts in language tests.....	377
<b>2</b>	<b>Research aims and questions .....</b>	<b>380</b>
2.1	Research aims .....	380
2.2	Research questions .....	380
<b>3</b>	<b>Research design and methodology .....</b>	<b>380</b>
3.1	Approach.....	380
3.2	Participants .....	381
3.3	Data collection .....	382
<b>4</b>	<b>Analyses and findings .....</b>	<b>385</b>
4.1	Participants' graph familiarity and English writing abilities.....	385
4.1.1	Participants' graph familiarity .....	385
4.1.2	Participants' writing performances under normal examination conditions .....	386
4.1.3	Participants' writing performances under think-aloud conditions .....	387
4.2	Developing a working model of cognitive processes for AWT1 tasks.....	388
4.2.1	Comprehending non-graphically presented task instructions .....	389
4.2.2	Comprehending graphic information.....	390
4.2.3	Re-presenting graphic and non-graphic information in continuous discourse .....	390
4.2.4	Summary of the working model .....	391
4.3	Addressing the four research questions .....	393
4.3.1	Research question 1 .....	393
4.3.2	Research question 2 .....	397
4.3.3	Research question 3 .....	400
4.3.4	Research question 4 .....	403
<b>5</b>	<b>Conclusion.....</b>	<b>406</b>
	References .....	410
	Appendices .....	414

## BACKGROUND AND RATIONALE

### 1.1 Introduction

This study relates to the first broad area of interest identified by the IELTS Joint Research Committee Call for Proposals 2007/2008 (Round 13), namely “test development and validation issues”.

In particular, this research investigated “the cognitive processes of IELTS test takers” when completing Academic Writing Task One (AWT1). The cognitive processes of taking AWT1 were examined at two different points – one before and one after special short training on how to achieve the best performance in AWT1. As such, this research is also linked to “issues in test impact” in relation to “test preparation practice” – another area of interest identified by the IELTS Joint Research Committee.

This section provides the theoretical rationale and background for this study with specific reference to the effects of the features of graphs and test takers’ “graphicacy” (Wainer 1992, p 16) on their AWT1 performances, as explained below. Details of the research design and methodology are provided from Section 2 onwards.

### 1.2 Dearth of research into test takers’ cognitive processes of completing AWT1

In IELTS AWT1 tasks candidates are asked to “describe some information (graph/chart/table/diagram), and to present the description in their own words”. It is suggested that candidates should spend 20 minutes on this and write at least 150 words. Candidates are assessed on their ability to organise, present and possibly compare data, describe the stages of a process or procedure, describe an object or event or sequence of events, or explain how something works (IELTS Handbook 2006, p 8). AWT1, therefore, can be considered as an integrated writing task, requiring candidates not only to comprehend the graph input, but also to re-present in written English the information accessible to them (various terms such as chart, graph and diagram have been used interchangeably in research; see Friel, Curcio, & Bright 2001, Fry 1981, Wainer 1992). The term “graph” is probably the most widely used in applied cognitive psychology, the key knowledgebase upon which this research will draw, and we will use “graph”, hereafter, to represent all the other three terms - table/chart/diagram - that the IELTS Handbook (2006) has used.

Graph comprehension is a *sine qua non* for successful performance of the writing task. As a result, the variability in the graph input and the candidates’ different familiarities and proficiencies in comprehending the graphs may pose a threat to the validity of AWT1 as a measure of the candidates’ academic writing abilities. Surprisingly, to the best of our knowledge, we notice that only two IELTS funded research projects (Mickan, Slater, & Gibson 2000; O’Loughlin & Wigglesworth 2003) have so far investigated some of these issues. As only a very small part of their research focus, Mickan et al. (2000) investigated how test takers interpreted AWT1 task prompts and planned their writings, but they did not examine the effects of characteristics of graphs on the process or the product of the AWT1 tasks. O’Loughlin and Wigglesworth (2003) examined the extent to which the difficulty of AWT1 was affected by the quantity and the manner of presentation of information in graphs. As a primarily product-oriented study via the analyses of the written scripts, they found that the writings produced in the tasks with less information in the graphs were more linguistically complex than those writings produced in the tasks with more information in the graphs, irrespective of the participants’ language proficiency level. The results also indicated that there were no substantial differences in the difficulty across the tasks which varied in terms of the quantity and the manner of information presented in the graphs. However, we should point out that only three types of graphs (i.e., statistical table, bar chart,

line graph) were used in O’Loughlin and Wigglesworth (2003). These graphs were also limited to two topics (i.e., number of people attending places of entertainment, women and men in postgraduate studies). All these put in question the generalizability of their finding that there are no substantial differences in the difficulty across the AWT1 tasks. Would similar findings be observed if other types of graphs (e.g., diagrams depicting the sequence of events and statistical tables in conjunction with other visual presentations) of various topics had been used in their study?

Furthermore, “we have no way of knowing what exactly was attended to by the participants while planning for and completing the tasks” (Xi 2005, pp 496-497), and how exactly the different quantity and manner of presentation of the information in the graphs affected the participants’ test taking processes. Compared to the enormous quantity of funded research into IELTS Academic Writing Task Two (e.g., Mickan & Slater 2003; Mickan, et al 2000; Moore 1993), the dearth of research into AWT1 is striking. Indeed, in the field of language testing, research into the use of graphs in writing or speaking tests is only recently emerging (i.e., Katz, Xi, Kim, & Cheng 2004; Xi 2005), although graphs have been as test prompts for years and a rich knowledgebase in applied cognitive and educational psychology has much to offer language testers to understand the relevant issues. As Xi (2005) commented (see above), these few studies in language testing are very much product-oriented and are not very helpful in understanding test takers’ cognitive processes when completing tasks using graphs as prompts. Therefore, the need to gain better understandings of the cognitive processes in taking AWT1 is also compelling, in order not only to understand the validity of the AWT1 *per se* but also to make important contribution to the current debate on the effects on language test performance of the features of graphs and test takers’ graphicacy, i.e., “proficiency in understanding quantitative phenomena that are presented in a graphical way” (Wainer 1992, p 16).

### 1.2.1 Theories of graph comprehension in cognitive psychology and their implications for research into integrated writing tasks using graph prompts in language tests

AWT1, as an integrated writing task, involves two basic processes – the comprehending of the information presented in graphs and the re-presentation of the information from graphs in continuous written discourse. We, thus, identify two main areas of research that inform this study – graph theories in cognitive psychology and second/foreign language writing processes. However, in this report we draw more on the findings in graph theories in cognitive psychology than on second language writing processes, because we believe graph theories have been somewhat neglected in studies about the processes and the assessment of second/foreign language writing. In this section, we briefly review the key factors in graph comprehension in psychology and statistics/mathematics education pursuant to a cognitive approach (as opposed to the view of graph comprehension as a social practice, in which graphs are seen not to have meaning *a priori*, but rather the meaning of graphs arises from the contexts of use [Roth 2002, 2003]). We will discuss the implications of these theories for language test development and validation, with specific reference to IELTS AWT1.

In cognitive psychology, several models or frameworks of graph comprehension (e.g., Carpenter & Shah 1998; Freedman & Shah 2002; Guthrie, Weber, & Kimmerly 1993; Hollands & Spence 1998, 2001; Körner, 2004; Lohse 1993; Peebles & Cheng 2002, 2003; Pinker 1990; Schnotz, Picard, & Hron 1993; Shah, Freedman, & Vekiri 2005) have been proposed. See also Pinker (1990) and Shah and Hoeffner (2002) for some reviews of graph comprehension models.

Among these models, we find that the knowledge-based construction-integration model proposed by Shah and associates (e.g., Carpenter & Shah 1998; Freedman & Shah 2002) is the most illuminative and probably most directly relevant to research on integrated writing assessment using graphs

as prompts, because the knowledge-based model is analogous to Kintsch's (1988) Construction-Integration model of text comprehension. Graphs are a special kind of text, and graph comprehension is subsumed into a more general activity of text comprehension. Carpenter and Shah (1998) consider graph comprehension to be an integrated sequence of interactions between conceptual and perceptual processes, e.g., "pattern-recognition processes that encode graphic patterns, interpretive processes that operate on those patterns to retrieve or construct qualitative and quantitative meanings, and integrative processes that relate these meanings to the referents inferred from labels and titles" (p 75). Freedman and Shah (2002) developed further this knowledge-based model of graph comprehension. According to the knowledge-based model, graph comprehension is influenced not only by the display characteristics of a graph, including, for example, different types of graphs, their cognitive demands (Lewandowsky & Spence 1989), and the relative importance and relationships of graphical elements (e.g., Feeney, Holo, Liversedge, Findlay, & Metcalf 2000; Shah & Carpenter 1995; Shah, Mayer, & Hegarty 1999; Simkin & Hastie 1987; Zacks & Tversky 1999), but also a viewer's domain knowledge (e.g., Curcio, 1987), graphical literacy skills, and explanatory and other scientific reasoning skills [e.g. spatial reasoning, (Feeney, Adams, Webber, & Ewbank 2004; Stewart, Hunter, & Best 2008; Trickett & Trafton 2004, 2006)], and incremental interactions between these factors.

Further empirical evidence supports this knowledge-based model of graph comprehension. Specifically, in terms of the effects of the features of graphs on information extraction and processing, Meyer, Shinar and Leiser (1997) compared the relative efficiency of line graphs, bar graphs and tables, in different types of information processing tasks, and found that bar graphs had the advantage for reading exact values and identifying maxima, whereas line graphs had the advantage for reading trends. Furthermore, participants also tended to describe concrete contrasts in data presented in bar graphs (e.g., higher, lower, greater than, less than); whereas when they saw line graphs, they tended to describe trends (e.g., rising, falling, increasing, decreasing). In other words, it seems that there may be certain innate features and cognitive naturalness or conventions of different types of graphs that govern how readers would interpret and extract information from graphs. Even the same type of graphs may lead to different cognitive demands, for example, horizontal bar graphs were found to require longer decision making times than vertical bar graphs, and similarly negative number graphs required longer decision making times than positive number graphs (Fischer, Dewulf, & Hill 2005). Furthermore, as Pinker (1990) pointed out: "different types of graphs are not easier or more difficult across the board, but are easier or more difficult depending on the particular class of information that is to be extracted" (p 111). However, overall, as Carpenter and Shah (1998) noted, "even relatively simple graphical displays require relatively complex cognitive processes" (pp 98-99). Pinker's model of graph comprehension predicts that it would be easier to make discrete comparisons between individual data points from bar graphs, using terms such as higher, lower, greater than, and less than; and easier to assess trends from line graphs, using terms such as rising, falling, increasing, and decreasing. Zacks and Tversky (1999) confirmed that readers had a strong tendency to describe discrete comparisons when they saw bar graphs, and describe trends when they saw line graphs - a phenomenon that Zacks and Tversky called "bar-line message correspondence". They also found that people produced bars to depict discrete comparisons and lines to depict trends - a phenomenon that they called "bar-line data correspondence" (p 1077). These two correspondences in comprehension and production of graphs conform to the principles of "cognitive naturalness" of using space to convey meaning in graphic communication (Tversky 1995).

This knowledge-based model of graph comprehension helps to explain the differences between novices and experts in comprehending graphs, and the relative ease or difficulty of graph comprehension encountered by these two groups. Similarly, Friel, Curcio and Bright (2001)

identified four critical factors that appeared to influence comprehension of statistical graphs: (a) the purposes for using graphs, e.g., whether for analyses or communication (Kosslyn 1989), (b) task characteristics (see Simkin & Hastie, 1987), (c) discipline characteristics, e.g., spread and variation with a dataset, the type and size of the data, and the way a representation provides a structure for data (i.e., graph complexity), and (d) characteristics of graph readers, including their prior knowledge or bias of the graphic information (Vernon 1946). These four critical factors are congruent to the key components of the knowledge-based construction-integration model described earlier. The important role that the characteristics of graph readers can play in graph comprehension has been widely supported with empirical evidence. For example, Carpenter and Shah (1998) noted that “individual differences in graphic knowledge should play as large a role in the comprehension process as does variation in the properties of the graph itself” (p 97). With reference to students’ academic achievements, Åberg-Bengtsson and Ottosson (2006) noted that graphic knowledge had the strongest correlation with mathematic/science achievements.

However, Roth (2002) suggested that graphic knowledge or expertise was more complex and we need to move beyond thinking about familiarity because not only knowledge but also experience and expertise of the phenomenon depicted in the graphs affects comprehension. He categorized graph comprehension along a continuum from “transparent”, “competent” to “problematic” readings of graphs, based on the familiarity and knowledge of the graph readers. In “transparent reading”, the familiar graph provides the readers with a “transparent window onto a familiar world”, the graph and the phenomenon have “fused” (p 5). “Readers no longer think of words, or parts of a line curve, but go directly to the things they know them to be about” (p 6). In “competent reading”, graphs and their topics are less familiar, “more work is required on the part of a reader” (p 6). In “problematic reading”, “people are unfamiliar with graphs, phenomena, or the translation between the two, problems in reading become apparent. Most of the reading activity is then concerned with structuring the graph (and accompanying text) itself [our emphasis] rather than with relating it to some phenomenon” (pp 8 9).

In the field of second or foreign language testing, there are very few studies that address the effects of visual input, such as graphs, pictures and diagrams, on test performance, although graphs have been used quite widely as test prompts. Furthermore, we noticed that among these few studies that addressed the effects of graphic inputs on test performance, they were often conducted in the context of listening (e.g., Ginther 2002; Gruba 1997) and speaking assessments (e.g., Katz et al 2004; Xi 2005), and much less in writing assessment, except for O’Loughlin and Wigglesworth (2003). The TOEFL Program reported that it discontinued temporarily the chart-graph format with explicit comparison statement in its Test of Written English pending further investigation in early 1990s (Golub-Smith, Reese, & Steinhaus 1993), due to the findings that this format (see *ibid.* 18) produced the highest mean scores, compared to other formats without graphs. However, the findings of these studies in the field of language testing have provided empirical evidence of the potential effects of test takers’ graphicacy on test performance. For example, Xi (2005) found that test takers’ graph familiarity affected their oral reproduction of the information accessible to them in graphs and that graph familiarity had differing effects on the holistic scores of the speaking test based on the bar and line graph tasks. The individual and probably idiosyncratic differences in terms of the effects of the features of graphs and test takers’ graph familiarity raise questions on the validity and fairness of such tasks. Although the analyses on the effects of the quantity and the manner of presentation of information in graphs, as done by O’Loughlin and Wigglesworth (2003), shed some light on the validity of the IELTS AWT1, such studies have not provided insights into the ways the characteristics of the graphs and the test takers’ graphicacy and the interactions of these factors contribute to the test

takers' performance. The validity of AWT1 tasks should be investigated not only by analysing the written scripts of the tasks but also test takers' cognitive processes when doing the tasks. This, then, has both informed and shaped the focus of this research.

## 6 RESEARCH AIMS AND QUESTIONS

### 2.1 Research aims

The overall aim of this study is to explore the cognitive processes of intending IELTS test takers completing AWT1 that uses graphs as test prompts. It aims to understand the validity of AWT1 in a dynamic, rather than “one-off” approach. In other words, this validation study aims to explore the cognitive processes at different time points - before and after test preparation activities. As such, this study will thus also develop insights into the impact of test preparation practice on task performance, so as to explore the validity and fairness issues of intensive preparation for AWT1 tasks. In addition, the possible differential effects of test preparation activities on test takers' cognitive processes will provide invaluable information for understanding the dynamics of the cognitive processes involved (Carpenter and Shah 1998).

### 2.2 Research questions

The study addressed the following research questions (RQ):

**RQ1:** To what extent are there differences in the candidates' cognitive processes due to different AWT1 prompts?

**RQ2:** To what extent are the candidates' cognitive processes affected by their graphicacy?

**RQ3:** To what extent are the candidates' cognitive processes related to their writing abilities?

**RQ4:** To what extent are the candidates' cognitive processes influenced by test preparation opportunities offered to them by the research team?

## 7 RESEARCH DESIGN AND METHODOLOGY

### 3.1 Approach

We employed a case study and grounded approach to this research, as this was congruent with the focus of the study and the importance of gaining in-depth understandings of the cognitive processes of taking AWT1. We consider the ‘case study’ approach to be a strength, not a limitation, as it afforded the opportunity, on the one hand, to collect rich and in-depth data on the cognitive processes that large scale product-oriented quantitative studies may not be able to provide, and on the other, to develop hypotheses for further research in a wider number of contexts (e.g., with participants whose first language is not Chinese, or/and who have different graph familiarity level). We also notice that the majority of studies investigating test-taking processes in the field of language testing have used a “one-off” approach, and has not – as we proposed – collected data from the same participants at different time points. This last design feature is what is unique about the current research to understand the cognitive processes of taking AWT1 tasks.

We followed the University of Bristol's research ethics and data protection guidelines and regulations during the whole process of this project. Written consent was obtained from each participant.

We set out our research methodology below.

### 3.2 Participants

This research was conducted in a leading university in China. It was chosen because a large number of its undergraduate and postgraduate students take the IELTS academic module each year. In this respect, this research will be more beneficial for IELTS partners to gain a better understanding of the validity of AWT1 than if a smaller university which has fewer prospective IELTS test takers was selected. With the co-ordination of its Foreign Affairs Office, the Graduate School and the student Society of International Communication and Exchange, calls for participation were sent to all departments across the university through its administrative intranet which all staff and students access on a daily basis, and to the university's public communication platform specifically for its students. As expected, there was a considerable amount of interest. Over 380 students (and 1 teaching staff member) emailed us their personal details to apply for a place. We envisaged that a large number of students would sign up because of the possible benefits of having free training on how to take AWT1 as offered by the project. However, due to the nature of the case study approach to collecting data on cognitive processes, we could select only a small number of them to participate, by taking into account their background characteristics such as gender, department (science, social sciences, or arts) and academic status (i.e., undergraduate or postgraduate including Master and Doctoral students). We applied the following criteria in the order indicated below, by examining the personal details that the applicants provided. These details of the selected participants were re-collected later in the graph familiarity questionnaire (see 3.3) to double-check their accuracy.

- 1 For undergraduates, they must be in their final year (or fourth year for medical students who are in 7-year programmes) of their degree programme. For postgraduates at Master level, they must be in their second year (some Master programmes in the university require 3-4 year study); and for postgraduates at doctoral level, they can be any year of their study.
- 2 The participants must have no experience of taking the actual IELTS test, but are planning to take IELTS at the end of 2008 or early 2009. We anticipated that these participants would be more engaged in this research project. Students who did not provide this information in the application form were not selected.

After applying the two criteria above, we still had 121 students remaining in our database: 21 doctoral students, 28 undergraduates and 72 M-level students. A follow-up email was then sent to these students asking them their availability between the end of October and early November. This information also helped the research team to have a sense of the students' commitment to the research project. This reduced the number to 53 students who could be available every day during that week. Among these students, we selected randomly 24 as participants. In case there were dropouts during the first session, we operated a waiting list. Prior to data collection, a few students informed us that they were not able to attend the first session for various reasons, so we phoned those in the waiting list and still managed to have 24 participants when the data collection commenced (see Table 1). There were 13 female and 11 male students from different departments across the university such as economics and finance, management, psychology, physics, chemistry, biology and food science, computer science, agriculture, civil and electrical engineering, medicine and pharmaceutical sciences. Among them, 12 were undergraduates, 12 postgraduates (including 5 PhD candidates); 17 in science and engineering subjects

and 7 in social science and arts. The subject areas of the selected participants reflected the overall picture of the specialisms of the 380 students who expressed their interest in participation. None of them had taken official IELTS tests, but all of them were planning to take IELTS at the end of 2008 or early 2009. For ease of reference in this report, we assigned each participant a code, from A to X.

		Status			
		MSc	PhD	UG	Total
GENDER	F	5	1	7	13
	M	2	4	5	11
TOTAL		7	5	12	24

**Table 1. Cross-tabulation of 24 participants by gender and academic status**

However, at the start of Stage 2 data collection (see 3.3), 5 dropped out (Students F, M, O, P, V). At the start of Stage 4 data collection, 1 more dropped out (Student A). Therefore the final complete dataset consisted of 18 participants (see Table 2): 11 female, 7 male; 11 undergraduate, 7 postgraduate (incl. 2 PhD candidates).

		Status			
		MSc	PhD	UG	Total
GENDER	F	4	1	6	11
	M	1	1	5	7
TOTAL		5	2	11	18

**Table 2. Cross-tabulation of the final set of 18 participants by gender and academic status**

### 3.3 Data collection

This research collected both qualitative and quantitative data at five stages with different research focuses and instruments, as summarized below in time order.

#### STAGE 1: Baseline data collection and think-aloud training

- 1 At the first meeting, the participants were briefed about the purpose of the project and the schedule of data collection. Written consent was obtained from each participant (see Appendix 1).
- 2 Administration of IELTS Academic Writing Tasks 1 and 2 (see Appendix 2), developed by the research team according to IELTS Handbook (2006), to measure the participants’ writing abilities without thinking aloud. The AWT1 task asked the participants to summarise, making comparisons where relevant, the main features of a line graph reporting the UK greenhouse gas emissions by different end users between 1990 and 2003. The AWT2 task asked the participants to write on the extent to which they agreed or disagreed with the following statement: “Once children start school, the teachers would have more influence in their intellectual and social development than parents”.



- 3 Administration of the graph familiarity questionnaire to understand the participants' graph familiarity and comprehension (see also Xi 2005). This self-evaluation questionnaire contained questions on: the participants' experience of using graphs in their academic study and reading; their familiarity with different types of graphs; their ability in comprehending and interpreting graphs; and their ability in using graphs in written communication and in assisting learning (see Appendix 3). The first 11 questions were used to re-collect the participants' personal information such as their gender, IELTS experience, specialism, and year group (see 3.2 for the participant recruitment procedure). As "graph" can potentially mean different types, we provided not only Chinese explanations on what "graph" means but also examples of graphs such as bar graph, line graph, pie chart, statistical table, and flow chart.
- 4 The participants were trained on how to think aloud (see Appendix 4). They were first asked to think aloud when solving some simple mathematics word problems, and then think aloud when doing an IELTS AWT1 task. An example of think aloud was then provided for them to compare it with their own think aloud protocols to raise their awareness of the importance of keeping thinking aloud. The participants were allowed to think aloud in English and/or Chinese as appropriate and comfortable.

*STAGE 2: First round of collection of the participants' cognitive processes (i.e., pre-training)*

- 1 At the beginning of Stage 2 data collection, the participants were given further brief training and practice in think-aloud using the first AWT1 task they did the other day. Only when they were comfortable with think-aloud did they start doing the AWT1 tasks (see Appendix 5). None of the graphs used in the AWT1 tasks is 3-D in order to avoid a potentially extra information processing load on the participants (Carswell, Frankenberger, & Bernhard 1991).
- 2 Administration of the first set of 4 AWT1 tasks, printed in colour (see Appendix 5, Set A). These graphic prompts included (i) a simplified diagram/drawing showing the changes of the landscape or layout of an area from 1937 to 1995 (hereafter A-layout), (ii) a numerical table showing the top countries from which the USA imported crude oil (hereafter A-oil), (iii) a stacked bar chart reporting the instruction hours of children in OECD countries (hereafter A-instruction), and (iv) a line graph describing the UK CO<sub>2</sub> emissions trend (hereafter A-UKCO<sub>2</sub>). The AWT1 tasks were assigned to each participant in random order, one at a time (25 minutes for each task, i.e., 5 minutes more than the IELTS recommended 20 minutes due to the extra requirement of thinking aloud). The participants were asked to think aloud while completing the writing tasks. Each was given a digital voice recorder to record their think-aloud process. In order to minimize the influence from each other and avoid recording other people's think-aloud in his/her own recorder, several measures were taken to ensure: (i) only four participants a time were allowed into a very big classroom, (ii) they were seated in the four corners of the classroom, and (iii) at no point did participants have the same writing prompt. Field notes on how the participants were engaging with the tasks were taken by the first author.

*STAGE 3: Training on AWT1 tasks*

In the short training session of just over two contact hours, the first author presented to the participants an overview of the IELTS test and the types of AWT1 task prompts and rating criteria. Examples of the participants' own recorded think-aloud protocols in Stage 2 were used to illustrate the cognitive

processes and strategies they had used to complete AWT1 tasks and the problems as shown in the think-aloud protocols that the participants should try to avoid. In particular, it was emphasized that they should describe only the information contained in the graphs and make comparisons where necessary but should try to avoid making unnecessary personal conclusions not based on the data presented in the graphs (see also 4.3.4). The students were also reminded of the dos and don'ts as suggested by Cambridge ESOL, and some frequently used words and phrases for describing the process and trend of decreasing and increasing, and those for describing comparisons. In addition, they were given 12 AWT1 tasks using a variety of graphs for them to practise after the training session or at least to make themselves familiar with the different types of graphs that might be used in AWT1 tasks. A formal training booklet (which is available from the authors upon request) including the key messages mentioned above was provided for every participant.

*STAGE 4: Second round of collection of the participants' cognitive processes (i.e., post-training)*

- 1 Following the same procedure of Stage 2, although no training on think-aloud was given this time as they were by then more familiar with this data collection tool, the participants were asked to think aloud while completing the second set of 4 AWT1 tasks (Appendix 5, Set B). The four tasks are: (i) a diagram showing the sequence of the events leading to the broadcast of a documentary by a TV programme (hereafter B-broadcast), (ii) a numerical table and the related coloured world map showing the amount of CO<sub>2</sub> emissions by the top eight countries (hereafter B-map), (iii) column and pie charts describing grams of CO<sub>2</sub> emissions per passenger/km of different transport methods and the EU funds spent on them (hereafter B-EUfund), and (iv) a line graph showing the individuals viewing share of the major TV channels in the UK (hereafter B-viewing).
- 2 Students were also asked to evaluate the training and their experience in this project and provide any suggestions on AWT1 preparation. This was not originally planned in our research proposal, but we feel it can serve the central purpose of understanding how the training may have affected their cognitive process of taking AWT1 tasks (i.e., RQ4). The students filled in the questionnaire (see Appendix 6) either right after the final AWT1 task, or completed it at home and returned it the next day when they came for the interviews.

*STAGE 5: interviews*

Due to 6 dropouts in total, we decided to modify the original plan of individually interviewing 50% of the 24 participants based on their gender, graphicacy and writing abilities. Instead, we interviewed all 18 participants. The participants were asked to comment on (i) the ways their cognitive processes may have been affected by the different graphic prompts, (ii) the relationship between their graph familiarity and comprehension and writing abilities, and (iii) whether their cognitive processes might have changed due to the training (see Appendix 7 for a list of the questions that guided the semi-structured interviews). The interviews ranged from 45-68 minutes each, conducted in English and/or Chinese as appropriate and audio-recorded.

In summary, this research comprised five distinct stages and adopted a layered and progressive approach to data collection. The data included the participants' academic writing performance without thinking aloud, graph familiarity and comprehension (Stage 1), AWT1 performances of different graph prompts while thinking aloud (Stages 2 & 4). Between Stages 2 and 4 (i.e., Stage 3), training on AWT1 was provided to the participants. At Stage 5, interviews with the participants were conducted, aiming to probe further their cognitive processes of taking AWT1 tasks.

## 4 ANALYSES AND FINDINGS

The data were analysed first of all to understand the participants' graph familiarity and their writing performances under the normal examination conditions and under think-aloud conditions (4.1).

Section 4.2 presents an overview of the participants' cognitive processes. Section 4.3 addresses the four research questions one by one.

### 4.1 Participants' graph familiarity and English writing abilities

#### 4.1.1 Participants' graph familiarity

Data on the participants' graph familiarity was collected via the questionnaire (Appendix 3) at Stage 1 from 24 participants (see Table 1). Except Questions 42-44 (see below), the participants' responses were used to calculate their graph familiarity score. Their responses to all the remaining questions, except Questions 18-20 and 40-41, were added together, with a bigger number representing higher graph familiarity. Responses to Questions 18-20 and 40-41 where a smaller number represented higher graph familiarity (i.e., negative questions), were recoded to be consistent with the majority of the questions. Questions 42-44 were quite neutral: responses to bigger/smaller numbers did not necessarily mean higher/lower graph familiarity. They were not used to calculate the total score of graph familiarity, rather they were analysed separately (see 4.3.2) from the other questions. As a result, there are 32 questions included in the calculation of the total score of graph familiarity. Reliability analyses of the scale (32 items) indicated an Alpha of 0.948. Overall, we trust the participants responded to the questions consistently and the measure of their graph familiarity using this questionnaire achieved satisfactory reliability. Using the 32 items to measure the participants' graph familiarity, the maximum possible score would be 192 (32 x 6) and the minimum 32 (32 x 1). Within this sample, we observed a minimum of 96, maximum of 182, mean of 143.8 (i.e., around 75% of the total maximum possible score), std. deviation=22.9. The graph familiarity is close to a normal distribution (Kolmogorov-Smirnov  $Z=0.65$ , n.s., see Figure 1). The data suggests that the participants were quite familiar with graphs. As part of their academic study, a third of the participants (8/24) *very often* (i.e., the highest of the 1-6 scale) used special computer software to produce graphs, nearly 46% (11/24) very often needed to produce graphs, and a quarter (6/24) very often needed to interpret graphs. For further details of these participants' graph familiarity, see Appendix 8 which reports the frequency statistics of their responses to each question.

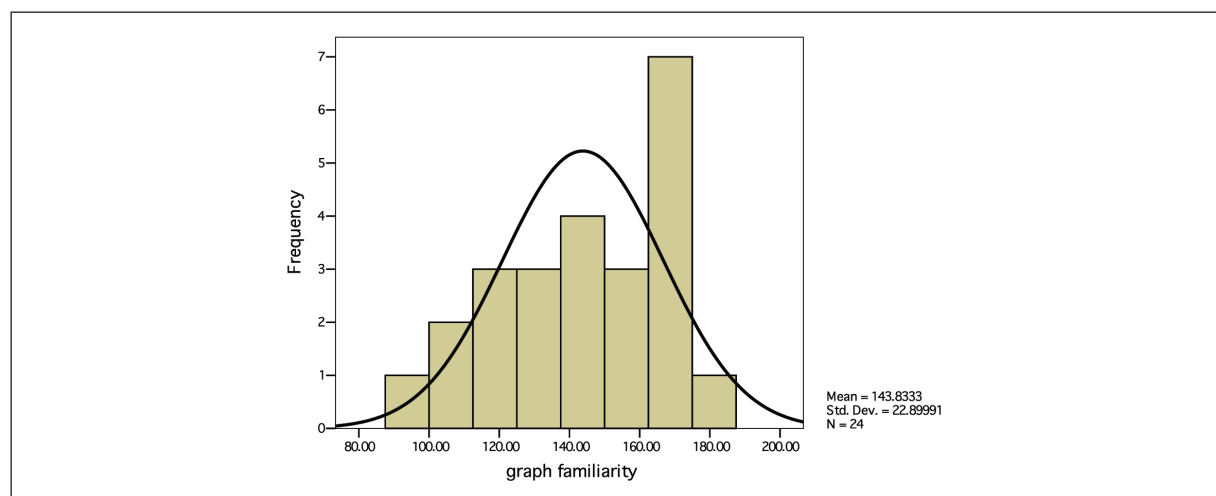
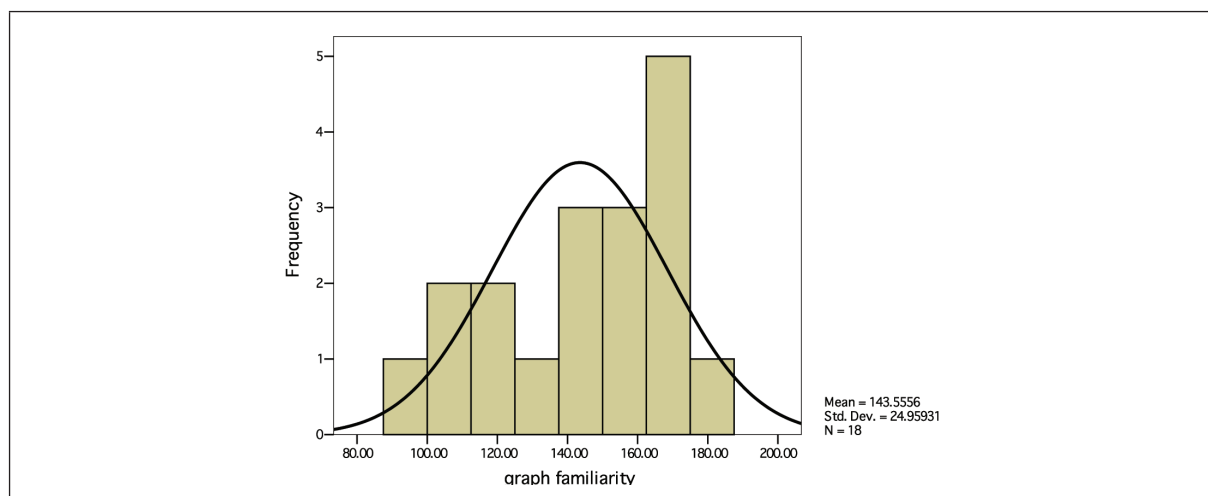


Figure 1. Graph familiarity of the 24 participants

Although we understand this is a small sample, we also tried to examine if there was any meaningful difference in graph familiarity between male and female participants. It was noted that there was no statistically significant difference in graph familiarity between male and female participants: Mean of 11 male participants = 151.4, std. deviation=19.0; mean of 13 female participants (n=13) =137.5, std. deviation=24.7 ( $t=1.52$ ,  $df=22$ , n.s.)

The above statistics included all the 24 participants. However, as 6 of them dropped out, we did separate analyses using the remaining 18 participants only: very similar findings were noted (see Figure 2). Mean = 143.6, std. deviation = 24.96, min=96, max=182; and the difference between female and male participants was not significant either.



**Figure 2. Graph familiarity of the 18 participants**

Below we report the participants' AWT1 writing performance under normal examination conditions and when thinking aloud.

#### **4.1.2 Participants' writing performances under normal examination conditions**

Under normal examination conditions without think-aloud, the participants completed two writing tasks: AWT1 and AWT2 (Appendix 2). Their written scripts were Word processed before being double marked by the research team according to the IELTS writing band descriptors (public versions). Each researcher independently marked 2/3 of the printed writings randomly assigned to them, to ensure double blind marking (see Table 3). If the difference in marks between two raters was bigger than 1, a third rater marked the writing in question. If the third mark was the average of the previous two marks, the third mark was reported as the final mark of the writing; otherwise, the average of the two most adjacent marks was reported as the final mark of the writing (Note: We retain  $\frac{1}{4}$  and  $\frac{3}{4}$  marks in the analysis). If there was still no agreement among the three raters, we then had face-to-face moderation discussions. The initial agreement between raters before moderation was not satisfactory, largely due to the different interpretations of the first set of rating criteria of AWT1 - "task achievement".

<b>SCRIPT ID</b>	<b>RATER 1</b>	<b>RATER 2</b>	<b>RATER 3</b>
1	x	x	
2		x	x
3	x		x
4	x	x	
5		x	x
6	x		x
7	x	x	

**Table 3. Double blind marking scheme**

At the moderation meetings, we discussed the scripts in question and marked them together. The moderation exercise was an essential step to improve rater reliability. The average inter-rater reliability Cronbach's Alpha was improved (above 0.83).

Marking the AWT2 scripts was much more straightforward and achieved high inter-rater consistency before moderation: Cronbach's Alpha = 0.843 between Rater 1 and Rater 2, 0.881 between Rater 1 and Rater 3, 0.793 between Rater 2 and Rater 3. Only two scripts had difference larger than 1 and were marked by a third rater. The participants' performances in these two writing tasks are reported.

	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Std. Deviation</b>
AWT1 no think-aloud	24	4.50	7.00	5.69	0.673
AWT2 no think-aloud	24	4.50	7.50	6.32	0.764

**Table 4. Performance in AWT1 and AWT2 without think-aloud**

As shown in Table 4, the mean of the participants' performance in AWT1 and AWT2 under normal examination conditions without think-aloud was 5.69 (std. deviation=0.67) and 6.32 (std. deviation=0.76) respectively. Below we report the participants' performances in the eight tasks under think-aloud conditions.

#### **4.1.3 Participants' writing performances under think-aloud conditions**

The writings of the eight tasks under think-aloud conditions were randomly assigned to and double blind marked by the research team, using the same procedure as for marking the first AWT1 task under normal examination conditions (see 4.1.1). Overall, the inter-rater reliability Cronbach's Alphas were satisfactory: 0.69 for A-instruction scripts, 0.86 for A-layout scripts, 0.85 for A-oil scripts, 0.90 for A-UKCO2 scripts, 0.78 for B-broadcast, 0.74 for B-EUfund, 0.71 for B-map, and 0.79 for B-viewing. The participants' performances in the eight tasks are reported in Table 5 below.

	Type of Graph(s)				Mean Statistic	Std. Error	
Task A: instruction	Bar graph	18	5.00	7.00	6.00	.151	.642
Task A: layout	Linear drawing	18	5.00	6.75	6.10	.115	.486
Task A: oil	Statistical table	18	4.75	7.00	5.93	.164	.696
Task A: UK CO2	Line graph	18	4.75	7.25	6.03	.150	.635
Task B: broadcast	Diagram of sequence of events	17	4.75	7.25	6.00	.165	.679
Task B: EU fund	Bar graph and pie chart	17	4.75	7.00	6.04	.144	.595
Task B: map	World map and statistical table	18	5.25	7.25	6.36	.107	.456
Task B: viewing	Line graph	18	5.00	7.50	6.50	.181	.770

**Table 5. Performance in the eight AWT1 tasks with think-aloud**

One student (# T) wrote down what she was thinking aloud, rather than a piece of writing as responding to the AWT1 tasks. Her data were not included in the analyses of the participants' performance, but included when analysing the cognitive processes.

As shown in Table 5, the participants' performance ranged from around 5 to 7.5, with mean scores from 6 to 6.5.

## 4.2 Developing a working model of cognitive processes for AWT1 tasks

The think aloud protocols and the post-task interviews were transcribed, coded and categorized in a qualitative data analysis computer programme – winMAX (Kuckartz 1998), to identify the patterns of the cognitive processes of completing AWT1 tasks. This grounded approach is appropriate to understand the complexity of the qualitative data of the cognitive processes (see 3.1). In order to develop a working model of cognitive processes for AWT1 tasks, we analysed the think-aloud protocols, cross-referencing to the other four main sources of data in this research project (i.e., AWT1 scripts, interviews, graphicacy, and English writing abilities) and the theories of graph comprehension, especially the knowledge construction-integration model (see 1.3). Below we explain in detail the working model which emerged from the data.

The AWT1 tasks seemed to require three key consecutive processes:

- 1 comprehending non-graphic task instructions
- 2 comprehending (and interpreting) the components of graphs (see Section 1.3 on the key factors that can influence graph comprehension)
- 3 re-presenting or re-producing the non-graphic and graphic information as continuous discourse in written form in English as a foreign language.

Although these three processes were found to be iterative and to vary among the participants in terms of, for example, the amount of time they spent on each process for different tasks and the differential interactions with and influences on test performance of other factors such as the participants' prior domain knowledge about the information presented in the graphs, the conventions of graph presentation, graphic skills and English writing abilities, it seemed to be very clear that these three processes happened, first and foremost, in time order. However, overall, the whole process for completing AWT1 tasks, as graph comprehension, is iterative and recursive, rather than serial (Carpenter & Shah, 1998).

#### 4.2.1 Comprehending non-graphically presented task instructions

All the participants started to read the non-graphically presented task instructions first, presented before the actual graph(s). However, not all text instructions received the same attention or effort from the participants. They seemed to be very strategic in allocating their time to different components of the task instructions. It is interesting to note that all of the participants skipped the first line (i.e., *You should spend about 20 minutes on this task*) and the last line (*Write at least 150 words*) of the task instructions, although both sentences were clearly marked in bold. This may be largely because the participants were already familiar with these two specific task requirements – time allowance and the expected length of their writings.

What the participants repeatedly read and re-read at different stages of completing the tasks were two sentences. The first was the summary-like sentence about the graph (e.g., *The following graph shows the UK CO2 emissions by end users from 1970 to 2004* in Task A:UKCO2, *The following map and table show the amount of CO2 emission by the top 8 countries* in Task B:Map). This introductory sentence gave the participants a clear lead as to what the following graph was all about (i.e., the topic or the theme of the graph), and therefore may have facilitated the participants to process the information contained in the graph (see 4.2.2). However, the summary-like sentence served only as an entry point, in other words, the participants still had to work out what else they should focus on when looking for the main features and making comparisons. For example, Participant K reflected in the interview on the summary-like sentence in Task A: Instruction:

##### Extract 1

*I read the direction, it says different age group. Now that it says different age group in the task direction, so I wondered if it is necessary to figure out and mention the differences with reference to different age groups. After writing about the total length of the bars, i.e., the total number of intended instruction hours, I didn't know if it is necessary also to write about the differences between countries in the different age groups. (Participant K)*

In Section 4.2.3 we report on how the non-graphic task instructions were re-presented or re-produced in the actual writings.

The second sentence – “*summarise the information by selecting and reporting the main features, and make comparisons where relevant*” – was in fact the same for all AWT1 tasks. Why didn't the participants skip this sentence as they skipped the sentences on time allowance and expected length because they had already known that this was exactly the same requirement for AWT1 tasks – summarise main features and make comparisons? Contrary to our anticipation that the participants would pay less attention to this sentence after they became more familiar with the AWT1 task requirement, we noticed that there were more occurrences of re-reading of this sentence in the think-aloud protocols of the second set of four tasks than the first four tasks. Data from the interviews and end-of-project evaluation (Appendix 6) indicated that this was largely attributable to the short training (see data collection in Stage 3). Before the training, the participants had a strong tendency to interpret what they observed in the graphs and try to find the reasons behind the data. At the training session, we explicitly pointed out that the main task was to describe rather than to interpret the data, in other words, the main task was to summarise and make comparisons rather than make personal interpretations based primarily on prior domain knowledge. Therefore, the participants were probably made more aware of this requirement in the second four tasks after the training, and they might have used re-reading this sentence as a kind of constant reminder to themselves that their major task was to summarise and make comparisons of what are in the graphs, but not to involve too much personal interpretation (see 4.3.4 on the coachability of AWT1 tasks).

#### 4.2.2 Comprehending graphic information

After comprehending the non-graphic task instructions which were placed before the graphs, the participants moved on to read the graphs, look for main features of the graphs, and make comparisons. It seems to be a quite natural cognitive process and transition from focusing on the non-graphic information to the graphic information. The comprehension of graphic information involved a series of iterative and internal processes (see 1.3), starting from reading, understanding, deciding on the main features, to interpreting the graphs, prior to re-presenting and re-producing them in continuous written discourse in English as a foreign language (see 4.2.3). The think-aloud protocols demonstrated a series of activities involved in graph comprehensions, corresponding to what Curcio's (1987) three levels of graph processing and comprehension: read the data, read between the data, and read beyond the data.

When reading the data or the information contained in a graph, the participants started to search and locate specific information and try to perceive trends and patterns, i.e., they started from local to global search of graphic information (see also Guthrie et al 1993), constantly checking the graphic information against their prior knowledge about the subject or the content of the graphs (e.g., CO<sub>2</sub> emissions, environmental protection, and crude oil importation) as well as their prior knowledge about the types and conventions of graphs. The discriminated symbols and conventions used in graphs and the ways they were configured were meaningful to graph readers, and anticipated certain type of information processing, e.g., fact-retrieval from line or bar graphs, trends from line and bar graphs, making proportion judgements from pie charts, making comparison judgements and determining the slope of the trends from a regression line. In other words, the types and conventions of graphs influenced the way how comparisons can and should be made, following the cognitive naturalness of graphic communication conventions. In addition to the prior knowledge about the content and conventions of graphs, the participants' graphicacy, explanatory skills and other scientific reasoning skills also influenced, in varying degrees, the cognitive processes of completing AWT1 tasks and subsequently their task performance (see 4.3.2 on the effects of graphicacy).

#### 4.2.3 Re-presenting graphic and non-graphic information in continuous discourse

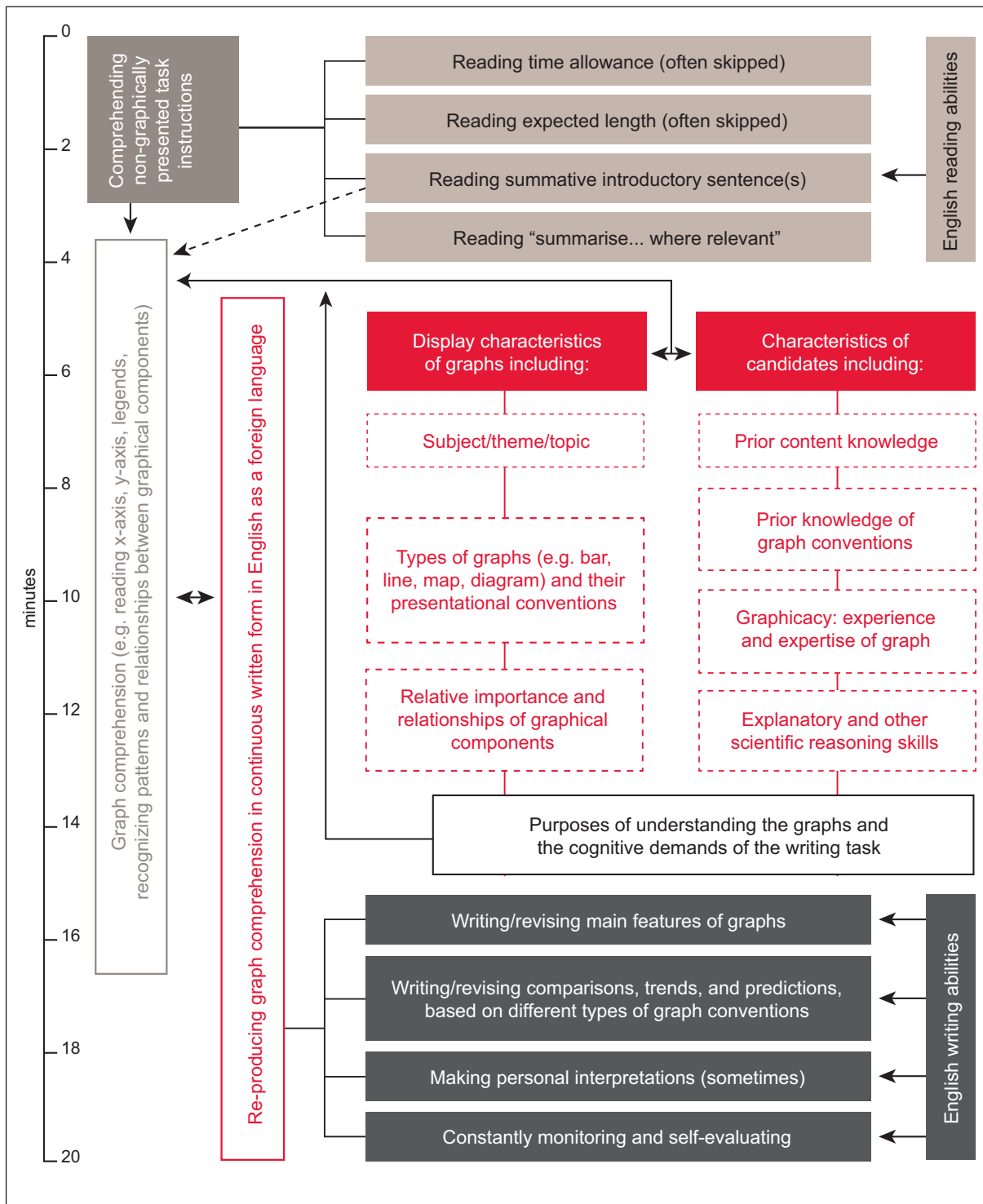
As described in Section 4.2.1 and 4.2.2, the iterative cognitive processes involved various factors including, for example, reading non-graphic task instructions, reading and interacting with the display characteristics of the graphs (including most prominently title/caption, x-axis, y-axis, source of data, legend, and colour schemes) and their conventions, understanding the graphs, extracting the key information or data points, making comparisons of main features, and trying to find out the underlying reasons for the phenomenon being described (although this is not required by the AWT1 tasks). As responses to the requirements of AWT1 – essentially a writing task, the participants were constantly planning and organizing content (including choosing specific words to make trend assessments, discrete comparisons, or predictions, following the conventions of graphic presentations), checking the accuracy of linguistic forms and content, self-monitoring and evaluation of writing (including re-reading what has been written and counting number of words of their writings). The knowledge construction-integration model of graph comprehension (see 1.3) seems to be fit for our purpose of explaining and evidencing some of the cognitive processes of completing AWT1 tasks, especially during the graph comprehension stage. However, AWT1 tasks are unique in the sense that they involve not only candidates' processing of graphs and but also more importantly, re-producing the graphic and the non-graphic information in a continuous written discourse in English as a foreign language. It is this writing process that makes AWT1 a unique integrated reading/writing task. It is this writing process that we noted that in effect facilitated the participants to better interact with graphic information and enhanced their graph comprehension through writing. In other words, the comprehending and the re-producing of graphic information were found to be mutually beneficial.



As AWT1 tasks are a kind of writing activity, primarily knowledge telling, to reproduce meanings already depicted in the task prompts, candidates are to some extent able to copy and deploy the lexical and syntactic components of the non-graphic prompts in their own writings. The importance and the value of the summary-like sentence in the non-graphic task instructions are clearly evidenced in the participants' actual writings. Taking the scripts of Task A:UKCO2 as an example, we noticed that all the 19 participants almost unanimously started their writing with this summary-like sentence or its slight variation, e.g., "The line chart illustrates the trend of the UK CO2 emissions by different end users between 1970 and 2004 in million tonnes carbon equivalent" (Participant A), "The graph in the paper show UK CO2 emissions by end user from 1970 to 2004" (Participant B), "The graph in the picture indicates UK carbon dioxide emissions by end user from 1970 to 2004" (Participant C), "From the graph above, we can see that the UK CO2 emissions by end user has fallen down in 34 years" (Participant D), "It's a graph which shows the UK CO2 emission by end user from 1970 to 2004" (Participant E). It is the same case for the other seven tasks too. Consistently, the participants started their writings, with almost verbatim copy of this summary-like sentence from the task instructions, and then provided further details in subsequent paragraphs. Asked why they started with this copying exercise at the interviews, all the participants said that there couldn't be a better introductory sentence than this authoritative one in the task instructions.

#### 4.2.4 Summary of the working model

The working model of the cognitive processes for AWT1 tasks is presented in Figure 3. The vertical timeline indicates the time allowed for AWT1 tasks, and the positions of the three key processes along the vertical timeline in the left column represent approximately the average duration of each process. The overlapping between "graph comprehension" and "re-producing graph comprehension in continuous written form in English" indicates that some students developed their understanding of the graphs through writing rather than following the linear process of graph comprehension and then writing. The participants may also change and improve their understanding of the graphs, when they are writing. At this first stage of "comprehending non-graphically presented task instructions" and its associated activities, especially the summative introductory sentence. Comprehension of this sentence directly affects the next stage, i.e., graph comprehension, which is further governed by the display characteristics of graphs, the characteristics of graph readers, and their interactions, plus "the purposes of understanding the graphs and the cognitive demands of the writing tasks". At the third process, i.e., re-producing graph comprehension in continuous written discourse in English, the key aspects include writing and revising the main features of graphs, writing and revising the comparisons where relevant, writing personal interpretations (sometimes), and constantly monitoring and self-evaluating their writings.



**Figure 3: A working model of cognitive processes for taking IELTS AWT1 tasks**

Overall, the participants’ English writing abilities should play a pivotal role, at least in theory; and at this stage, the more generic second language writing models become more applicable for AWT1 tasks.

This working model will be used to guide our analyses of the empirical data to address the four research questions in Section 4.3.

### 4.3 Addressing the four research questions

#### 4.3.1 Research question 1

*To what extent are there differences in the candidates' cognitive processes due to different AWT1 prompts?*

In terms of cognitive processes of completing AWT1 tasks, we identified three key stages: comprehending non-graphic task instructions, comprehending graphic information, and re-producing graphic and non-graphic information in continuous discourse (see Figure 3). Think-aloud protocols, interviews and AWT1 scripts form the three main data sources to investigate the extent to which the cognitive processes were due to the use of different graphic prompts.

As shown in Table 5 (see 4.1.3) and Appendix 5, we used different types of graphic prompts, including bar graph (Task A:Instruction, Task B:EUfund), linear drawing of the layout of an area (Task A:Layout), statistical table (Task A:Oil, Task B:Map), line graph (Task A:UKCO2, Task B:Viewing), diagram of the sequence of events (Task B:Broadcast), pie chart (Task B:EUfund), and world map (Task B:Map). Two of the tasks used two types of graphic prompts. In Task B:EUfund both bar graph and pie chart were used, and in Task B:Map, world map and statistical table.

As reported in 4.2.1, there does not seem to be any significant difference in the way the participants processed the non-graphic task instructions. All the participants (except Participant T) started with the non-graphic task instructions. When doing Task A:Layout, Participant T started by analysing the graphic information first, according to her think-aloud. Overall, it is only when comprehending graphic prompts and re-producing the main features of the graphic information in written discourse that the participants demonstrated various differences in their cognitive processes.

We noticed that the participants' performance in the eight tasks varied significantly, in terms of the final scores they received (Chi-square=19.12, df=7, n=17, p<0.01). As shown in Table 5 (section 4.1.3), scripts of Task B:Viewing which used a line graph received the highest mean score (6.5) among the eight tasks, and those of Task A:Oil which used a statistical table the lowest mean score (5.93). The participants considered line graphs the easiest as the trends and the slopes are readily perceivable from the visual displays. Line graphs were also the most familiar to them (see 4.3.2). This may explain why scripts of Task B:Viewing received the highest mean score. However, it may not be necessarily the case that all AWT1 tasks using line graphs are always the easiest. Task A:UKCO2 which also used a line graph presented a different situation. The average performance of Task A:UKCO2 was not much better than tasks using other types of graphs. The lowest mean score of the scripts of Task A:Oil may well be due to the fact that it used a statistical table which is different from and probably more cognitively demanding than line graphs, as Fry (1981) explained:

*Graphs pack a high density of information into a small area... are more globally visible than they are detailed, symbolic, and sequential... tend to show the 'big picture' or gestalt.... Often relationships can be seen better with a graph than with a purely verbal or numerical information." (Fry 1981: 388).*

Perhaps also as the two 19th century economists commented: "Getting information from a table is like extracting sunlight from a cucumber" (Farquhar & Farquhar 1891), cited in Wainer (1992: 18), statistical tables are challenging to understand. However, not all tables are born equal. As Lohse (1993) pointed out, the cognitive demands of processing tabular information depended on tasks.

Tables do not require a complex mapping of syntactic symbols to semantic information. The semantic information is read directly from the table. This greatly reduces the information load in STM (i.e. short term memory, added by the authors) when the queries are simple. However, for complex comparisons of trends, tables require the reader to look at several entries in the table and make mental calculations and comparisons among these values. These intermediate values are held in STM during the comparison. In contrast, line graphs ... show trend information without the scanning and interpolation of intermediate data. Slopes are perceived readily from a line graph (Lohse 1993: 360).

Similarly, Vessey (1991) argued that graphic representations emphasize spatial information, while tabular representations emphasize symbolic information, hence it is more time-consuming to process tabular information. However, Vessey (1991) further suggested that so long as there is cognitive fit (i.e. matching between the nature of the representation, the process and the nature of the task) each representation, whether graphic or tabular, would lead to both quicker and more accurate problem solving. In case of AWT1 tasks, as all participants were required to describe whatever graphic and tabular information they were given, therefore the level of cognitive fit varied from participant to participant, and from task to task. The cognitive demands for processing tabular information may be particularly high for writing tasks such as AWT1 as it requires not only locating and extracting specific information but also integrating and identifying the overall patterns and making comparisons from the tabular information which may not be readily available from the table per se. The participants had to work out the patterns by themselves, through a series of calculations and comparisons in rows and columns of the statistical table. They felt that the statistical table on USA oil imports was not straightforward, as several participants commented, for example:

### Extract 2

*The oil statistics, it takes a long time to understand the information, to understand what it means....I was really puzzled, trying to work out what it means. It is quite exercising.*  
(Participant T)

### Extract 3

*The statistics look very much the same, and I found it very difficult to present the key information from the table. But the line graphs, bar graphs and pie graphs are relatively straightforward and vivid.* (Participant U)

When the tabular information is presented in conjunction with other graphic prompts, e.g., Task B:Map which used the world map and a statistical table with fewer data points than the statistical table in Task A:Oil, the cognitive demands for processing tabular numerical information seemed to be reduced. On average, the scripts of Task B:Map received the second highest mean score (6.36) among the eight tasks. The visual assistance of the world map in colour might have assisted the reading and the interpretation of the numerical data contained in the table. In addition, the fewer data points of this table might have rendered itself more manageable than the statistical table about US oil imports from 15 countries in June and July of 2008 and July 2007 and Year to Date of 2007 and 2008. The difficulty of this complex statistical table in Task A:Oil was further increased because of one particular phrase – Year to Date. All the participants knew these three words separately, but none of them knew exactly the statistical meaning of YTD. Similarly in Task B:Viewing, some participants seemed to have difficulty in understanding what MCH (multi-channel) meant. It seems that a single phrase in the accompanying text of a graph could increase substantially the difficulty level of the graph. Coupled with the higher cognitive demands of statistical tables per se, the participants' lower familiarity with processing tables of numerical data

had probably also made AWT1 tasks using statistical tables more challenging than other tasks. The participants reported that they were less familiar with tables of numerical data than other graphs, with the exception of diagrams (see 4.3.2). In relation to diagrams, however, the use of a simple and linear diagram reporting the sequence of the events leading to the broadcast of a TV documentary in Task B:Broadcast did not necessarily lead to the lowest performance; with a mean score of 6.00, it was higher than the lowest - Task A:Oil (5.93). The substantial accompanying text in Task B:Broadcast might have mitigated the innate high cognitive demands of diagrams.

In addition to the participants' average performance data which provided some insights into understanding the effects of different graphic prompts on the cognitive processes of completing AWT1 tasks, we investigated which words were used frequently in each task as another means of examining the effects of graphic prompts on the participants' lexical decision-making. Are the words used in the writings predictable? We ran word frequency using Wordsmith Tools (the full word list is available upon request). Appendix 9 reports the frequently used words specific to each task. A clear pattern is observed in the content words used for each task. The most frequently used words (i.e., excluding non-content words such as *the*, *have* and *has*) for each task matched very well, as anticipated, with the theme or topic of graphs. In other words, the graphs determine directly and probably naturally the words test takers use; the content of the writing is largely predictable. As Participant W commented:

#### Extract 4

*Different graphs would require the use of different words and you may have different level of knowledge of different words... As I said, different tasks would stimulate your vocabulary knowledge, for example, the line graph activates your vocabulary such as increase and decrease. (Participant W)*

At the same time, we also noticed that certain words are used frequently anyway, regardless of the topics of the graphs, for example, “difference”, “more”, “decrease”, “increase”, and their variations. The use of these four words demonstrates that the participants were constantly making comparisons as required by the task instructions. Another word that also appeared frequently in the writings is “show”. Two factors may explain the high frequency of this particular word: firstly because it appeared in all the task instructions and secondly because this is such a neutral word that it can be applied to all tasks, for example, “This diagram shows...”, “This table shows...”, “This line graph shows...”. However, some participants were looking for synonyms of “show”, e.g. “reveal”, “demonstrate”, “display” and so on. Therefore, content-wise, the use of different graphic prompts apparently affected which words were likely to be used and consequently tested; however, process-wise, the participants were doing the similar activity – constantly trying to make comparisons in all the tasks except B:Broadcast where none of the four comparison words were used with any frequency.

The primarily product-oriented analyses above on the average test performance and the lexical features of the writings of each task provided a useful perspective to examine the potential differential effects of different types of graphs on the cognitive processes. Next we examine the think-aloud protocols and interviews to explore further the effects of different graphic prompts on the cognitive processes of completing AWT1 tasks. We report below some of the prominent features of graphs (e.g., the display characteristics and conventions of graphs, the prior background knowledge required to understand the graphs) that the participants thought made the tasks particularly challenging or easy for them. It emerged that three types of graphic prompts received the most comments or attention from the participants: Task A: Instruction using stacked bar graph, Task A:Oil using statistical table, and tasks using line graphs.

Task A: Instruction used a stacked bar graph. The participants found it particularly “*difficult to find out the length of the second and the third bars, as they are not from the same starting point*” (Participant N). The density of the data points of this graph caused further problems as Participant N commented:

**Extract 5**

*There are so many bars. It would be better if there were only 5 bars, for example, and it would be ideal if within every part of a bar is indicated with its percentage of the hours of a particular country.* (Participant N)

Indeed several participants tried to work out the length of each section of a bar so that they could get a better sense of the number of instruction hours for different stage of schooling. For example, Participant T tried to estimate the percentage of each part of a bar. To her, the difficulty of describing this stacked graph was not just “because there are too many bars”; the main challenge of this graph also came from the fact that “the differences are not easily observable”. Similarly, Participant K compared this stacked graph with the line graph in Task B: Viewing, and said:

**Extract 6**

*Not like the task about TV individuals viewing share, you can see the change there. It is impossible that I would measure the differences by using some rulers.* (Participant K)

Or, indeed as Participant W said:

**Extract 7**

*because the trend of all the countries is all the same, only slight difference between the countries, but these differences are not obvious to bare eyes. It does not tell us the exact number of hours.* (Participant W)

When comparing Task A: Oil which used statistical table with other tasks using line or bar graphs, the comments made by Participants H and K are typical. They clearly demonstrated the easier accessibility of the visual displays of line and bar graphs, while extra efforts were required to make trend assessments out of information from statistical tables: more calculation and judgements were required to make trend assessment possible.

**Extract 8**

*I think the pie or bar graphs are more straightforward, while the statistics tables contain a lot more information. In my study of physics, we often have to process a lot of data from experiment; we can take a couple of hours to interpret data. At first glance of the statistics table, I was a bit lost, not sure what to do with so much information.* (Participant H)

**Extract 9**

*The line graph task is relatively easier, because I don't have to work out the information, I only need to follow the lines, the trends; I just need to describe what is there, there seems to be set procedure, you just follow the steps, one by one, this seems to be a right template you should follow. But for this type of tasks (Referring to the task with tables), I need to select the key statistics, it is therefore demanding. At first glance of the table, I tried to compare which was bigger or smaller, and also tended to find out the changes, for example, I did some calculations.* (Participant K)

The following eloquent arguments, from Participants R and T, on the effects of the different graphic prompts on the processes and the products of AWT1 tasks summarize nicely the views expressed by all the participants on such effects.

#### Extract 10

*If it is a line graph, I would first of all look for the changes of patterns, the trends of the lines. If it is a bar graph, I would compare the ones that are in the same category, to compare which number is bigger or smaller, higher or lower. If it is a table with statistics, it won't be as straightforward as the line or bar graphs. When you see a table, you will first of all get a sense of the numbers in the table. However, from the graphs, you don't have a straightforward sense of numbers, but graphs give you a clearer overall picture of the data. (Participant R)*

*For example, bar graphs, you would first of all talk about the x and y axis, and they also have some statistics, you would put them into different categories according to their number; e.g. these are below 50, and those are above. (Participant R)*

*If it were a table, like this one, I would do in the same way. For example, the USA is high, and China is also high. Both have different colours, and the other countries listed however have almost the same kind of colour; they all look green. I put them into three categories, the highest, the lowest and those in the middle. (Participant R)*

#### Extract 11

*Each type of graph has its own pattern or convention. For example, line graphs would demonstrate trends of development, for example. Bar chart normally compares the amount of two or three things. When you know the underlying convention and meaning, it is less likely that we may deviate the focus or the main message of the graphs. (Participant T)*

In summary, the types and conventions of graphic prompts did matter. They affected how the participants processed the graphic information and how they followed the graphic conventions to re-produce their comprehension in written discourse in English. Such effects of different AWT1 graphic prompts on the cognitive processes were clearly manifested in the mean scores, in the use of vocabulary, and in whether and how they would make comparisons or trend assessments, following the graphic conventions in presentation, interpretation and re-presentation.

#### 4.3.2 Research question 2

*To what extent are the candidates' cognitive processes affected by their graphicacy?*

In 4.3.1, we discussed the effects of different graphic prompts on the cognitive processes and performance of AWT1 tasks, and noted that there might be some dubious interactions between types of graphs and the participants' familiarity with different types and conventions of graphs, as some students may be more familiar with one type of graph than another. In this section, we will focus specifically on how the participants' graphicacy level might affect their cognitive processes and performance, from two perspectives – the participants' views as reflected in the questionnaire and interview data, and their actual writing performance.

In the questionnaire (Appendix 3) we asked the participants about their familiarity with different types of graphs: bar, line, pie, diagram, table with numerical data respectively (Q21-25). The Friedman test

indicated that there was a significant difference in students' familiarity among these five different types of graphs (Chi-square=11.33, df=4,  $p<.02$ ). Further analyses showed that the difference was mainly attributable to students' higher level of familiarity with line graphs than other types of graphs. Their familiarity with the different types of graphs is in the order of line, pie, bar, table with numerical data, and diagram. In other words, they were most familiar with line graphs, and least familiar with diagrams. This finding was in line with the participants' comments on the effects of different types and conventions of graphs on their cognitive processes and performance (see 4.3.1). In the questionnaire we also asked the participants if they think they “*may do better in IELTS Academic Writing Task One using familiar graphs than unfamiliar ones*” (Question 42) and if they “*would prefer one type of graph to be used in IELTS Academic Writing Task One*” (Question 43). As anticipated, the majority of the participants thought they would do better in tasks using familiar graphs than unfamiliar ones; 95.8% chose 4 to 6 (“strongly agree”). However, the specific ways their familiarity with different graphs influenced their actual AWT1 writing performance may be another matter. Indeed, as shown in their response to Question 43, their belief in the effects of graph familiarity on the processes and performances of AWT1 tasks seemed to be weaker, as fewer students (79.1%) had chosen 4 to 6 (“strongly agree”). As for their belief on whether special training in interpreting graphs would be helpful for getting a higher score in their AWT1 writing, 91.7% had chosen 4 to 6 (“strongly agree”). However, like the perceived effects of graph familiarity on test performance, the actual usefulness and effects of special training on test performance may be another matter.

At the interviews, we specifically asked the participants to give their opinions about the ways their graph familiarity affected how they processed and wrote about the graphs. The interview data confirmed that these students were more familiar with line graphs than other types. For example, Participant C commented:

#### Extract 12

*As for the other types of graphs, e.g. line graph, probably because we use these types of graphs often, whether in study or other activities, therefore we are more familiar with these graphs, and know better how to write about these graphs. (Participant C)*

In addition, it was noticed that graph familiarity was considered helpful, at least, psychologically, to boost the participants' confidence that they were doing something they were familiar with.

#### Extract 13

*When you are given a task with a familiar graph, you will feel more confident. When you are given a task with an unfamiliar graph, you may feel less secure when writing. (Participant I)*

#### Extract 14

*I like to read the line graphs; it is probably because I am familiar with this type of graphs. (Participant J)*

#### Extract 15

*Line, bar and pie chart graphs are more familiar to me; I was a bit surprised when I was given the map task. (Participant S)*

Beside the potential psychological boost, the familiarity with graphs was considered to be helpful for guiding the writing process too, as the participants would be able to follow a kind of writing



convention or template that they already know or are comfortable with. Unfamiliarity however could probably create a kind of anxiety and insecurity (see also Participant I's comment above).

#### Extract 16

*For example, the map task, as I haven't come across this type of graph, I found it difficult. For familiar tasks, you would know how to write the beginning paragraph, the body paragraph and the conclusion paragraph. For unfamiliar graphs, you don't know how to start, how to develop your writing and how to end your writing. You don't know where the entry point for this kind of unfamiliar tasks is. ... [Researcher: How would you start then?] If I had not come across a particular type of graph, I would not have a model to follow, and then I would have to follow my instinct to write... (Participant R)*

Furthermore unfamiliarity may also present test takers with some real challenges in processing and writing about the graphic information, particularly when they are trying to integrate their prior knowledge with the graphic information presented. For example, Participant W commented that:

#### Extract 17

*Some graphs (e.g. the world map) may also present challenges, especially as I may be lack of some knowledge of geography and history. I was thinking if the different geographical and economic positions of these countries affect their amount of CO2 emissions. As I'm not familiar with the geographical positions of these countries in the world map, this could be a challenge to me. [Researcher: If you had not known Alaska belongs to America, would you think it is another country in the world map?] Yes. In fact, I did ask X [Participant R] yesterday what country is this one (referring to Alaska). To me, when I received the tasks with the map, I was a bit nervous. But apart from that, it was OK. (Participant W)*

Participant W further explained her belief that test takers may have different levels of graph familiarity, but it is their writing ability that matters the most in the AWT1 tasks.

#### Extract 18

*It is inevitable that some students may be highly familiar with certain graphs, although his or her writing may not be strong. It is possible that this student may do well because he or she is highly familiar with the graph; this kind of situation is possible, and I also think this kind of situation is inevitable. However, generally speaking, his or her writing is a positive reflection of his or her writing ability. (Participant W)*

After all, these largely positive attitudes towards the potential impact of graph familiarity on test performance may be due to the fact these participants had high graph familiarity (see Figure 3). Indeed, as Participant T acknowledged: “I don't think there are many graphs which are unfamiliar to us. In the textbooks we study there are graphs”. To what extent will this view on the potential relationship between graph familiarity and test performance be evidenced in the actual test performance?

Next we ran a series of regression analyses on graph familiarity and performance of each AWT1 task (including the first one without think-aloud, see Appendix 2), the average performance of the first four tasks, the average performance of the second four tasks, and the average performance of all eight tasks under think aloud conditions. No significant correlation was found between graph familiarity and test performance. We also ran ANOVA to test if each AWT1 task was affected by the participants'

familiarity with the particular type of graph used in that task. For example, we ran one-way ANOVA to examine if there was a significant difference in performance in Task A:UKCO2 between participants of different familiarity with line graphs (i.e. factor = question No. 22, see Appendix 3 graphicacy questionnaire), a difference in performance in Task B:Viewing between participants of different familiarity with line graphs (i.e. factor = question No. 22, see Appendix 3), and a difference in performance in Task A:Oil between participants of different familiarity with statistical tables (i.e. factor = question No. 25, see Appendix 3). No significant differences were found either.

In summary, it seems that graph familiarity did not affect these participants' AWT1 task performance in terms of the marks that their writings received, although some potential psychological impact of graph familiarity on task performance was expressed clearly by the participants. This may be largely due to the fact that these participants had a high level of graph familiarity. However, as shown in 4.3.1, the participants' understanding and knowledge of the conventions of different types of graphs may affect the ways that they processed and wrote about the graphs. In other words, graph familiarity should be examined from two perspectives: one is the quantified graph familiarity as measured by the graphicacy questionnaire (Appendix 3), as discussed in this section, and the other is their familiarity with and understanding of the conventions of different types of graphs (see 4.3.1).

### 4.3.3 Research question 3

*To what extent are the candidates' cognitive processes related to their writing abilities?*

As described in the working model of cognitive processes (Figure 3), test takers' writing abilities played a role when they were re-producing their comprehension of the graphs in written discourse in English as a foreign language. As lexical knowledge is an important aspect of a test taker's writing ability, the use of different vocabulary is clearly an essential indication of the relationship between the cognitive processes of completing the AWT1 tasks and the test taker's writing ability. As reported in 4.3.1, different graphs activated the use of different vocabulary, in other words, different graphs had different lexical demands (see Appendix 9 and Extract 4). In this section, we report further on (a) the relationships in performance between different AWT1 tasks and between AWT1 and AWT2 tasks, and on (b) how test takers' English writing abilities, their expectations and experiences of academic writing, whether in English or Chinese, might have shaped the way that the comprehended graphic information from AWT1 tasks was re-produced in written discourse in English as a foreign language.

Table 4 (see 4.1.2) reported the participants' performance in AWT1 and AWT2 tasks in normal examination conditions, that is, without thinking aloud (Appendix 2). It is interesting to note that the correlation between the AWT1 and the AWT2 performances was not significant ( $r=0.33$ , n.s.). The difference in performance between the AWT1 and AWT2 tasks was statistically significant ( $t=-3.73$ ,  $p<0.005$ ). On the one hand, this insignificant correlation between AWT1 and AWT2 performances raises the question about the extent to which these two writing tasks share an underlying construct; and on the other hand, it demonstrates the necessity of using not only AWT1 but also AWT2 tasks to measure candidates' academic writing abilities (as in IELTS), and also the necessity of using both test scores to analyse the relationships between the so-called English writing abilities and the cognitive processes of doing AWT1 tasks.

The correlation between the participants' performance of the AWT1 task under normal examination conditions and their average performance of the eight AWT1 tasks under think-aloud conditions was statistically significant ( $r=0.53$ ,  $p<0.05$ ,  $n=17$ ), which indicates that the overall effects of think-aloud as a data collection procedure on test performance may be minimal. However, it should be noted that

the correlation between the AWT1 task under normal examination conditions and the first four AWT1 tasks under think-aloud conditions before the special training was not significant ( $r=0.38$ , n.s.); the overall significant correlation is therefore largely attributable to the more significant correlation between the second four AWT1 tasks and the AWT1 task completed under normal examination conditions ( $r=0.60$ ,  $p<0.05$ ). The correlation between the first four and the second four tasks is more significant ( $r=0.74$ ,  $p<0.01$ ). In other words, when the tasks were all under think-aloud conditions, the correlation between AWT1 tasks is stronger than if one was under think-aloud conditions and the other was not. This stronger correlation in performance between the AWT1 without think aloud and the second 4 tasks under think-aloud conditions, and the better performance in the second than the first 4 AWT1 tasks under think-aloud conditions (mean difference =0.24,  $t=2.9$ ,  $df=16$ ,  $p<0.05$ ) might be attributable to three factors. Firstly, the participants might have got used to think-aloud towards the end of data collection. Secondly, the training before the administration of the second set of tasks (see Stage 3 in section 3.3) might have helped not only to mitigate the effects of think-aloud on test performance but also improve test performance. Thirdly, think-aloud itself might have helped the participants to concentrate on their tasks and improve their understanding of the graphic information and hence their test performance. Therefore we would argue that it is imprudent to ignore completely the potential effects on the writing performance of think-aloud through which the cognitive processes of these participants were elicited (see 4.3.4 about the participants' comments on the use of think-aloud).

As indicated earlier, the correlation between the AWT1 and the AWT2 tasks under normal examination conditions (i.e., without think aloud) was not statistically significant ( $r=0.33$ , n.s.). In other words, AWT1 and AWT2 may not share the same construct. We therefore also used AWT2 performance as another indicator of the participants' writing abilities to predict their performance in AWT1 tasks under think-aloud conditions. It was found that the correlations between AWT2 and the first set of four AWT1 tasks ( $r=0.59$ ,  $p<0.01$ ), between AWT2 and the second set of four AWT1 tasks ( $r=0.78$ ,  $p<0.001$ ), and between AWT2 and the mean score of all eight tasks under think aloud conditions ( $r=0.74$ ,  $p<0.001$ ) were all statistically significant. However, it is puzzling that the participants' performances of the two AWT1 and AWT2 tasks under normal examination conditions were not significantly correlated. We speculate that the score of the AWT1 task using only one type of graph (here a line graph) was less capable, than the mean score of the eight AWT1 tasks using different graphic prompts, of measuring these participants' AWT1 writing abilities. This speculation is in line with our findings regarding the effects of different graphs on the cognitive processes (see 4.3.1), because different types of graphs may activate different lexical demands and therefore involve different English writing abilities. Below we report how test takers' English writing abilities (in terms of grammar and word choice), their expectations and experiences of academic writing might have shaped the way they comprehended the graphic information and re-produced their comprehension in written discourse.

As demonstrated in the think aloud protocols, the participants were constantly monitoring their grammar and word choice. For example, when Participant S was considering which tense to use to describe the sequence of the events leading to the broadcast of a documentary (i.e. Task B:Broadcast), she said: "To complete their things it always take tense will..., information, just use the past tense". Similarly Participant G was also constantly monitoring her use of grammar, whether in present or past tense.

#### Extract 19

*Domestic use is less than that, so, how should I say about domestic carbon emission, domestic use is, much, is much less than, no, was, was, was much less than that for industry... (Participant G)*

Word choice is another area that the participants often had to make decisions when completing AWT1 tasks. For example, Participant K must have mis-read the word “disused” as “discussed” in Task A:Layout, he then wrote: “In the South the lake, a railway line which was under discussing in 1937 has already been built up”, which of course led to a completely different interpretation from what was intended. Other examples of the participants’ decision-makings on word choice are less dramatic. For example, Participant L tried to figure out whether “proportion” would be a more appropriate word than “amount” when describing the biggest source of UK CO2 emissions in 1970; whether to use “sources”, “uses” or “approaches”, and whether to use “steadily”, “steady”, “stead”, or “stable” when describing the trend of CO2 emissions. Finally she decided to use “proportion”, “approaches” and “steady increase”.

### Extract 20

*Industry in 1970 accounted for the largest amount, no, accounted for the largest proportion in different end users, the largest proportion in the UK CO2 emissions.*

*Other, UK emitted about 30 million carbon in other sources, in other uses, in other approaches.*

*Transport, however, transportation, transportation, transportation emitted more and more and more during the 34 years steady, more and more showing a steadily, showing a trend of stead, steadily, stable, steady increase in 2004. (Participant L)*

However, when the participants were much less certain about the meanings of unfamiliar words, this can be a long and struggling process. Often their final decision was to “just copy” if these happened to be the words in the accompanying texts of the graphs. For example, Participant G was not quite sure “what is the meaning of million tonnes carbon equivalent”. Although she adopted the strategy of “*I just copy this*”, later on, she was still repeatedly asking herself “*what carbon equivalent mean*”. Similarly she kept asking herself “*what’s the meaning of Year to Date, year to date, year to date, I don’t know what is year to date*”. A “just copy” strategy was also used to “solve” this problem. YTD is clearly written in her script.

When asked about the contribution of their writing ability on the AWT1 task performance, all the participants agreed that it is the writing ability, in particular, lexical knowledge, that matters most. The following comment is typical:

### Extract 21

*This is definitely so, for example, your knowledge of certain words and some sentence structures and so on, these are essential for the completion of the AWT1 tasks. For example, if you only know the words such as decline, increase, but you don’t have other more powerful words, your writing will become a bit boring and plain. (Participant J)*

### Extract 22

*I think the most challenging inhibitors may be that I could not find the right words straightaway, or even if I can remember of a word vaguely, but I may not be able to spell it correctly, then I will have to use a simpler word instead to be safe. (Participant S)*

Another aspect of writing, although not necessarily synonymous to English writing ability proper, the participants’ experience and expectation of academic writing, whether in English or Chinese,

did seem to affect the way how they interpreted the data and whether they would think it essential to include interpretations in AWT1 writings. Their comments on the purpose of describing graphs (see Figure 3, a key component of the working model: *Purposes of understanding the graphs and the cognitive demands of the writing task*) are thought provoking and raise some fundamental issues of IELTS AWT1 tasks – what are the purposes of the writing, and how are these communicated to and interpreted by test takers?

### Extract 23

*Usually when we write and describe data, we have a purpose, have an aim, why do we do this, but these tasks, we don't have a clear purpose. I don't know where to start and how to start to describe the information in the graphs, so I feel these tasks do not have a clear purpose, so I feel I can't have a clear idea or the logics of understanding what information is more important and how to summarise and organize them. I don't know what the key messages to summarise are. (Participant L)*

Her comment was fully supported by other participants. The following conversation between them on the importance of including interpretations as a natural part of academic writing is interesting.

### Extract 24

*Participant E: I still think we should write interpretations because these are for academic tasks.*

*Participant U: It is always like that in academic writing that you present the data and explain and interpret why, the reasons why there is such data.*

*Participant E: Yes, like that. We should write like that. In fact, some of the facts, we don't need to present them, while the interpretation is important.*

*Participant L: Yes, this is the sort of structure of academic papers, you first present data and then interpret the data. The purpose of presenting the data is to serve the interpretation and discussions. This is the pattern. It is always like that for academic writing. ... Why bother to present data alone?*

In fact, all the participants in this project insisted that they should be allowed to include their interpretation in the AWT1 writings as this is something expected in academic writing. This explains why there were a lot of personal interpretations and comments in the writings of the first four tasks before the special training session which emphasized that they should “describe” rather than “interpret” data.

In 4.3.4 below we report how the short test preparation training changed the ways that the participants described or interpreted the main features of graphs.

#### 4.3.4 Research question 4

*To what extent are the candidates' cognitive processes influenced by test preparation opportunities offered to them by the research team?*

In 4.2.1 we reported that the participants were constantly reminding themselves the second sentence of the task instruction (summarise the information by selecting and reporting the main features, and

make comparisons where relevant) to make sure that they keep focusing on the task, rather than being carried away too much by personal interpretations, predictions and comments based on prior knowledge which was not necessarily depicted in the graphs. We view this as the most discernable influence of the test preparation training on the cognitive processes. As presented in Appendix 10, we can see that the participants in the first round of AWT1 tasks made extensive interpretations, predictions and comments. Some were more reasonable than others, while some were simply wrong or irrelevant. In some scripts, more than 1/2 of the space were devoted to explaining the potential underlying reasons why a particular phenomenon existed and what actions should be taken. In the second round of AWT1 tasks, i.e., after the short test preparation training, the participants tended to refrain from making too many personal interpretations and comments. Although there are still interpretations which were not based primarily on the data of the graphs, it is obvious that the interpretations became more restrained, reasonable and reduced. There are fewer personal interpretations, comments and generalizations in the texts of Tasks B answers than of Tasks A. Overall, the participants had better understanding about the task requirements after the training. For example, Participant B acknowledged in the interview:

**Extract 25**

*At the beginning, I did not really understand the task requirements, but after the Wednesday training session, I have a better understanding of the tasks. ... For example, as for bar graph, there may be many bars in the graph, and a lot of information, but there is a limit of time, and your choice of words is also limited, you will have to select the most important information, not necessarily to include every bit of information. I think it is particularly useful for me to know this. .... Sometimes we read graphs when doing literature review. And quite often I try to use as much as possible to describe graphs, and try to explore and extract as much information as possible from graphs. As for AWT1 tasks, the first impression I had was I should try my best to interpret what's in the graphs. For example this graph, I try to find out why this is low and why that is high, trying to solve this problem. But I can get it wrong, and it would lead me to a wrong route. (Participant B)*

Similarly, Participant U agreed that “After training, it is unlikely that I will still do too much interpretation of the data as I did in the first set of tasks”. Participants E and W also held the same view:

**Extract 26**

*When I write, e.g. about the line graph, I always think about why there is such a change in the lines, why, this is the question I always ask myself, and add my own understanding/ interpretation in my writing. Only when you pointed out in the tutorial that it is not necessary to include my own interpretations of the data, the main thing is to describe the information in the graph, that's it, did I realize that I was not right to always try to interpret the data to find the underlying reasons for the changes in a graph, for example, the lines. (Participant E)*

**Extract 27**

*Before the Wednesday training, when I wrote the tasks, I tried to explain the reasons, the underlying reasons. For example, there are more and more population, and there are more houses, and the man-made facilities were becoming more and more useful for human beings and the environmental impacts of these are also getting bigger.... If it is your subject specialist area, it is quite natural that you have a tendency to reason and argue this way. (Participant W)*

This short test preparation training also seemed to clarify a confusion that the participants had about a similar type of writing tasks in the College English Test in China which in fact required test takers to explain why a particular phenomenon depicted in a graph happened. The explanations by Participants W and Q on their previous experiences shed light on why these students always had a strong tendency to interpret the data when doing IELTS AWT1 tasks.

#### Extract 28

*In AWT1 tasks, no matter what kind of graphs or diagrams, you are only asked to describe. However, in the Chinese tests, you are given a picture or a graph, you are not only asked to describe some kind of information at surface level, however, the most important part is to find out and write about the intended message and meaning. However, in AWT1 tasks, you are only asked to describe.... in CET and Gaokao (Note: Chinese university entrance examination) composition writing, you only use 20-30 words max to describe the graphs or pictures, but the following analyses are more important. (Participant W)*

#### Extract 29

*In essays, normally we present and describe a problem briefly and what follows is more important, you interpret and give some arguments and provides solutions to the problem. It is normally this kind of pattern for argumentative essays. This is the kind of training we have received since secondary education. (Participant Q)*

However useful the short test preparation training may have been to help the participants focus on describing rather than interpreting the graphs, we still noticed in think-aloud protocols that interpretations still feature, maybe as a natural process of understanding the graphs. As Participant T commented, although interpretations may not be totally avoidable, she would not include the interpretations in her actual writings.

#### Extract 30

*Before the training, I always had the desire to interpret the reasons, e.g. why is there an increase, why is there a decrease? After the training, I had to curb my desire of interpreting, but I still quite naturally, very naturally, or unconsciously try to interpret, though I will not include my interpretations in my writings, as I understand this is not what the task requires, this is not what the markers would be looking for; I have to refrain from writing too much, it has to be within 150 words. I don't need to bother anything not required. (Participant T)*

Data were also collected from the participants about their views on the research project and the short training (Appendix 6 Student evaluation questionnaire). Table 6 reports the descriptive statistics of their evaluation on the overall training, learning support, handout, content, teaching quality, and their learning from the training and their own contribution. As can be seen from Table 6, the participants overall had very positive comments on the research project and the training provided.

Score	Overall training		Learning support		Handouts		Content		Teaching quality		Learning from training		Own contribution	
	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
1														
2													1	7.1
3			1	7.1							3	21.4		
4	3	21.4	3	21.4	1	7.1	5	35.7	2	14.3	4	28.6	5	35.7
5	8	57.1	7	50.0	4	28.6	6	42.9	9	64.3	6	42.9	4	28.6
6	3	21.4	3	21.4	9	64.3	3	21.4	3	21.4	1	7.1	4	28.6
Mean	5.0	4.9	5.6	4.9	5.1	4.4	4.7							
STD	0.68	0.86	0.65	0.77	0.62	0.93	1.14							

**Table 6. Student evaluation of the project (n=14)**

We received very positive comments from the participants regarding the quality of the training and their genuine interest and commitment in participating in the project. The qualitative comments (see Appendix 11) provided further empirical evidence on the effects of the short training on the cognitive processes of completing AWT1 tasks. In particular, the participants noted the benefits of learning special vocabulary for making comparisons and trend assessments (the training booklet is available from the authors upon request). Analyses of the writing performances by the same students on different graphs indicate that there is some re-occurrence of formulaic phrases and verbs; and this is particularly so in the four tasks completed after the short training. This clearly demonstrates the coachability of the AWT1 tasks. However, this coachability may not on its own lead to a high level performance in such tasks, there is a possibility that test takers may produce some formulaic and rigid phrases without necessarily involving true understanding of the phrases or the graphic information.

## 5 CONCLUSION

This research aimed to understand the cognitive processes of candidates taking IELTS AWT1 with different graphic prompts at two different time points – before short training on how to do AWT1 tasks and post-training. In particular, it explored the extent to which candidates' cognitive processes were affected by the use of different graphs, their “graphicacy” (Wainer 1992, p 16) and English writing abilities, and the short training. A grounded and multi-layered case study approach was employed to collect data on candidates' cognitive processes. Eighteen intending IELTS candidates were recruited from a large Chinese university, and completed eight AWT1 tasks while thinking aloud their processes of doing the tasks (4 before training and 4 after training) under examination conditions. In addition, their English writing abilities and graphicacy were also collected, and post-task interviews were conducted with all the participants.

The think-aloud protocols were analysed using a qualitative computer programme winMAX to identify the common patterns in the cognitive processes. The patterns which emerged from the think-aloud



protocols were interpreted with reference to the other four main sources of data – AWT1 scripts, the interviews, graphicacy questionnaire and English writing abilities. From these data sources, a model of cognitive process was developed, consisting of three interrelated stages – comprehending non-graphically presented task instructions, comprehending graphs, and re-producing graph comprehension in written discourse in English as a foreign language. We used this model to guide our analyses to address the four research questions. Below we summarize the main findings of each research question.

- 1 With regard to the *effects of types of graphs on cognitive processes*, it was found that the types and conventions of graphic prompts did matter. They affected how the participants processed the graphic information and how they followed the graphic conventions to re-produce their graph comprehension in written discourse in English as a foreign language. Such effects of different AWT1 graphic prompts on cognitive processes were clearly evidenced in the mean scores of the writing performances, in the use of vocabulary, and in whether and how they would make comparisons or trend assessments. The graphic conventions or “cognitive naturalness” of graphs (Tversky 1995) affected the processes of comprehending and re-producing graphic information. Candidates had a strong tendency to make trend assessments when describing line graphs, and make discrete comparisons when describing bar and pie charts. When describing a statistical table - the most challenging of the AWT1 tasks in this study, candidates tended to do some calculations of the numbers to develop their reasoning. Different types of graphs also activated the use of different vocabulary types.
- 2 With regard to the *effects of graphicacy on cognitive processes*, we found that: although graph familiarity as measured via the graphicacy questionnaire did not seem to affect AWT1 task performance in terms of the scores for the writing performances, the participants clearly expressed some potential psychological impact of graph familiarity on their task performance. In addition, the effects of the participants’ familiarity with and understanding of graphic conventions also influenced the way they processed and re-produced the graphic information (see 1 above).
- 3 The effects of the writing abilities on cognitive processes were manifested in the use of different vocabulary choices for different graphic prompts. There is also a strong correlation between the mean performances of the AWT1 tasks under think-aloud conditions and the AWT2 task of topic-based argumentative writing. Besides the participants’ English writing abilities, their expectations and experiences of academic writing also shaped the way that they interpreted and re-produced their graph comprehension in written discourse in English as a foreign language.
- 4 The influence of the special training was strong; this clearly demonstrated the coachability of the AWT1 tasks. Whether or not the candidates tried to make personal interpretations and comments by linking the graphic information and their domain knowledge about the graphs were clearly influenced by the short test preparation training they received, although they did not necessarily agree with the AWT1 task requirements on “describe”.

As we discussed in Section 1.1, this research project addressed two broad areas of interest identified by the IELTS Joint Research Committee – (a) “test development and validation issues” in relation to “the cognitive processes of IELTS test takers”, and (b) “issues in test impact” in relation to “test preparation practice”. The findings of this study have implications for the validation and development of AWT1 tasks from the perspective of test takers’ cognitive processes. The working model of cognitive processes

(Figure 3) will be a useful framework for designing AWT1 tasks, considering the three interrelated stages of AWT1 cognitive processes – comprehending the non-graphically presented task instructions, comprehending graphic information, and re-producing graph comprehension in written discourse in English. The findings of this research project suggest that when designing AWT1 tasks, we need to consider what accompanying instructions should be provided with the graphs and whether the summary-like introductory sentence should serve as the entry point for levelling the playing field for test takers with varying background knowledge, graphic skills and reading abilities. As AWT1 tasks are essentially reading/writing integrated tasks, it is essential to consider not only the graphic skills of the test takers but also their reading abilities. Although the tasks may require just reading one summary-like sentence in the instructions, the sentence can actually serve as a guide for test takers to grasp the key information of the graphs, and it may become more and more critical when test takers move further down the line to re-produce the main information embedded in the graphic prompts.

It is also important to consider and compare the potential differential effects of different graphic prompts on the cognitive processes of test taking as well as on test performance, because different graphs have different conventions in presentation and interpretation. Beside the types of graphs, we also need to consider the information density, or data points of the graphs, as information density clearly has impact on how and what test takers can extract from the graphs. Other display characteristics of graphs also need to be taken into account when test designers consider the information density of the graphs, e.g. whether to include caption or title. As display characteristics of graphs can affect task performance, we would be very cautious about O’Loughlin and Wigglesworth’s (2003) recommendation: “a variety of presentation types can be encouraged and manipulated” (p 114). Although the final scores that two pieces of writings may receive may be similar, regardless of which graph type they are based on, it is very clear, as shown in our study, that the use of different graphic prompts can activate different forms of “cognitive naturalness” in comprehending and re-presenting graphic information, and can invite candidates to produce different writings, particularly in terms of the use of different vocabulary choices. The quality of a graph is not just innate in its display features (Bertin 1983; Saint-Martin 1990; Tufte 1983); it is in fact determined in interaction between these features and the characteristics of graph readers (e.g., their familiarity with and experience in using graphs). These interactions are the starting points for investigating the cognitive processes of AWT1 tasks.

We noticed that AWT1 tasks seemed to be highly trainable and the writings were highly predictable: the participants were easily influenced in their ways of re-producing graph comprehension by the short training. The high coachability of the tasks and the rigid use of the formulaic phrases and words in AWT1 writings without necessarily understanding them bring up the ongoing question of how far AWT1 fosters rote-learning of fixed phrases. They also raise an equity issue. If test takers have not attended intensive training courses and are not aware of the task requirement of “describing” only, they could be disadvantaged if they tried to integrate the information depicted in the graphs with their prior content and graphic knowledge. Whether and to what extent reasonable “interpretation” should be allowed in AWT1 writings have implications for comparability in marking. The participants’ strong view that interpretation of data is an integral and natural part of academic writing also oblige us to consider the extent to which the AWT1 tasks mirror academic writing in target language use domains, for example, when describing and interpreting lab results.

The findings of this research study are useful not only to language test providers and language testing researchers, but also to intending IELTS candidates and English language professionals and teachers to develop a greater understanding of the AWT1 tasks, as well as other language tests (listening, speaking and writing assessments) using graphs as prompts. It contributes to the development of theories and

practices in foreign/second language writing, in particular, in relation to our understanding of the roles that non-language knowledge and skills (i.e., graphicacy) can play in language assessments using visual input or data. In order to further develop our understandings of tasks using graphic prompts, we acknowledge several limitations of this research study and present two pointers for future studies.

Firstly, although sufficient training for think-aloud was provided to the participants in this research project, the effects of think aloud on test performance may never be removed completely, it is important to examine in detail the extent to which think-aloud may have affected the participants' cognitive processes and their actual writing performances. Detailed discourse analyses on the changes between think-aloud protocols and what the participants finally put down in their writings would be a useful perspective to analyse the effects of think aloud on composition and more importantly on the cognitive processes of re-producing graph comprehension. Investigation of content coverage would be another useful perspective to understand the cognitive processes of AWT1 tasks, for example, what is included and what is not included, how and why, and the extent to which the inclusion or exclusion is related to the test takers' English writing abilities and graph familiarity.

Secondly, our analyses on the participants' comprehension of graphic information did not focus specifically on the sequence that the different graphic components were noticed and comprehended. Instead we focused on the participants' overall comprehension of graphs. As Kennedy (1974) argued, "sometimes we read a label or caption before looking at the picture, but more often, probably, we notice the picture first and recognise the pictured object without any help from the accompanying words" (p 7). Although our data indicated that the participants read or skipped the accompanying textual instructions before focusing on the graphs, it would be useful to systematically analyse which graphic components were noticed and comprehended first, and whether the comprehension of different graphic components was affected by the participants' graphic skills and types of graphs. In addition, it would be useful to investigate more precisely the timeline of AWT1 task completion, for example, the time spent in reading and comprehending the graphs, time spend in re-producing the graphic information in writing, checking and re-checking understandings, and re-writing. Future studies may examine the sequence of comprehending graphic components, as the findings could have clear implications for AWT1 task design.

## **ACKNOWLEDGEMENTS**

The authors acknowledge the role of the IELTS Partners in making this study possible: The British Council provided the research grant which enabled us to conduct this study.

Any opinions, findings, conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the British Council, its related bodies or its partners.

The authors acknowledge the administrative support of Zhejiang University. Special thanks are also due to the student participants who wish to be acknowledged by name, including Hu Zhinan, Li Kangkang, Li Wenting, Liu Yuguo, Ni Shuang, Zhang Kejun, plus others who wish to remain anonymous. Their commitment and enthusiasm made this project possible.

## REFERENCES

- Åberg-Bengtsson, L, and Ottosson, T, 2006, 'What lies behind graphicacy? Relating students' results on a test of graphically represented quantitative information to formal academic achievement' *Journal of Research in Science Teaching*, vol 43, no 1, pp 43-62
- Bertin, J, 1983, *Semiology of graphics: Diagrams, networks, maps* University of Wisconsin Press, Madison, WI
- Carpenter, PA, and Shah, P, 1998, 'A model of the perceptual and conceptual processes in graph comprehension' *Journal of Experimental Psychology: Applied*, vol 4, no 2, pp 75-100
- Carswell, CM, Frankenberger, S, and Bernhard, D, 1991, 'Graphing in depth: Perspectives on the use of three-dimensional graphs to represent lower-dimensional data', *Behaviour and Information Technology*, vol 10, no 6, pp 459-474
- Curcio, FR, 1987, 'Comprehension of mathematical relationships expressed in graphs' *Journal for Research in Mathematics Education*, vol 18, no 5, pp 382-393
- Farquhar, AB, and Farquhar, H, 1891, *Economic and industrial delusions: A discourse of the case for protection* Putnam, New York
- Feeney, A, Adams, J, Webber, L, and Ewbank, M, 2004, 'Individual differences in graphical reasoning' *Lecture Notes in Computer Science*, 2980, pp 271-285
- Feeney, A, Hola, A, Liversedge, S, Findlay, J, and Metcalfe, R, 2000, 'How people extract information from graphs: Evidence from a sentence-graph verification paradigm' in *Theory and application of diagrams*, eds M Anderson, P Cheng & H Haarslev, pp 149-161
- Feeney, A, & Webber, L, 2003, 'Analogical representation and graph comprehension' in *Smart graphics*, eds A Butz, A Krüger & P Olivier, pp 212-221
- Fischer, MH, Dewulf, N, and Hill, RL, 2005, 'Designing bar graphs: Orientation matters' *Applied Cognitive Psychology*, vol 19, no 7, pp 953-962
- Freedman, EG, and Shah, P 2002, 'Toward a model of knowledge-based graph comprehension' *Diagrams*, pp 18-30
- Friel, SN, Curcio, FR, and Bright, GW. 2001, 'Making sense of graphs: Critical factors influencing comprehension and instructional implications' *Journal for Research in Mathematics Education*, vol 32, no 2, pp 124-158
- Fry, E, 1981, 'Graphical literacy', *Journal of Reading*, vol 24, no 5, pp 389-392
- Ginther, A, 2002, 'Context and content visuals and performance on listening comprehension stimuli' *Language Testing*, vol 19, no 2, pp 133-167
- Golub-Smith, M, Reese, C, & Steinhaus, K, 1993, *Topic and topic type comparability on the test of written English*, Educational Testing Service, Princeton, NJ

- Gruba, P 1997, 'The role of video media in listening assessment' *System*, vol 25, pp 335-346
- Guthrie, JT , Weber, S , and Kimmerly, N 1993, 'Searching documents: Cognitive processes and deficits in understanding graphs, tables, and illustrations' *Contemporary Educational Psychology*, vol 18, no 2, pp 186-221
- Hollands, JG , & Spence, I, 1998, 'Judging proportion with graphs: The summation model' *Applied Cognitive Psychology*, vol 12, no 2, pp 173-190
- Hollands, JG , & Spence, I, 2001, 'The discrimination of graphical elements' *Applied Cognitive Psychology*, vol 15, no 4, pp 413-431
- Katz, IR , Xi, X , Kim, H -J , and Cheng, PCH, 2004, *Elicited speech from graph items on the test of spoken English*, ETS Research Report 74, Educational Testing Service, Princeton, NJ
- Kennedy, JM, 1974, *Psychology of picture perception*, Jossey-Bass, San Francisco
- Kintsch, W, 1988, 'The role of knowledge in discourse comprehension: A construction-integration model', *Psychological Review*, 95, no 2, pp 163-182
- Körner, C, 2004, 'Sequential processing in comprehension of hierarchical graphs' *Applied Cognitive Psychology*, vol 18, no 4, pp 467-480
- Kosslyn, SM, 1989, 'Understanding charts and graphs' *Applied Cognitive Psychology*, vol 3, no 3, pp 185-225
- Kuckartz, U, 1998, *Winmax: Scientific text analysis for the social sciences* (Version 98 Pro) VERBI Software-Consult-Research GmbH (Germany), Berlin
- Lewandowsky, S, and Spence, I, 1989, 'The perception of statistical graphs' *Sociological Methods Research*, vol 18, nos 2 and 3, pp 200-242
- Lohse, GL, 1993, 'A cognitive model for understanding graphical perception', *Human-Computer Interaction*, vol 8, no 4, pp 353-388
- Meyer, J, Shinar, D, and Leiser, D, 1997, 'Multiple factors that determine performance with tables and graphs', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol 39, no 2, pp 268-286
- Mickan, P, and Slater, S, 2003, 'Text analysis and the assessment of academic writing' in *IELTS Research Reports Volume 4*, ed R Tulloh, IELTS Australia Pty Limited, Canberra, pp 59-88
- Mickan, P, Slater, S, and Gibson, C, 2000, 'Study of response validity of the IELTS writing subtest' in *IELTS Research Reports Volume 3*, ed R Tulloh, IELTS Australia Pty Limited, Canberra, pp 29-48
- Moore, PJ 1993, 'Metacognitive processing of diagrams, maps and graphs' *Learning and Instruction*, vol 3, no 3, pp 215-226

- O'Loughlin, K, and Wigglesworth, G, 2003, 'Task design in IELTS academic writing task 1: The effect of quantity and manner of presentation of information on candidate writing' in *IELTS Research Reports Volume 4*, ed R Tulloh, IELTS Australia Pty Limited, Canberra, pp 89-130
- Peebles, D, and Cheng, PCH, 2002, 'Extending task analytic models of graph-based reasoning: A cognitive model of problem solving with cartesian graphs in act-r/pm' *Cognitive Systems Research*, vol 3, no 1, pp 77-86
- Peebles, D, and Cheng, PCH, 2003, 'Modeling the effect of task and graphical representation on response latency in a graph reading task', *Human Factors*, vol 45, no 1, pp 28-47
- Pinker, S, 1990, 'A theory of graph comprehension', in *Artificial intelligence and the future of testing*, ed R Freedle, Lawrence Erlbaum Associates, Hillsdale, NJ, pp 73-126
- Roth, WM, 2002, 'Reading graphs: Contributions to an integrative concept of literacy' *Journal of Curriculum Studies*, vol 34, no 1, pp 1-24
- Roth, WM, 2003, *Toward an anthropology of graphing: Semiotic and activity-theoretic perspectives*, Kluwer Academic Publishers, Amsterdam
- Saint-Martin, F, 1990, *Semiotics of visual language*, Indiana University Press, Bloomington, IN
- Schnotz, W, Picard, E, and Hron, A, 1993, 'How do successful and unsuccessful learners use texts and graphics?' *Learning and Instruction*, vol 3, no 3, pp 181-199
- Shah, P, and Carpenter, P A, 1995, 'Conceptual limitations in comprehending line graphs' *Journal of Experimental Psychology: General*, vol 124, no 1, 43-61
- Shah, P, Freedman, E G, and Vekiri, I, 2005, 'The comprehension of quantitative information in graphical displays' in *The cambridge handbook of visuospatial thinking*, eds P Shah and A Miyake, Cambridge University Press, New York, pp 426-476
- Shah, P, and Hoeffner, J, 2002, 'Review of graph comprehension research: Implications for instruction' *Educational Psychology Review*, vol 14, no 1, pp 47-69
- Shah, P, Mayer, RE, and Hegarty, M, 1999, 'Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension' *Journal of Educational Psychology*, vol 91, no 4, pp 690-702
- Simkin, D, and Hastie, R, 1987, 'An information-processing analysis of graph perception' *Journal of the American Statistical Association*, vol 82, no 398, pp 454-465
- Stewart, B, Hunter, A, and Best, L, 2008, 'The relationship between graph comprehension and spatial imagery: Support for an integrative theory of graph cognition' *Diagrams*, pp 415-418
- Trickett, SB, and Trafton, JG 2004, 'Spatial transformations in graph comprehension' *Diagrams*, pp 1-19
- Trickett, SB, and Trafton, JG, 2006, 'Toward a comprehensive model of graph comprehension: Making the case for spatial cognition' *Diagrams*, pp 286-300

Tufte, ER, 1983, *The visual display of quantitative information* Graphics Press, Cheshire, CT

Tversky, B 1995, 'Cognitive origins of graphic conventions' in *Understanding images*, ed FT Marchese, Springer-Verlag, New York, pp 29-53

Vernon, MD, 1946, 'Learning from graphical material' *British Journal of Psychology*, vol 36, no 3, pp 145-158

Vessey, I, 1991, 'Cognitive fit: A theory-based analysis of the graphs versus tables literature' *Decision Sciences*, vol 22, no 2, pp 219-240

Wainer, H 1992, 'Understanding graphs and tables' *Educational Researcher*, vol 21, no 1, pp 14-23

Xi, X, 2005, 'Do visual chunks and planning impact performance on the graph description task in the speaking exam?' *Language Testing*, vol 22, no 4, pp 463-508

Zacks, J, and Tversky, B, 1999, 'Bars and lines: A study of graphic communication' *Memory and Cognition*, vol 27, no 6, pp 1073-1079

## APPENDIX 1: CONSENT FORM

Dear Participant,

Thank you for agreeing to participate in the Cog-Pro research project funded by British Council through its IELTS Research Programmes ([www.ielts.org](http://www.ielts.org)) and carried out by the consultants from the University of Bristol ([www.bristol.ac.uk/education](http://www.bristol.ac.uk/education)) in October and November 2008. This project aims to gain better understanding of IELTS Academic Writing Task One (AWT1) that uses graphs as test prompts. The data collection for this research would involve you (in the time order):

- taking IELTS academic writing tasks 1 and 2,
- answering a questionnaire measuring your graph familiarity and comprehension,
- taking IELTS AWT1 tasks, while thinking aloud your test taking process,
- having free training on how to take IELTS AWT1 tasks provided by the consultants,
- taking IELTS AWT1 tasks again, while thinking aloud your test taking process,
- (some of you) being interviewed on a one-to-one basis on how you took the AWT1 tasks

Your think-aloud and interviews will be audio-recorded. Your participation is voluntary and will not be paid, but we will provide you with free training and assessment on AWT1 during data collection (see above). As a potential IELTS test takers, you will benefit from participating in this research. You have the right to withdraw your participation any time if so you wish without any consequences, but we would like to encourage you to work your best until the end of the project to maximize your learning benefits.

We would like to ask for your consent formally, as recommended by the ethical guidelines for the conduct of research of International Language Testing Association ([www.iltaonline.com](http://www.iltaonline.com)) and British Association for Applied Linguistics ([www.baal.org.uk](http://www.baal.org.uk)). All data collected for this research (including your age, gender, graph familiarity, test performance, audio-recorded think-aloud protocols and interviews) will be anonymised and used solely for this research in a fair and respectful manner, in its research report and subsequent academic publications and disseminations. Your data will be protected in accordance with the Data Protection Act 1998.

We would be very grateful if you could read this consent form carefully and sign below, and if you do sign them, to indicate the manner in which you would like your contribution to be acknowledged in the research report and any publications and disseminations based on this.

Your Chinese name [in print] \_\_\_\_\_ Signature \_\_\_\_\_ Date \_\_\_/\_\_\_/\_\_\_

Please select either A or B for acknowledgement of contribution to this research.

- I would like acknowledgement and thanks expressed generically, i.e. to the students at Zhejiang University. OR [please tick here ]
- I would like acknowledgement and thanks expressed to mention me explicitly, i.e. to the students at Zhejiang University, which includes (my name). [please tick here ]

If you have any queries about the Cog-Pro project or this consent form, please get in touch.

Best wishes

Guoxing Yu, Graduate School of Education, University of Bristol



## APPENDIX 2: IELTS AWT1 TASKS WITHOUT THINK-ALOUD (STAGE 1)

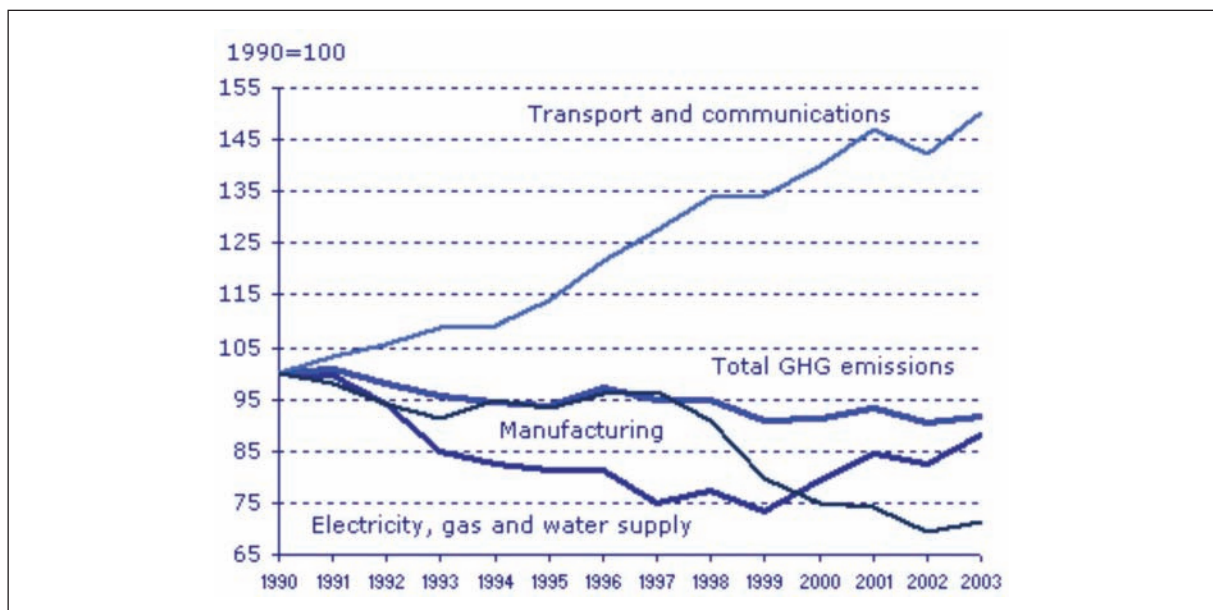
### WRITING TASK ONE

You should spend about 20 minutes on this task.

The following graph shows the total UK greenhouse gas (GHG) emissions between 1990 and 2003 in comparison to 1990 as 100 in different end users

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.



### WRITING TASK TWO

You should spend about 40 minutes on this task.

Write about the following topic:

Once children start school, the teachers would have more influence in their intellectual and social development than parents.

To what extent do you agree or disagree?

Give reasons for your answer and include any relevant examples from your own knowledge or experience.

Write at least 250 words.

**APPENDIX 3: GRAPHICACY QUESTIONNAIRE**

This questionnaire will collect your personal information and your experience, familiarity and understanding of graphs (图表、数字统计图、数字统计表格、示意图、流程图等, including bar, line, chart, diagram, and table with numerical data). You are asked to provide **ONE** answer by ticking the relevant box or filling the blank which describes best your **OWN** situation. Please answer them **independently and honestly**. There is no right or wrong answer.

For example: Male []

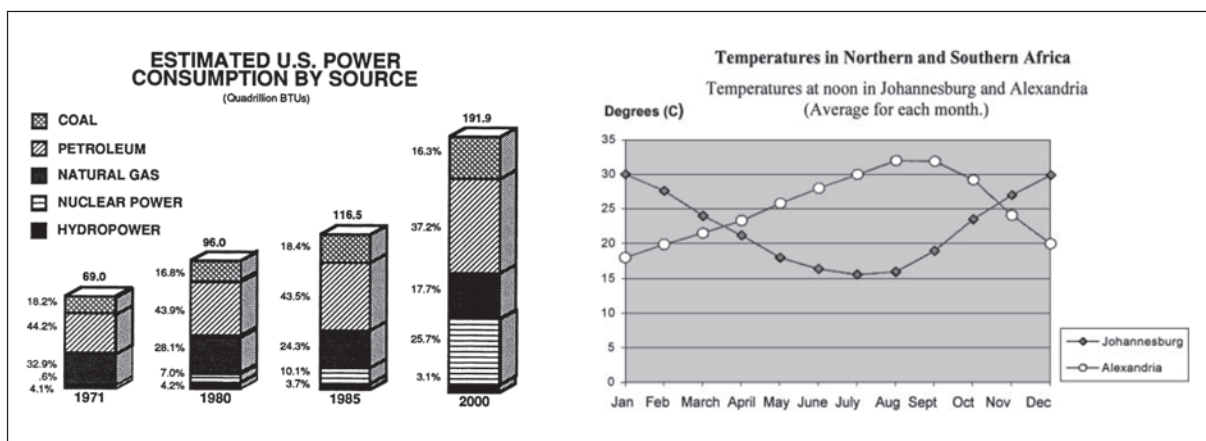
If you don't fully understand a question, please ask the tutor for an explanation.

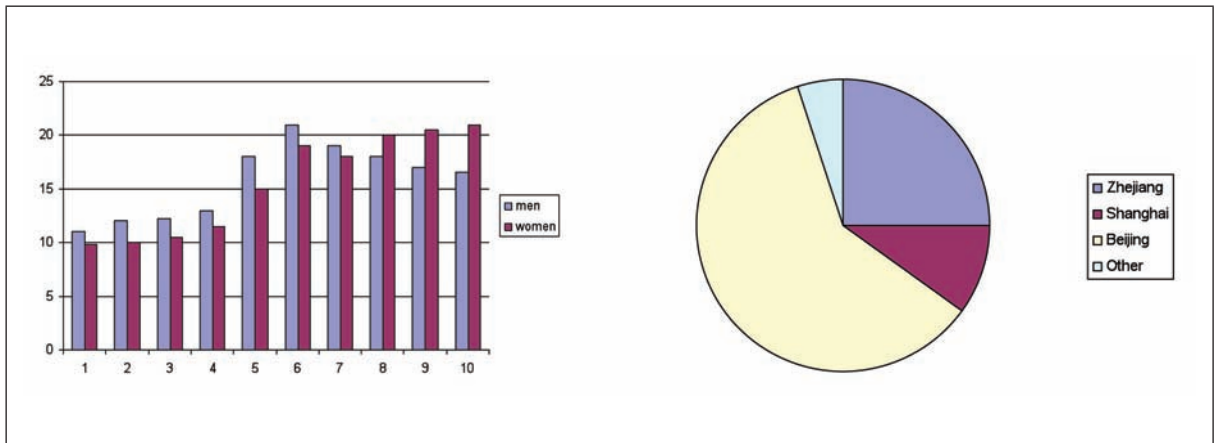
Personal information

- 1 Your contact mobile phone number \_\_\_\_\_
- 2 Your email address \_\_\_\_\_ (Please in print)
- 3 Your CHINESE Name \_\_\_\_\_ (Please in print)
- 4 Gender: Male [] Female []
- 5 Faculty/Department/Specialism \_\_\_\_\_
- 6 Year Group: [1] [2] [3] [4] [5] [6] [7]
- 7 Status: Undergraduate []  
Postgraduate [] If postgraduate, Master [] or PhD []
- 8 Have you taken an IELTS test? Yes [] No []
- 9 If YES to Question 8, when did you take your last IELTS test? [yyyy/mm] \_\_\_\_\_
- 10 If NO to Question 8, are you planning to take IELTS test? Yes [] No []
- 11 If YES to Question 10, when are you planning to take IELTS test? [yyyy/mm] \_\_\_\_\_

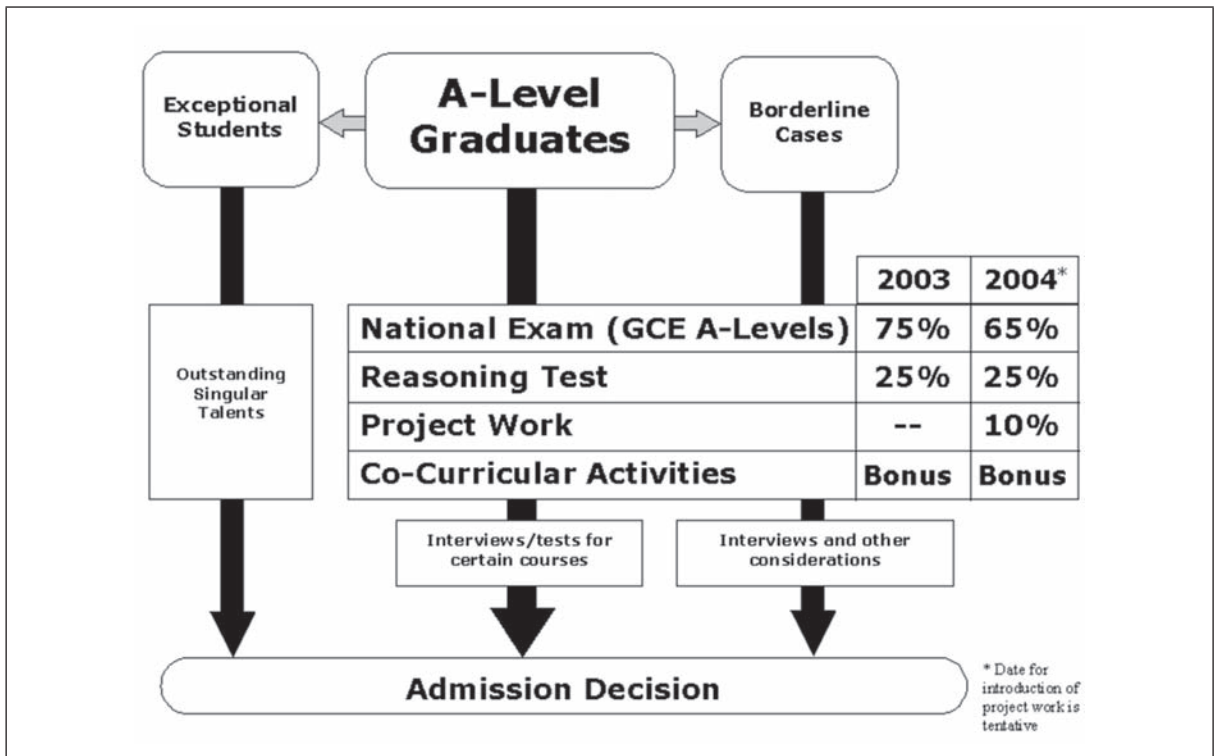
Questions on your graphicacy

Below are several statements concerning your experience, familiarity and preference in using graphs (图表、数字统计图、数字统计表格、示意图、流程图等, including bar, line, chart, diagram, and table with numerical data). Six examples of these graphs are given below. We will use GRAPHS as a generic term covering all these different types of graphs in this questionnaire, so your answer should reflect the AVERAGE of using these different types of graphs, unless otherwise stated.





Countries	packaging	
	Tonnes exported in bags	Tonnes exported in containers
China	652	2001
India	4361	5002
New Zealand	82	44032



**QUESTIONS START HERE**

Please tick ONE number which best describes your own situation. There is no right or wrong answer.

	Never <span style="float:right">→</span> Very Often					
	[1]	[2]	[3]	[4]	[5]	[6]
I use special computer software to produce graphs.	[1]	[2]	[3]	[4]	[5]	[6]
As part of my academic study, I need to produce graphs.	[1]	[2]	[3]	[4]	[5]	[6]
As part of my academic study, I need to interpret graphs.	[1]	[2]	[3]	[4]	[5]	[6]
I read graphs in the popular press (e.g. magazines, newspapers).	[1]	[2]	[3]	[4]	[5]	[6]
When I read a graph, I try to identify the main trend or the overall pattern that the graph is trying to convey.	[1]	[2]	[3]	[4]	[5]	[6]
When I read a graph, I try to think about the possible underlying reasons for the main trend or the overall pattern of the data presented.	[1]	[2]	[3]	[4]	[5]	[6]
When I read a graph, I do not study it in detail.	[1]	[2]	[3]	[4]	[5]	[6]
When I encounter a graph in a text in popular press (e.g. magazines, newspapers), I tend to ignore/skip it.	[1]	[2]	[3]	[4]	[5]	[6]
When I encounter a graph in an academic paper, I tend to ignore/skip it.	[1]	[2]	[3]	[4]	[5]	[6]

	Strongly disagree <span style="float:right">→</span> Strongly agree					
	[1]	[2]	[3]	[4]	[5]	[6]
I am familiar with reading bar graphs.	[1]	[2]	[3]	[4]	[5]	[6]
I am familiar with reading line graphs.	[1]	[2]	[3]	[4]	[5]	[6]
I am familiar with reading pie charts.	[1]	[2]	[3]	[4]	[5]	[6]
I am familiar with reading diagrams representing a process.	[1]	[2]	[3]	[4]	[5]	[6]
I am familiar with reading tables with numerical data.	[1]	[2]	[3]	[4]	[5]	[6]
I notice errors or misinterpretations in graphs presented in the popular press.	[1]	[2]	[3]	[4]	[5]	[6]
I notice errors or misinterpretations in graphs presented in academic papers in my field.	[1]	[2]	[3]	[4]	[5]	[6]
I recognize the different components of a graph (e.g. X and Y axes, legends, colours).	[1]	[2]	[3]	[4]	[5]	[6]
I recognize how the different components of a graph (e.g. X and Y axes, legends, colours) are combined to represent the data.	[1]	[2]	[3]	[4]	[5]	[6]
I understand the relationships between a graph and the numerical data it represents.	[1]	[2]	[3]	[4]	[5]	[6]

I can identify the relationships or the patterns displayed in one graph.	[1]	[2]	[3]	[4]	[5]	[6]
I can identify the relationships or the patterns displayed in a few graphs about one similar theme	[1]	[2]	[3]	[4]	[5]	[6]
I can tell when one type of graph is a better representation of the data than another.	[1]	[2]	[3]	[4]	[5]	[6]
I can identify a poorly constructed graph.	[1]	[2]	[3]	[4]	[5]	[6]
I can revise and improve a poorly constructed graph.	[1]	[2]	[3]	[4]	[5]	[6]
I can describe the general trend or overall pattern of a graph in words.	[1]	[2]	[3]	[4]	[5]	[6]
I can use a graph to describe/convey the general trend or overall pattern of numerical data.	[1]	[2]	[3]	[4]	[5]	[6]
I find graphs useful to vividly represent the numerical data.	[1]	[2]	[3]	[4]	[5]	[6]
I find graphs helpful for me to remember the key information in the numerical data.	[1]	[2]	[3]	[4]	[5]	[6]
Graphs are a waste of space in a text.	[1]	[2]	[3]	[4]	[5]	[6]
I am concerned that I can not fully demonstrate my writing ability in IELTS Academic Writing Task One because I am not good at interpreting graphs.	[1]	[2]	[3]	[4]	[5]	[6]
I may do better in IELTS Academic Writing Task One using familiar graphs than unfamiliar ones.	[1]	[2]	[3]	[4]	[5]	[6]
I would prefer one type of graph to be used in IELTS Academic Writing Task One.	[1]	[2]	[3]	[4]	[5]	[6]
Special training on how to interpret graphs would be helpful for me to get a higher score in IELTS Academic Writing Task One.	[1]	[2]	[3]	[4]	[5]	[6]
Overall, on a scale of 1-6, how would you rate your own experience in using graphs?	[1]	[2]	[3]	[4]	[5]	[6]



Overall, on a scale of 1-6, how would you rate your own ability in interpreting graphs?	[1]	[2]	[3]	[4]	[5]	[6]
---	-----	-----	-----	-----	-----	-----



ADDITIONAL COMMENTS you want to make about your experience, familiarity and proficiency of using graphs. You can respond in English and/or Chinese.

---



---

Thank you for completing this questionnaire.

## APPENDIX 4: THINK-ALOUD TRAINING DOCUMENT

Purpose of collecting think-aloud protocols:

In the Cog-Pro project, you will be asked to think-aloud while doing the IELTS Academic Writing Task One. The main purpose of collecting your think-aloud protocols is to understand your test-taking process. It will also help you to identify where you could and should improve through listening to your own think-aloud protocols later; and we are also going to use some think-aloud protocols as examples in the group training session to demonstrate what strategies you have used in completing the AWT1 tasks and how best to complete such tasks.

Practise thinking aloud

- The most important thing is to keep talking, i.e., verbalizing what you are doing during the whole process including:
  - what you are reading,
  - what you are thinking and
  - what you are writing.
- You can use English and/or Chinese.
- Don't worry about grammar or sentence structure at all.
- The tutor will remain silent unless you stop talking for more than 10 seconds. In that case, the tutor will show you a white paper/card with TALK written on it.
- You are required to start recording with:

My name is \_\_\_\_\_ .

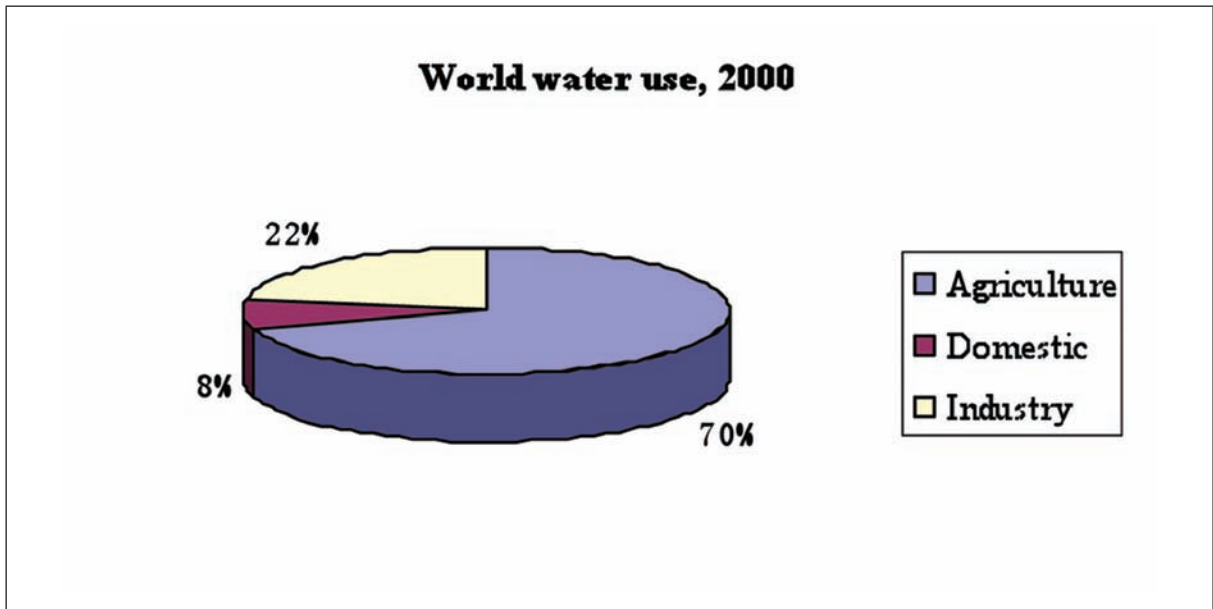
It is now \_\_\_\_\_ o'clock, on \_\_\_\_\_ of October 2008.

The task is \_\_\_\_\_ (Read the first paragraph of the task: e.g., The following graph shows \_\_\_\_\_)

Then think-aloud when you do the AWT1 task.

**Examples:**

- $124+3546 = ?$
- $124 \times 378 = ?$
- The average mass of 3 parcels is 6 kg. Two of the parcels have a mass of 4.6 kg and 6 kg. Find the mass of the third parcel.
- Try to think-aloud when doing the following AWT1 task.



	<i>China</i>	<i>India</i>	<i>New Zealand</i>	<i>Canada</i>
Agriculture	69%	92%	44%	8%
Domestic	9%	5%	46%	12%
Industry	22%	3%	10%	80%

**Water use, selected countries, 2000**

**Below is an example of a think-aloud protocol when doing the task above.**

*(Taken from Li (2006) a MSc dissertation supervised by the lead author at the University of Bristol)*

... The chart and table above give information about the way in which water was used in different countries in 2002, 2000. So we have a chart and a table. And then I have to summarise the information by selecting and reporting the main features, and make comparisons where relevant. So the tasks are the same. But we have different graphs. One is a pie chart, the other is a table. Now I am looking at the pie chart. The title of the pie chart is World Water Use, 2000. And we have 3 parts, agriculture, domestic and industry. And... ok, and each has different proportions. And then I am looking at the second graph which is a table. It is about water use in selected countries in the year 2000. So we have China, India, New Zealand, and Canada in the agriculture, domestic, industry. Ok. So it seems that there is connection between these two graphs. One is about the world water use in 3 categories. The other one is about eh, water use in selecting countries in these 3 areas. Yes. Now we have got a rough picture of what picture like. Now I am trying to find the main features.

In the first one, obviously, agriculture took the main proportion, 70%, and then it is industry and then domestic use. In the second graph, I found that in different countries, the proportion of the three categories are different. Like in India, agriculture took about 92% while Canada is only 8%. Now I can compare these figures. But there is much more information in the

table than in the pie chart. So I will concentrate more on the table. Ok, I think I am going to give, make a very very brief draft. I will firstly give an opening paragraph, and then I am going to talk about, yeah, the first pie chart and then second table. If I have time, I will draw a conclusion. If I don't have time, just forget it. Ok, now start.

Opening paragraph, eh... we have some sentence pattern like report. Ok, it is reported that, ok, it is reported that in the pie chart that, oh, no no I made a mistake. Because I just directly to the second paragraph. I should have the opening paragraph. So, I should firstly say the chart and table, eh... I see, chart and table, below, yes, ok, below, describe, we use describe instead of giving information about, describe, how water was used in the whole world. I am changing the wording and paraphrasing in the world as well as in five different countries in the year 2000. Because it is in year 2000, actually it is past tense. This is very important... past tense. And then, describing the first graph. As something shows, as the pie chart shows, eh, agriculture, I should have a phrase here, take, account for is better. Agriculture accounted, past tense, accounted for 70% of the world water use in 2000. Now I am comparing so I can use the link word while, while domestic, while industry, industry took 22% and domestic use, here use is a noun, domestic use, 8%. Because I only give the figure, and then I should give some comments, so the amount of the water used by agriculture was, say, was more than twice. Because 70% and together, eh, agriculture took 70% but together domestic and industry only took 30%. So you can say the amount of water used by agriculture was more than twice the amount of industrial and domestic use. Ok, seems enough.

Now, I am moving on to the table. This is more difficult and more complex. Therefore, I should pay more attention to it. Let's say. According to the table, I think I can, Ok, I think I can firstly give a brief account of the main information given in the table like in the five countries. No, four countries, one two three four, make a mistake, four countries, not five countries, just now miscounted. One of the four countries, yes, China and India used more water in agriculture while New Zealand and Canada not. And ok, so in China and India, agriculture accounted for the most water user while in New Zealand and Canada, it is not the case. Ok, ok, we will start with this. According to the table, India and China, agriculture, oh, maybe I should just refine wording in the last sentence of the second paragraph. I should say, the amount of the water used by agriculture was more than twice the amount of industrial and domestic use together. I can add together to make it more accurate. Now I am continuing. According to the table of India and China, agriculture, eh, took the largest proportion of water use in the year 2000. eh, yes, ok, yes, right...ok...took the largest proportion of water use in the year 2000. India used 92% while China, I am comparing, while China 69%. Eh, however, in New Zealand and Canada, because Canada use the smallest amount of water in agriculture, I can put Canada first. In Canada and New Zealand, oh, I should change, it didn't mention, because it is different, in New Zealand, domestic use accounts for biggest proportion in Canada, it's industry. I should change. I should describe it one by one. However, in Canada, industry took the, I should, I should change the wording, paraphrase, industry took the, took the most amount of water use. The most amount, the largest amount. Industry took the largest amount of water use reaching 80%. While in New Zealand, I can use the sentence pattern to emphasize. It was, while in New Zealand, it was domestic use that accounted for the biggest water use, ok, so about largest water user in the four countries.



And now, I should discuss, I think, I can discuss the smallest water user. Yeah, ok, Eh, In India, industry only used 3%, thus, eh, thus becoming the smallest water user. And domestic only, only, a kind of, this is a kind of redundant. Use the same phrase again and again. In India industry only used 3%, thus becoming the smallest water user and domestic just 5%. China, China dedicated 22% of this water to industry while only 90% in domestic. New Zealand, New Zealand, eh, actually, I think I should talk about, more about the water use in New Zealand. Because the amount of water use in agriculture and domestic were all the same. I think, I can add information, ah, I can add information here to New Zealand. That is water use in domestic is ok. While in New Zealand it was domestic use that accounted for the biggest water use, making 46%...making 46%. Eh, I can say, It is noted that the water use in agriculture was 44% and close to industry in New Zealand. It is good. New Zealand... only... I am continuing writing about the smallest water user. New Zealand only use 10% of water in its industry. Ok, last and least turn to Canada. Canada, Canada gave a very small proportion of its water to agriculture. As low as 8% and only 12% for domestic use. Ok, Ok. So I am almost done. I still have four minutes. I can just give a conclusion. Eh, in conclusion, the proportions of different water uses in the world vary from country to country. It appears that in developing countries, more water was used in agriculture while in developed countries, more in industry in the year 2000. Ok, I think, that is... I think I add one more sentence, as I still have one more minute. In conclusion, the proportions of different water uses in the world vary from country to country. Although, agriculture account, accounted for a significant majority in world water use. it appears that in developing countries, more water was used in agriculture while in developed countries, more in industry in 2000. Ok.

## APPENDIX 5: THE EIGHT AWT1 TASKS

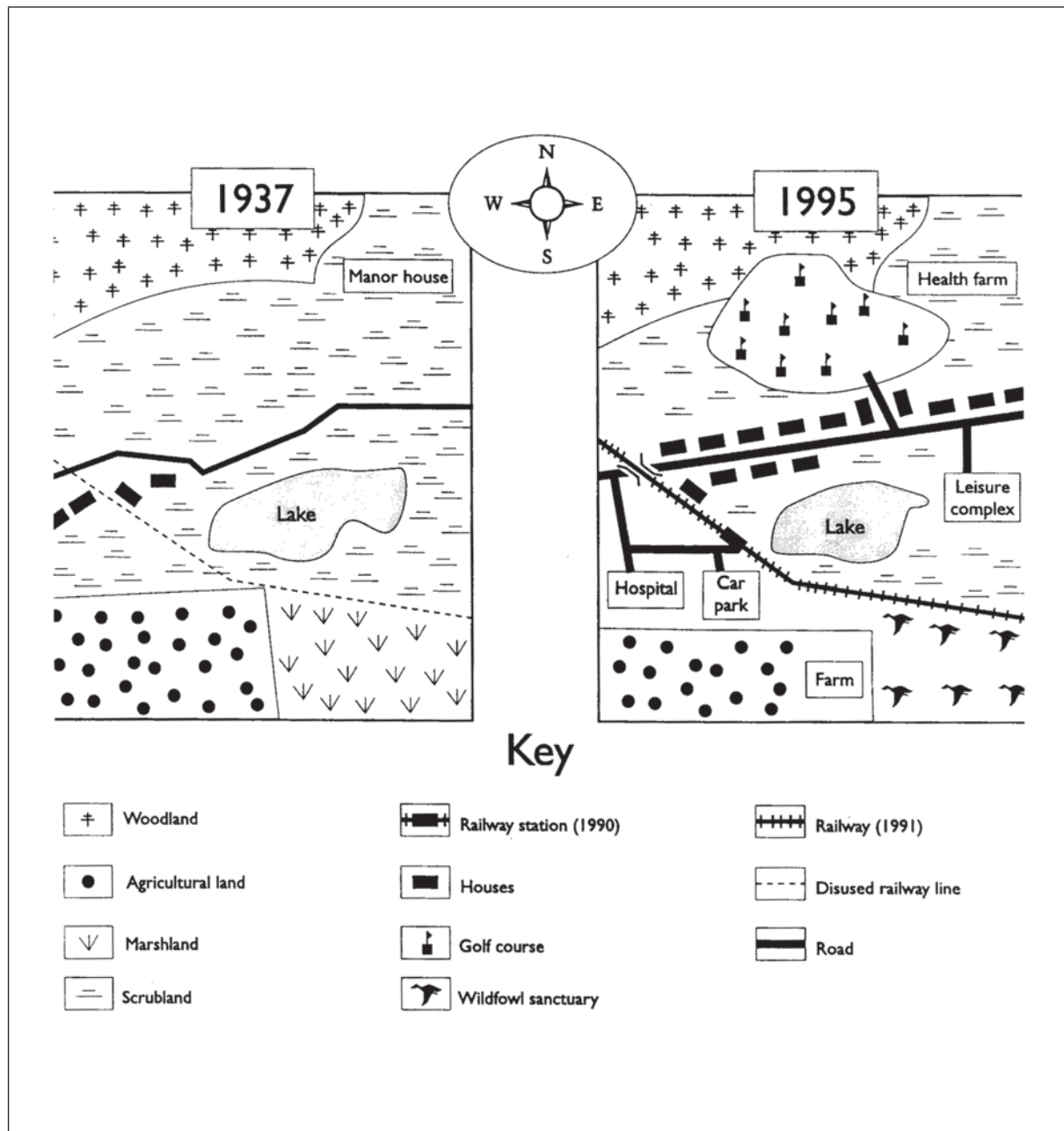
### Set A (four tasks)

You should spend about 20 minutes on this task.

The following graph shows the layout of an area in 1937 and 1995.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.



You should spend about 20 minutes on this task.

The following table shows the latest statistics of the top 15 countries that the US imports crude oil per day on average.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.

<b>Crude Oil Imports (Top 15 Countries), Thousands Barrels per day on average</b>					
Country	Jul-08	Jun-08	YTD 2008	Jul-07	YTD 2007
Canada	1,960	1,883	1,899	1,818	1,872
Saudi Arabia	1,661	1,479	1,543	1,434	1,411
Mexico	1,200	1,124	1,194	1,469	1,457
Venezuela	1,187	1,085	1,038	1,167	1,117
Nigeria	741	946	993	890	1,003
Iraq	696	693	677	460	473
Angola	640	636	517	392	542
Brazil	241	280	224	147	156
Algeria	232	269	306	537	500
Ecuador	226	178	197	159	189
Russia	202	228	123	99	130
Colombia	178	177	182	207	122
Azerbaijan	134	53	57	68	49
Kuwait	122	179	205	197	194
Chad	108	107	102	61	68

**Source: US Energy Information Administration**

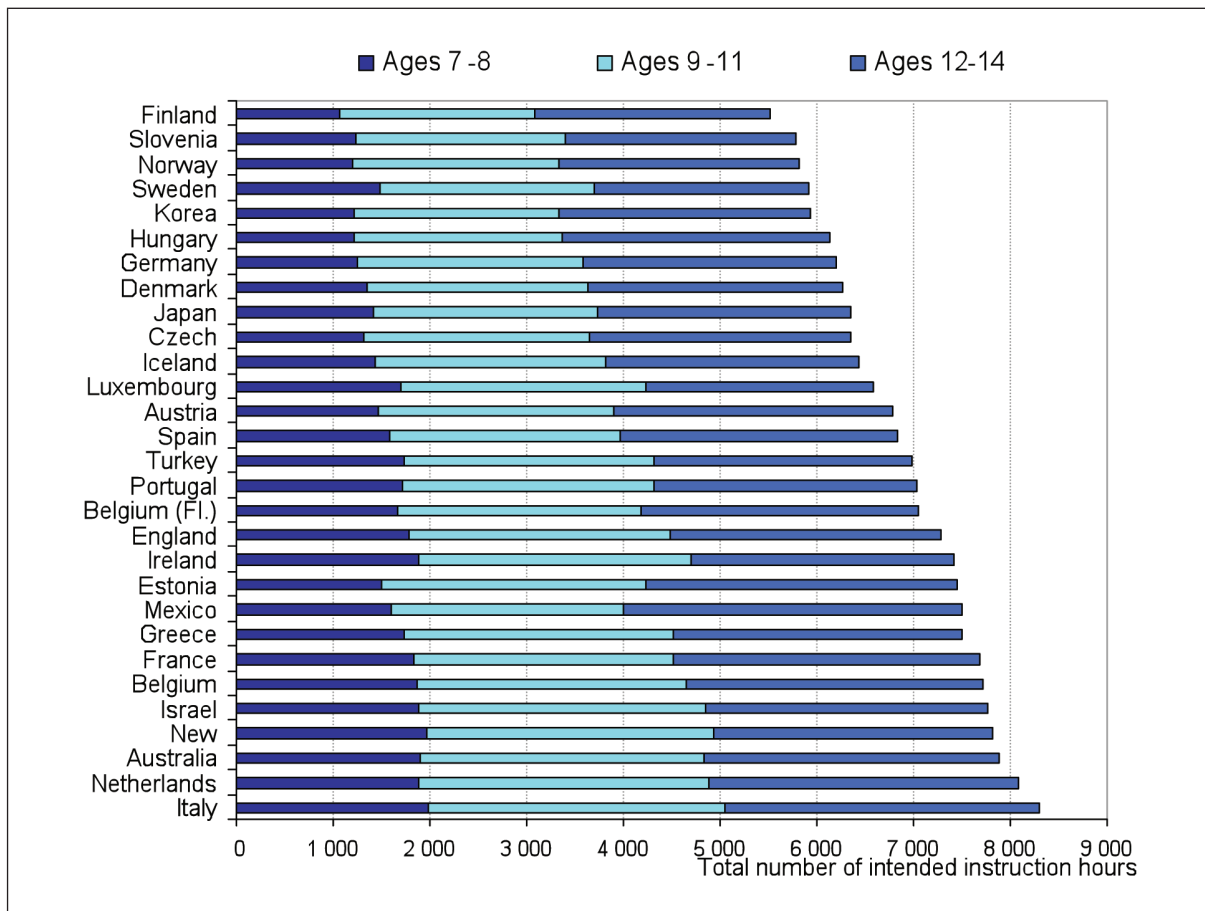
Notes:  
YTD=Year to date.

You should spend about 20 minutes on this task.

The chart below shows total number of intended instruction hours in public institutions between the ages of 7 and 14 (2005). Countries are ranked in ascending order of total number of intended instruction hours.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.



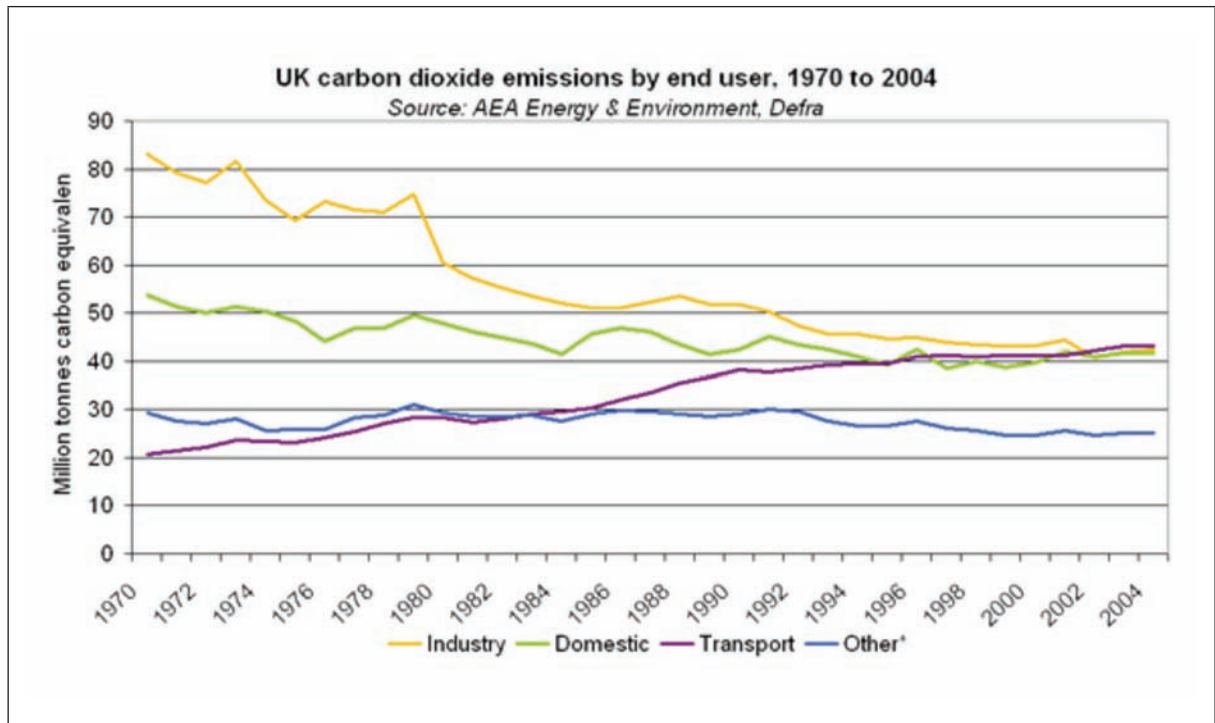
Source: Organisation for Economic Co-operation and Development

You should spend about 20 minutes on this task.

The graph shows the UK CO2 emissions by end user from 1970 to 2004.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.



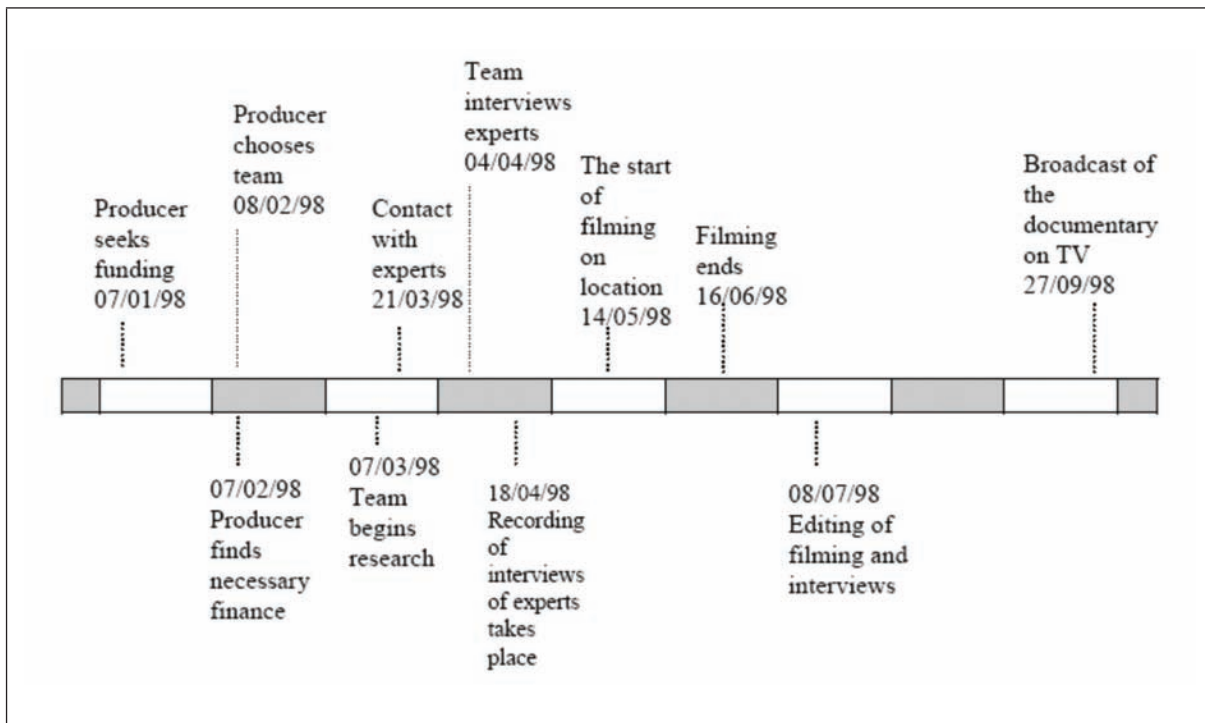
**Set B (four tasks)**

You should spend about 20 minutes on this task.

The following diagram shows the sequence of the events that led to the broadcast of a documentary by a TV programme.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.

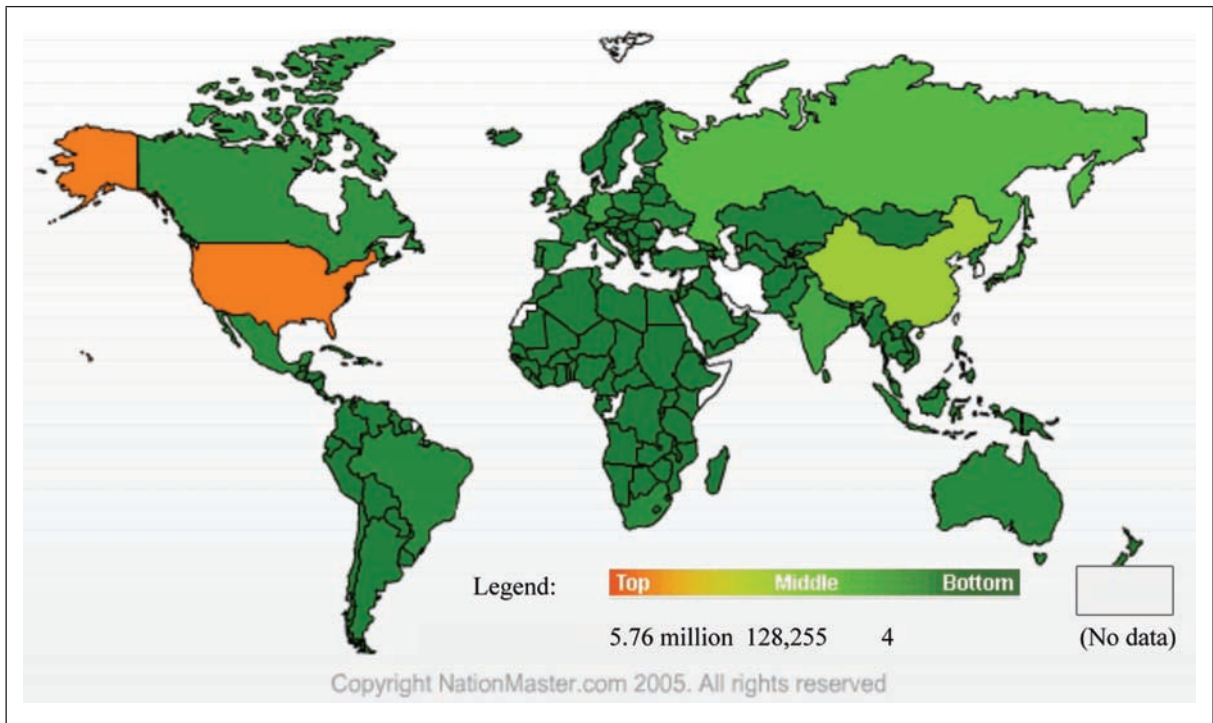


You should spend about 20 minutes on this task.

The following map and table show the amount of CO2 emission by the top 8 countries.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.



Rank	Countries	Amount (top to bottom) Units: thousand metric tonnes of carbon dioxide
#1	United States	5,762,050
#2	China	3,473,600
#3	Russia	1,540,360
#4	Japan	1,224,740
#5	India	1,007,980
#6	Germany	837,425
#7	United Kingdom	558,225
#8	Canada	521,404

**CO2 emissions**

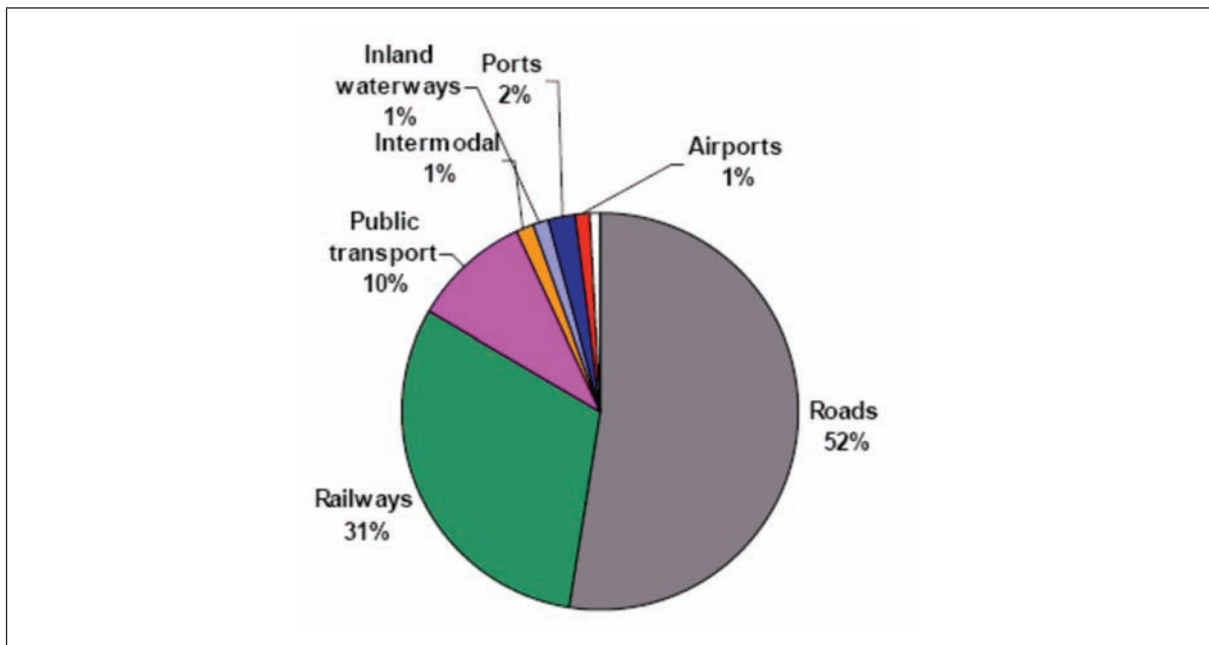
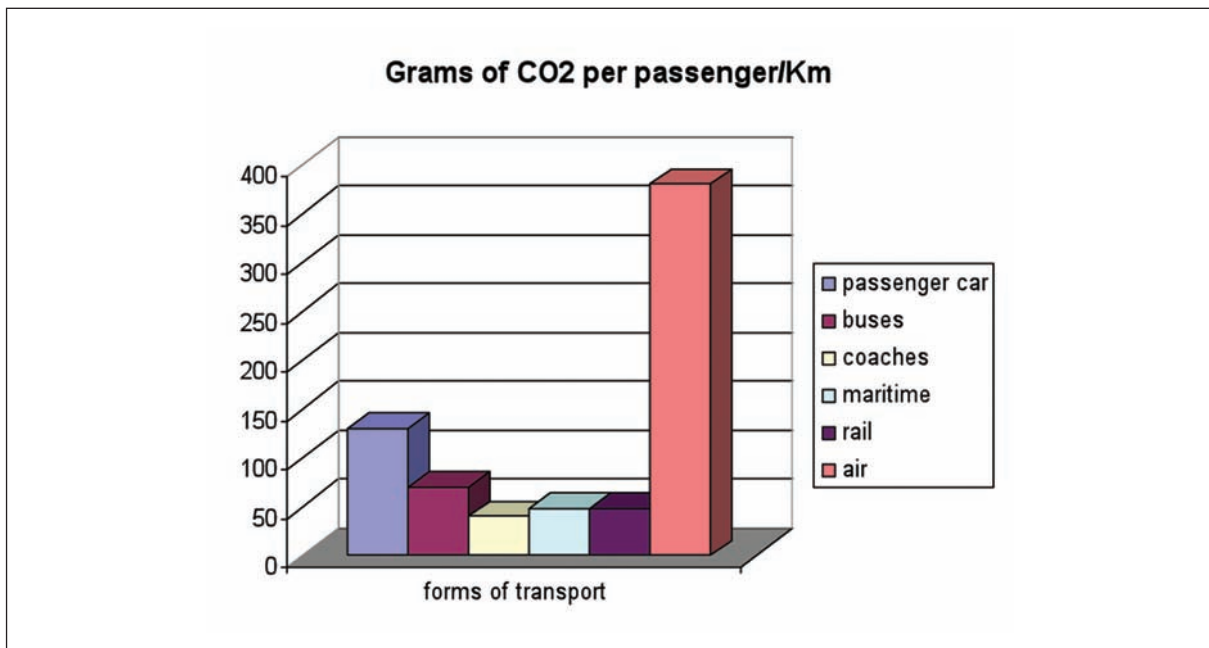
**Source: World Resources Institute (2003)**

You should spend about 20 minutes on this task.

The column chart below shows CO2 emissions for different forms of transport in the European Union. The pie chart shows the percentage of European Union funds being spent on different forms of transport.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.



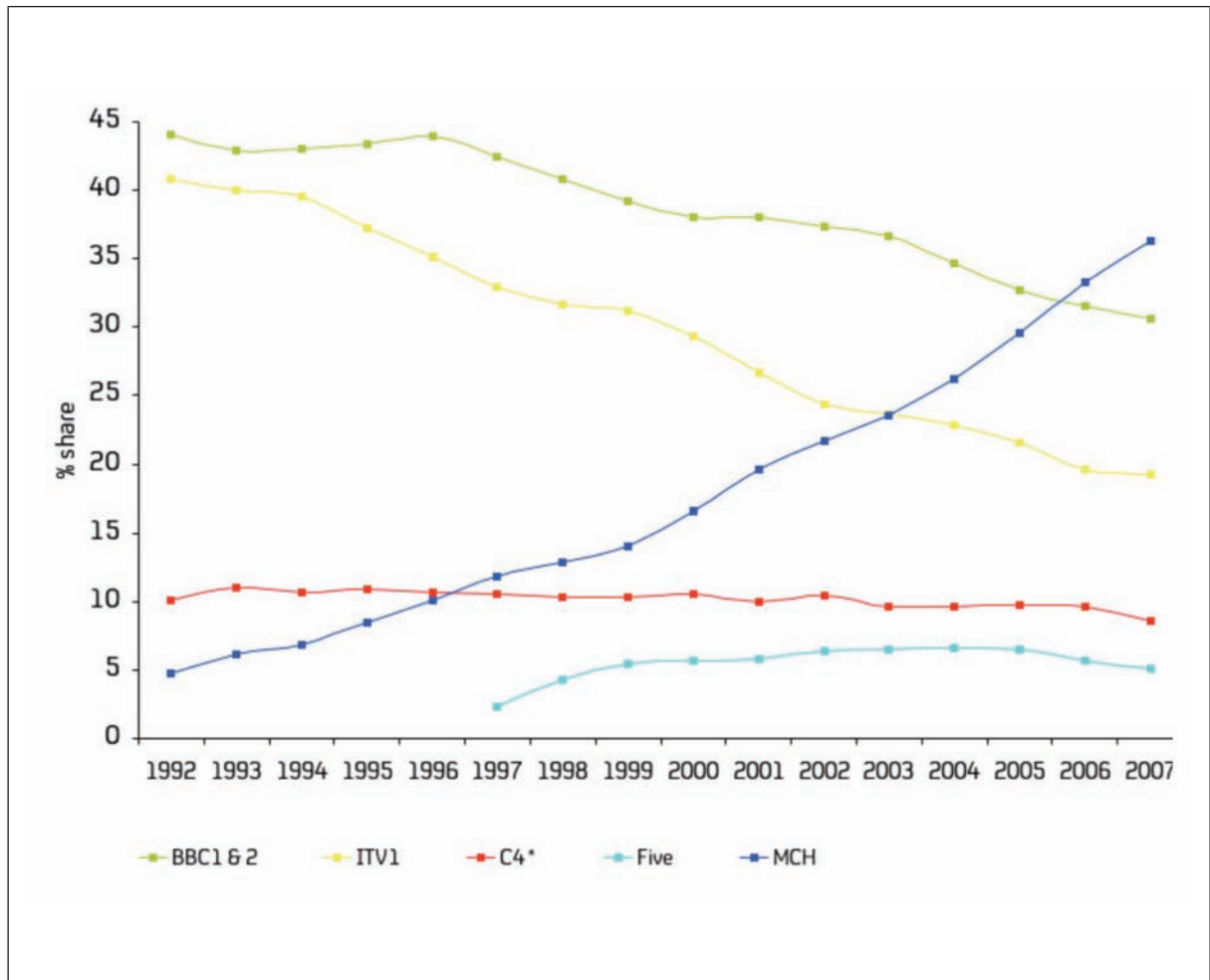


You should spend about 20 minutes on this task.

The following chart shows the individuals viewing share of the major TV channels in the UK.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.



Source: BARB/TNS Ratings Analyser and InfosysTV, Network Homes

## APPENDIX 6: STUDENT EVALUATION OF THE COG-PRO AWT1 TRAINING

This questionnaire is an opportunity for you to record your assessment of our teaching and your learning experience on the IELTS AWT1 training provided by the Cog-Pro project funded by British Council. Please return it to the tutor. Thank you.

Please rate, by ticking the appropriate box on the right and use the space below each question to provide more information in CHINESE and/or ENGLISH.

Excellent / Really good	→	Disappointing								
[1]	→	[2]	→	[3]	→	[4]	→	[5]	→	[6]

### 1. THE TRAINING OVERALL

[1] [2] [3] [4] [5] [6]

Overall, the training was useful for my preparation for IELTS test.

*Comment:*

### 2. LEARNING SUPPORT OVERALL

[1] [2] [3] [4] [5] [6]

Overall, the learning support was helpful for my preparation for IELTS test.

*Comment:*

#### 2.1 LEARNING SUPPORT: handouts

[1] [2] [3] [4] [5] [6]

The handouts provided were useful for my preparation for IELTS test.

*Comment:*

#### 2.2 LEARNING SUPPORT: content

[1] [2] [3] [4] [5] [6]

The content of the training covered was useful and relevant.

*Comment:*

e.g., Which areas are MOST useful?

Which areas are LEAST useful?

What areas do you wish the training should have covered?



## APPENDIX 7: INTERVIEW QUESTIONS

- 1 Briefing the purpose of the interview: to better understand your thinking process when doing IELTS AWT1 tasks.
- 2 Asking the students to talk about their experience of doing the AWT1 tasks, in particular, what is their general impression of the tasks, which task(s) do they find more challenging and why?
- 3 In what ways, do you think your AWT1 writing process may be affected by different graphs/prompts? Did you work differently for different graphs?
- 4 In what ways, do you think your AWT1 writing process may be affected by your familiarity and comprehension of graphs?
- 5 In what ways, do you think your AWT1 writing process may be affected by your writing ability?
- 6 In what ways, do you think your AWT1 writing process was changed due to the group training we had the other day? Did you do the tasks differently before and after the group training? What do you want/think the test preparation/training for AWT1 tasks should look like?
- 7 Any other comments

### Notes:

- All participants to be interviewed.
- The recorded think-aloud protocols may be revisited if necessary at the interviews.
- The interviews are to be recorded.
- An interview will last around 45-60 minutes.

## APPENDIX 8: FREQUENCY OF THE RESPONSES TO THE GRAPH FAMILIARITY QUESTIONS (12-46)

Question	Statement	Frequency						mean	Std. deviation
		1	2	3	4	5	6		
12	I use special computer software to produce graphs.		4	4	5	3	8	4.29	1.52
13	As part of my academic study, I need to produce graphs.		1	2	4	6	11	5.00	1.18
14	As part of my academic study, I need to interpret graphs.	1		4	4	9	6	4.58	1.28
15	I read graphs in the popular press (e.g. magazines, newspapers).	1	5	2	9	2	5	3.88	1.51
16	When I read a graph, I try to identify the main trend or the overall pattern that the graph is trying to convey.		1	2	6	7	8	4.79	1.14
17	When I read a graph, I try to think about the possible underlying reasons for the main trend or the overall pattern of the data presented.		2	2	8	10	2	4.33	1.05
18	When I read a graph, I do not study it in detail (data recoded)		2	1	5	13	3	4.58	1.06
19	When I encounter a graph in a text in popular press (e.g. magazines, newspapers), I tend to ignore/skip it. (data recoded)	3		2	6	9	4	4.25	1.51
20	When I encounter a graph in an academic paper, I tend to ignore/skip it. (data recoded)		1		3	7	13	5.29	.999
21	I am familiar with reading bar graphs		1	4	5	7	7	4.63	1.21
22	I am familiar with reading line graphs		1	1	7	4	11	4.96	1.16
23	I am familiar with reading pie charts			1	10	4	9	4.88	.99
24	I am familiar with reading diagrams representing a process		1	5	6	8	4	4.38	1.14
25	I am familiar with reading tables with numerical data		1	1	10	7	5	4.58	1.02
26	I notice errors or misinterpretations in graphs presented in the popular press		9	3	7	3	2	3.42	1.35
27	I notice errors or misinterpretations in graphs presented in academic papers in my field	2	4	5	3	9	1	3.67	1.47
28	I recognize the different components of a graph (e.g. X and Y axes, legends, colours)				7	9	8	5.04	.81

Question	Statement	Frequency						mean	Std. deviation
		1	2	3	4	5	6		
29	I recognize how the different components of a graph (e.g. X and Y axes, legends, colours) are combined to represent the data		1	3	5	9	6	4.67	1.13
30	I understand the relationships between a graph and the numerical data it represents			1	6	10	7	4.96	.86
31	I can identify the relationships or the patterns displayed in one graph			3	6	10	5	4.71	.96
32	I can identify the relationships or the patterns displayed in a few graphs about one similar theme	1		4	6	11	2	4.33	1.13
33	I can tell when one type of graph is a better representation of the data than another		2	3	3	9	7	4.67	1.27
34	I can identify a poorly constructed graph	1	3	7	6	6	1	3.67	1.24
35	I can revise and improve a poorly constructed graph		2	5	10	6	1	3.96	.999
36	I can describe the general trend or overall pattern of a graph in words		2	3	7	9	3	4.33	1.13
37	I can use a graph to describe/convey the general trend or overall pattern of numerical data			3	6	11	4	4.67	.92
38	I find graphs useful to vividly represent the numerical data			4	2	7	11	5.04	1.12
39	I find graphs helpful for me to remember the key information in the numerical data		2		5	5	12	5.04	1.23
40	Graphs are a waste of space in a text (data recoded)					7	17	5.71	.46
41	I am concerned that I can not fully demonstrate my writing ability in IELTS Academic Writing Task One because I am not good at interpreting graphs		6	5	5	5	3	3.75	1.39
42	I may do better in IELTS Academic Writing Task One using familiar graphs than unfamiliar ones			1	8	9	6	4.83	.87
43	I would prefer one type of graph to be used in IELTS Academic Writing Task One	1	3	1	11	5	3	4.04	1.30
44	Special training on how to interpret graphs would be helpful for me to get a higher score in IELTS Academic Writing Task One			2	1	13	8	5.13	.85
45	Overall, on a scale of 1-6, how would you rate your own experience in using graphs?		4	4	5	11		3.96	1.16
46	Overall, on a scale of 1-6, how would you rate your own ability in interpreting graphs?		1	7	11	5		3.83	.82

N=24.

**APPENDIX 9: THE MOST FREQUENTLY USED WORDS IN EACH TASK**

<i>Task</i>	<i>No. of tokens</i>	<i>No. of types</i>	<i>Average word length</i>	<i>Frequently used words specific to the task</i>	<i>Frequency</i>	<i>%</i>
A: Instruction	3263	402	4.46	Age/ages/aged Between/betweens chart children countries/country difference/differences/different Finland Group/groups Hours Institutions/institutions/institution Instruct/instruction/instructions/ instructing Intend/intended/intending Italy More Number Public Show/shows/shown/showed Spend/spent Total We	121 38 29 34 73 49 28 15 116 33 83 73 38 37 63 36 16 29 37 22	3.71 1.16 0.89 1.04 2.24 1.51 0.86 0.46 3.56 1.01 2.54 2.23 1.16 1.13 1.93 1.10 0.49 0.89 1.13 0.67
A: Layout	3510	493	4.23	Agriculture/agricultural Area Becomes/become/becoming Build/builed/building/built Change/changed/changes Complex Course Farm Hospital Lake Land/lands Layout/layouts Leisure Marshland More Park People Picture/pictures Railway Road Scrubland Smaller/small Wildfowl Woodland Year/years	16 64 13 38 36 20 24 26 22 27 33 16 20 20 49 19 21 17 30 42 35 17 17 17 17	0.46 1.82 0.37 1.08 1.03 0.57 0.68 0.74 0.63 0.77 0.94 0.46 0.57 0.57 1.40 0.54 0.60 0.48 0.85 1.20 1.00 0.48 0.48 0.48 0.48

<b>Task</b>	<b>No. of tokens</b>	<b>No. of types</b>	<b>Average word length</b>	<b>Frequently used words specific to the task</b>	<b>Frequency</b>	<b>%</b>
A: Oil	3178	406	4.32	Algeria Average Azerbaijan Barrel/barrels Canada Countries/country Crude Day Decrease/decreases/decreased Diffent/different/difference/differently Export/exportation/exports/exporter/ exported Import/importation/imports/imported Increase/increases/increased/ increasing July More Per Saudi Table Thousand/thousands Top US Year/years YTD	14 43 14 32 31 93 72 52 12 19 24 77 21 29 25 51 17 27 27 27 62 22 14	0.44 1.35 0.44 1.00 0.98 2.93 2.27 1.64 0.38 0.59 0.75 2.42 0.65 0.91 0.79 1.60 0.53 0.85 0.85 0.85 1.95 0.69 0.44
A: UKCO2	3461	506	4.42	By Carbon Change/changed/changes Decrease/decreases/decreased/ decreasing Different/difference Domestic During Emissied/emission/emissions/emitted End Fluctuation/fluctuations/fluctuate Graph/graphs Increase/increasing/increased/ increases Million More Other/others Same Show/shows/showed/shown Transport/transportation/transport UK Use/user/users/using Year/Years	58 15 12 31 13 50 17 123 31 10 22 22 33 23 39 17 18 58 36 45 47	1.68 0.43 0.35 0.90 0.38 1.44 0.49 3.56 0.90 0.30 0.64 0.64 0.95 0.66 1.19 0.49 0.53 1.68 1.04 1.30 1.36



<i>Task</i>	<i>No. of tokens</i>	<i>No. of types</i>	<i>Average word length</i>	<i>Frequently used words specific to the task</i>	<i>Frequency</i>	<i>%</i>
B: Broadcast	3046	385	4.37	After	28	0.92
				Began/begin/beginning/begins	32	1.05
				Broadcast/broadcasted/broadcasting	47	1.54
				Choose/chose/choosing/chooses/ chosen	20	0.66
				Contacts/contact/contacting	19	0.62
				Day/days	12	0.40
				Diagram	18	0.59
				Documentary	52	1.71
				Edit/editing/edits	23	0.76
				End/ends/ended	13	0.42
				Event/events	28	0.92
				Experts/expert	43	1.41
				Film/filming	53	1.74
				Finance	16	0.53
				First/firstly	14	0.46
				Funding/fundings	19	0.62
				Interview/interviewing/interviews/ interviewed	52	1.84
				Led	15	0.49
				Location	16	0.53
				Make/making	12	0.39
				Months/month	64	2.10
				Necessary	16	0.53
				Produce/producer/production/ producing	37	1.20
				Program/programmes/programme	33	1.09
				Record/recording/recorded	16	0.52
				Research	17	0.56
				Seek/seeked/seek/seeking	19	0.63
				Sequence	18	0.59
				Show/showed/shows/shown	13	0.42
				Start/started/starts	23	0.76
				Step/steps	16	0.53
				Take/takes	15	0.49
				Team	43	1.41
				TV	44	1.44

<i>Task</i>	<i>No. of tokens</i>	<i>No. of types</i>	<i>Average word length</i>	<i>Frequently used words specific to the task</i>	<i>Frequency</i>	<i>%</i>
B: EU Fund	3511	374	4.40	Air	29	0.83
				Airport/airports	14	0.39
				Buses	21	0.60
				Cars/car	24	0.69
				Chart/charts	77	2.19
				Coach	28	0.80
				Column	23	0.66
				Different	45	1.28
				Emit/emission/emitted/emissions	58	1.65
				EU/Europe/European	59	1.68
				Forms	63	1.79
				Fund/funds	71	2.02
				Gram/grams	52	1.48
				Inland	15	0.43
				Intermodal	14	0.40
				Km	41	1.17
				Large/larger/largest	22	0.63
				Maritime	20	0.57
				More	18	0.51
				One	20	0.57
				Only	17	0.48
				Other/others	24	0.69
				Passenger/passengers	71	2.02
				Per	49	1.40
				Percent/percentage/percentages	34	0.97
				Pie	31	0.88
				Public	17	0.48
				Rail/railway/railways	40	1.13
				Road/roads	22	0.63
				Second	23	0.66
				Show/showing/shows/showed	26	0.74
				Spend/spends/spent	52	1.48
				Take/taken/takes	27	0.77
				Transport/transport	100	2.84
				Union	27	0.77
				Waterway/waterways	15	0.43

<i>Task</i>	<i>No. of tokens</i>	<i>No. of types</i>	<i>Average word length</i>	<i>Frequently used words specific to the task</i>	<i>Frequency</i>	<i>%</i>
B: Map	3132	409	4.31	Amount Canada Carbon China Color/colour/colours/coloured Countries Differ/different/difference/differences Dioxide Emission/emissions/emit German/Germany Green India Japan Large/larger/largest Map Metric Million More Russia States Table Thousand Tonnes Top/tops/toppest United World/worlds	56 22 18 40 25 106 19 18 97 16 14 20 19 20 45 44 32 16 28 22 32 36 44 42 31 30	1.79 0.70 0.57 1.28 0.80 3.38 0.61 0.57 3.09 0.51 0.45 0.64 0.61 0.64 1.44 1.40 1.02 0.51 0.89 0.70 1.02 1.15 1.40 1.34 0.99 0.96
B: Viewing	3140	434	4.12	BBC Change/changes/changing Channel/channels Chart Decrease/decreased Drop/drops/dropped Five Increase/increased/increasing Individual/individuals Major MCH (multi-channel) More Only Remain/remained/remains Rise/rised/rises/rising Share Shown/showed/shows Some Stable/stabilized Trend/trends TV UK View/viewing Year/Years	33 12 85 24 18 11 34 11 43 26 38 16 13 13 12 106 19 12 12 14 38 20 54 61	1.05 0.39 2.71 0.76 0.57 0.35 1.08 0.35 1.37 0.83 1.21 0.51 0.41 0.41 0.38 3.38 0.60 0.38 0.38 0.45 1.21 0.64 1.72 1.94

Note: Words spelt wrong are also included in this analysis

## APPENDIX 10: PERSONAL INTERPRETATIONS AND COMMENTS IN TASK B AND TASK A WRITINGS

<b>Task</b>	<b>Personal interpretations and comments</b>	<b>Participant</b>
B: Broadcast	This report suggested that the prepared work and the support work were the most important job in the broadcast.	C
	The hole [whole] task of finishing the documentary was difficult and need a lot of money and labours, and also with complex process.	
	It would be benefitful for them to do some research first so they finished it in early March.	H
	Interestingly, we find that most of the process of the broadcast had been done comparatively long before it was broadcasted on TV.	K
	In conclusion, the broadcast of a documentary by a TV programme takes a lot of time and work, and it also needs financial support to make all of these happen.	L
	Finally, the greatest moment came. The documentary was broadcasted on TV on 27/09/98. The diagram gives us a complete and clear picture about the sequence of the events that led to the broadcast of a documentary by a TV programme.	Q
	Doing an broadcast of the documentary on TV is an hard work. We need a plan and do just as the plan.	X
B: EU Fund		
	From the two charts, we can conclude that the transport receiving the largest proportion of EU funds is not the one that emits the most CO <sub>2</sub> for different forms of transport.	G
	So it may be possible to speculate that we would promote the development of buses, coaches, rail and so on.	J
	When comparing the bar chart and pie chart, we may find that, generally, the form make more grams of CO <sub>2</sub> , the less percent of fund it would receive.	K
	The conclusion of my report is: The amount of CO <sub>2</sub> in different forms of transport are quite different, and the pecentage [percentage] of funds be used in different kinds of transport are also very different from each other, and the coast of money in one kind of transport may not be linked with its CO <sub>2</sub> emission.	N
	(think-aloud): I don't actually know the meaning about the forms of the transport, I can't find the corresponding parts in the pie chat for each form of transport. ... Then I make some conclusion. The emission cause the funds increase or the funds down cause the emission increase. What is the relationship between them? I can't decide it up. And I don't know the efficiency about the funds. I can only get the trends that if pay more money on the transport, the emission will go down. Actually, the deep relationship between them, I can't say much clearly.	T
	Maybe the government need to expend more funds on the air transport to reduce the CO <sub>2</sub> emissions of air to control the pollution.	X
B: Viewing	we can draw a conclusion that, the five major TV channels got different individuals viewing share and had different dropping or growing trends year by year.	B
B: Map	To sum up, the amount of CO <sub>2</sub> emission is different among countries all over the world. And, the difference is a significance.	B
	In a word, North America and Asia were the largest CO <sub>2</sub> emission countries". (Note: summary based on previous knowledge)	C
	The square of Japan is much smaller than other countries in the table, but it has the rank of 4 in CO <sub>2</sub> emission, even more than India's 1,007,980. (involving prior knowledge of the size of Japan)	D
	There are two countries in North America, three countries in Asia and three countries in Europe. The Middle East countries, like Iraq, just have no data, which means that these countries have rarely no CO <sub>2</sub> emission" (involving prior knowledge of continents. It misinterpreted the Iraq information)	E

<b>Task</b>	<b>Personal interpretations and comments</b>	<b>Participant</b>
	Unfortunately, China is the second country after America, over 3.47 million thousand metric tonnes of CO2 emissions. In the map the area of China is described in light green. .... Hopefully the color of China would turn into dark green in the future.	G
	It is astonishing that Japan ranks 4th, since its total area of the country is not very large, compared to the countries rank at the tops.	H
	... and all of the top 8 countries of CO2 emission are developed countries except China and India (involving knowledge about developed and developing countries)	J
	The most contribution of CO2 output countries are located in north America and Asia. (involving prior knowledge about continent)	N
A: Instructions	In the chart, we can see that the countries which have fewer instruction hours like in North Europe, such as Finland, Norway, Sweden and so on". (involving prior knowledge about geography)	C
	From the chart, the top three countries Finland, Norway and Sweden are all from the North Europe. These countries all do very well in instructing the children in public institutions. So the children can meet less problem when they are in these public institutions with no adults. (involving prior knowledge, but wrong interpretation re: with no adults)	E
	So I think if we want the children to do more instructions, we should encourage them to do it when they are at the earlier ages from 7-8, since maybe earlier will be better. (personal judgement and evaluation)	H
	So the hours shows a steady increase in these 3 age groups, which shows a quick development in the ability of acceptable learning time of children.	L
A: Oil	Last, I want talk about the Iraq, because of the Iraq War. There is a big growth from 2007 to 2008. I think, after the Iraq War, US imported more oil from Iraq.	B
	First, I see there are both developed and developing countries above.	D
	Canada is US's neibough, so the US imports the largest crude oil from Canada... The US imports more crude oil from American countries than those from other continents, especially those from Middle East, or Africa. There are not many European countries which have crude oil exports to the US, China, neither. Maybe the transport cost is very high, if the US wants to import crude oil from these countries or because these countries have no crude oil production.	E
	However, the earlier the data is the bigger the disturbances are, which means economies in differences changes differently, some developed fast and so they need more oil and some developed slow, so they don't need as much oil as others did. (wrong interpretation of the data: considering countries importing oil)	I
	There are some countries like Colombia, Azerbaijan, Kuwait, Chad import crule oils even less than 200, I think there're some reasons. One is that they are not very developed, as those countrys like Canada, and the other, they also can not provide enough crude oil with their industry. They're not rich, they don't have enough money to import the crude oil. I just skip the Venezula, I'm not familiar with it. And its statits is similar with Mexio, so I conclude that, there are three or four groups. One is that they import larger than 1000, one is that they import less than 200, the other is that they import between 200 and 100. Different reasons and facets in different countries cause the differences. For example, the Canada and Saudi Arabia import large every year, because they are well-developed country, and they have modern industry, which need more crude oil. And another example is Russia, it is also the developed country, but it has its own crude oil field and it doesn't need to import as much as Canada. And the third example is, I find that country, the Chad, the small country with poor industry, they don't have to import many crude oil, its industry can't consume as many oil as Canada. (Wrong interpretation)	T (from think-aloud protocol)

<b>Task</b>	<b>Personal interpretations and comments</b>	<b>Participant</b>
A: UK CO2	Furthmore, in my point of view, the measure to control the amount of CO2 emissions is successful these years, especially to control the CO2 emissions by industry. The measure can be done in the future. However the measure to control the transport CO2 emissions is not so successful, I think better measure may be taken to solve the problem. And I think the problem may be because of the number of cars owned by personal user becomes larger and larger year by year. So we may use some way to enough people go out to work or to have great time by bus, by train or other public transport. Another way to solve the problem is to use new energy which produce less emmison than the fuel we use nowadays. (a large proportion of her writing is about finding out the underlying reasons)	B
	The reasons of the graphs reveals that more and more people began to own their private cars and transport make big contribution in CO2 emissions in the UK. The industry and domestic graphs decrease because we use a lot of energy like electricity or nuclear power than 1970s. (a large proportion of her writing is about finding out the underlying reasons)	C
	The industry has been always developing, so I think the UK government must have took some methods and give a policy to cut down CO2 emission in industry. And the CO2 emissions by transport grows slowly during the years. It's because that the UK is using more transport, maybe its population grows or it has more travellers from the world. So the Government has to do well in his job to offer enough transport. I see that CO2 emission by domestic changes a little but it seems to be periodic, that is it decrease in a year or a period, and in the next years, it increases. So is there relationship between domestic and economic because the economic development is also periodic. The industry produces less and less CO2 while the transport produces more and more, I think the UK government was trying to change its economic..., to develop industry as a smaller part and to develop its ... as a bigger part. (almost the whole piece of writing is about explaining the reasons)	E
	In conclusion, CO2 emissions by industry is less and less than before, while we have seen an increasing emissions by transportation in the last 30 years. That's what we should care about.	G
	Interestingly, I think now the transport should be taken consideration carefully. I think it will be the big problem to our living environment. Because in this graph it is the only line climbing high step by step. We can see from the graph, in 1970 the CO2 emission is about 20, but in 2004, the CO2 emission excess the 40 million tonnes, just increased 2 times as before. All in all, I think the transport problem we should cope with earlier before it deteriorating.	R
	Now I think it must be the reason why the government restrain the personal car and set up more rules go the licences. It may be used, I thought; Oh. It's not relation to the graph. I got back to it. Plus them, distinguish the difference among them and what will I do next. Maybe I'll find them a reason . The industry cut down its emission for the process and technology moderate, maybe the transport can cut down it emission at the same way. Like enwide the road, redesign the program of traffic lights, transport emission may go down by those measures.	T (from think aloud protocol)
	There is no end for developments of industry and domestic because the technology is still developing. So I predict that the UK CO2 emissions of industry and domestic will increase in the future when CO2 emission of transport increase to some level.	W
A: Layout	In my point of view, the development of world make great changes in the area of the picture. Firstly, in modern world, there's not enough space in cities for people to do sport, and we also like the woodland and scrubland which is hardly seen in the cities, so a golf course is built. Secondly, just I stated above to the space in the cities is not enough. As a result, more people build their houses here, as well as new hospital, car park, leisure complex and so on. In order ot go the places easily, people broaden the road and build the railway. Because of the use of land, our lake area and agricultural land become smaller. Furthermore, people now take more care of wild animals, so I think that's why the marshland in 1937 is replaced by the wildfowl sanctuary.	B
	Although there are a lot of buildings in 1995, the lake and the farm was becoming smaller and smaller. It's the cost of economic and social development.	C
	The economy level should have a great improve.	D

<b>Task</b>	<b>Personal interpretations and comments</b>	<b>Participant</b>
	In a word, I think that the land has been used to its maximum and more comfortable for people to work and live. Transport, entertainment and medical care are all getting better. The road also reaches the leisure complex, which adds more happiness to the living.	E
	The precious manor house had become a health farm in 1995. ....We can concluded that more people moved to this area in 1995 than before. It's certain that development would cost something.	G
	As the span of the time was about 60 years, and it was normally that there're so many changes in the same area, as the society was improved day by day. Then we can get the conclusion that. As the society improves day by day, there are much more modern buildings and constructions in the area. But as the scrubland and the woodland or the agricultural land are all important to our environment and will be benifutl to the atmosphere which we rely to survive. And according to the changes during the past several years I think the trend of the changes in the future in this area will be benifutl to the citizens reside in this area.	H
	Newly built roads would lead people to the hospital, car park, golf course and the leisure complex. Compared with the old road, the new is more straight, so that the trip would be shortened.	I
	And the land has become more convient; comfortable and beautiful than it was in 1937. I think a lot of people may come to this area for leisure activities. And it will become a famous land for wildlife.	J
	In conclusion, this area has changed a lot during the last sixty years and it is hard for us to reveal its original face.	K
	In conclusion, the traffic system in this area has greatly improved, and also several facilities have built to enrich the life of this area, which all turn this area into a more suitable place for people living and working.	L
	The changes of the layout also show the change of the society. The population increased and then the public organization.	Q
	All in all, from 1937 to 1995, this land become more suitable for people living. But I prefer the land of 1937, since of the more nature. For example, more lake, more farm and so on. .... The hospital, I think, is more advanced than the clinic in 1937, so it do good to our health. ... The people in 1995 became more wealthy than in 1937.	R
	And 1995, there is a new hospital, car park and golf course, seems that there are some better life for the people who live on the land. The houses are along the road and they are 4 times as the numbers in 1937. The lake is shrinked, maybe caused by construction. The main change of the land, may have some reasons. Transport develop the agriculture and leisure consummation, so they may have the golf course and son on. The farmland didn't shrink, or cut down, means they may need more supplement for the new town. May call them the town, because there are some road and hospitals, car park, railway, it have developed to a town. The house are more, and the facilities is constructed to satisfied the need of the people in there: with land redesigned and rebuild, the people live in there have to adapt the change and make it more fitable to live there.	T (from think-aloud protocol)
	Although more buildings have been built, people are concerned more on the environment of this area, the plants in the woodland are well protected, so is the farm. In a word, this area is like a leisure place for tourists to enjoy their time, the great nature in 1995.	U
	We may see from those two picture that the layout of 1995 is more colorful and reasonable. It utilize land greatly. What's more, this layout offer wildfowl some better land, which can improve the balance development of the nature. At the same time, this layout make our life better. We have many and more different land to live.	W
	When the health of the people become more and more important, the hospital become. As the economy developed, more and more people have their own cars, so car park is needed on the north area...	X

## APPENDIX 11: STUDENT EVALUATION OF THE PROJECT AND ITS TRAINING: GRADE AND QUALITATIVE COMMENTS

Note: blank means no additional comments were provided by the student.

ID	Grade	Comments on the training overall
1	5	if there are some good examples to analyses, cause what we see here is written by Chinese, that would be better
2	5	I now have an overview of the requirements of IELTS AWT1 tasks, how to analyze graphs and what I should pay attention to, what I should avoid, e.g. it's necessary to describe what is in graphs, but not to further reasoning
3	4	partly, I haven't take a view of the whole test
4	6	Through this training, I know a lot more about AWT1
5	4	
6	4	I'm not quite sure about the long-term effect yet, but think-aloud is very useful, it helps us to pay attention to the problems in our writing process and find some solutions to these problems.
7	5	To me this training is extremely useful, this kind of training is much better and effective than other intensive preparation courses and self-study
8	5	Now I have systematic understanding of AWT1
9	6	The lecture was very useful
10	5	
11	6	I think it is beneficial to both my writing and speaking ability
12	5	I have learned some special words and idioms for describing the trend or extent
13	5	The training made me aware of the way I think when doing the graph writing, and helped me develop myself. It also gave us some useful tips
14	5	Maybe it's a bit short. If the period can be arranged a little longer, such as 2 weeks, and the training be more scattered, the effects will be much better

ID	Grade	Comments on learning support
1	6	
2	5	I now know a lot more useful words to describe graphs
3	4	partly, I wasn't ready psychologically, it was quite new to me, I need time to adjust myself and make the best use of the resources provided
4	5	It let me know what is AWT1
5	5	
6	4	The tutor listened to our think-aloud and pointed out some problems we had; this is really useful for individuals to notice our problems.
7	5	(same as for Q1: To me this training is extremely useful, this kind of training is much better and effective than other intensive preparation courses and self-study)
8	5	
9	6	The only pity is that there could be more feedback so that we know what our weaknesses are
10	4	
11	6	
12	5	
13	5	It gave us much better understanding of all types of graphs, background knowledge, it is eye opening, broaden my horizon.
14	3	The support can be given more individually



<b>ID</b>	<b>Grade</b>	<b>Comments on Handouts</b>
1	6	It is well organized and there are a lot of useful information
2	6	Excellent content coverage and well delivered lecture: I believe it will be very useful for my preparation for IELTS writings
3	4	Some of the information contained in the training package can be found online; but I will continue using think-aloud to train myself for the test
4	6	I want to be provided much more
5	6	
6	5	It explained in much detail about all aspects of AWT1, provided constructive feedback and suggestions, provided some samples and exercises; it is really helpful for us to understand and grasp AWT1
7	6	The materials are well prepared, with detailed and rich content, very good for test preparation
8	5	
9	6	
10	6	
11	6	
12	5	
13	6	These materials tell us what AWT1 tests, and the constructive and analytical tutorials, which are all very helpful preparation for AWT1 tasks
14	5	Although I've only read part of the handouts, I find it really useful. The tips were helpful to me.

<b>ID</b>	<b>Grade</b>	<b>Most useful content</b>	<b>Least useful content</b>	<b>Wish list of content</b>
1	6	Don'ts and dos based on my action		feedback of the essays we wrote
2	5	Especially the lecture on the third night	We think faster than speak/write. When thinking and write, we have time to re-organize ideas; but at think-aloud & write, we don't have time to re-organize, we say what we think, and sometimes our thinking can be interrupted, maybe it's because of my low English ability	I think it might be better for training if we wrote without think-aloud first, then think-aloud or re-visit our thinking
3	5	think-aloud training, it helps me to find out my weaknesses	I've already known some information contained in the handouts	It would be good to have more comments from the tutor on each individual writing
4	4	How to think aloud and describe a chart/table/diagram...		For specific graph or chart, how to describe and in what sequence we should organize our ideas and thoughts
5	4			
6	4	the third training session	Although think-aloud has great potential, some of the problems detected through think-aloud are the problems I'm already aware of, therefore I don't think it is the most efficient strategy to improve writing	how to analyse all sorts of different graphs, provide exemplars, and explain good reasoning strategies
7	5	interaction during the training	NONE	It seems that we have already had all.
8	4	interaction		the training could be longer

<b>ID</b>	<b>Grade</b>	<b>Most useful content</b>	<b>Least useful content</b>	<b>Wish list of content</b>
9	5	wish the project is longer, with more examples of writings		
10	6			
11	5	Think-aloud training		analysis of every task
12	5	Magic words	the history of IELTS	the samples of the given titles
13	6	the dos and don'ts, magic and useful words, think-aloud	NA	more detailed guidance on how to analyze different types of graphs
14	4	It provides me a new way to look at IELTS, especially AWT1. The task is new and up to date	No.	We may want more feedback from the tutor, as we didn't receive individual help from the tutor

<b>ID</b>	<b>Grade</b>	<b>Comments on teaching quality</b>
1	5	
2	5	It isn't necessarily interesting, but really attractive as it helps to improve my writing or to know more about IELTS writing test
3	5	It is interesting and effective, but a little rush; we need more time to digest what is there the helpful information
4	6	Good! Excellent!
5	5	
6	5	YES! The tutor is very patient and attentive; the training was very detailed and complete
7	6	the interaction
8	5	I haven't done similar training, I feel it is interesting.
9	5	
10	5	
11	6	The teacher broadened my horizon of the IELTS AWT1 and I felt this training unique and useful
12	4	
13	5	helps to know different types of graphs, background knowledge, very interesting. The tutor is passionate about what he teaches and humorous, which encouraged us to do the best in the project
14	4	Due to the long hours, I really felt a little tired when I needed to record what I'm thinking and writing

<b>ID</b>	<b>Grade</b>	<b>Comments on Learning from the training</b>
1	6	Through the handout and the training class I can describe the graphs much better than before
2	4	The training is too short, but it is specifically for AWT1, helpful to know how to do AWT1. I also got a chance to practise my writing
3	3	At the beginning, I haven't adjusted myself well; but later it was ok. I was there on time every day and finished all the tasks; gradually I am in the tasks now.
4	3	A bit improved, because the training time is so short
5	4	
6	4	it is a long process to learn English well, although in such a short time we know well AWT1 now ,it still needs more time and practise to integrate what we've learned; however, this is a very good direction, absolutely essential good direction and beginning
7	5	I feel I improved
8	5	
9	5	
10	3	
11	5	
12	5	
13	5	helped me to understand how I think when doing the tasks, my weaknesses and strength, as a result improving how I do AWT1 tasks, and my way of thinking
14	4	Before taking a test, I need to figure out why the examiner designed the question like this. By figuring it out, I can get a better score more easily.









# IELTS Research Reports

Volume 11

## An impact study into the use of IELTS by professional associations and registration entities: The United Kingdom, Ireland and Canada

Examines the history and rationale for selection of IELTS as a language benchmarking system for professional associations, and explores the main competitors to IELTS in global testing for professionals, in Canada, the UK and Ireland.

*Glenys Merrifield, GBM Associates*

## Learning to play the 'classroom tennis' well: IELTS and international students in teacher education

Addresses the question of an appropriate IELTS score for graduate entry teacher education courses by investigating the extent to which current IELTS scores into graduate entry teacher education courses are considered adequate: by the lecturers of such students, by prospective teacher registration authorities and by the students themselves.

*Wayne Sawyer and Michael Singh, University of Western Sydney*

## A multiple case study of the relationship between the indicators of students' English language competence on entry and students' academic progress at an international postgraduate university

Investigates selection practices and decision making rationales of admissions personnel in an international, postgraduate UK setting and the consequences for borderline non-native English speaking students' academic progress.

*Gaynor Lloyd-Jones, Charles Neame and Simon Medaney, Cranfield University*

## Construct validity in the IELTS Academic Reading test: a comparison of reading requirements in IELTS test items and in university study

Investigates suitability of items on the IELTS Academic Reading test in relation to the reading and general literacy requirements of university study, through a survey of reading tasks in both domains, and interviews with academic staff from a range of disciplines.

*Tim Moore, Swinburne University, Janne Morton, Melbourne University and Steve Price, Swinburne University*

## An investigation of the process of writing IELTS Academic Reading test items

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.

*Anthony Green and Roger Hawkey, University of Bedfordshire*

## The cognitive processes of taking IELTS Academic Writing Task 1

Investigates cognitive processes of candidates taking IELTS Academic Writing Task One with different graphic prompts at two time points: before AWT1 task training, and after.

*Guoxing Yu, Pauline Rea-Dickins and Richard Kiely, University of Bristol*

[www.ielts.org](http://www.ielts.org)

**IELTS™**

 **BRITISH  
COUNCIL**

 **idp**  
IELTS AUSTRALIA

 **UNIVERSITY of CAMBRIDGE**  
ESOL Examinations