Research Notes

Contents

Editorial Notes

Welcome to issue 28 of *Research Notes*, our quarterly publication reporting on matters relating to research, test development and validation within Cambridge ESOL.

In this issue we focus on the theme of testing young learners, an especially timely theme given the launch in January 2007 of our Young Learners English (YLE) tests in their revised format. Juliet Wilson's introductory article outlines the background to the review of the YLE tests conducted over the past 3–4 years; she goes on to consider the modifications and trialling of three different tasks and describes the research which was carried out to update the vocabulary lists.

Among other aspects, the YLE review and revision project addressed the policy on marking candidate responses, specifically the role of spelling within this. Helen Spillett explains the reasons for considering a change in existing policy and she describes the procedures and the outcomes of a consultation exercise and small-scale research study which informed the new policy implemented from 2007. Linked to this, we are pleased to include a guest article from Shelagh Rixon exploring the relationship between Cambridge ESOL's YLE tests and current understanding of children's first steps in reading and writing in English. As Senior Lecturer at the Centre for English Language Teacher Education, University of Warwick, and as someone with considerable experience in the teaching of young learners, Shelagh provided invaluable input to our research project investigating spelling issues in the YLE marking policy.

Our suite of language tests for young learners of English aged 7 to 12 are offered at three levels – Starters, Movers and Flyers, thus providing a 'ladder' to encourage and support progression in language learning steadily up the proficiency continuum. Fiona Barker and Stuart Shaw's article reports on an ongoing and long-term study to locate the three levels of YLE onto a common scale, thereby providing empirical validation for the vertical equating of levels. The following short article discusses the relevance of Can Do statements to provide stakeholders with transparent descriptions of young learner proficiency; the construction of a set of Can Do descriptors for YLE test takers will be a focus of research activity for 2007–8.

In order to provide appropriate support to candidates for the revised YLE tests, their teachers and other stakeholders, a range of support materials have been developed for use from 2007. These resources are listed on page 19 and can be used to learn more about the revised YLE tests.

Issue 28 includes an update on recent research and development activities relating to other Cambridge examinations and also a summary of some recent publications of interest, including the latest volume published in the Studies in Language Testing series focusing on IELTS research. Finally, we include news of recent and future conference events along with the call for proposals for Round 13 (2007/8) of the IELTS Joint-funded Research Program.

Editorial team for Issue 28: Lynda Taylor and Louise Maycock.

Reviewing the Cambridge Young Learners English (YLE) tests

JULIET WILSON ASSESSMENT AND OPERATIONS GROUP

Introduction

This article outlines the background to the recent review of the Cambridge YLE tests, considers the modifications and trialling of three different tasks and describes the research which was carried out to update the vocabulary lists.

The development of the Cambridge YLE tests

The YLE test development project started in 1993 and involved extensive research and consultation. Since at that time relatively little research had been carried out into the assessment of second language learning in children, a review of the available literature focused on work done in three related fields: children's socio-psychological and cognitive development, second language teaching and learning, and second language assessment. In addition to this, a wide range of course books used in primary language classrooms was reviewed in order to identify the main content areas and topics to include in the tests. Another important aspect of the test development project was collaboration with staff at Homerton College, then the teacher training college of Cambridge University. The Cambridge ESOL YLE development team were able to draw on the experience of their colleagues in Homerton of working with primary age children. Versions of the YLE tests were trialled in 1995 and 1996 with over 3000 children in Europe, South America and South East Asia. The results of these trials, including feedback from teachers and statistical analysis of candidates' responses, were used to construct test versions. The YLE tests went live in 1997 and since then have been taken in over 65 countries and have enjoyed a growing candidature.

Reviewing the tests

As part of Cambridge ESOL's ongoing commitment to improving the quality of its tests, all Cambridge ESOL tests are regularly reviewed. In 2003, 10 years after the initial development, we began a review of the YLE tests using the model of test development and revision described in Saville (2001). The project began with a consultation stage in which questionnaires were sent out to test centres, teachers and examiners. Burrow and Wilson (2004) detail this initial consultation stage and the results of the stakeholder questionnaires. Careful analysis of the questionnaires showed a high level of satisfaction with the YLE Tests – although a few tasks were identified where possible changes could usefully be trialled. The trialling process involved trying out the new tasks in a number of centres across the world including China, Portugal, Japan, Argentina, UK, Thailand, Libya and Spain. A range of different L1s was one of the key criteria in choosing the participating centres. Two or three rounds of trialling were carried out, depending on the outcomes. A total of 4000 trial tests were taken overall. Statistical, as well as qualitative feedback from teachers and children, was collected. A retired test version with calibrated Rasch difficulty values was chosen for the trials and the revised tasks were used in place of the original tasks. The unrevised tasks thus acted as anchors.

Starters Listening Part 2

Feedback on Starters Listening Part 2 showed us that some of the pictures did not always make it entirely clear to the candidates what they had to do. The picture has to indicate what kind of response is required without 'giving away' the answer.

For example in Figure 1 opposite, the picture of the cake is the prompt for the question, *How old is Sue?* In the original construct of the test, it was considered of paramount importance to keep the amount of reading and writing to a minimum in the Listening paper which is why the prompts are pictures.

Two variations of this task were trialled. For both variants, instead of individual pictures for each question, a global illustration to set a context was included at the beginning of the task. The first variant used the prompt *Name* or *Number* to indicate what kind of response the child needed to write. The second variant gave a full question as a prompt.

The expectation that having to read and process full questions would raise the difficulty of the task was not borne out. In fact with the full question variant, the task's overall Rasch difficulty dropped a little. Clearly processing questions is a very familiar task in course books. Moreover at this level the set of question types is very limited and the length of questions used is constrained. Improving the clarity of the task meant that in fact it became a little easier. In contrast the *Name* or *Number* task had a much higher Rasch difficulty.

Overall, there was a preference for the full questions version. For example, one teacher in the UK commented:

> I very much prefer the version which gives the candidates the questions. At present the candidates don't really know the significance of what they are expected to write. I think the face validity of the task is much better like this. *(Teacher, UK)*

The revised task for Starters Listening Part 2 is now as shown on the opposite page as Figure 2.

Movers Reading and Writing Part 4

In the original Movers Reading and Writing Part 4 task, children completed a text by writing words in gaps

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

(see Figure 3 above). Small pictures above each gap illustrated the target words.

The task was trialled in two different forms – in one the words with pictures appear in a box at the end of the text and children have to choose the correct word and copy it into the appropriate gap. In the other, the picture and the first letter of the word was included.

Trialling showed that the variant with pictures and first letter only raised the difficulty of the task whereas the variant where children chose the correct word from a box was much more in line with the original difficulty level, so the format shown in Figure 4 was agreed.

The aim of this revision was to change the focus of the task from writing isolated words based on picture clues to reading the whole text for meaning. The new version of the task ensures that children correctly identify the missing words by looking for clues in the text before and after the gap (e.g. does the context require a noun or a verb in the gap?). The task is also well scaffolded as the written forms are supplied for children to copy. This task is similar to the Part 4 Reading/Writing task at Starters, but is more

supported than the Part 4 task at Flyers which includes words without pictures.

Starters Reading and Writing Part 3

In Part 3 of the Starters Reading and Writing paper, spelling and written production of vocabulary are tested through the use of anagrams. Children are asked to write five words from a given lexical set. For each item they are supplied with a picture, an appropriate number of spaces and a set of jumbled letters.

Some teachers commented that the jumbled letters might hinder rather than help children. There was also a feeling that the concept of an anagram was beyond the cognitive development of younger children. So, a variant of this task was trialled in which candidates were given a picture and the first letter of the word with the correct number of spaces in which to write the word. This task is similar to a task in KET. The revised task was trialled with two different lexical sets (furniture and clothes) and in both cases the difficulty of the task increased significantly. It was noted that children did not attempt to write the words in many cases. From this

we could infer that children may find it hard to produce a word with very limited support. Furthermore the new task also seemed to discourage risk taking. Qualitative feedback corroborated these findings:

> This task seems to be more difficult than when the candidates were given…. the jumbled letters. They get less help if they do not remember the name of the object and they might also make more spelling mistakes. *(Teacher, Argentina)*

For these reasons it was decided to retain the original task (see Figure 5).

Updating the vocabulary lists

One of the key areas of investigation during the review of the Cambridge YLE tests was the vocabulary lists. The vocabulary lists are an essential element of the tests: they form the basis of the syllabus, and are made publicly available in the handbook. Initial suggested changes to the YLE vocabulary lists were made based on feedback from the teams of test writers. A comprehensive study of primary English Language Teaching coursebooks was then carried out. The suggested words were mapped onto those in the coursebooks to establish areas of overlap. Alongside this, Cambridge ESOL's Research and Validation Group carried out research using the YLE Spoken Corpus. Over 100 YLE Speaking tests which had already been transcribed were included. The research used a corpus-based methodology (manual analysis and Wordsmith Tools software) to compare the existing vocabulary lists and the suggested inclusions to see whether children produced any of the suggested additions to the list at the appropriate level or produced any words or structures not in the list.

Updating the vocabulary lists based on the results of this consultation and research involved the addition of some words at each level and the movement of words from one level to another. The main reasons for adding words were:

- to extend an existing lexical set, e.g. *sand* and *shell* have been added at Starters to join *beach* and *sea*; *puppy* and *kitten* are new words at Movers and *fire engine* has been added at Flyers

- to update the lists to ensure they include words which have entered common usage and are relevant for children today, e.g. *great* at Starters and *DVD* at Movers.

The main reasons for moving words from one level to another were:

- to extend a lexical set at a particular level, e.g. *doctor* and *nurse* have moved from Flyers to Movers as other words and expressions in the lexical set of health such as *hospital*, *cough*, and *What's the matter?* already appear at Movers

- to reflect more closely the level at which these words are normally introduced in course books, e.g. *shop*, *park* and *playground* have moved from Movers to Starters and *make* has moved from Flyers to Starters.

The revised vocabulary lists for each level are included in the handbook and are arranged alphabetically, thematically and grammatically. This enables teachers to prepare candidates fully for all the language they will encounter in the tests.

What next?

What have we learnt so far from the YLE test revision process? The revised tests went live in January 2007 and we have already received positive feedback from teachers. There is an appreciation of the clearer test focus for each task and the new words in the vocabulary lists. In addition the new task types have led to clearer guidelines for markers and this enhances marker standardisation. Clearly any revision is an iterative process and the trialling and resulting new tasks have led to new issues and insights and there is a need for continued analysis and reflection. Further questions which have arisen out of the process are: What is the optimum balance between reading and writing in the Reading and Writing tests at each level? How closely do the tasks and texts reflect language and content which is interesting and relevant for children of this age group? These will be areas for future research and we will of course continue to monitor and evaluate the impact of the revision of the Cambridge YLE tests.

References and further reading

Burrow, T and Wilson, J (2004) The YLE Review: findings from a stakeholder survey, *Research Notes* 15, 16–18.

Saville, N (2001) Test development and revision, *Research Notes* 4, 5–8.

The marking of spelling for the revised YLE tests from January 2007

HELEN SPILLETT ASSESSMENT AND OPERATIONS GROUP

Introduction

As a part of Cambridge ESOL's ongoing commitment to improving the quality of its tests, the Cambridge YLE tests (YLE) underwent a process of review and revision which began in 2003 and led to the launch of the revised tests in January 2007. In the course of the review the existing policy for marking candidate responses which were not spelled correctly was reconsidered. A consultation exercise and small-scale research study were carried out to develop a new policy on spelling and this was implemented to coincide with the launch of the revised tests. The main change has been the introduction of a requirement for 100% correct spelling in the YLE Reading and Writing papers. Some misspellings of words targeted in YLE

Listening Movers and Flyers Part 2 are accepted. This article explains the reasons for considering a change in policy and describes the procedures and the outcomes of the consultation exercise and research study.

Why change the policy?

The YLE tests were introduced in 1997 for a target candidature of children between the ages of seven and twelve. The tests are pre-graded and taken on demand at centres in a wide range of countries and then marked by a team of trained markers in Cambridge. YLE tests are also administered on a fixed dates basis in China and marked by teams of trained markers at YLE centres there. It has always been an important aspect of the YLE tests that they aim to give children a positive first experience of the formal assessment of English. This is reflected in the fact that there is no pass or fail for these tests and that they are designed to be of high facility. All children attempting the tests receive an award which focuses on what they know, rather than penalises them for what they do not know. It is important that children in this age range should know what to expect when they sit down to take the test, or the unfamiliarity could cause anxiety. The Cambridge YLE handbook therefore provides a clear syllabus which contains task specifications and lists of topics, words and structures used at each level (Starters, Movers and Flyers). Examples of each task type are available in the sample papers. The issue of spelling does not arise for some parts of the Reading and Writing and Listening papers in which children respond by drawing a line, ticking a box, or following an instruction to colour or draw.

Various approaches are possible to the policy for the marking of spelling in the YLE tests, from a strict requirement of 100% correct spelling, to a policy of extreme leniency based on the idea already mentioned that the tests should be a positive and enjoyable early experience of an international test. The policy in operation prior to the review steered a middle course between these two extremes: judgements as to whether a particular misspelling should be deemed acceptable were made taking a variety of different factors into account (e.g. whether the misspelling was phonetically plausible, whether the error was the omission or addition of just one letter, the over-generalisation of an English spelling rule, etc.). The implementation of this policy generated theoretical and practical difficulties ranging from the identification of agreed criteria for the acceptance of misspellings to the consistent management of marker queries. Fairness to candidates and reliability of test scores were the key motivations for a systematic review of this aspect of the YLE marking policy. A secondary but important factor was the need to clarify practical and efficient procedures for marking.

The consultation exercise

In order to establish whether a list of clear criteria for the acceptability of particular misspellings could be drawn up, it was agreed that there should be a process of consultation involving scholars and practitioners with appropriate expertise in the teaching and assessment of young second

language learners and knowledge of relevant research in applied linguistics. Six participants were chosen, three external consultants who are involved in the production of YLE test papers and three applied linguists with an interest in second language acquisition and the teaching and assessment of young learners. There were two stages of this consultation process.

In the first stage, which took place in February 2005, a list of preliminary questions was sent to each participant:

1. What types of misspelling should be marked as correct at each level of the tests, i.e. in:
   - Starters RW – Parts 4 and 5
   - Movers RW – Parts 4 and 5
   - Flyers RW – Parts 4, 5 and 7
   - Starters L – Part 2
   - Movers L – Part 2
   - Flyers L – Part 2?
2. What criteria underlie your responses to the first question?
3. What differences should there be in the policy on spelling for Reading/Writing and the policy for Listening?
4. Are you aware of any relevant research studies e.g. into the relationship between spelling and the development of literacy in ESL/EFL young learners?

Participants were sent examples of each task type in the YLE tests which requires a written response from candidates and, with reference to the mark schemes for these tasks, they were asked questions 1–3 above. Question 4 was included to elicit information about any research studies into the relationship between spelling and the development of literacy in ESL/EFL young learners. It was hoped that insights from such research could then inform the next stage of the consultation exercise.

The results of this first stage were analysed to identify any consensus about particular criteria. None of the participants at this stage proposed a requirement of 100% correct spelling for all tasks. Equally, none of them suggested that any vaguely recognisable attempt at the target word should be marked correct for any task – everyone took the view that some criteria to limit acceptability must be applied. An extensive range of criteria were identified but there was very little consensus about which should be used. In Table 1 all the criteria suggested are listed in the left-hand column and the views of the six numbered participants are indicated. A tick means that the participant firmly selected the criterion, a cross means that the participant specifically ruled the criterion out and a blank reflects the fact that the participant did not include a discussion of the criterion in their response. A question mark indicates that the criterion was tentatively considered.

The second stage of the exercise was carried out in July 2005. A list of 308 actual candidate misspellings was compiled, of a variety of words from different parts of the test at each of the YLE levels. These misspellings were taken from common wrong answers compiled during the calibration of new test versions, and care was taken to ensure inclusion of examples of all types of misspelling which might meet the criteria for acceptability listed in

Table 1: Suggested criteria for marking decision

	1	2	3	4	5	6
Consideration of L1 phonetic systems			?			
'Copying' task so should be 100% correct		✓	✓			✓
Different criteria for S/M/F	x			✓	✓	
Different criteria for different ages	?	?				
Phonetically plausible for English sound/spelling system		✓			✓	
More leniency for 'lexical' rather than 'grammatical' items					✓	
More leniency for more cognitively challenging and for 'freer' writing tasks	✓	✓				
'Recognisable' by non TEFL native speaker			✓	✓	✓	✓
Ease of marking	✓					✓
Stricter for higher frequency words restrict # of errors per word by # of letters				✓		
					✓	
Restrict # of errors per word by syllables						✓
Do not allow misspellings that could be a different English word		✓		✓		
Leniency for errors native speaker children make				✓		
Stricter/more lenient with certain types of error, e.g. –ing/ed, double letters, vowels	✓	✓	✓			
Need to distinguish between test of reading and test of writing	✓			✓		
Allow text messaging spelling				?		
Listening more lenient than R/W (except Starters Pt 2 where words are spelt out)	✓	✓		✓	✓	differences should be based on aims/ demands of tasks
US/UK – both allowed in markscheme	✓	✓		✓	✓	
Need to allow for variations based on US spelling/sound system		✓			✓	

Table 2: Example of misspellings

Level	L/RW	Part	Word	Misspelling	✓x	Comment
S	L	2	eight	eigh		
S	L	2	eight	eigt		
S	L	2	eight	eihte		
S	L	4	rooms	roms		

Table 1. This list was sent to participants, with the information presented as shown in Table 2.

Participants were asked not to look through the list in advance or to refer back to the criteria for marking misspellings which they had suggested as part of the first stage of the consultation exercise. They were asked to mark each misspelling right or wrong with a tick or a cross, depending on the part of the test in which it occurred and to add a comment, as they marked, about the reason for their decision for each case. They were then asked, on completion of the marking of this list of misspellings, to refer back to their own suggested criteria and look again at their ticks and crosses, but not to make any changes to their marking decisions. They were invited to make some observations about their original proposed criteria as a result of their experience of marking in this second stage. There was a very low degree of consensus as to the marking of the items on the list. Of the 308 misspellings, there was agreement between all six experts on only **nine** of these words as being acceptable. In the observations made by the experts a consensus emerged that there are strong arguments for being reasonably strict with regard to spelling, requiring 100% correct spelling in Reading and Writing and also for Listening with the exception of Movers and Flyers Part 2, in which candidates have to write words which they hear but which do not appear on the question paper. The subjectivity demonstrated by the variations in the marking exercise gave further support to this view.

Small-scale research study

As a result of the outcome of the consultation exercise, a small-scale study was carried out to quantify the impact of a requirement of 100% correct spelling (in Reading and Writing tasks in which this was not already a requirement) on candidate scores and shields. Only a restricted number of parts of the tests are affected by this change. Of these, as a part of the revision, Reading and Writing Part 4[1] (all levels), which is a gap-fill task, now has written forms of the target words supplied on the question paper. Reading and Writing Part 5 (Movers and Flyers) which was previously a question and answer comprehension task, has been changed to a sentence gap-fill task in which the words candidates must supply can be found in the text. For Reading and Writing Part 5 (Starters), which is a picture story comprehension task requiring a written response, a sample of 71 scripts, from candidates at 5 different centres, for two different versions of a Starters paper, was remarked. The number of candidates whose mark would have been reduced if correct spelling had been obligatory was seven. These results therefore suggest that insisting on correct spelling in Part 5 would reduce the overall mark in a small minority of cases. In only two of these cases would there have been a change in the shield awarded. The number of cases could be expected to decrease when the revised test is live, because teachers preparing children taking the tests will have been made aware that spelling must be 100% correct. The effects of the change will also be closely monitored.

---

1. Examples of each part of the revised YLE Tests can be seen in the sample papers which can be downloaded from the Cambridge ESOL website at www.cambridgeesol.org/support/dloads/yle

For Listening Starters Part 2, candidates have to write numbers (1–20) and words which they hear spelled out in the recording. Teachers are advised to encourage candidates to use numbers rather than spelling out number words. In line with the decision to take a stricter view on the marking of misspelling in the interests of fairness, it has been decided that candidates should be expected to spell number words (which are on the Starters wordlist) correctly and that therefore 100% correct spelling should be required for Listening Starters Part 2. In the research study it was found that 9.8% of candidates chose to spell rather than write the number they hear. Of these, all either spelled the number correctly or so badly that the misspelling did not feature on the list of acceptable misspellings on the markscheme. Insisting on correct spelling of the number would have made no difference to the mark of those candidates looked at, and therefore also not to the shields awarded.

Listening Part 2, Movers and Flyers

For Listening Part 2 (Movers and Flyers), in which candidates must write words they hear, with no support on the question paper, the view of all participants in the consultation exercise was that some acceptable misspellings should be included on the markscheme for some targeted words, and so the difficult issue of which criteria to select for determining inclusion as an acceptable misspelling remained to be resolved for these parts of the test. On the basis of majority agreement on criteria for this part of the test after the second stage of the consultation exercise, and ease of applicability with a minimum of subjective judgement, a list of criteria has been agreed. For example, 'recognisable by a non TEFL native speaker' is not included in the list but 'single consonants for double and vice versa in words of more than one syllable' is included. No allowance is made for L1-specific problems as, given the global candidature for YLE, it would not be possible to cover

all of these and this would therefore risk discriminating unfairly in favour of certain L1 groups. Before new test versions are calibrated, the chair and subject officer will agree a list of acceptable spellings for each targeted word in the light of the agreed criteria and with reference to frequent KET/PET misspellings in the Cambridge Learner Corpus (see Barker 2007). Answers queried by markers in the calibration process will be considered for inclusion in the final key in the same way. A new database will be compiled of accepted misspellings for all words targeted in Listening Part 2, Movers and Flyers.

Conclusion

A requirement of 100% correct spelling is not in itself in conflict with the ethos of a test aimed at supporting and encouraging young learners: the constraint that tests at each level can target only those words included in the relevant vocabulary list, and the accessibility of those vocabulary lists to teachers of YLE preparation courses, provide a natural limitation to this requirement. A part of what is being tested in the YLE tests is knowledge of the lexical items on the vocabulary lists and it is perhaps not unreasonable to include knowing how to spell a word as part of what it means to know a word (Nation 1990:31).
The lack of consensus in the first stage of the consultation exercise and in the practical exercise in the second stage confirmed the need to establish a clear and consistent policy on spelling. The removal of any element of subjectivity in the marking of Reading and Writing and its constraint in the marking of Listening Part 2 seems a positive step forward in ensuring that all YLE candidates are fairly treated.

References and further reading

Barker, F (2007) Corpora and language assessment: trends and prospects, *Research Notes* 26, 2–4.

Nation, I S P (1990) *Teaching and Learning Vocabulary*, Boston: Heinle and Heinle.

Cambridge ESOL YLE tests and children's first steps in reading and writing in English

SHELAGH RIXON CENTRE FOR ENGLISH LANGUAGE TEACHER EDUCATION, UNIVERSITY OF WARWICK

Introduction

This contribution springs from the recent research project by Cambridge ESOL to inform policy over the marking of spelling in the Cambridge YLE tests (see Spillett in this issue). It also reflects a long standing interest of my own in the origins, composition and uses of the vocabulary lists for the Starters, Movers and Flyers tests (see Rixon, 1999). The analysis that follows is based on the new lists created for the updated 2007 tests.

The language specifications for the YLE tests were created with the interests and communication needs of children as

their starting point, that is, principally with meaning rather than matters of form in mind. This is not to say that the vocabulary lists, my main focus, were compiled without regard for the degree of formal challenge that the words contained might pose in terms of spelling and recognition. However, the intention is for topics and their associated vocabulary items to drive the syllabus in a way that reflects the priority to allow Young Learners (YLs) from a very wide range of countries and cultures to use English meaningfully.

The YLE handbook for 2007 (p. 3) states: 'particular attention is paid to the educational consequences of using a language test with young learners' and that clearly refers

to the type of activities that are chosen for testing purposes, the linguistic content of the tests and the types of answer that gain rewards in terms of marks. All are going to have a washback effect on what takes place in preparation courses. In the area of linguistic content, I should like to examine the potential that the contents of the vocabulary syllabuses have for furthering principled approaches to handling the written word by teachers preparing children for the tests. This is particularly appropriate now that there is a policy of requiring 100% accuracy in the spelling of words that form answers to Reading and Writing components of the tests.

Teaching and learning vocabulary

Even at the Starters level children are required to decode and manipulate a certain amount of written text with some accuracy. It seemed interesting and useful to look at the word lists from the point of view of the particular decoding and spelling difficulties their contents might present to children for whom English is a foreign language. From such a study it might be possible to draw insights to help teachers to decide on what they should focus. In addition, it seemed appropriate to analyse the lists and to discuss how they match up with what is expected of native users of English also taking their first steps towards reading and writing proficiency. Of course there are significant differences between the two types of learner, and this is not at all to suggest that the stages of early literacy building for native speakers should strongly influence the content or sequence of items for YL foreign language courses which usually have a largely speaking and listening focus. However, it seems a sensible procedure at least to check that learners of English as a Foreign Language (EFL) are not being expected to operate at a higher level of challenge where formal correctness is involved than their native speaker counterparts.

It seems worthwhile to treat the Cambridge ESOL YLE vocabulary lists as a given, and, taking them in their new 2007 form, to reflect on ways in which teachers preparing children for the exams could group and focus on vocabulary items from the point of view of how challenging they will be to spell correctly. Test designers and teachers alike might find it useful to have procedures by which to predict which words are likely to be difficult for children to link in their spoken and written forms or to spell when required by one of the tests. For this study, my original plan was to analyse the vocabulary specifications for all three levels of Cambridge YLE tests. However, in this article I have limited myself to the vocabulary list for the Starters tests. This can be justified on two counts. Firstly there are pragmatic reasons of time and space. This study is an experiment to see if this type of analysis has any practical pedagogical value, and so it makes sense to expend time and effort only on a small manageable sample of language in the first instance. Secondly, and more cogently, the stage of learning in which I am particularly interested – the children's early contacts with the written language – is most likely to be represented by candidates preparing for Starters. It would be very useful, however, to receive feedback and comments from readers on whether

continuing the study to Movers and Flyers levels would seem to be worthwhile.

Results from a recent survey of YL teachers (see Rixon, 2007) suggest that pedagogical ideas such as these could be useful, in that YL teachers have often not received special training in the teaching of initial literacy in English. In addition, published YL materials used in many contexts, including those in which the native language uses a different writing system, tend to present the written word in English on the page from the very start as if this were an unproblematic support to learning rather than a challenge in itself. Where systematic reading instruction is attempted in YL materials, it is often through much reduced and impoverished processes compared with those that have been developed for native speaker learners. A common practice with YL, for example, is to focus mainly on the symbol/sound correspondences of the onset sections of words and to teach these correspondences following alphabetical order. So, lessons move from 'a' for apple' to 'b for boy' etc. Some English for Young Learners (EYL) materials I have analysed stop at 'z' and do not seem to acknowledge that there are 43 phonemes in modern British Received Pronunciation (44 with the conservative pronunciation of the diphthong in 'poor') while there are only 26 letters with which to represent them. Furthermore, many phonemes in English are commonly realised by more than one letter-symbol. Many EYL materials never even reach common digraphs such as 'ch', 'th' and 'sh' or only treat them very late. The teaching of native speaker initial literacy, of course, is based on a very different fundamental condition – the fact that native speaking children already have a secure oral/aural grasp of the language involved – but the teaching that native speaking children receive follows a sequence and involves a concern for detail that acknowledges that there are significant challenges even for them. In addition, it is recognised that some letters and letter combinations are more 'trustworthy' and consistent in their symbol/sound relationships than others, and the ideal teaching sequence for focusing upon these does not follow the a, b, c, order that we see in many YL materials. Many authorities on early reading (e.g. Goswami and Bryant 1990, Goswami and East 2000) advocate helping children to work out words by analogy, especially through paying attention to the end-element of syllables – the so-called 'rime', and suggest that grouping English words by rime (e.g. *the fat cat sat on a mat*) is much more accessible and useful for children when they are learning to read than grouping them by initial letter (*boy*, *ball*, *boat*, *beach*). This insight is rarely made use of in YL materials, although the widespread use of rhymes and chants – often used only for fun purposes – offers an opportunity to build awareness of rime and analogy in a child-friendly way.

As mentioned above, there is the issue that while with native speaking children there is acceptance by their teachers that full control over the written forms of words will only gradually emerge over a number of years, in the case of young EFL learners it is often true that their teachers are concerned if rapid full accurate control of words which have been 'taught' is not evidenced. We need to be vigilant

that we are not requiring more of an 8 or 9 year old beginner in English in terms of accurate spelling and decoding than we would of his or her native speaking counterpart.

The study

The vocabulary list

When working with the Starters list, I made a number of adjustments that affected the total count of words on which I carried out calculations. For example, I took the singular and plural of nouns with irregular plural forms, e.g. *foot/feet,* as separate items and I incorporated the numbers 1–20 in word form into the list. The word count on which the results are based is therefore my own. Not all of my decisions about how to count and group words may meet agreement by all readers, but I think that the global picture that emerges will remain unaffected by small differences in allocation.

Measures used

A number of measures were applied to determine the challenges that particular words could present for YLs. Most are derived from work involving young native speaker children in their initial stages of building literacy and they therefore need to be modified or considered in a different way to suit EFL contexts. I have explained my decisions under each heading.

Orthographic depth and transparency

Orthographic depth is a term whose coinage is attributed to Katz and Frost (1992). It denotes the degree to which the symbols used to write any particular language give a reliable representation of its phonemes and whether, conversely, any given phoneme is consistently represented by just one symbol or combination of symbols. Languages in which the symbol/sound symmetry is perfect or near-perfect are said to be *orthographically shallow or transparent* and those in which the relationship is loose or complex are said to be *orthographically deep*. English is one of the 'deep' languages as can be seen in Table 1, adapted from the account of the cross-linguistic project 'Literacy Acquisition in European Orthographies' reported on www.dundee.ac.uk/psychology/collesrc

Table 1: Comparison of languages in terms of orthographic depth

Shallow	Orthographic Depth ››››				Deep
Finnish	Greek	Spanish	Portuguese	French	
	Icelandic	Norwegian	Swedish	Danish	English

Measures of orthographic depth are applied across languages to arrive at inter-language comparisons such as in Table 1, but they are also applicable *intra-linguistically*, within a language, in order to determine which words present less challenge from the symbol/phoneme point of view and which words present more or extreme challenge. Examples of least challenging items for English are the 'fat', 'cat', 'mat', 'hat' set with their consistent consonant-vowel-consonant (CVC) structure with a short vowel consistently represented by a single letter. Examples of very challenging (although also rather frequent) items are 'eight', 'laugh' and 'two'. It has been posited that learners need more than one route to cope with decoding words in a language like English where, although for some words rules for conversions between symbol and sound can reliably be applied, in other cases a visually-based 'look up' recognition system needs to be used for words where symbol/sound relationships are weak or uncertain. For children working towards Starters, words like 'two' and 'eight' are good examples of when a visual route is needed for rapid recognition or correct spelling of the word. Older readers and spellers may remember struggling with the correct pronunciation of words like 'chaos', 'Leicester' and 'misled' – words where visual recognition and being informed about the pronunciation that goes with each word is the only sure way.

This duality within English perhaps explains the constant debate among early reading specialists between advocates of phonics (which emphasises sound/symbol correspondences in the teaching of reading) and of 'look and say' (which emphasises visual recognition). It may also account for the quiet decision by most practical teachers of early reading to give children training in both approaches.

A measure which indicates the orthographic transparency or otherwise of individual words concerns whether the number of letters (graphemes) in a word is equal or not to the number of phonemes. This phoneme/grapheme difference is discussed and applied in Spencer (1999). English is highly challenging in this way, in that many words employ a significantly different number of symbols to represent sounds – usually more symbols than sounds. The more 'transparent' words like 'cat' have exact correspondence in numbers of symbols and phonemes (so a difference of 0) but other words such as 'eight' (pronounced /eɪt/) have a large difference – 5 symbols for 2 phonemes in this case (so a difference of 3). A word such as 'six' has a difference of 1, but in a minus direction since the symbol 'x' here represents two phonemes /ks/ so there are more phonemes than symbols. Applying this formula to a list of words such as those for Starters highlights ways in which reading and spelling may also present cognitive challenges to young beginners that go beyond merely learning to recognise the different phonemic values that some letters may have and the different ways in which phonemes may be encoded in letters. The challenges apply to both native and non native English users and lie in coming to grips with the very fact that there are several different ways in which phonemes and symbols may match up and be distributed. This is an area that may be overtly taught (as it is in the UK National Literacy Strategy where children play 'count the phonemes and count the letters' games) or acquired (or not) through trial and error and hard experience. Some overt teaching may reduce the burden on YLs to work out the complexities for themselves. Table 2 indicates some of the ways in which phonemes and graphemes can relate with regard to the words on the Starters list. I have divided the list so that information is given on words that go from one syllable up to the few four-syllable words that can be found in Starters.

From Table 2 it may readily be seen that a scale of difficulty can be estimated in which single-syllable words with transparent 1-to-1 relationships between letters and phonemes could be seen as probably easier to manage, while words like 'throw' and 'choose' may need a special teaching focus if learners are to cope with their spelling. A zero difference may not always mean that a word is easy, as we can see in the cases of 'new' and 'one'. In British RP there is an 'invisible' phoneme not represented in letter form in the /j/ glide in 'new' /nju/ and an unseen /w/ at the beginning of 'one'. It may be hypothesised that words with more syllables may present more challenge, with differences in challenge within these groups as with the 3-syllable but very transparent 'coconut' (7 letters and the

Table 2: Phoneme-grapheme differences in one-, two-, three- and four-syllable words
(Total number of words in the Starters list = 460)

Type of phoneme-grapheme difference (British RP pronunciation is assumed)	Number of words	Examples
One-syllable words		
0 difference: and a 1 to 1 relationship of phonemes and letters	82	'bag'. 'from'
0 difference but with a complex relationship between phonemes and letters	2	'new' 'one'
+ 1 difference: 1 more symbol than phonemes (includes double letters e.g. 'dd', 'silent r', and common digraphs)	161	'add' 'bath' 'skirt'
+ 2 differences: 2 more symbols than phonemes	45	'beach' 'shirt' 'throw''clothes'
+ 3 differences: 3 more symbols than phonemes	2	'choose' 'eight'
Minus 1 difference: 1 more phoneme than symbols	1	'six'
	Total 293	
Two-syllable words and phrases		
0 difference: and a 1 to 1 relationship of phonemes and letters	31	'Alex' 'baby' 'lemon' 'seven'
0 difference but with a complex relationship between phonemes and letters	1	'sixteen'
+ 1 difference: 1 more symbol than phonemes (includes double letters e.g. 'dd', 'silent r', and common digraphs)	51	'correct' 'fifteen' 'monkey' 'tick' 'tiger'
+ 2 differences: 2 more symbols than phonemes	39	'answer' 'classroom' 'nineteen' 'sausage' 'TV'
+ 3 differences: 3 more symbols than phonemes	9	'armchair' 'thirteen'
Minus 1 difference: 1 more phoneme than symbols	2	'next to' 'OK'
	Total 133	
Three-syllable words and phrases		
0 difference: and a 1 to 1 relationship of phonemes and letters	9	'animal' 'badminton' 'camera' 'eleven'
0 difference but with a complex relationship between phonemes and letters	2	'beautiful' 'example'
+ 1 difference: 1 more symbol than phonemes (includes double letters e.g. 'dd', 'silent r', and common digraphs) 'seventeen'	12	'alphabet' 'basketball' 'crocodile' 'elephant'
+ 2 differences: 2 more symbols than phonemes	6	'afternoon' 'dining room' 'motorbike'
+ 3 differences: 3 more symbols than phonemes	1	'pineapple'
Minus 1 difference: 1 more phoneme than symbols	0	
	Total 30	
Four-syllable words and phrases		
0 difference: and a 1 to 1 relationship of phonemes and letters	0	
0 difference but with a complex relationship between phonemes and letters	0	
+ 1 difference: 1 more symbol than phonemes (includes double letters e.g. 'dd', 'silent r', and common digraphs)	2	'helicopter' 'watermelon'
+ 1 difference but with a complex relationship between phonemes and letters	1	'television'
+ 2 differences: 2 more symbols than phonemes	1	'table tennis'
+ 3 differences: 3 more symbols than phonemes	0	
Minus 1 difference: 1 more phoneme than symbols	0	
	Total 4	

same number of phonemes) compared with the 3-syllable but less transparent 'motorbike' in which 9 letters represent 7 phonemes.

It would be interesting to compare these predictions with the data on actual YLE test misspellings that were collected for the study carried out by Spillett and reported in this issue.

Orthographic regularity

Cook (1997) discusses an additional measure which could be very relevant to the young spellers taking the Cambridge YLE tests – that of orthographic regularity. It concerns language users' awareness of permissible positions of letters in words when these seem to be determined purely on a visual basis. For example, the letter 'k' representing the phoneme /k/ is flexible in its permitted positions and can appear both at the beginnings and the ends of syllables, as in 'kite' and 'look'. The letters 'ck', on the other hand, represent the same /k/ phoneme, but can appear only at the ends of syllables, as in 'black' or 'tick'. Similarly, the letter combination 'tch' can only appear at the ends of syllables, as in 'catch' whereas 'ch' which often represents the same phoneme can appear both at the beginnings and ends, as in 'chips' and 'beach'.

There are relatively few words in the Starters list to pay attention to on this basis and of course it is not recommended that the rules above are taught overtly as rules to young children. The information here is for teachers and for test compilers to help their estimates of how difficult different items may be for YLE to produce 100% correctly in a test.

Frequency

The measures above may have some value in themselves, but an additional, more obvious measure that has long (e.g. Elley 1969) been felt useful with regard to native speaker beginning readers is that of *frequency*. For these learners, frequency (in speech as well as the written form) may be seen to interact with the depth or transparency of words, perhaps to render them less of a challenge if a word is likely to be encountered very frequently, or more of a challenge if it is rarely met. Frequency is often in this context linked with *familiarity*. The idea of familiarity is reasonable to apply with native speakers since they are expected to 'know' a large number of words from their day-to-day experiences with the language and thus to be able to apply this knowledge in their reading. For foreign language readers or spellers, however, it is not so appropriate to think in terms of familiarity when judging how difficult a word is likely to be to recognise or to spell. YLs of English as a Foreign Language are likely to be 'familiar' with very few if any words outside the context of their language courses. On the other hand, frequency may be applied as a means of judging how useful an item on a syllabus is likely to be to the children for their future learning or English-using experiences. If children learn frequently-used words early this may contribute to a more confident progression to more challenging reading and writing tasks.

The basis on which to judge frequency presents further problems when we are concerned with young foreign language learners (see Rixon 1999:61–63). In the literature on reading and spelling development in native speaking children, many studies (e.g. Spencer, 1999) have been carried out on the basis of modified word lists derived from corpora of adult language. This seems a curious choice. However, corpora of child language such as that developed by Raban (1988) from Wells's (1985) study might also offer problems for our purposes. They are often based mainly or only on oral language and that language naturally contains many culture-specific references, and we are presumably not trying to turn our YLs into close linguistic replicas of British or American children. Some, like the Raban (1988) corpus of five-year-old children's speech, are derived from the speech of children rather younger than our age group, so although they can offer very useful general insights into aspects such as the rich array of verbs used by native speaking children compared with the contents of many YL courses which tend to be noun-rich and verb-poor, I decided not to use a corpus for this particular study.

It seemed appropriate to use word lists that were frequency-based and directly relevant to the reading/writing mode. The UK National Literacy Strategy resources file gives, as List 1, 209 words (by my method of counting) which it is recommended that children should be able to recognise on sight by the end of Year 2 (age 6–7). The reason given is that they are 'high frequency words which pupils will need, even to tackle very simple texts' (DfEE 1998:60–61). Although no information is given about the basis on which frequency is judged, this list seems a justifiable choice for simple comparison with the Starters word list. A further set of 121 words (List 2) that could be considered contains 'Medium frequency words to be taught through Years 4 and 5' (DfEE 1998:62) and here spelling difficulties are specifically mentioned. These National Literacy Strategy lists include transparent words as well as words whose sound/symbol relationships are more complex. The contents of the lists should not of course be taken a priori as essential or desirable content for a course in English as a *foreign* language. There could well be words on a reading/spelling list for native speaking children that it is not vital to teach on a YL course, especially where the major focus is speaking and listening. Conversely, there are words that may not figure in the lists intended for native speaking children that are nonetheless appropriate for a YL course. Examples might be language needed for classroom management, such as 'book' and 'teacher' or needed in order to 'do' a particular topic like animals. However, since all the words on the Cambridge YL list could potentially form the basis for written responses in a Reading, Writing or Listening test component, it seems legitimate to investigate how many of them are listed as important for native speaker learners in their early stages of literacy development.

Results of a comparison with the National Literacy Strategy lists

There was a considerable overlap with regard to the National Literacy Strategy List 1, with 107 words (51%) from that list found on the Starters list. These 107 words represent 23% of the total Starters list. Word groups in common are days of the week, months of the year and

Table 3: Vocabulary list analysis

Word	In NLS list 1 or 2?			Total
	0	**1**	**2**	
a det	0	1	0	1
a lot ad	1	0	0	1
a lot of	1	0	0	1
about pr	0	1	0	1
add v	1	0	0	1
afternoon	1	0	0	1
again ad	0	1	0	1
Alex n	1	0	0	1
alphabet	1	0	0	1
an det	0	1	0	1
and conj	0	1	0	1
angry ad	1	0	0	1
animal n	0	0	1	1
Ann	1	0	0	1
Anna	1	0	0	1
answer n	1	0	0	1
apartment n	1	0	0	1
apple n	1	0	0	1
arm n	1	0	0	1
armchair	1	0	0	1
ask v	0	0	1	1
at prep	0	1	0	1
baby n	0	0	1	1
badminton	0	0	1	1
bag n	1	0	0	1
ball n	0	1	0	1
banana n	1	0	0	1
baseball	1	0	0	1
basketball	1	0	0	1
bath n	1	0	0	1
bathroom n	1	0	0	1
be v	0	1	0	1
beach n	1	0	0	1
bean n	1	0	0	1
beautiful	1	0	0	1
bed n	0	1	0	1
bedroom n	1	0	0	1
behind p	1	0	0	1
Ben n	1	0	0	1
between	0	0	1	1
big adj	0	1	0	1
bike n	1	0	0	1
Bill n	1	0	0	1
bird n	1	0	0	1
birthday	0	0	1	1
black ad	0	1	0	1
blue adj	0	1	0	1
board n	1	0	0	1
boat n	1	0	0	1
body n	1	0	0	1
book n	1	0	0	1
bookcase	1	0	0	1
bounce v	1	0	0	1
box n	1	0	0	1
boy n	0	1	0	1
bread n	1	0	0	1
breakfast n	1	0	0	1
brother n	0	1	0	1
brown ad	0	1	0	1
burger n	1	0	0	1
bus n	1	0	0	1
but conj	0	1	0	1
Bye	1	0	0	1
cake n	1	0	0	1
camera n	1	0	0	1
can v	0	1	0	1
car n	1	0	0	1
carrot n	1	0	0	1
cat n	0	1	0	1
catch v	1	0	0	1
chair n	1	0	0	1
chicken	1	0	0	1
child	1	0	0	1
children	1	0	0	1
chips n	1	0	0	1
choose v	1	0	0	1
class n	1	0	0	1

Word	In NLS list 1 or 2?			Total
	0	**1**	**2**	
classroom n	1	0	0	1
clean ad	1	0	0	1
clock n	1	0	0	1
close v	1	0	0	1
closed adj	1	0	0	1
clothes	1	0	0	1
coconut	1	0	0	1
colour n	1	0	0	1
come v	0	1	0	1
complete	1	0	0	1
computer	1	0	0	1
correct	1	0	0	1
cousin n	1	0	0	1
cow n	1	0	0	1
crocodile	1	0	0	1
cross n	1	0	0	1
cupboard n	1	0	0	1
dad n	0	1	0	1
daddy n	0	1	0	1
day n	0	1	0	1
desk n	1	0	0	1
dining r	1	0	0	1
dinner n	1	0	0	1
dirty ad	1	0	0	1
do v	0	1	0	1
dog n	0	1	0	1
doll n	1	0	0	1
door n	0	0	1	1
double a	1	0	0	1
draw v	1	0	0	1
drawing n	1	0	0	1
dress n	1	0	0	1
drink n	1	0	0	1
drive v	1	0	0	1
duck n	1	0	0	1
ear n	1	0	0	1
eat v	1	0	0	1
egg n	1	0	0	1
eight numeral	0	1	0	1
eighteen numeral	0	1	0	1
elephant	1	0	0	1
eleven n	1	0	0	1
end n	1	0	0	1
English	1	0	0	1
enjoy v	1	0	0	1
eraser n	1	0	0	1
evening	1	0	0	1
example	1	0	0	1
eye n	1	0	0	1
face n	1	0	0	1
family n	1	0	0	1
father n	0	0	1	1
favourite adj	1	0	0	1
feet n	1	0	0	1
fifteen numeral	0	1	0	1
find v	1	0	0	1
fish s	1	0	0	1
fishing	1	0	0	1
five numeral	0	1	0	1
flat n	1	0	0	1
floor n	1	0	0	1
flower n	1	0	0	1
fly v	1	0	0	1
food n	1	0	0	1
foot n	1	0	0	1
football n	1	0	0	1
for prep	0	1	0	1
four numeral	1	0	0	1
fourteen numeral	0	1	0	1
friend n	0	0	1	1
fries n	1	0	0	1
frog n	1	0	0	1
from pre	0	1	0	1
front n	1	0	0	1
fruit n	1	0	0	1
funny ad	1	0	0	1
game n	1	0	0	1

Word	In NLS list 1 or 2?			Total
	0	**1**	**2**	
garden n	0	0	1	1
get v	0	1	0	1
giraffe n	1	0	0	1
girl n	0	1	0	1
give v	1	0	0	1
glasses n	1	0	0	1
go v	1	0	0	1
goat n	1	0	0	1
good adj	0	1	0	1
goodbye	1	0	0	1
grandfather	1	0	0	1
grandma n	1	0	0	1
grandmother	1	0	0	1
grandpa	1	0	0	1
grape n	1	0	0	1
gray adj	1	0	0	1
great ad	0	0	1	1
green ad	0	1	0	1
grey adj	0	1	0	1
guitar n	1	0	0	1
hair n	1	0	0	1
hall n	1	0	0	1
hand n	1	0	0	1
handbag	1	0	0	1
happy ad	0	0	1	1
hat n	1	0	0	1
have v	0	1	0	1
he pron	0	1	0	1
head n	0	0	1	1
helicopter	1	0	0	1
Hello	1	0	0	1
her poss	0	1	0	1
here adv	1	0	0	1
hers pron	1	0	0	1
him pron	0	1	0	1
hippo n	1	0	0	1
his poss	1	0	0	1
hit v	1	0	0	1
hobby n	1	0	0	1
hockey n	1	0	0	1
hold v	1	0	0	1
horse n	1	0	0	1
house n	0	1	0	1
how int	0	1	0	1
how many	1	0	0	1
how old	0	1	0	1
I pron	1	0	0	1
ice cream	1	0	0	1
in prep	0	1	0	1
it pron	0	1	0	1
its poss	1	0	0	1
jacket n	1	0	0	1
jeans n	1	0	0	1
Jill	1	0	0	1
juice n	1	0	0	1
jump v	0	1	0	1
kick v	1	0	0	1
Kim n	1	0	0	1
kitchen	1	0	0	1
kite n	1	0	0	1
know v	0	0	1	1
lamp n	1	0	0	1
learn v	1	0	0	1
leg n	1	0	0	1
lemon n	1	0	0	1
lemonade	1	0	0	1
lesson n	1	0	0	1
let's v	1	0	0	1
letter n	1	0	0	1
like pre	1	0	0	1
lime n	1	0	0	1
line n	1	0	0	1
listen v	1	0	0	1
live v	0	1	0	1
living r	1	0	0	1
lizard n	1	0	0	1
long adj	1	0	0	1

Table 3: Vocabulary list analysis (continued)

Word	In NLS list 1 or 2?			Total
	0	**1**	**2**	
look v	0	1	0	1
lorry n	1	0	0	1
lots adv	1	0	0	1
love v	0	1	0	1
Lucy n	1	0	0	1
lunch n	1	0	0	1
make v	0	1	0	1
man n	0	1	0	1
mango n	1	0	0	1
many det	0	1	0	1
mat n	1	0	0	1
May (gir	1	0	0	1
me pron	0	1	0	1
meat n	1	0	0	1
men n	0	1	0	1
mice n	1	0	0	1
milk n	1	0	0	1
mine pro	1	0	0	1
mirror n	1	0	0	1
Miss title	1	0	0	1
monkey n	1	0	0	1
monster n	1	0	0	1
morning n	0	0	1	1
mother n	0	0	1	1
motorbike	1	0	0	1
mouse n	1	0	0	1
mouth n	1	0	0	1
Mr title	1	0	0	1
Mrs title	1	0	0	1
mum	0	1	0	1
mummy n	1	0	0	1
my poss	0	1	0	1
name n	0	1	0	1
new adj	0	1	0	1
next to	1	0	0	1
nice adj	1	0	0	1
Nick n	1	0	0	1
night n	1	0	0	1
nine numeral	0	1	0	1
nineteen numeral	1	0	0	1
no	0	1	0	1
nose	1	0	0	1
not	0	1	0	1
now adv	0	1	0	1
number n	0	0	1	1
of prep	0	1	0	1
oh dear excl	1	0	0	1
oh dis	1	0	0	1
OK dis +	1	0	0	1
old adj	0	1	0	1
on prep	0	1	0	1
one numeral	0	1	0	1
onion n	1	0	0	1
open adj	0	0	1	1
or conj	0	1	0	1
orange a	0	1	0	1
our poss	0	1	0	1
ours pro	1	0	0	1
page n	1	0	0	1
paint n	1	0	0	1
painting n	1	0	0	1
pardon i	1	0	0	1
park n	1	0	0	1
part n	1	0	0	1
Pat n	1	0	0	1
pea n	1	0	0	1
pear n	1	0	0	1
pen n	1	0	0	1
pencil n	1	0	0	1
people n	1	0	0	1
person n	1	0	0	1
phone n	1	0	0	1
photo n	1	0	0	1
piano n	1	0	0	1
pick up v	1	0	0	1
picture	1	0	0	1
pineapple n	1	0	0	1

Word	In NLS list 1 or 2?			Total
	0	**1**	**2**	
pink adj	1	0	0	1
plane n	1	0	0	1
play v	0	1	0	1
playground	1	0	0	1
please d	1	0	0	1
point v	1	0	0	1
potato n	1	0	0	1
purple a	0	1	0	1
put v	0	1	0	1
question	1	0	0	1
radio n	1	0	0	1
read v	1	0	0	1
red adj	0	1	0	1
rice n	1	0	0	1
ride v	1	0	0	1
right ad	0	0	1	1
robot n	1	0	0	1
room n	1	0	0	1
rubber n	1	0	0	1
ruler n	1	0	0	1
run v	1	0	0	1
sad adj	1	0	0	1
Sam n	1	0	0	1
sand n	1	0	0	1
sausage	1	0	0	1
say v	1	0	0	1
school n	1	0	0	1
sea n	1	0	0	1
see v	0	1	0	1
sentence n	1	0	0	1
seven numeral	0	1	0	1
seventeen numeral	1	0	0	1
she pron	0	1	0	1
sheep n	1	0	0	1
shell n	1	0	0	1
shirt n	1	0	0	1
shoe n	1	0	0	1
shop n	1	0	0	1
short adj	1	0	0	1
show v	1	0	0	1
sing v	1	0	0	1
sister n	0	1	0	1
sit v	1	0	0	1
six numeral	0	1	0	1
sixteen numeral	0	1	0	1
skirt n	1	0	0	1
sleep v	1	0	0	1
small ad	0	0	1	1
smile v	1	0	0	1
snake n	1	0	0	1
so dis	0	1	0	1
soccer n	1	0	0	1
sock n	1	0	0	1
sofa n	1	0	0	1
some det	0	1	0	1
song n	1	0	0	1
sorry adj	1	0	0	1
spell v	1	0	0	1
spider n	1	0	0	1
sport n	1	0	0	1
stand v	1	0	0	1
start v	0	0	1	1
stop v	0	0	1	1
store n	1	0	0	1
story n	1	0	0	1
street n	1	0	0	1
Sue n	1	0	0	1
sun n	1	0	0	1
supper n	1	0	0	1
swim v	0	0	1	1
T shirt	1	0	0	1
table n	1	0	0	1
table te	1	0	0	1
tail n	1	0	0	1
take v	1	0	0	1
talk v	1	0	0	1
teacher	1	0	0	1

Word	In NLS list 1 or 2?			Total
	0	**1**	**2**	
television	1	0	0	1
tell v	1	0	0	1
ten numeral	0	1	0	1
tennis n	1	0	0	1
test n +	1	0	0	1
Thank you	1	0	0	1
thanks d	1	0	0	1
that det	0	1	0	1
the det	0	1	0	1
their poss	1	0	0	1
theirs pron	1	0	0	1
them pron	0	1	0	1
then adv	0	1	0	1
there adv	0	1	0	1
these det	1	0	0	1
they pro	0	1	0	1
thirteen numeral	1	0	0	1
this det	0	1	0	1
those det	1	0	0	1
three numeral	0	0	1	1
throw v	1	0	0	1
tick n +	1	0	0	1
tiger n	1	0	0	1
to prep	0	1	0	1
today adv	0	0	1	1
Tom n	1	0	0	1
tomato n	1	0	0	1
Tony n	1	0	0	1
too adv	1	0	0	1
toy n	1	0	0	1
train n	1	0	0	1
tree n	0	1	0	1
trousers	1	0	0	1
truck n	1	0	0	1
try n +	1	0	0	1
TV n	1	0	0	1
twelve numeral	0	1	0	1
twenty numeral	0	1	0	1
two numeral	0	1	0	1
ugly adj	1	0	0	1
under prep	0	0	1	1
understand v	1	0	0	1
us pron	0	1	0	1
very adv	0	1	0	1
walk v	0	0	1	1
wall n	1	0	0	1
want v	0	1	0	1
watch n + v	1	0	0	1
water n	0	1	0	1
watermelon	1	0	0	1
wave v	1	0	0	1
we pron	0	1	0	1
wear v	1	0	0	1
well dis	1	0	0	1
well done excl	1	0	0	1
what int	0	1	0	1
where in	0	1	0	1
which in	1	0	0	1
white adj	1	0	0	1
who int	0	1	0	1
whose in	1	0	0	1
window n	0	0	1	1
with pre	0	1	0	1
woman n	1	0	0	1
women n	1	0	0	1
word n	0	0	1	1
Wow excl	1	0	0	1
write v	0	0	1	1
yellow a	0	1	0	1
yes	0	1	0	1
you pron	0	1	0	1
young adj	0	0	1	1
your pos	0	1	0	1
yours pron	1	0	0	1
zoo n	1	0	0	1
Total	**322**	**107**	**31**	**460**

colour words. Main differences concern the fact that a number of YL class management and topic content words such as 'book' and 'teacher', as discussed above, are not on National Literacy Strategy List 1. On the other hand, words like 'laugh' which are likely to appear in the story books devoured by young native speaker readers in class time and private reading time appear on List 1 but do not appear in Starters. In addition, there are past tense forms in the National Literacy Strategy list but none on the Starters list because the grammar syllabus does not include them at this level. Taking the 121 words of National Literacy Strategy List 2 into consideration added 31 more words to the overlap. Taking List 2 and List 1 together, we have 138 words from this source represented in the Starters vocabulary syllabus. That is 30% of the Starters vocabulary list. Detailed results can be seen in Table 3 where the table shows if a Starters word appears in List 1, List 2 or not at all. It will be interesting to see how the Movers and Flyers lists match up with the National Literacy Strategy lists.

Conclusion

It is in the positive washback effects on teaching of analysing this aspect of the linguistic components of the test that I am chiefly interested. Looking at possible sources of difficulty for children in the grapheme/phoneme differences of certain words has been illuminating to me and I hope that it goes some way towards helping teachers to empathise with conceptual difficulties that children may face but that they themselves have long overcome and forgotten. Grouping words with similar rimes and vowel sounds into short verses and chants seems practical and could lead to more enjoyable child-friendly practice. For example, the Starters list offers us groups such as: 'cat', 'hat', 'mat', 'Pat'; 'bed', 'red'; 'dog', 'frog'; and 'Ben', 'men', 'pen' and 'ten' . Working with additional visual means such

as special flash-cards (with 'red for danger' borders?) with the 'tricky' less transparent words may be another idea to help young test takers perform better in spelling – an area of English that it is recognised is easy for no-one, whether native user or language learner.

References and further reading

Department for Education and Employment (1998) *The National Literacy Strategy for Teaching*, Sudbury: DfEE publications.

Cook, V (1997) *Orthographic Processes in L2 Users*, unpublished paper: http://homepage.ntlworld.com.vivian.c/Writings/Papers/OrthProcesses..htm

Elley, W B (1969) The Assessment of Readability by Noun Frequency Counts, *Reading Research Quarterly*, 4 (3), 411–427.

Goswami, U and Bryant, P E (1990) *Phonological Skills and Learning to Read*, East Sussex: Psychology Press Ltd.

Goswami, U and East, M (2000) Rhyme and Analogy in Beginning Reading: Conceptual and Methodological Issues, *Applied Psycholinguistics* 21 (1), 63–93.

Katz, L and Frost, R (1992) The reading process is different for different orthographies: the orthographic depth hypothesis, in Frost, R and Katz, L (Eds) *Orthography, Phonology, Morphology, and Meaning*, Amsterdam: Elsevier North Holland Press, 67–84.

Raban, B (1988) *The Spoken Vocabulary of Five-Year Old Children*, Reading: The Reading and Language Information Centre, School of Education, University of Reading.

Rixon, S (1999) Where do the words in EYL Textbooks come from? in Rixon, S (Ed.) *Young Learners of English: Some Research Perspectives*, Harlow: Pearson, 55–71.

Rixon, S (2007) *Forward to Basics? YL teachers' practices in early reading instruction*, paper presented at the Young Learners Special Interest Group Pre-Conference Event, IATEFL Annual Conference, Aberdeen, 18 April 2007

Spencer, K A (1999) Predicting word-spelling difficulty in 7- to 11-year olds, *Journal of Research in Reading*, 22 (3), 283–292.

Wells, G (1985) *Language Development in the Pre-School Years*, Cambridge: Cambridge University Press.

Linking language assessments for younger learners across proficiency levels (Phase 1)

FIONA BARKER AND STUART SHAW RESEARCH AND VALIDATION GROUP

Background

Cambridge ESOL offers a suite of language tests designed for young learners of English between the ages of 7 and 12 called the Cambridge Young Learners English Tests (YLE). The tests cover all four language skills – Reading and Writing (combined), Listening, and Speaking. Three key levels of assessment – Starters, Movers and Flyers – provide a 'ladder' to support and encourage progression in language learning steadily up the proficiency continuum.

Levels-based exams are particularly attractive in the context of language study within an educational setting.

The exam can select material and elicit performance which is appropriate to the level, and which can impact positively on the learning process. However, this can also raise issues for the vertical comparison of lower with higher levels. In the case of the YLE, for example, the setting of different kinds of task, with different degrees of challenge, may complicate the comparison of performance across levels (although it is true to say that some of the YLE tasks are the same or similar across levels).

In validating the YLE tests it is important to verify that horizontally each language skill is tested at a similar level,

and that vertically the three levels are related systematically. This article reports on a small-scale study that sought to vertically equate the three levels of YLE on a common scale, thereby providing empirical validation for the vertical equating of levels. It was envisioned that a rational progression across levels would be discerned.

Design of the study

In total, 162 candidates (53 Starters, 75 Movers, 34 Flyers candidates) from six participating centres across all three YLE levels took a set of 33 common Movers items as part of one of two hybrid tests (Version A and Version B). Test Version A comprised Starters/Movers Listening and Reading/Writing tasks and Test Version B comprised Movers/Flyers Listening and Reading/Writing tasks, with the common Movers tasks used to form a crossover design (shown in Figure 1). Items for the hybrid tests were taken from a retired version of YLE.

Figure 1: Hybrid test cross-over design
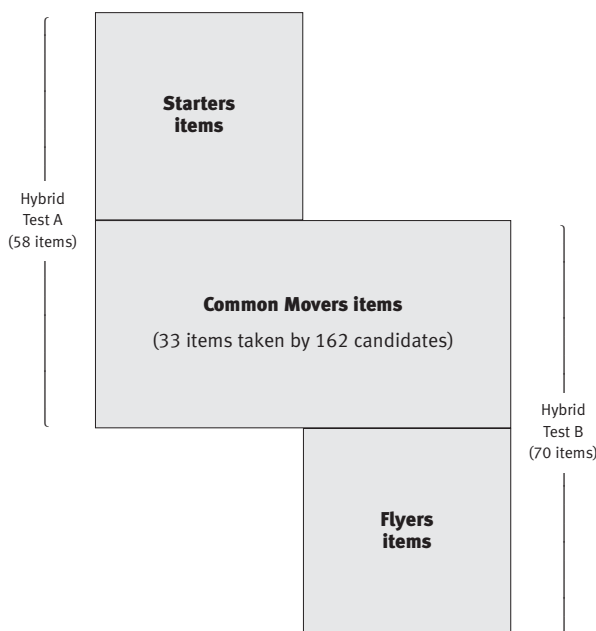
In accordance with the design matrix, no Flyers candidates took Starters items and no Starters candidates took Flyers items.

Both hybrid tests were marked according to a specially constructed mark scheme based on the three mark schemes employed for the retired version. The trial data consisted of candidate response data: two text files per centre (Test A and Test B). In addition to candidate response data, two class registers were developed to record the name, age, gender and approximate YLE level as evaluated by centre teachers or actual YLE level (weak/medium/strong at Starters/Medium/Flyers level) or shield already awarded at a particular level for each candidate who sat the tests. It was necessary, however, to extract certain candidate response strings from data sets as information pertaining to candidate level was incomplete or ambiguous (i.e. the class

teacher had not indicated a level for a candidate, or had filled in both approximate level and a shield awarded which did not match up). The final data sets, therefore, comprised 77 candidates (Version A) and 93 candidates (Version B).

The proportion of levels of candidates taking Version A is given in Figure 2. Of the 77 candidates, 52 (68%) were at the Starters Level and 25 (32%) were at the Movers Level.

Figure 2: Version A spread of levels
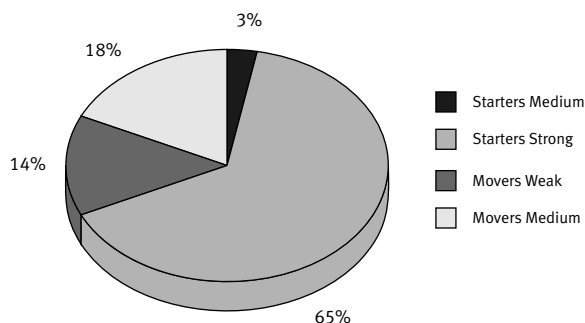
The proportion of levels of candidates taking Version B is given in Figure 3. In total, 93 candidates took Version B: 59 (64%) at Movers Level, 34 (36%) at Flyers Level.

Figure 3: Version B spread of levels
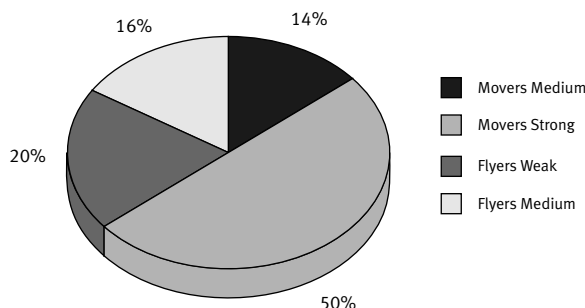
The 34 Flyers candidates that comprise this data set are, on the basis of candidate registers completed by centre staff, predominantly weak: 19 'weak' candidates (56%) compared to 15 'medium' candidates (44%). Close scrutiny of the candidate registers indicates a complete absence of strong Flyers candidates. Further proof that the Flyers candidates are weak is garnered from comparison with historical data of all candidate performance on this version in live administrations.

Findings from the study

Two analyses were undertaken on the candidate response data using a Rasch modelling methodology. One analysis involved identifying candidates by proficiency level (based on teacher evaluations). The other entailed analysis by item. Whilst both methodological approaches yielded a similar picture, only the findings from the item analyses are presented here.

Rasch modelling indicated no real evidence of unexpected candidate performance at any level in this data,

i.e. candidates were performing well within the parameters of the Rasch model and there were no anomalies. Similarly, there was no evidence of unexpected item performance at any level. Item bias analysis indicated no significant bias for any group on any item, i.e. all three groups of candidates performed well on familiar items and again were well within the parameters of the Rasch model.

Table 1 shows item facilities (the degree of difficulty of the test items calculated on the basis of the group's performance) for the sets of items at each level. The pattern which emerges from these figures for candidates at all levels is strong performance on familiar items. A logical progression in facility with increasing level of proficiency is observed: Flyers candidates on Flyers items (0.75); Movers candidates on Movers items (0.78); and Starters candidates on Starters items (0.85).

Table 1: Item facilities

Candidates	Items		
	F	M	S
F	**0.75**	0.76	
M	0.75	**0.78**	0.84
S		0.67	**0.85**

The trial seems to suggest evidence of a 'familiarity' effect, that is, candidates anticipating both item content and level of difficulty at all three levels. Familiarity with YLE test items has been enhanced through candidate coursebooks and practice materials, past papers and examination reports, an on-line resource for teachers provided by Cambridge ESOL and a wide range of seminars for teachers.

Flyers candidates perform as expected on both Flyers and Movers items. As expected, Starters candidates perform well on Starters level items but less well on the more challenging Movers level items. Movers candidates perform well on Starters items; perform as well as their Flyers counterparts on Flyers items (this could be symptomatic of the number of candidates and the distribution of their proficiencies within each level); and out-perform Flyers on the Movers items.

Rasch analyses reflect item facilities. The Rasch model used here is a strict mathematical expression of the theoretical relation that would hold between all items and all persons along a single underlying continuum; no items and/or persons will ever perfectly fit the model. Of interest is the identification of those items and/or persons whose pattern of responses deviate more than expected. Because there are too many deviations (or residuals) from the model, the fit diagnosis (i.e. model fit) typically is summarised in a fit statistic, expressed either as a mean square fit statistic or a standardised fit statistic (usually a z or t distribution). In the tables which follow, the Mean Square (Msq) statistic has been selected. The infit Msq shown in Tables 2–4 tends to place more emphasis on unexpected responses near a person's or item's measure (infit statistics) as opposed to those unexpected responses (outliers) far from a person's or an item's measure

(commonly referred to as outfit statistics). Table 2 shows the infit mean square statistic for persons, and Table 3 the infit mean square statistic for items.

Table 2: Person infit Msq statistic

Candidates	Items		
	F	M	S
F	1.02	1.02	
M	0.97	0.97	1.05
S		1	1

Infit in Table 2 is a measure of how a group of *persons* perform against that predicted by the Rasch model. Any deviations are from the ideal scenario of '1'. McNamara (1996:173) suggests a range of approximately 0.75–1.3 as being an acceptable general rule. Observed values greater than 1 would suggest greater variation than might be expected, whereas values less than 1 would demonstrate less than expected variation. There is no evidence of unexpected candidate performance at any level in this data, i.e. candidates are performing well within the parameters of the Rasch model and there are no anomalies.

Table 3: Item infit Msq statistic

Candidates	Items		
	F	M	S
F	0.97	0.99	
M	0.97	0.99	1.06
S		0.99	1.06

Infit in Table 3 is a measure of how a group of *items* perform against that predicted by the Rasch model. Again, the values are within the tolerable range.

Table 4 shows the bias statistics for the items taken by 'group'. There is no evidence of significant bias for any group on any item: all three groups of candidates perform well on familiar items (corroborated by facilities in Table 1).

Table 4: Bias statistics

Candidates	Items		
	F	M	S
F	0.1	-0.13	
M	-0.07	0.1	-0.26
S		-0.05	0.11

Starters candidates are performing well on familiar Starters items and, not unexpectedly, very slightly underperforming on Movers items. Centre evidence (teacher approximations of candidates' levels) supports the notion that this Starters group is strong, hence their good performance on those items for which they are prepared. Whilst Movers candidates exemplify the full range of ability (strong, medium, weak), there is a tendency to

under-perform on both Flyers and Starters items in this data set. Given the preponderance of 'strong' Movers in the design this might be considered a somewhat surprising finding. The pattern which emerges for Flyers is not dissimilar to that for the Starters i.e. strong performance on familiar items and underperformance on 'unfamiliar' items.

Vertical progression

In terms of vertical progression between levels on Listening tasks, the 'distance' between the Starters Level and the Movers/Flyers Levels is significant across all tasks. Out of the six common parts which all candidates encountered in the hybrid tests, only in Part 3 is there clear evidence of expected, logical vertical progression through the levels.

Interestingly, in Parts 2 and 5 (Table 5), Movers candidates outperform Flyers candidates, probably due to the skill focus of the two parts being identical and the Movers being a potentially stronger group than the Flyers. In terms of main skill focus the equivalent Part 2 Flyers task is identical: recording key information on the basis of a gapped input text and dialogue is a shared skill across the upper two levels. Likewise, the skills required for Part 5 (listening for specified information) based on an input picture and dialogue is identical across both Movers and Flyers levels.

Table 5: Common Movers Listening tasks

Test Part	2	3	5
Main Skill Focus	Listening for specified information	Listening for lexical items and verb phrases (past tense)	Listening for specified information
Input	Gapped text and dialogue	Pictures and days of the week and dialogue	Picture and and dialogue
Expected Response/ Item Type	Record words or numbers	Match days of week to pictures by copying name of day	Colour and draw or write
Starters Facility	0.555	0.823	0.698
Movers Facility	0.685	0.939	0.837
Flyers Facility	0.624	0.976	0.782
Vertical Progression (relative distance)	M F S	F M S	M F S

For the Reading and Writing tasks, again there is clear 'distance' between the Starters Level and the Movers/Flyers Levels. Part 7 is the only part where all three groups of candidates perform equally well, probably due to the nature of the

of the expected candidate response (i.e. yes/no). The remaining Reading and Writing parts (Parts 2, 5, 9, 11) all demonstrate evidence of Movers candidates outperforming their Flyers counterparts.

In Parts 9 and 11 (shown in Table 6) Movers candidates outperform the Flyers despite the fact that the skills required to successfully complete Part 9 are identical, i.e. the principal skills focus is to complete a gapped text with one word (noun or verb).

Table 6: Common Movers Reading/Writing tasks

Test Part	7	9	11
Main Skill Focus	Understanding short texts	Completing a gapped text with one word (noun or verb)	Completing a gapped text with one word (1 word from 3)
Input	Picture and short text	Gapped text with picture cues	Gapped text and word sets
Expected Response/ Item Type	Write 'yes' or 'no' next to the texts	Write words in gaps	Complete text by selecting the best word and copying
Starters Facility	0.929	0.563	0.472
Movers Facility	0.940	0.701	0.557
Flyers Facility	0.951	0.685	0.488
Vertical Progression (relative distance)	F M S	M F S	M F S

Out of the six common parts, only in Part 3 (a Listening task) is there clear evidence of expected vertical progression up the levels (Figure 4). Only in Part 7 (a Reading and Writing task) is there an indication of all three groups of candidates performing equally well, probably due to the nature of the expected candidate response (i.e. yes/no). The other 4 parts (Parts 2, 5, 9, 11) all demonstrate evidence of Movers candidates outperforming their Flyers counterparts.
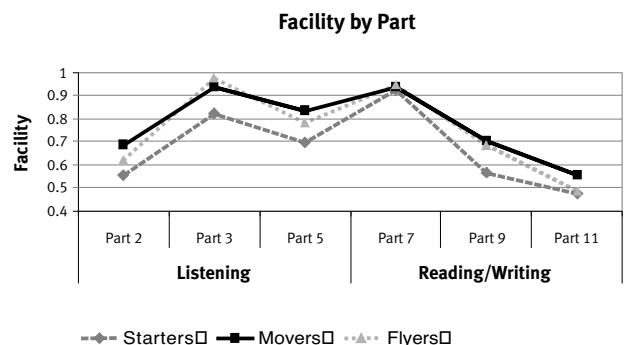

Figure 4: Facility by part across Listening and Reading & Writing tests

Conclusion

The study described in this article sought to vertically equate the three levels of YLE on a common scale thereby providing insights into how candidates at one level perform on tasks unfamiliar to them. Evidence of a 'familiarity effect', where candidates anticipate both item content and level of difficulty at all three levels, was manifest in the data.

Findings from the study provide further support for measurement-based YLE grading activities. Outcomes should inform operational grading of the recently revised YLE tests and could also inform a Young Learners common scale where it could be stated, for example, that 1 shield increase in YLE equates to a certain amount of teaching/learning or that a shield 3 on Flyers is equivalent to a shield 4 on Movers. This could also form the beginnings of work into linking YLE to CEFR at A1/A2 levels.

Considering possible developments for YLE there are issues concerning test design arising from this study. For example, could a common task be included in tests at adjacent levels? This would have face appeal as an empirical basis for vertically linking the YLE levels. Extension of the research described here might entail bias analysis by gender, age and first language on a larger data set. Centres involved in a replication study could also be asked whether they use an ELP (European Language Portfolio) and whether qualitative data on what Young Learners can do from the teachers of these candidates is collated. This would further help inform YLE Can Do statements, a likely research focus for 2007.

References and further reading

McNamara, T (1996) *Measuring Second Language Performance*, Harlow: Addison Wesley Longman Limited.

Developing Can Do statements for young learners

A focus of research activity for 2007 in the context of the YLE tests is the construction of a set of Can Do descriptors for young learners. Can Do statements will provide relevant stakeholders with transparent descriptions of the proficiency of young learners.

Recent research has explored the appropriateness of existing Can Do statements for young learners. The projects have involved collaboration with teachers, and sometimes students, to modify existing CEFR descriptors. For example, in the Bergen Can Do project (Hasselgreen, 2003) descriptors were adapted for young learners (11–12 year-olds). The primary aim of this project was 'to develop a set of materials for systematic ongoing assessment/portfolio assessment in lower secondary school' in a way that 'provides concrete, positive criteria for assessing' young learners (Hasselgreen, 2003:9). The Can Do statements developed have been used as a self-assessment tool for learners of English in eight countries in the Nordic-Baltic region. The University of Fribourg has employed a more empirical approach to validate Can Do statements for young learners at levels A1 to B2 (www.unifr.ch/cerle/). The Cambridge ESOL ALTE group has also contributed qualitative expertise by investigating the progression of young learners at Breakthrough (A1) level.

The development of these descriptors raises a number of issues which can be conceived of as a series of questions relating to the cognitive development of young learners. These include:

- Is the route of proficiency progression for young learners the same as for their adult counterparts?
- Are the cognitive demands of B2 and above too challenging for young learners?
- Is the wording of existing statements comprehensible to young learners?
- Can the wording of these statements be modified in such a way as to retain the salient features, in terms of progression, of the original?

It is envisioned that as part of the development of Can Dos for young learners these, and other issues, will be addressed.

References and further reading

Hasselgreen, A (2003) *Bergen 'Can Do' Project*, Strasbourg: Council of Europe.

Université de Fribourg (2007) Centre d'Enseignement et de Recherche en Langues Etrangères (CERLE): www.unifr.ch/cerle/

YLE update

In order to support candidates preparing for the revised YLE tests and their teachers, a range of support materials were developed for use from January 2007. These can be used to learn more about the revised YLE tests and to prepare children to take them.

What's new on the YLE website

In the run-up to the launch of the revised tests, sample papers showing revised tasks and a handbook describing the syllabus at all 3 levels of the YLE tests were produced, as well as a summary of the revised test specifications. In addition to this, information leaflets for candidates were written to help familiarise candidates, parents and teachers with the revised tasks. The leaflets give a simple description of the tasks at each level and show example questions. All these materials can be accessed by clicking on the downloads link on www.cambridgeesol.org/exams/yle.htm

The updated awards

The YLE awards were redesigned to give them a fresher look whilst retaining all the key information about candidates and the shields they achieve. On the reverse of the award, a table now shows how the different levels of the YLE tests relate to the Council of Europe's Common European Framework of Reference for Languages at A1 and A2 levels and some information about what particular shield scores mean.

Teacher seminars

Three new seminars, *Introduction to the YLE update, Introduction to YLE* and *Understanding YLE*, were developed to prepare teachers for the introduction of the revised YLE

tests. The last of these gives a detailed introduction to the tests with samples of tasks and classroom activities. Two seminars are currently being updated: *Get ready for YLE*, which focuses on classroom activities and gives test tips, and *Speaking Skills for YLE*, which looks in detail at tasks, candidate performance and classroom preparation ideas. Finally, there are plans to complete the range of seminars available with two new seminars: *Listening Skills for YLE* and *Reading & Writing Skills for YLE*.

Teaching Resources website

Revisions to the YLE pages of the Teaching Resources website went live in December 2006. As well as describing the changes to the revised tasks, new classroom activities for teachers were added and continue to grow. The YLE picture bank, which contains colour pictures that teachers can download and print off, was expanded and now includes pictures of new words on the vocabulary lists, such as 'grapes' and 'kitten'. The YLE home page is at www.cambridgeesol.org/teach/yle2007/index.htm

Sample papers

Sample papers containing examples of revised tasks and language from the revised vocabulary and structure lists were made available from summer 2006. A new feature of the sample papers is the inclusion of an audio CD which contains a full Listening test at each of the 3 levels of YLE.

Handbook

A revised handbook, containing descriptions of the revised tasks and the revised vocabulary and structure lists was made available in summer 2006 in order to prepare teachers and candidates who will take the revised tests.

Research and development update

Reviewing FCE and CAE

Cambridge ESOL is committed to keeping its examinations up to date with best practice in language assessment, with technological developments and with the changing needs of teachers and learners. To this end, we regularly carry out a formal review of all the examinations we produce which includes extensive consultation with teachers, candidates, and exam centres, as well as expert input from Cambridge ESOL's subject specialists and research teams, external consultants and senior examiners. Such a review often leads to changes to the test being explored using a well-established project-based methodology which is fully described in the account of the CPE Revision Project (Weir

and Milanovic, 2003). Since it was first published in 2000, *Research Notes* has been an important communication channel for reporting on our research relating to changes in CPE, KET and PET, BEC and BULATS, CELS, YLE, as well as the IELTS Speaking and Writing tests.

Our latest review has focused on the First Certificate in English (FCE) and the Certificate in Advanced English (CAE). The process began in 2004 with the aim of looking at several key questions in relation to these exams, particularly the potential to reduce the overall length of the exams and make them more accessible to candidates without making them less challenging or rigorous. We also wanted to review the individual test components within FCE

and CAE to ensure that task types and their test focus continue to reflect the needs of candidates, that we are making the most effective use of resources such as corpora and new technology, and that we are keeping up to date with changes in teaching methodology.

Early phases of the review involved a series of face-to-face meetings and survey questionnaires with a range of stakeholders: 1,900 candidates, about 100 local secretaries, and over 600 teachers, directors of studies, examiners and school owners. Reports reviewing current issues to do with writing and marking the tasks were also commissioned from external consultants and principal examiners.

Following this initial consultation stage, extensive trialling of new and modified FCE and CAE tasks was undertaken with representative populations at schools in 30 countries. Detailed research was conducted into the likely effect of any changes on candidates or their performance and proposed changes were extensively discussed with teachers and other stakeholders at meetings worldwide. April 2006 saw a further round of consultation via an invitational meeting held in Cambridge for representatives of ELT organisations such as publishers, the British Council, English UK, test centre personnel and others. A series of international seminars was also held. As each cycle of trialling and consultation was completed throughout 2005–2006, the results were analysed and compiled by the Research and Validation Group in collaboration with colleagues in the Assessment and Operations Group as well as external language teaching/testing specialists.

A final decision on changes to the tests was ratified in the second half of 2006 with a view to introducing these from the December 2008 test administration. Accordingly, new test specifications for FCE and CAE were introduced in December 2006, allowing a full 2 years for publishers, schools and teachers to prepare for the updated exams. Key changes to the two tests can be summarised as follows:

- the tests will focus on a wider range of skills and functions to improve content and domain coverage
- the overall length of the FCE and CAE exams will be reduced by approximately one hour to make them more accessible and practical for users
- some components of the exams will include new and/or improved task types to reflect changes in language and contexts of use
- some components will be more standardised in content, making them user-friendly for students preparing for the exams.

Since May 2005, Cambridge ESOL has issued regular bulletins to provide updates on progress. These report in more detail the activities and outcomes of the review and consultation process; together with an overview of the full FCE and CAE December 2008 specifications, they are available at: www.cambridgeesol.org/exams/fce.htm and www.cambridgeesol.org/exams/cae.htm

Recent issues of *Research Notes* have also reported briefly on some of the research conducted as part of the process of reviewing FCE and CAE. From now on we hope to write up and report more of this work in *Research Notes* as part of the lead-up to the introduction of changes to FCE and CAE in December 2008.

References and further reading

Weir, C J and Milanovic, M (2003) (Eds) *Continuity and Innovation: Revising the Cambridge Proficiency in English examination 1913–2002* (Studies in Language Testing, volume 15), Cambridge: UCLES/Cambridge University Press.

Recent publications of interest

Studies in Language Testing

March 2007 saw the publication of another title in the *Studies in Language Testing* series, published jointly by Cambridge ESOL and Cambridge University Press. Volume 19 in the series is edited by Lynda Taylor and Peter Falvey and is entitled *IELTS Collected Papers: Research in speaking and writing assessment*. The book brings together a collection of research studies conducted between 1995 and 2001 under the auspices of the British Council/IELTS Australia Joint-funded Research Program which promotes independent research among IELTS stakeholders worldwide. The ten studies – four on speaking and six on writing assessment – provide valuable test validity evidence for IELTS and directly informed the revised IELTS Speaking and Writing tests introduced in 2001 and 2005. As well as the research studies, the volume contains additional chapters and sections intended to achieve various aims: chronicle the evolution of the Writing and Speaking tests in ELTS/IELTS from 1980 to the present day; explain the rationale for revisions to the IELTS Speaking test (2001) and the IELTS Writing test (2005), and the role played by research findings; discuss and evaluate a variety of research methods to provide helpful guidance for novice and less experienced researchers. This collection of studies will be of particular value to language testing researchers interested in IELTS as well as to institutions and professional bodies who make use of IELTS test scores; it is also relevant to students, lecturers and researchers working more broadly in the field of English for Academic Purposes. More information is available at: www.cambridgeesol.org/research/silt.htm

IELTS Research Reports, Volume 6

Between 1998 and 2003 IELTS Australia published five volumes of IELTS Research Reports covering projects

completed under the joint-funded research program. *IELTS Research Reports*, Volume 6 – published towards the end of 2006 – is the first of two volumes to be published jointly by IELTS Australia and British Council. It contains seven reports of research studies focusing on the IELTS Speaking Test conducted between 2002 and 2004 by applied linguists and language testers in Australia, New Zealand, the UK, and Denmark. Volume 6 is available in both hard copy and CD-ROM versions. For more information, visit the IELTS website: www.ielts.org

Publications by ESOL research staff

In 2003–4 UCLES, in collaboration with the National Foundation for Educational Research (NFER) and the Faculty of Education at Nottingham University, organised a series of seminars to discuss issues in reading and assessment. Colleagues from Cambridge ESOL, from OCR and from the Research Division joined with fellow reading and assessment specialists from the UK, France and the USA to participate in the seminars and later contributed chapters to a book based on the series. *Assessing Reading: from theories to classrooms* was edited by Marian Sainsbury, Colin Harrison and Andrew Watts and was published by NFER in 2006. The book explores the theories, practices and conflicts that surround the idea of reading at the start of the 21st century and includes a chapter on cognitive psychology and reading assessment by Alastair Pollitt and Lynda Taylor. For more information go to: www.nfer.ac.uk/bookshop.

Cambridge Assessment Network followed up a seminar held in Cambridge in September 2006 by publishing a special supplement *Excellence in Assessment: Assessment for Learning*. This included an article by Neil Jones – 'Assessment for Learning: the challenge for an examination board' – looking at ways in which assessment expertise might contribute to supporting learning. The supplement is available on-line at: www.cambridgeassessment.org.uk/research/confproceedingsetc/ExcellenceInAss1/file

We often receive email requests to suggest useful introductory texts to the field of language testing and assessment. An invited book review by Lynda Taylor was recently published in the *International Journal of Applied Linguistics* (16/3, Nov 2006) and provides a useful overview of two such titles – Tim McNamara's *Language Testing* (2000, OUP) and Arthur Hughes' *Testing for Language Teachers* (2nd ed., 2003, CUP). Both these titles will help anyone working in language education and/or research to gain an understanding of the basic principles and practice of assessment and to participate in the discourse of our field.

Other publications

As fellow departments within the same university, Cambridge ESOL and Cambridge University Press continue to collaborate in producing a wide variety of English language materials for teachers and learners. These include the recent publication of *The CELTA Course* by Scott Thornbury and Peter Watkins; this new course book is designed to support trainers and trainees preparing candidates for Cambridge ESOL's Certificate in English Language Teaching for Adults (CELTA). The popular series *Common Mistakes… and how to avoid them* has recently been completed with the addition of *Common mistakes at KET* and *Common mistakes at IELTS*. All the titles in this series draw on evidence from the Cambridge Learner Corpus to show where students regularly trip up and what they can do to help themselves. Still on the subject of IELTS, the *Cambridge Grammar for IELTS* provides a useful self-study book to support students in their preparation for the test. A new advanced level of Business Benchmark has also recently appeared offering a course for general Business English training as well as exam preparation for BEC and BULATS. For more information on all these new titles see the Cambridge English Language Teaching catalogue for 2007, which is available online at: www.cambridge.org/uk/catalogue

Finally, a new report from think tank Demos on the impact of English language varieties has drawn on Cambridge ESOL expertise and support. *As You Like It: Catching up in an age of Global English* calls for radical changes in Government policy in response to the challenges of the rise of Global English and the lack of language skills in the UK. The report illustrates the issues by citing examples from Cambridge ESOL projects such as the Beijing Speaks English campaign (part of the city's preparation for the 2008 Olympics) and the Asset Languages initiative. A full pdf copy of the report can be downloaded from www.demos.co.uk

Conference reports

BALEAP Professional Issues Meeting, February 2007

The recent BALEAP PIM was hosted by the University of Salford and was on testing and assessment. The plenary session given by Professor Barry O'Sullivan (Roehampton University) was on the validation of in-house testing. He began with a cautionary tale of the Chinese Imperial Examination System, describing a situation where the test 'ate' the system: the test came to dominate the education system to such an extent that the focus of the curriculum became solely passing the test. A brief history of EAP testing was then outlined, describing the move within the UK towards tests reflecting the needs of the learner. Test validity was discussed in the light of a socio-cognitive framework, highlighting the importance of collecting validity evidence from all aspects of the framework rather than from one or two selected areas. Finally the idea of an integrated

learning system was introduced in which the elements of curriculum, assessment and delivery all interact with each other. It was stressed that the assessment system needs to be designed into the learning system and should include as many assessment perspectives as possible, for example summative, formative, teacher, self and peer assessment.

A series of informative parallel sessions given by delegates from a number of universities followed. Topics included pre-sessional testing, online testing using a Virtual Learning Environment (VLE), concurrent formative and summative assessment, diagnostic testing of writing and the training and standardisation of marking. The day finished with an information and discussion session on the BALEAP Can Do Statements project, the development of a set of Can Do statements for use with pre-sessional courses. A brief overview of the project, preliminary findings from the literature review and intended methodology was given. This was followed by a discussion of the goals and desired outcomes of the project.

Advance notice: Association of Language Testers in Europe (ALTE) 3rd International Conference, Cambridge 2008

The ALTE 3rd International Conference will be held from 10–12 April 2008 at the University of Cambridge and hosted by Cambridge ESOL. The theme of this multilingual conference will be *The Social and Educational Impact of Language Assessment,* and papers are invited on the topics of *Assessment for Teaching and Learning, Intercultural Dialogue,* and *Impact and Stakeholders*. ALTE Cambridge 2008 is a unique event of unparalleled scope, range, depth and reach for all professionals with an interest in language assessment and associated issues. For further details visit www.alte.org/2008

IELTS Joint-funded Research Program

As part of their ongoing commitment to IELTS-related validation and research, IDP: IELTS Australia and the British Council are once again making available funding for research projects in 2007/8.

For several years now the partners have issued a joint call for research proposals that reflect current concerns and issues relating to the IELTS test in the international context. Such research makes an important contribution to the monitoring and test development process for IELTS; it also helps IELTS stakeholders (e.g. English language professionals and teachers) to develop a greater understanding of the test. All IELTS research is managed by a Joint Research Committee which agrees research priorities and oversees the tendering process. In determining the quality of the proposals and the research carried out, the committee may call on a panel of external reviewers. The committee also oversees the publication and/or presentation of research findings. A full list of joint-funded and published or forthcoming research studies can be found on the following pages.

Details of the call for proposals for Round 13 (2007/8) including application forms and guidance on topics and resources can be found on the IELTS website: www.ielts.org

IELTS Joint-funded and published (or forthcoming) research: Rounds 1 to 12: 1995 to 2006

Rounds and Researchers	Research Title	IELTS publications
Round 1: 1995		
Annie Brown and Kathryn Hill, *The University of Melbourne, Australia*	Interviewer style and candidate performance in the IELTS oral interview	IELTS RR[1] 1 SiLT[2] 19
Clare McDowell and Brent Merrylees, *LTC Language & Testing Consultants Pty Ltd, Australia*	Survey of receiving institutions' use and attitude towards IELTS	IELTS RR 1
Fiona Cotton and Frank Conrow, *University of Tasmania, Australia*	An investigation of the predictive validity of IELTS amongst a sample of international students at the University of Tasmania	IELTS RR 1
James D H Brown, *The University of Melbourne, Australia*	An investigation into approaches to IELTS preparation with a particular focus on the Academic Writing component of the test	IELTS RR 1
Magdalena Mok, Nick Parr, Tony Lee and Elaine Wylie, *Griffith University, Australia*	A comparative study of IELTS and Access Test results	IELTS RR 1
Gayle Coleman and Stephen Heap, *The University of Queensland, Australia*	The misinterpretation of directions for questions in the Academic Reading and Listening sub-tests of the IELTS test	IELTS RR 1
Round 2: 1996		
Kathryn Hill, Neomy Storch and Brian Lynch, *The University of Melbourne, Australia*	A comparison of IELTS and TOEFL as predictors of academic success	IELTS RR 2
Brent Merrylees, *LTC Language and Testing Consultants Pty Ltd, Australia*	A survey of examiner attitudes and behaviour in the IELTS oral interview	IELTS RR 2
Cynthia Celestine and Cheah Su Ming, *IDP Education Australia*	The effect of background disciplines on IELTS scores	IELTS RR 2
Tim Moore, *Monash University, Australia* and Janne Morton, *The University of Melbourne, Australia*	Authenticity in the IELTS Academic Module Writing test: A comparative study of Task 2 items and university assignments	IELTS RR2 SiLT 19
Round 3: 1997		
Annie Brown, *The University of Melbourne, Australia*	An investigation of raters' orientation in awarding scores in the IELTS oral interview	IELTS RR 3 SiLT 19
Clare McDowell, *Australian Research and Testing Services, Australia*	Monitoring IELTS examiner training effectiveness	IELTS RR 3
Carol Gibson and Stephen Slater, *University of South Australia* and Peter Mickan, *The University of Adelaide, Australia*	A study of the response validity of the IELTS Writing Module	IELTS RR 3
Kieran O'Loughlin, *The University of Melbourne, Australia*	An investigation of the role of gender in the IELTS oral interview	IELTS RR 3 SiLT 19
Mary Kerstjens and Caryn Nery, *RMIT University, Australia*	Predictive validity in the IELTS Test	IELTS RR 3
Round 4: 1998		
Barbara Mayor, Ann Hewings and Joan Swann with Caroline Coffin, *The Open University, UK*	A linguistic analysis of Chinese and Greek L1 scripts for IELTS Academic Writing Task 2	SiLT 19
Brent Merrylees, *LTC Language & Testing Consultants Pty Ltd, Australia*	An impact study of two IELTS user groups: candidates who sit the test for immigration purposes and candidates who sit the test for secondary education purposes	IELTS RR 4
Rae Everett and Judy Coleman, *University of New England, Australia*	A critical analysis of selected IELTS preparation materials	IELTS RR 5
Peter Mickan, *The University of Adelaide, Australia* and Stephen Slater, *Heian Jogakuin University, Japan*	Text analysis and the assessment of Academic Writing	IELTS RR 4
Round 5: 1999		
Annie Brown, *The University of Melbourne, Australia*	Legibility and the rating of second language writing: An investigation of the rating of handwritten and word-processed IELTS Task 2 essays	IELTS RR 4
Chris Kennedy and Dilys Thorp, *University of Birmingham, UK*	A corpus-based investigation of linguistic responses to an IELTS Academic Writing task	SiLT 19
John Read, *Victoria University of Wellington* and Belinda Hayes, *Auckland University of Technology, New Zealand*	The impact of IELTS on preparation for academic study in New Zealand	IELTS RR 4
Kieran O'Loughlin, *The University of Melbourne* and Gillian Wigglesworth, *Macquarie University, Australia*	Task design in IELTS Academic Writing Task 1: The effect of quantity and manner of presentation of information on candidate writing	IELTS RR 4 SiLT 19
Clare Furneaux and Mark Rignall, *University of Reading, UK*	The effect of standardisation- training on rater judgements for the IELTS Writing Module	SiLT 19
Round 6: 2000		
Barry O'Sullivan and Mark Rignall, *University of Reading, UK*	Assessing the value of bias analysis feedback to raters for the IELTS Writing Module	SiLT 19
Catherine Elder, *The University of Auckland* and Kieran O'Loughlin, *The University of Melbourne, Australia*	Investigating the relationship between intensive English language study and band score gain on IELTS	IELTS RR 4
Chris Kennedy and Dilys Thorp, *University of Birmingham, UK*	Investigation of linguistic output of General Training Writing Task 2	
Cynthia Celestine, Cheah Su Ming, Geetha Rajaratnam, *IDP Education Australia, Malaysia* and Norazina Ismail, *The University of Malaysia*	A comparative study of IELTS to ascertain its viability for the Malaysia private secondary school market	IELTS RR 5
Cyril Weir and Anthony Green, *University of Surrey, Roehampton, UK*	Monitoring score gain on the IELTS Academic Writing Module in EAP programmes of varying duration	
Peter Mickan, *The University of Adelaide, Australia*	What's your score? An investigation into language descriptors for rating written performance	IELTS RR 5
David Coleman, Visiting Research Fellow, *The University of Sydney*, Sue Starfield, *The University of New South Wales* and Anne Hagan, *University of Ulster, UK*	The attitudes of IELTS stakeholders: Student and staff perceptions of IELTS in Australian, UK and Chinese tertiary institutions	IELTS RR 5

[1] IELTS Research Reports www.ielts.org/teachersandresearchers/research   [2] Studies in Language Testing www.cambridge.org/uk

Rounds and Researchers	Research Title	IELTS publications
Round 7: 2001		
Annie Brown, *University of Melbourne, Australia*	A cross sectional and longitudinal study of examiner behaviour in the revised IELTS Speaking test	
Barry O'Sullivan and Mark Rignall, *University of Reading, UK*	Assessing the value of bias analysis feedback to raters for the IELTS Writing Module	SiLT 19
Chandra Rao, Kate McPherson, Rajni Chand and Veena Khan, *USP Solutions, The University of the South Pacific*	Assessing the impact of IELTS preparation programs on candidates' performance on the General Training Reading and Writing test modules	IELTS RR 5
Cyril Weir, *University of Luton, UK* and Anthony Green, *Research and Validation, Cambridge ESOL, UK*	The impact of IELTS on the preparation classroom: stakeholder attitudes and practices as a response to test task demands	
Russell Whitehead, *ESOL consultant*	Issues in the assessment of pen and paper and computer-based IELTS Writing Tasks	
Round 8: 2002		
John Read, *The University of Auckland* and Paul Nation, *Victoria University of Wellington, New Zealand*	An investigation of the lexical dimension of the IELTS Speaking Test by measuring lexical output, variation and sophistication, and the use of formulaic language	IELTS RR 6
Barry O'Sullivan, *Roehampton University, UK* and Yang Lu, *University of Reading, UK*	The impact of candidate language of examiner deviation from a set interlocutor frames in the IELTS Speaking test	IELTS RR 6
Cyril Weir, *University of Bedfordshire, UK*, Barry O'Sullivan, *University of Reading, UK*, Jin Yan, *Jiao Tong University, China* and Steven Bax, *Canterbury University, Christchurch, UK*	Does the computer make a difference? Reactions of candidates to a CBT versus traditional hand-written forms of the IELTS Writing component: Effects and Impact	IELTS RR 7
Round 9: 2003		
Catherine Elder and Gillian Wigglesworth, *The University of Melbourne, Australia*	An investigation of the effectiveness and validity of planning time in Part 2 of the IELTS Speaking Test	IELTS RR 6
Michael Carey, Robert Mannell and Geoff Brindley, *Macquarie University, Australia*	Inter-examination variation due to inter-language phonology accommodation in the pronunciation section of the IELTS Speaking test	
Cyril Weir, *University of Bedfordshire, UK*, Barry O'Sullivan, *Roehampton University, UK* and Tomoko Horai, *Roehampton University, UK*	Exploring difficulty in Speaking tasks An intra-task perspective	IELTS RR6
M.A. Yadugiri, *Bangalore University, India*	A study of the linguistic and discourse features in the output from IELTS academic writing tasks	
Annie Brown, *The University of Melbourne, Australia*	An examination of the rating process in the revised IELTS Speaking test	IELTS RR 6
Christopher Hampton, *British Consulate, Shanghai, China*	What makes a good Speaking test? Perceptions of candidates and examiners	
Hilary Smith, Stephen Haslett, *Systemetrics Research Limited, New Zealand*	Attitudes of tertiary key decision-makers towards English language tests in Aotearoa New Zealand	IELTS RR 7
Pauline Rea-Dickens, Richard Kiely and Guoxing Yu, *University of Bristol, UK*	Student identity, learning and progression: The affective and academic impact of IELTS on 'successful' candidates	IELTS RR 7
Round 10: 2004		
Paul Seedhouse, *The University of Newcastle upon Tyne, UK* and Mary Egbert, *The University of Southern Denmark*	The interactional organisation of the IELTS Speaking test	IELTS RR 6
Jayanti Banerjee, Florencia Franceschina and Anne Margaret Smith, *Lancaster University, UK*	Documenting features of written language production typical at different IELTS band levels	IELTS RR 7
David Ingram and Amanda Bayliss, *Melbourne University Private, Australia*	IELTS as a predictor of academic language performance. Part 1: The view from participants	IELTS RR 7
Andrea Paul, *Melbourne University Private, Australia*	IELTS as a predictor of academic language performance. Part 2: Case studies of learner language	IELTS RR 7
Kieran O'Loughlin, *The University of Melbourne, Australia*	The use of IELTS for university selection in Australia: A case study	IELTS RR 8
Peter Mickan, *The University of Adelaide, Australia*	An ethnographic study of classroom instruction in an IELTS preparation program	IELTS RR 8
Carole Gibson and Anne Swan, *The University of South Australia*	The significance of socio-linguistic backgrounds of teachers of IELTS preparation courses in selected Malaysian institutions	IELTS RR 8
Round 11: 2005		
Glenys Merrifield, *GBM & Associates, Australia*	An impact study into the use of IELTS as an entry criterion for professional associations: USA, New Zealand and Australia	
Hilary Smith and Stephen Haslett, *Systemetrics Pty Ltd, New Zealand*	The IELTS General Training Module as a predictor of performance in practical tertiary programs	
Kieran O'Loughlin and Sophie Arkoudis, *The University of Melbourne, Australia*	Investigating IELTS exit score gains in higher education	
John Field, *Independent Researcher, UK*	A cognitive validation of the lecture-listening component of the IELTS listening paper	
Cyril Weir, *University of Luton, UK*	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	
Christine Pegg and Alison Wray, *Cardiff University, UK*	The effect of memorised learning on the Writing scores of Chinese IELTS test takers	
Round 12: 2006		
John Read and Rosemary Wette, *The University of Auckland, New Zealand*	Assessing the English proficiency of medical professionals using IELTS and other means	
Peter Mickan, *The University of Adelaide, Australia*	Learners' experiences preparing for the IELTS examinations	
Roger Hawkey and Cyril Weir, *University of Bedfordshire, UK*	The cognitive processes underlying the Academic Reading construct as measured by IELTS	
Richard Badger and Oscar Yan; *University of Leeds, UK*	The use of tactics and strategies by Chinese students in the Listening component of IELTS	
David Hyatt and Greg Brooks, *The University of Sheffield, UK*	Investigating stakeholders perceptions of IELTS as an entry requirement for higher education in the UK	
Katherine Golder, Kenneth Reeder and Sarah Fleming, *The British Columbia Institute of Technology, Canada*	Determination of the appropriate IELTS band score for the admission into a program at a post-secondary polytechnic institution	