

7

ResearchNotes

UNIVERSITY OF CAMBRIDGE LOCAL EXAMINATIONS SYNDICATE
ENGLISH AS A FOREIGN LANGUAGE (EFL)

FEBRUARY 2002



UNIVERSITY of CAMBRIDGE
Local Examinations Syndicate

ResearchNotes

Editorial Notes

Welcome to Issue 7 of *Research Notes*.

This is our first issue to carry a special focus. On this occasion we turn the spotlight on our suite of tests for children – the Cambridge Young Learners English (YLE) Tests, introduced in 1997 and now among our most popular tests. Lynda Taylor and Nick Saville describe the development of the tests from the early 1990s and discuss their nature and role in the context of children's language learning. A follow-up article reports on the analysis of YLE candidate performance carried out by the Research and Validation Group for 2000. Fiona Ball and Juliet Wilson describe work on a number of key research projects related specifically to the YLE Speaking Tests, and we also report on the BAAL/CUP seminar on the young language learner research agenda, held at the University of Manchester in June 2001.

In Issue 6 of *Research Notes*, Stuart Shaw outlined a range of issues in the field of assessing second language writing ability. Barry O'Sullivan picks up on this theme as he reports on an experimental study recently commissioned by UCLES using Multi-Faceted Rasch (MFR) to investigate sources of variability in the IELTS General Training Writing module. We also have news of the forthcoming review of the KET/PET examinations.

The issue of fairness continues to attract much attention within the language testing community and the topic is addressed in this issue in relation to two projects. Nick Saville outlines the development of the ALTE Code of Practice (adopted by UCLES EFL and other ALTE members since the early 1990s) and the recent activities of the Code of Practice Working Group; the latest phase of this work was the focus of workshops and discussions at the ALTE Conference in Budapest in November 2001. Professor Antony Kunnan of California State University presented his recently-developed Test Fairness Model at both the ALTE Conference and an internal EFL staff seminar; Ardeshir Geranpayeh summarises the nature and purpose of the model. In addition he reports on the Language Testing Forum held in Nottingham in November 2001 and on the contribution of UCLES staff to this event.

We are also pleased to include in this issue an announcement of the winner of the IELTS MA Dissertation Award for 2001, together with a call for submissions for the 2002 award.

Finally, it is now almost 2 years since we launched *Research Notes* in its current format and we have been pleased with the interest shown in it. For 2002 we plan to increase the number of issues from 3 to 4 (to appear in February, May, August and November). We are always happy to receive feedback from readers and this issue includes a short questionnaire to help us canvas your views. Please complete the questionnaire on page 24 and return it to us by 31 May 2002. We hope to publish the findings in Issue 9 and your responses will be invaluable in helping to inform our decisions about content and format in the future.

Contents

Editorial Notes	1
Developing English language tests for young learners	2
Candidate performance in the Young Learners English Tests in 2000	5
Research projects related to YLE Speaking Tests	8
Other YLE-related news	11
Striving for fairness – the ALTE Code of Practice and quality management systems	12
Investigating variability in a test of second language writing ability	14
Review of KET and PET Examinations	18
The UCLES EFL staff seminar programme	19
BAAL/CUP Seminar report	20
Studies in Language Testing	21
IELTS MA dissertation award 2001	22
IELTS MA dissertation award 2002	23
Readers' questionnaire	24

The URL for reading/downloading previous issues of *Research Notes* is:
http://www.cambridge-efl.org/rs_notes/

The URL for subscribing to *Research Notes* is:
http://www.cambridge-efl.org/rs_notes/inform.cfm

Developing English language tests for young learners

LYNDA TAYLOR, SENIOR RESEARCH AND VALIDATION CO-ORDINATOR, UCLES EFL
NICK SAVILLE, ASSISTANT DIRECTOR, UCLES EFL

Introduction

The steady increase in the teaching of English to young learners (i.e. approximately 6 to 12 years old) has led to a growing demand for various forms of assessment both within the state and the private education sectors. (Useful background articles by Shelagh Rixon on the development and growth of young learner English can be found in *Language Teaching*, 25/2, 1992, and in *Modern English Teacher*, 9/4, 2000.) In turn this growth has led to a debate within the language teaching and testing community on the issues associated with assessing children's second/foreign language skills (see the special issue in April 2000 of *Language Testing on Assessing Young Language Learners*, Volume 17/2). The debate has centred upon a number of key questions, e.g.

- what are appropriate approaches for assessing children?
- is it a good thing to have formal language tests for children?
- what is the relationship between testing and the classroom?
- how does assessment influence children's motivation and future learning?

Decisions made at national level in the state sector often have an influence on private schools providing "after school" support; they may perceive a need for external examinations (with certification) to provide grade objectives and evidence of progress for students. It is this kind of interaction between the state and private sectors that has given rise to the request for tests and examinations provided by external bodies. For an international examination board such as UCLES, already well-known for its range of high-quality EFL examinations for adults and teenagers, the key questions referred to above can be summarised as follows:

Is it possible to create an international English language test for children that makes an accurate and fair assessment and has a positive impact on a child's future language learning?

In responding to requests for young learner tests or certification, the starting point is the same as for any other test; the test developer must adopt a principled approach which allows the major considerations and constraints that are specific to the testing context to be taken into account during the development process.

In her recent book *Teaching English to Young Learners* (2000), Lynne Cameron includes a chapter on assessment and language learning. She begins by acknowledging the social and classroom realities of language assessment, and then goes on to suggest several key principles for assessing children's language learning:

- assessment should be seen from a learning-centred perspective
- assessment should support learning and teaching

- assessment should be congruent with learning
- children and parents should understand assessment issues
- assessment is more than testing

It was consideration of key principles such as these which informed the development of UCLES international tests for young learners of English at three levels – Cambridge *Starters*, *Movers* and *Flyers*. Development work began in 1993 and the tests were introduced in May 1997; by the end of 2001 the worldwide candidature had reached nearly 200,000, with large numbers of candidates in countries such as China, Spain, Argentina and Italy. This article charts the development of the Cambridge YLE tests and situates them and their role within the wider context of what we currently understand about children's language learning.

What makes a test "useful"?

Previous articles in *Research Notes* have described the general test development principles and practice adopted by UCLES EFL (see Issues 3 and 4) and have highlighted the notion of *utility* or *usefulness*, a concept which is well established in the testing literature (Messick, 1988, 1989; Bachman and Palmer, 1996). Underlying this concept is the view that a test should perform a *useful* function within an educational and social context through having an appropriate balance of essential test qualities. This means that the test must be *valid* in terms of the constructs represented and content covered. It must be *reliable* in that the results produced should be accurate/consistent and as free as possible from errors of measurement. It must be *practical* in terms of the demands it makes on the resources of the test developer, administrator and test taker. Finally, the test should be designed to have a *positive impact* on individuals, classroom practice and society in general (see previous articles on impact in *Research Notes*, Issues 2 and 6). In short, the usefulness of any test is achieved by obtaining a balance between these four qualities in relation to specifically defined testing contexts.

What are the implications in designing tests for young learners?

If tests for young learners are to be useful, then we might expect them to have certain features. For example, they should:

- take account of children's cognitive and social development
- be consistent with good practice in primary school teaching (materials and methodology)
- support language use with clear contexts and accessible tasks

- avoid any deficit model description of the children's language development (i.e. the concept of failure)
- be relevant and look interesting (e.g. by making use of colour and graphics)
- report meaningful results in order to encourage further learning.

The test features listed above reflect the principles suggested by Cameron (2000) for assessment in general, and each one has implications for the relative weighting assigned to the four main test qualities previously mentioned – validity, reliability, practicality and impact. Clearly, positive impact is a primary consideration in creating tests for young learners and will be largely achieved by emphasising validity (especially content validity) and practicality. It is the specific characteristics of children as language learners and test takers which tend to make young learners' tests a controversial issue, so careful account must be taken of children's socio-cognitive skills, as well as of pedagogic aspects related to the skills to be tested and the types of materials/tasks to be included in the test. The views of those who are currently engaged in teaching children and the materials and methods that they employ must also be noted. Although reliability cannot be ignored if the test is to function usefully, it is likely to be less heavily weighted than it might be in a test designed for older language learners.

How were the Cambridge Young Learners Tests developed?

The Cambridge Young Learners English Tests (YLE) were designed to cover three key levels of assessment: *Starters, Movers and Flyers*. Taken together the three tests aimed to build a bridge from beginner to Waystage level, and to achieve the following:

- sample relevant and meaningful language use
- measure accurately and fairly
- promote effective learning and teaching
- encourage future learning and teaching
- present a positive first impression of international tests

The planning phase of the YLE development project involved extensive research and consultation. Since 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. The literature on child social and cognitive development was surveyed in some depth (Elliott, 1995; Hughes and Green, 1994) and this helped to inform the choice and treatment of topics and tasks for children of different ages. Research has shown that children perform best on simple, clearly-focused tasks based within their immediate experience; instructions must be easily understood and should not require extensive processing or memory load. Tasks involving scanning, for example, were rejected since children only demonstrate search and stop strategies around age 11. Creating a

low-anxiety situation, free from risk of confusion or fear of failure, was regarded as a priority.

It is widely recognised that children are motivated by and perform best on tasks which directly reflect their own experiences of teaching/learning, i.e. test tasks must take account of the organisation and presentation of knowledge familiar through course materials and classroom activities. For this reason, a wide range of coursebooks and teaching materials used in primary language classrooms throughout the world was reviewed; this helped to identify the main content areas (topic, vocabulary, etc.) which frequently occur in primary syllabuses; it also highlighted the importance of presenting material in a lively and attractive manner, taking into account the age and background of the intended candidates. Another important aspect of the project was collaboration with staff at Homerton College (Cambridge University's teacher training college) who were developing a multimedia, CD-ROM based resource for young learners. The UCLES YLE test development team was able to draw on the research conducted by Homerton and their experience of working with primary age children.

Draft specifications and sample materials were drawn up in the design phase of the project. The Cambridge YLE Tests include coverage of all four skills – speaking, listening, reading and writing; greater emphasis is given to oral/aural skills because of the primacy of spoken language over written language among children; in addition, any writing activity is largely at the word/phrase (enabling skills) level since young children have generally not yet developed the imaginative and organisational skills needed to produce extended writing. The test specifications are "topic-led" (like many popular course books for young learners today) and, given the international nature of the tests, they aim as far as possible to avoid cultural bias; they also set out the notions, structures, lexis and tasks on which all the tests are based and these are graded across the three levels. For convenient reference, language is broken down into various categories; this also provides a checklist for test writing to ensure equivalence between different test versions. Test formats include frequent changes of activity or task-type and use tasks which are brief and 'active' or 'game-like', e.g. colouring activities. Tasks focus on meaning rather than form and are intended to test the meaningful use of language in clear, relevant, accessible contexts; they are based on the kinds of task-based communicative activity, often interactive in nature, which are already used in many primary classrooms around the world. A guiding principle for the project has been a desire to close the distance between children's experiences of learning and testing. Considerable care was taken to reflect the language covered in a wide survey of primary English courses and materials.

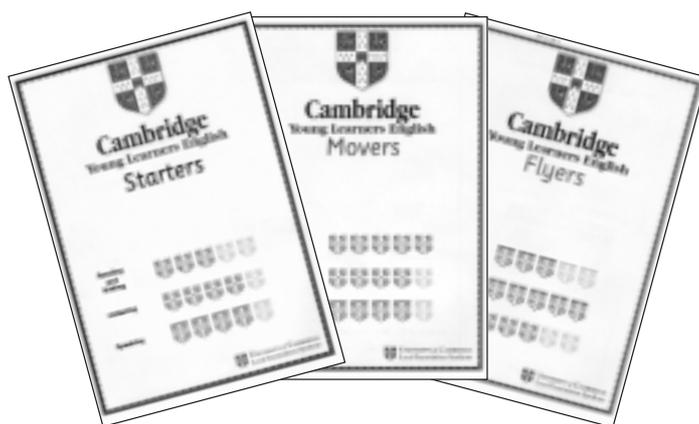
Versions of the tests were trialled in 1995/6 with over 3000 children in Europe, South America and South East Asia. The results of these trials, including feedback from teachers and statistical analysis of the children's answers, were used to construct the live test versions. In particular, trialling and feedback from potential users confirmed that question papers should be in colour.

The table below indicates the common characteristics and variations across the three test levels.

	Overall length	No of tasks/parts	No of items
Cambridge Starters			
Listening	Approx 20 minutes	4	20
Reading and Writing	20 minutes	5	25
Speaking	Approx 5 minutes	2	–
Total	45 minutes		
Cambridge Movers			
Listening	Approx 25 minutes	5	25
Reading and Writing	30 minutes	6	40
Speaking	Approx 7 minutes	2	–
Total	62 minutes		
Cambridge Flyers			
Listening	Approx 20 minutes	5	25
Reading and Writing	40 minutes	7	50
Speaking	Approx 9 minutes	2	–
Total	69 minutes		

The specifications provide a testing syllabus, i.e. a helpful description of the test format, tasks and sample items. 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. This does not mean, however, that a teacher preparing children for Cambridge YLE Tests should concentrate on teaching a list of structures or vocabulary. Apart from familiarising the children with the test format, teachers can continue their normal teaching programmes and concentrate on learning through a focus on meaning and context.

Most of the design and development work reported so far has related to issues of validity and impact; however, as noted above, issues of practicality and reliability were also important. For example, schools need to be able to enter children when the children are ready to take the test. This could be at any time in the year, so instead of a fixed timetable of administration a flexible system is available which fits in with different regional requirements; approved Cambridge Centres can order versions of the test by arrangement with UCLES. This means that the tests can be taken in familiar surroundings in a child's school and can be administered to fit in with local conditions (e.g. school terms, holiday periods). The face-to-face speaking component is marked on-site by a specially trained and standardised oral examiner, while test booklets for the Reading/Writing and Listening components are marked in Cambridge by trained markers. Results are issued as soon as possible, normally within about two weeks of receipt of the scripts by UCLES. The reporting of results is designed to provide positive encouragement to the learners: all candidates



receive an award (see above) which focuses on what they can do (rather than what they can't do), and gives every child credit for having taken part in the test.

The Cambridge YLE tests became operational in May 1997 and monitoring and evaluation activities, which are essential for test validation, have continued routinely since then. Given the importance of assessing test impact in the case of tests for young learners, the process of consultation with test users has been a priority. For example, questionnaires despatched in 1999 and 2001 gathered largely positive feedback from test users on the content and administration of the tests. This information, together with findings from internal studies of the tests' performance (see article on page 5), feeds into the ongoing process of evaluation and review. A formal review of the YLE tests is about to begin, which may well lead to minor changes in the test at a later stage. Other articles in this issue describe some of the ongoing research and validation projects relating to the YLE tests (see pages 8 and 11).

Conclusion

Specialists in the field of young learner English have sometimes expressed reservations about the role of language tests for children (Cameron, 2000; Rixon, 1992); they nevertheless accept the social reality of this form of assessment in today's world and, more importantly, they acknowledge that testing can have positive effects, e.g. the process and outcomes can motivate learners and support further learning. Lynne Cameron also reminds us that assessment practices carry messages for children about what parents and teachers (and perhaps wider society) consider important in language learning and in life. She ends her chapter on assessment as follows: "For young children, what matters is a solid base in spoken language, confidence and enjoyment in working with the spoken and written language, and a good foundation in learning skills. We should be searching out assessment practices that will reinforce the value of these to learners and to their parents." (240)

This article has described how, in developing the YLE tests, UCLES EFL sought to take account of these factors and to create a suite of international tests for children which convey the right messages to children, their parents and their teachers. We remain committed to giving children a positive experience of testing and to encouraging a positive impact on their future language learning.

Further information:

A Handbook for teachers (containing the specifications for *Starters*, *Movers* and *Flyers*) and a Sample Papers booklet (in colour) can be ordered direct from UCLES or downloaded from the UCLES EFL website: www.cambridge-eFL.org. A short information video for use by teachers and parents is also available, as well as an annual Examination Report. For information on presentations and seminars about the Cambridge YLE tests please email: eflseminars@ucles.org.uk

References and further reading

- Alderson, J C and Banerjee, J (2001): State of the Art Review – Language testing and assessment (Part 1), *Language Teaching*, 34/4, 213–236.
- Bachman, L F and Palmer, A (1996): *Language Testing in Practice*, Oxford: Oxford University Press
- Cameron, L (2000): *Teaching English to Young Learners*, Cambridge: Cambridge University Press
- Elliott, M (1995): *Developmental cognitive-psycholinguistics and cognitive determinants in linguistic performance among young L1 and L2 language learners (7 to 14 years)*, Internal report by the EFL Validation Group, UCLES
- Hughes, S and Green, A (1994): *A review of research on first and second language skills among school age children (8–14 years)*, Internal report by the Cognitive Psychometrics Section of the Research and Evaluation Division, UCLES

- Messick, S A (1988): The once and future issue of validity: Assessing the meaning and consequences of measurement. In H Wainer and H I Braun (eds): *Test Validity*, Hillsdale, NJ: Erlbaum
- Messick, S A (1989): Validity. In R L Linn (ed) *Educational Measurement*, third edition, New York: Macmillan
- Moon, J (2000): *Children Learning English*, Oxford: Heinemann Macmillan
- Rea-Dickins, P (2000): Special issue of *Language Testing: Assessing Young Language Learners*, 17/2
- Rixon, S (1992): State of the art article: English and other languages for younger children – practice and theory in a rapidly changing world, *Language Teaching*, 25/2, 73–93
- Rixon, S (1999): *Young Learners of English – some research perspectives*, Harlow: Longman
- Rixon, S (2000): Young Learners of English – background and issues, *Modern English Teacher*, 9/4, 5–10

IATEFL Young Learners Special Interest Group (SIG)

Membership is open to members of IATEFL and brings information on local and international events and conferences, a twice-yearly newsletter and access to the YL SIG email discussion group. Contact address: IATEFL, 3 Kingsdown Chambers, Whitstable, Kent, CT5 2FL, UK.

YL SIG website: www.countryschool.com/younglearners.htm

British Council

web page with information from the worldwide survey on young learners: www.britcoun.org/english/eyl

Candidate performance in the Young Learners English Tests in 2000

HELEN MARSHALL, VALIDATION OFFICER, UCLES EFL
MIKE GUTTERIDGE, CONSULTANT, UCLES EFL

UCLES EFL publishes annual reports on many of its examinations; these reports are intended to provide test users with an overview of candidate performance during a given year and to offer guidance on the preparation of candidates. This article provides a summary of the information contained in the recently published report on the Cambridge Young Learners English Tests – *YLE Examination Report 2000*.

The Cambridge YLE Tests were introduced in 1997 (see the article on page 2 for a description of their development); since then they have rapidly established their popularity with teachers, parents and children around the world and in 2000 the tests were taken by 128,000 candidates in 53 different countries.

Age of test-takers

The tests are aimed at children aged between 7 and 12 (although 13-year-olds in the same class as 12-year-olds who are taking the

exam are not excluded). Average ages of candidates at each level in 2000 were as follows:

Starters	9.8 years
Movers	10.9 years
Flyers	11.8 years

Marking

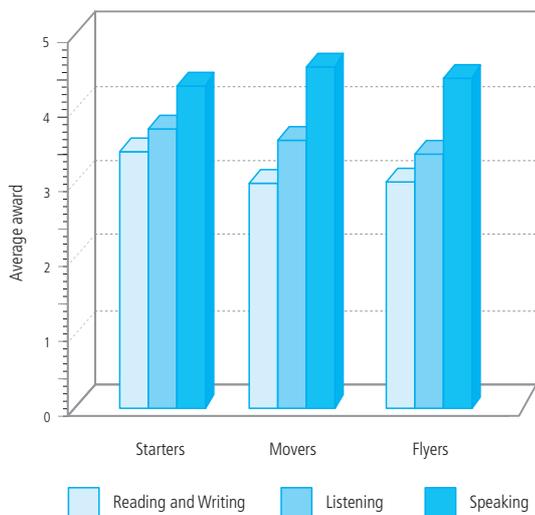
For the *Reading and Writing* and *Listening* tests, candidates record their answers on the question papers. These papers are then marked by a team of carefully selected and trained markers, according to a markscheme. Each question on the paper carries equal weight. Results are then issued as soon as possible, normally within two weeks of receipt of the scripts by UCLES.

For the Speaking tests, test centres make arrangements for oral examining locally, using approved examiners. In selecting people for training as examiners, emphasis is placed on experience of working with children as well as in EFL and all examiners receive special training in the needs of this age group. Examiners conduct the tests on a one-to-one basis and are trained to use encouraging language during the tests.

Grading

The YLE tests are not pass/fail exams. Candidates are given an 'award' which shows them how well they have done in each of the three components – *Reading and Writing*, *Listening* and *Speaking*. They can be awarded a maximum of five shields for each paper. Five shields are awarded only to candidates who gain full (or almost full) marks. Overall, the average candidate receives three shields for *Reading and Writing* and *Listening* and four shields for *Speaking*. Provided they attempt all parts of the tests, even the weakest candidate can achieve an award of one shield in each skill, thus gaining a total of three shields.

Average award in shields for all three levels



During 2000 five test versions were used at each of the three levels – *Starters*, *Movers* and *Flyers*. The remainder of this article reports in detail on candidate performance in one of these versions – Version 13.

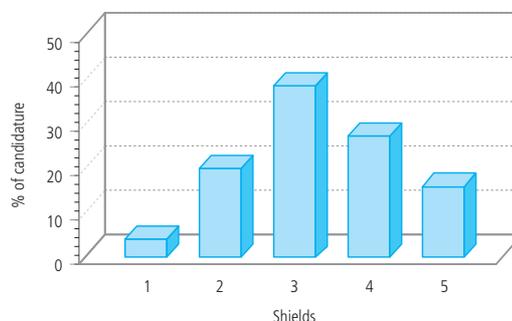
Candidate Performance in Reading and Writing (Component 1)

Starters: The *Starters Reading and Writing* paper was taken by 16,955 candidates. The average award over the year was 3.3 shields.

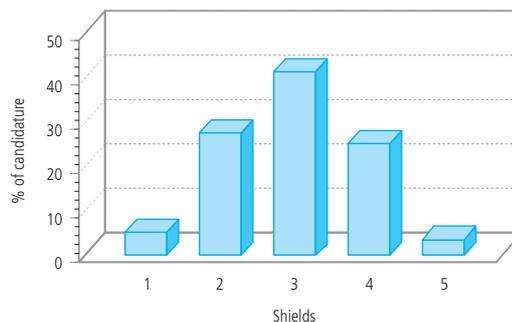
Movers: The *Movers Reading and Writing* paper was taken by 13,543 candidates. The average award over the year was 2.9 shields.

Flyers: The *Flyers Reading and Writing* paper was taken by 9,049 candidates. The average award over the year was 3.0 shields.

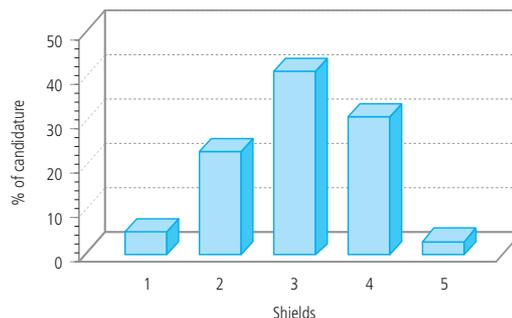
Percentages of candidates obtaining shields at each level are indicated below.



Starters: Reading and Writing - Achievement of Shields, 2000



Movers: Reading and Writing - Achievement of Shields, 2000



Flyers: Reading and Writing - Achievement of Shields, 2000

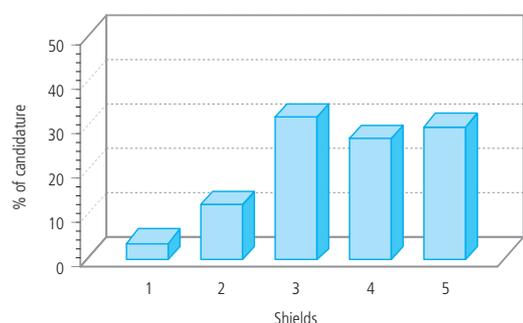
Candidate Performance in Listening (Component 2)

Starters: The Starters *Listening* paper was taken by 16,955 candidates. The average award over the year was 3.6 shields.

Movers: The Movers *Listening* paper was taken by 13,543 candidates. The average award over the year was 3.45 shields.

Flyers: The Flyers *Listening* paper was taken by 9,049 candidates. The average award over the year was 3.3 shields.

Percentages of candidates obtaining shields at each level are indicated below.



Starters: Listening - Achievement of Shields, 2000

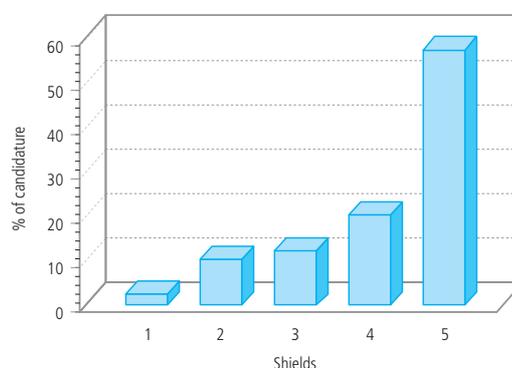
Candidate Performance in Speaking (Component 3)

Starters: The Starters *Speaking* paper was taken by 16,954 candidates. The average award over the year was 4.2 shields.

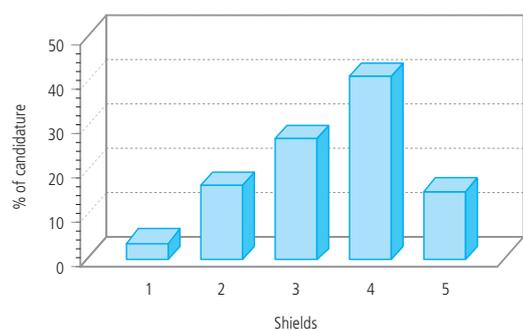
Movers: The Movers *Speaking* paper was taken by 13,543 candidates. The average award over the year was 4.4 shields.

Flyers: The Flyers *Speaking* paper was taken by 9,048 candidates. The average award over the year was 4.3 shields.

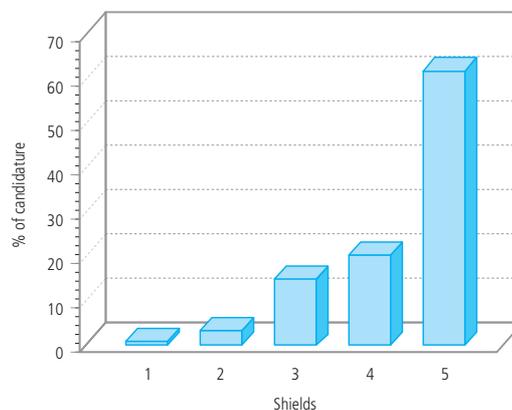
Percentages of candidates obtaining shields at each level are indicated below.



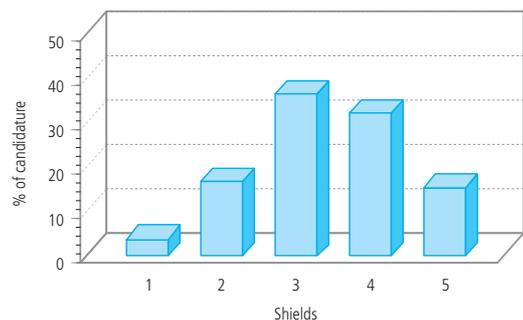
Starters: Speaking - Achievement of Shields, 2000



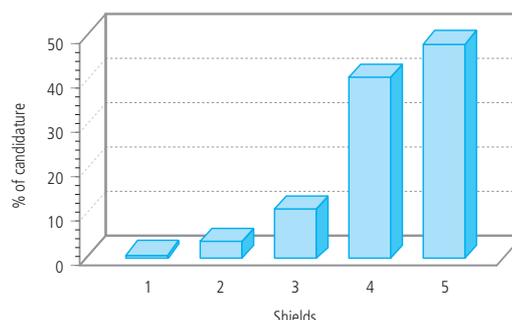
Movers: Listening - Achievement of Shields, 2000



Movers: Speaking - Achievement of Shields, 2000



Flyers: Listening - Achievement of Shields, 2000



Flyers: Speaking - Achievement of Shields, 2000

General comments

Overall, candidates performed better on the *Listening* and *Speaking* components than they did on the *Reading and Writing* component. Performance on the *Speaking* component was particularly good at all levels and this was the case for all countries where the tests were taken.

Candidate performance varied to a minor extent from country to country. The full YLE Examination Report gives the average performance for candidates from some of the main countries where the tests were taken in 2000.

Copies of the full *YLE Examination Report 2000* upon which this article is based are available free of charge from the UCLES EFL web-site (www.cambridge-efl.org.uk) or at a charge of £2.50 from:

UCLES Publications Unit
1 Hills Road
Cambridge CB1 2EU
UK
Tel: +44 1223 553930
Fax: +44 1223 460278

Research projects relating to YLE Speaking Tests

FIONA BALL, VALIDATION OFFICER, UCLES EFL,
JULIET WILSON, YLE SUBJECT OFFICER, UCLES EFL

Overview

This article focuses on the *Speaking* component of the YLE tests and describes some of the research projects that UCLES EFL is currently involved in. A description of how *Speaking Tests* are stored and analysed is provided and future avenues of research in this area are outlined.

The YLE Speaking Tests

Like the other Cambridge EFL examinations, the *Speaking* component of the YLE tests is a face-to-face test with a trained examiner, involving a range of tasks and test material designed to elicit a rich enough sample of language for assessment purposes. But unlike most of the other Cambridge tests, the YLE *Speaking Test* is conducted by one examiner with one candidate. This 'singleton' format was adopted because children of this age range may not display the interactive communicative functions e.g. turn-taking, which the collaborative element in the other Cambridge *Speaking Tests* focuses on. The examiner might therefore find it difficult to elicit an adequate sample of language from each candidate. It was also considered preferable to only have one examiner as it might be anxiety-inducing for the children to be in the test room with two unknown adults.

Obviously the impact of this 1:1 format on the examiner is quite significant. The examiner not only has to deliver the interlocutor frame, but also has to participate in the tasks with the child, as well as provide an accurate assessment. This is a challenging role for the examiner.

The following tables outline the *Speaking Test* tasks at the three YLE levels.

Starters

Part	Main Skill Focus	Input	Response
1	Listening and responding to spoken instructions (lexis)	Scene card	Point to items on Scene card
2	Listening and responding to spoken instructions (lexis and prepositions)	Scene card and object cards	Identify object cards and place on scene card
3	Listening and responding to simple picture-based questions	Scene card	1-word answers
4	Listening and responding to questions	Object cards and Examiner's questions	1-word answers
5	Listening and responding to personalised questions	Examiner's questions	1 word answers/ short phrases

Movers

Part	Main Skill Focus	Input	Response
1	Identifying and describing differences between 2 pictures	2 pictures	Words/ short phrases
2	Narrating a story	4-picture story	Extended narrative
3	Identifying 'odd one out' and explaining	Picture sets	Words/phrases
4	Listening and responding to personalised questions	Examiner's questions	Answer personalised questions (words/short phrases)

Flyers

Part	Main Skill Focus	Input	Response
1	Listening to statements and describing differences in pictures	Picture and Examiner's questions	Words/short statements phrases
2	Speaking (question forming) Reading for specific information Listening	Cue cards	Read appropriate answers from cue cards. Formulate questions
3	Narrating a story	5-picture story	Extended narrative
4	Listening and responding to personalised questions	Examiner's questions	Answering personalised questions (words/short phrases)

A research agenda for YLE

UCLES EFL's current research agenda includes the following areas relevant to the YLE Speaking Tests:

- Developing a corpus (electronic database) of Speaking Tests
- Qualitative analyses through transcriptions and checklists
- Rating scale development and validation
- Revision processes

The specific research projects related to the YLE Speaking Tests are detailed below.

Building a corpus of Young Learner Speaking Test performances

This project involves collecting a representative sample of YLE Speaking Tests from all over the world and developing these into a corpus which can be analysed in various ways. YLE Speaking Tests are currently being audio-recorded and transcribed, so that both sound and text versions are available for analysis. The resulting corpus of Starters, Movers and Flyers Speaking Tests will be linked to a database of information about the candidates which will be searchable by a range of variables including the age or first language of the candidate or the marks awarded. It is envisaged that the corpus will enable researchers to listen to the original Speaking Test and read a written interpretation of events at the same time, therefore providing two different ways of understanding what happens within the test. There are no plans to video YLE Speaking Tests to provide an additional medium through which to analyse this part of the examination, as it would be too intrusive for both candidate and examiner in this context.

Over one hundred Speaking Tests have been transcribed to date using standard orthography, although additional features of the interaction have been coded, including false starts, repetition and

pauses. An extract from the final part of a YLE Speaking Test is given below:

INFORMATION GIVING IN A YLE SPEAKING TEST

- E* now let's talk about your house X how many rooms are there in your house
- C* erm
- E* how many rooms are there in your house
- C* (.) six six rooms
- E* mmhm and what do you like doing at home
- C* I'm doing the play with my gameboy
- E* mm you like playing with your gameboy who is the youngest in your family
- C* it's me
- E* it's you and tell me about your bedroom
- C* my bedroom my bedroom has got my bed um it's got the cupboard the
- E* table
- C* the table erm the wardrobe the computer er that's it

Key: E examiner
C candidate
(.) short pause
(2) 2 second pause

We envisage that the corpus will provide data for a range of other YLE research projects so it has been designed and annotated with a range of potential uses in mind. The Young Learners corpus will also act as a pilot project for a larger electronic corpus of spoken English to be developed in the future for other UCLES EFL exams.

Investigating the story telling task

It has been noted by experts in children's language (Cameron, 2001), and also reported by some YLE examiners, that telling a story in L2 is a very difficult task for children. By transcribing and analysing samples of the YLE speaking tests, we will be able to investigate questions such as:

- How do candidates perform in the story-telling task compared to other parts of the test?
- Are there qualitative and/or quantitative differences between the language produced in the story-telling task and other parts of the speaking test?
- Do candidates hesitate or display uncertainty or nervousness in the story-telling task?
- How do examiners use back-up questions in the story-telling task?

Extracts from two YLE Speaking Tests are given below which show how two candidates performed differently on the same story telling task. The scenario presented to each candidate was:

'Now look at these pictures. They show a story. Look at the first picture. David and Gill are playing with their pet parrot. They are happy. Now you tell the story.'

CANDIDATE A'S STORY

- E ...now you tell the story
 C erm what are the names
 E the names David and Gill
 C the parrot later the parrot erm flying erm out the c their house
 E yes
 C and erm David and Gill er are sad and they are um (1) um
 E mmhm where are they (.) where are they
 C um are um in the garden
 E and what are they looking for
 C a parrot
 E yes and then the fourth picture
 C in the fourth picture they are um they saw parrot the parrot erm the parrot er is watching tv
 E mmhm the parrot likes television

CANDIDATE B'S STORY

- E ... now you tell the story
 C er the parrot no um David (2) the parrot er went to the garden and er David erm David look behind erm behind the tree and what is the name
 E um Gill
 C and Gill in the street but erm they not look the parrot (.) umm later er the parrot er is in the room on the sofa looked look watch the tv and the programmer is a is a parrot and in the television there is two parrots on the tree
 E good thank you

It is clear from these two extracts that there are differences in the amount and nature of the language produced by these two candidates in response to the same task. Additionally, the examiners provided different amounts of support to the candidates in the form of prompting and providing feedback to the candidates. These and other features of the interaction will be analysed in due course. Findings from this particular investigation will be reported in a future issue of *Research Notes*.

As well as these ongoing projects, there are a number of projects that will commence once sufficient data has been collected and analysed. These projects are described below.

Investigating the assessment criteria

In this project we will investigate how appropriate the assessment criteria for the YLE Speaking tests are, considering each part of the

speaking test separately. Results of a special re-rating project and questionnaires/protocol analysis with examiners will help us to evaluate the effectiveness of the assessment criteria and rating scales.

Investigating examiner talk

This project will investigate the use of the Examiner Frame, particularly in terms of how it affects the rating of candidates' performance. Despite all examiners using the same materials and having a 'frame' to follow, examiners differ with respect to whether and how they use the back-up prompts suggested in the interlocutor frame, how long a pause they allow before using a back-up prompt and other variations as indicated in the extracts provided in this article.

Investigating a paired format

There is some evidence (Johnstone, 2000) that children can produce a sample of language in a co-operative format and that this might be an appropriate format to use in the YLE Speaking Tests. We would therefore like to investigate to what extent a paired format speaking test would produce language which is distinct from that produced in a singleton format and whether this sample would be more representative of a construct of spoken language for Young Learners.

Conclusion

UCLES EFL is involved in a wide range of research projects in support of its English language examinations. The YLE projects reported in this article are unique in that every project is based on a growing corpus of YLE Speaking Tests that will form the basis of a larger corpus of Speaking Tests for all of UCLES EFL examinations. The research described here will not only provide greater insights into candidate and examiner behaviour in the YLE Speaking Tests, but will also suggest new ways of analysing and investigating the Speaking Test component of other examinations.

References and further reading

- Cameron, L (2001): *Teaching Languages to Young Learners*, Cambridge University Press
- Johnstone, R (2000): Context-sensitive assessment of modern languages in primary (elementary) and early secondary education: Scotland and European experience, *Language Testing* 17/2, 123–143

Other YLE-related news

Plans to review the Cambridge YLE tests

The YLE tests have been operational now for almost 5 years, and monitoring and evaluation activities have been ongoing since their introduction. At its meeting in December 2000 the YLE Steering Group agreed to undertake a formal review of the YLE tests with a view to introducing some minor adjustments where necessary. Although much of the tests will remain unchanged, some adjustments are likely to certain task-types. For example, an investigation is already under way into how the story-telling task functions in the *Movers* and *Flyers* Speaking Tests (see article on page 8). The vocabulary and structure lists for the three levels are also likely to be updated and expanded. Attention will also be paid to the 'look and feel' of the papers, including the way pictures and graphics are used in the tests.

Investigating perceptions of tests for Young Learners

In the latter part of 2001 a pilot survey was undertaken to explore perceptions of the Cambridge YLE tests in 13 of the countries where YLE is taken. A range of key personnel (local primary English specialists, UCLES local secretaries and development managers, YLE team leaders and examiners) with direct experience of the YLE tests in those countries were invited to give their responses to a series of questions about: local YL teaching methods and curriculum design; in-country attitudes towards YL testing; and the perceived purpose and benefits of the Cambridge YLE tests. The results of this exercise will feed directly into the process of reviewing and revising the YLE tests over the next year or so and may form the basis for a future survey on a much larger scale.

Training and certification for teachers of Young Learners

Throughout their history the Cambridge EFL tests have maintained a strong link with ELT pedagogy and since 1988 UCLES has been providing certification for English language teachers (as well as learners) in the form of the Cambridge English Language Teaching suite of qualifications (e.g. CELTA and DELTA). The number of

young learners studying English has led to a need for specialised training/certification for those teaching English to children and young adolescents. Since 1996 UCLES has offered the Certificate in English Language Teaching to Young Learners (CELTYL), an award which enables people to specialise in the teaching of Young Learners in language schools. In addition, a YL Extension allows holders of a CELTA (Certificate in English Language Teaching to Adults) to extend their previous training to include work with young learners.

Since the introduction of these qualifications, approximately 600 teachers have taken either the CELTYL or the YL extension in over 15 different countries; approximately 30 courses are now on offer each year in 21 centres around the world. For more information on CELTYL and on the other teaching awards, please visit our website: www.cambridge-efl.org/teaching

YLE and the wider framework of reference

Even if the YLE tests are not as high-stakes as some other Cambridge examinations, it is recognised that they need to be as reliable and interpretable as possible. Although the three levels – *Starters*, *Movers* and *Flyers* – are clearly linked within a logical and, to some extent, a measurement framework, it is important that we understand in more detail how they fit into a coherent framework of reference in relation to one another and also to other UCLES EFL tests. From a research and measurement perspective, this presents a number of interesting challenges. The YLE tests are not ideally suited to a latent-trait based anchoring and equating exercise since they are, by design, high-facility tests; this naturally places limitations upon the quality of measurement information they can provide for linking purposes. Nevertheless, constructing a coherent framework for YLE is important because there is increasing demand from users for more information to help interpret performance. For this reason, studies are ongoing within the Research and Validation Unit to help us gain a greater understanding of the relationship between the three levels, and of the link with the lower end of the main suite tests (i.e. KET/PET). Findings will be reported in future *Research Notes*.

LTRC 2002 – Advance notice

The 24th International Language Testing Research Colloquium (LTRC 2002) will be held at the Hong Kong Polytechnic University December 12–15 2002, immediately before the 13th World

Congress of Applied Linguistics (AILA 2002) in Singapore (16–21 December 2002), The Call for Papers has now been put onto the following website: <http://www.engl.polyu.edu.hk/ACLAR/ltrc.htm>

Striving for fairness – the ALTE Code of Practice and quality management systems

NICK SAVILLE, ASSISTANT DIRECTOR, UCLES EFL

All those who are affected by examinations, and above all the examination candidates themselves, should be concerned with fairness. In a general way, and without reference to technical jargon, a candidate is likely to have a good idea of what fairness means; you might, for example, hear the following types of positive comment:

It was easy to get information about the exam

The examination seemed relevant to what I was learning

The cost of the exam was not too expensive

On the day, I knew what I had to do and had enough time to do it

All candidates were treated equally

I felt I could trust the examiners

The result seemed a fair reflection of what I know

My teacher was pleased with the result

I was able to use the certificate to get a job

Negative versions of the same comments might be heard as the result of a bad experience where the candidate felt he/she had been treated unfairly!

For the language tester, these aspects of the candidate's experience are normally assigned to the technical features of the test; for example questions of validity or reliability. However, in recent years, the concept of fairness in its own right has also been discussed by language testing professionals alongside familiar psychometric concepts, such as validity and reliability. Varying views of what fairness means have emerged but typically they relate to concerns such as access, equity, lack of bias, trustworthiness, dependability and "justice for all".

In the fields of psychological and educational assessment, the USA has a long tradition of setting standards as reflected in the latest edition of the *Standards for educational and psychological testing* (AERA/APA/NCME, 1999). The *Standards* have influenced both language test developers and test users and are regularly referred to in the language testing literature (see for example Bachman, 1990 and Bachman and Palmer, 1996). The latest edition includes several new sections which were not in the earlier (1985) version, including Part II of the volume which is entitled *Fairness in Testing*. The sub-sections in this part cover: fairness in testing and test use, the rights (and responsibilities) of test takers, and the testing of individuals from different backgrounds or with disabilities.

More specifically in the field of language assessment there has

been a growing interest in this topic, including the development of codes of practice and an interest in ethical concerns. ILTA – the International Language Testing Association – conducted a review of international testing standards (1995) and in 2000 published its *Code of Ethics* (2000); this document presents a set of nine principles with annotations which "*draws upon moral philosophy and serves to guide good professional conduct.*"

In 2000, a volume of papers edited by Antony Kunnan was published in the Cambridge Studies in Language Testing Series based on the 19th LTRC (Orlando, 1997). This volume, entitled *Fairness and validation in language assessment*, focuses on fairness in two sections: Section One – Concept and Context, and Section Two – Standards, Criteria and Bias. Kunnan himself has been developing a "fairness model" which seeks to integrate the traditional concepts of validity and reliability with newer concepts such as absence of bias, fair access and fair administration. His recent work in this area was presented at the ALTE Conference in Budapest (November 2001) and formed part of a debate within ALTE on fairness and the ALTE Code of Practice.

Within the European context the work of ALTE itself began to exert an influence in relation to fairness and professional standards as early as 1991, when ALTE Members began developing their Code of Practice. At that time it was agreed by the founder members of the association that it was important for both *examination developers* and the *examination users* to follow an established Code of Practice which would ensure that the assessment procedures are of high quality and that all stakeholders are treated fairly. A code of practice of this kind must be based on sound principles of good practice in assessment which allow high standards of quality and fairness to be achieved.

The discussion of what constitutes good practice within ALTE has continued since then and reflects a concern for accountability in all areas of assessment which are undertaken by the ALTE members. In this respect it recognizes the importance of *validation* and the role of *research and development* in examination processes. Over the years ALTE has provided a forum for this work to be carried on.

In 1994, ALTE published its first *Code of Practice* which set out the standards that members of the association aimed to meet in producing their language exams. It drew on *The Code of Fair Testing Practices in Education* produced by the Washington D.C. Joint Committee on Testing Practices (1988) and was intended to be a broad statement of what the users of the examinations should expect and the roles and responsibilities of stakeholders in striving for *fairness*.

The *Code of Practice* identifies three major groups of stakeholders in the testing process:

- **the examination developers**, (i.e. examination boards and other institutions which are members of ALTE)
- **the primary users** – who take the examination by choice, direction or necessity
- **the secondary users**, who are the sponsors of the candidates or who require the examination for some decision-making or other purpose.

In addition the Code of Practice lays down responsibilities of the stakeholder groups in *four broad areas*:

- developing examinations
- interpreting examination results
- striving for fairness
- informing examination takers

An important feature is that it emphasises the joint responsibility of the stakeholders and focuses on the responsibilities of the examination users as well as the *examination developers* in striving for fairness.

While the ALTE Code of Practice outlines the principles in general terms, it provides no practical guidance to the practitioner on the implementation of these principles or how standards can be set and guaranteed. In attempting to address this issue, a supplementary document entitled *Principles of Good Practice for ALTE Examinations* was drafted (Milanovic & Saville, 1991 and 1993) and discussed at ALTE meetings (Alcalà de Henares, 1992, Paris and Munich, 1993). This document was intended to set out in more detail the *principles* which ALTE members should adopt in order to achieve their goals of high professional standards.

The approach to achieving good practice was influenced by a number of sources from within the ALTE membership (e.g. work being carried out by UCLES) and from the field of assessment at large, (e.g. the work of Bachman, Messick and the *AERA/APA/NCME Standards*, 1985). ALTE members sought feedback on the document from external experts in the field (Jan–Mar 1994) and it was discussed again in Arnhem in 1994. While it was not published in its entirety, parts of the document were later incorporated into the *Users Guide for Examiners* (1997)* produced by ALTE on behalf of the Council of Europe.

In 1999/2000 ALTE re-established a Code of Practice working group to take this project forward. In 2001 the working group met three times in January, May and October and reported on progress at the ALTE meetings in Perugia (May 2001) and Budapest (November 2001). So far two substantive outcomes have been achieved and these were discussed by the Membership at the Budapest meeting:

- a) a revised version of *Principles of Good Practice for ALTE Examinations* (2001) has been produced;
- b) the Code of Practice itself has been redesigned and expanded as a *checklist* which is being used by Members as part of a *Quality Management System* (QMS).

The revised version of *Principles of Good Practice for ALTE Examinations* (2001) is based on the earlier version but has been updated and reworked in many parts. It addresses in more detail the central issues of *validity* and *reliability* and looks at the related issues surrounding the *impact* of examinations on individuals and on society. This version, like the earlier drafts, draws on the revised *AERA/APA/NCME Standards* document (1999) – especially in the sections on validity and reliability – as well as the work of Bachman, 1990, and Bachman and Palmer, 1996.

The new dimension which has been introduced into the work at this stage is the concept of QMS – Quality Management System – and related to this is the reworking of the Code of Practice to reflect the *practical* aspects of assessment work within the ALTE membership. The aim of the QMS is to establish workable procedures and programmes of improvement which ultimately will be able to guarantee minimum quality standards based on the Code of Practice.

The approach to Quality Management which is being implemented is based on the following key concepts taken from the literature on QM:

- the organisation
- self-assessment and peer monitoring

In this approach it is important to identify the roles and responsibilities of *key stakeholders* and to apply the system with flexibility according to the *specific features* of each organisation (i.e. the different ALTE members and their stakeholder groups). In the current phase of the project, the working group has been focusing on the different organisational factors within the ALTE Membership and on the range of diversity that exists. In seeking to establish standards it is not the aim to make all ALTE Members conform to the same models of assessment for all 24 languages represented, and it is important to recognise the varied linguistic, educational and cultural contexts within which the examinations are being developed and used. An appropriate balance is required between the need to guarantee the professional standards to users and the need to take into account the differing organisational features of the ALTE institutions and the contexts in which their exams are used.

As a next step, the working group recommended that all members should attempt to identify the current strengths and the areas in need of immediate improvement within their own organisation. On this basis, it will then be possible to establish the desired outcomes for both short- and long-term developments. The aim should be:

- to set *minimum* acceptable standards
- to establish “best practice” models

* This document has been renamed *Language Examining and Test Development* and is being published by the Council of Europe to accompany the Common European Framework marking the European Year of Languages 2001.

- to aim at continuous improvement (move towards best practice)

It is likely that no ALTE member will meet best practice in *all* areas and improvements will always be possible. The aim for all Members should be to raise standards over time, i.e. to aim at the best practice models through an on-going process of development.

In a QM system of this kind, standards are not imposed from "outside" but are established through the mechanism of the system itself and the procedures to monitor standards are based on awareness raising and self-assessment in the first instance. External (peer) monitoring is introduced at a later stage to confirm that the minimum standards are being met. In its current form the Code of Practice has been reworked to function as an awareness raising tool at this stage of the project. In its re-designed format it now reflects the four aspects of the test development cycle with which all ALTE members are familiar:

- Development
- Administration
- Processing (marking, grading and results)
- Analysis and Review

Following the meeting in Budapest in November 2001 it was agreed that all members would complete the procedures to fill in the Code of Practice checklists and will return these to the working group by the end of January 2002. The working group will carry out an analysis and report back to the members at their next meeting in April.

For information on the ALTE five-level scale and the examinations which it covers, visit the ALTE website www.alte.org

or contact

The ALTE Secretariat

1 Hills Road

Cambridge CB1 2EU

United Kingdom

Tel: +44 1223 553925

Fax: +44 1223 553036

e-mail: alte@ucles.org.uk

Investigating variability in a test of second language writing ability

BARRY O'SULLIVAN, UNIVERSITY OF READING, AND CONSULTANT, UCLES EFL

Introduction

One measure of the value of a test is the degree to which construct-irrelevant variance intrudes on reported performance. This variance constitutes a serious threat to the validity of any inferences we wish to draw from test performance, in that it means that the test score is not wholly representative of the ability being measured but is an amalgam of variance that can be attributed to the ability being tested as well as any 'noise' caused by outside factors.

This study was designed to investigate various sources of variability – candidates, examiners, tasks and the rating criteria – within the context of the IELTS General Training Writing Module.

Method

Scripts produced by 180 candidates from two test centres whose candidates represent a typical range of performance were used in this study. The sample was based on a single 'live' version of the IELTS General Training Writing (i.e. 2 writing tasks of 150 and 250 words each). The range of performances was from 2 to 9 on the IELTS scale. 26 trained, experienced and currently accredited IELTS examiners participated in the project. By collecting demographic data from a group of volunteer examiners, a profile was created.

This was used to ensure that the examiners who participated in the project were representative of the typical IELTS examiner population. Examiners awarded separate scores for each of the two tasks using the current IELTS rating criteria. Rating was carried out to a predetermined matrix, to ensure that there was adequate connectivity between raters – a prerequisite for using Multi-Faceted Rasch (MFR) Analysis. This study adopted an MFR Analytic approach using FACETS 3.22 (Linacre 1999). MFR is a probabilistic statistical procedure, which allows the tester to include different factors, or facets, in the calculation of a 'fair average' score for each test-taker. Thus, it is possible to base a candidate's score on their own ability; on the relative harshness and consistency of the examiners whose ratings are being combined to generate this score; on the relative difficulty level of the tasks they attempted; and on any other variables which may be intervening in the procedure (included as facets in the analysis).

Analyses and results

In order to ensure that the data in the final analysis were adequate, a preliminary analysis was first studied using the two 'fit' indices contained in the Facets output. Since the results of 24 candidates were found to be misfitting, in that their indices were outside the

Table 2 Bias Interaction Analysis: reported by rater, task and criteria

Rater	Task 1			Task 2		
	TF	CC	VSS	VSS	AIE	CQ
1						
2				VSS		CQ
3	TF					
4						
5	TF				AIE	
6	TF			VSS	AIE	CQ
7	TF	CC				CQ
8				VSS		
9	TF	CC		VSS	AIE	
10						
11		CC			AIE	
12			VSS			CQ
13						
14				VSS		CQ
15						
16	TF					
17	TF	CC	VSS	VSS		CQ
18			VSS		AIE	
19					AIE	
20	TF			VSS		
21	TF		VSS	VSS	AIE	
22				VSS	AIE	
23					AIE	
24	TF					
25	TF					CQ
26						
Instances of bias	11	4	4	9	9	7
Total possible	26	26	26	26	26	26
% of total	42	15	15	35	35	27
Total potential	156					
Total cases	44					
% of total	28					

Bold = lower than expected *Italic* = Higher than expected

related pattern. However, 11 of the 26 raters (or 46%) seemed to be having a degree of difficulty with the first criterion TF (task fulfilment) in Task 1, while the criterion VSS (vocabulary and sentence structure) appears to have been treated differently in the two tasks to which it applied.

The Tasks

This analysis indicated that there is no significant difference in the scores awarded for performance on the two tasks. This finding raises some interesting implications; it might be expected that the second task (Task 2) would result in lower raw scores since it is actually weighted in the final calculation of the overall writing score. Review of the raw data shows that the difference between the two tasks was one or less bands on 90% of the 1,374 rating occasions included in this study.

The Criteria

In the data used here, examples of 'flat' profiles (i.e. the same score awarded for all criteria) were observed with 22% of the scores awarded for Task 1 and with 29% of the scores for Task 2. In the original dataset the figures were 22% and 38% for Tasks 1 and 2 respectively. It is difficult to make a comparison however, as the original examiners were only instructed to use the profile scoring procedure when there was some problem assigning a score to a script, so it was only done on approximately 30% of the occasions.

One comparison that can be made is to correlate the fair average scores suggested by this analysis with the original scores awarded by the IELTS examiners. While the overall correlation is a very respectable 0.77, the breakdown of correlations by the initial rating approach is of real interest. Here, the correlations are 0.7, 0.76 and 0.81 for the scripts that have received no profile scores, a single profile score and a full profile score respectively. When we consider that these latter scores will have been awarded where the examiners saw some difficulty with awarding a single holistic score, the result is certainly of real interest, as it implies that a move to profile scoring may bring with it a higher degree of consistency.

The Rating Scale

The scale appears to be performing well. Using the Step Calibrations (calibration of the step up to a category) we can plot a graph of how the scale is working, see Figure 1. This shows that the steps seem to have a logical, linear progression through the

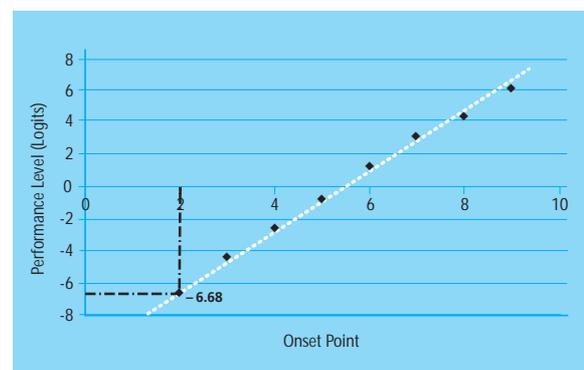


Figure 1 : Graphical representation of scale functioning

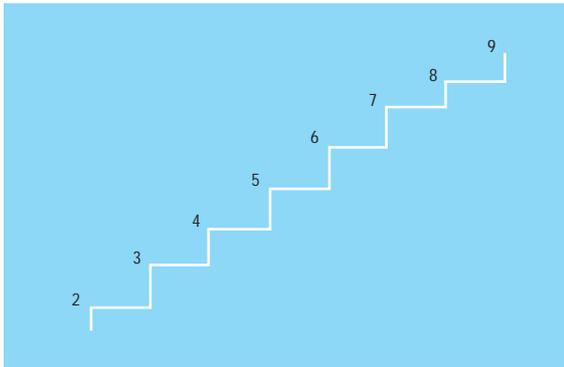


Figure 2 : Step diagram of scale functioning

performance levels from 2 to 9. In order to demonstrate how the graph should be read, one transition point has been highlighted.

From this same graphic, it is possible to create a step diagram of the scale's functioning (Figure 2). Here again we can see clearly that there is a basic linearity (demonstrated by the relative similarity of steps as the performance attracts a higher score). It is also clear that the top end of the scale is not as obviously linear as the lower end, with the scale performing at its best in the middle range of 4 to 7. Since this is the vital range for IELTS (as most decisions are based on criterion levels in the region of 6 to 7), this is a very satisfactory result.

Conclusions

The conclusions will be dealt with in terms of the four sources of variability investigated.

The Candidate

- the test is successful in sorting the candidates into distinct levels of proficiency
- the test is 95% accurate to within less than half of one band either side of the reported score
- error (or construct-irrelevant) variance accounts for only a small proportion of the reported proficiency measure

The Examiner

- the impact of examiner severity on candidate scores is generally insignificant
- 94% of the original IELTS ratings were within one band of the fair average scores suggested by the FACETS programme (based on the multiple ratings in this study)
- the Task Fulfilment criterion on Task 1 appears to cause examiners some difficulty
- no extensive 'flat profile' effect observed

The Tasks

- the levels of performance on the two tasks are very similar
- the tasks themselves seemed to result in consistent performances

The Rating Criteria

- the criterion related to vocabulary and grammar was most harshly interpreted
- the differences across the criteria were not significant
- the scale functioned well (the only area for concern was at the 7.0 level, where the probability of achieving this score was relatively low when compared with the other scores)

Limitations

This study has limitations in relation to the generalizability of the results, with just two centres and 26 examiners. Nevertheless, the fact that the scripts used here were from a genuine IELTS administration and involved the use of an active test version goes some way to lessening the impact of these limitations, as does the fact that all examiners were trained and accredited and a range of candidate ability was represented. Another limitation is that this has been a quantitative study, and as such tells only part of the story. In order to learn more about how the test is working, we need to engage the participants (candidates and examiners) and the test designers to gain insights into what was expected to happen and what actually happened in the test event.

This study highlights the need to think about test reliability/validity from the perspective of quality assurance. The potential of the variables investigated here to result in unacceptably high levels of score variance is certainly a validity issue that needs to be more fully understood, as is the issue of how these variables can affect the reliability of our test scores. This study and other similar studies commissioned by UCLES EFL help to increase our understanding of these important issues, and the findings also feed directly into the ongoing test development and revision process for key tests such as IELTS.

References

- Linacre, M (1999): FACETS computer programme, Chicago: MESA
- Lunz, M E, and Wright, B D (1997): Latent Trait Models for Performance Examinations. In Jürgen Rost and Rolf Langeheine (Eds) *Applications of Latent Trait and Latent Class Models in the Social Sciences* <http://www.ipn.uni-kiel.de/aktuell/buecher/rostbuch/ltlc.htm>
- Myford, C M and Wolfe, E W (2000): *Monitoring Sources of Variability Within the Test of Spoken English Assessment System*, TOEFL Research Report 65, Princeton: ETS

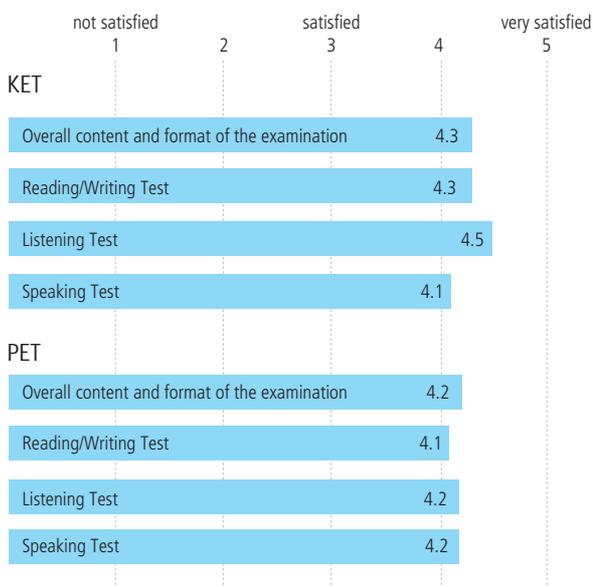
Review of KET & PET Examinations

NIGEL PIKE, SUBJECT MANAGER, UCLES EFL
LIZ GALLIVAN, SUBJECT OFFICER, UCLES EFL

The team responsible for the Key English Test (KET) and Preliminary English Test (PET) have been reviewing these two examinations. This is a standard procedure not only to ensure the tests remain amongst the fairest and most accurate tests within today's global testing market but also to ensure their continued success. (For a fuller description of the UCLES EFL test development and revision process see a previous article in *Research Notes 4*, February 2001.)

As part of the detailed and comprehensive review of KET and PET, stakeholders were asked what they thought about the tests. Teachers, students, Local Secretaries and Senior Team Leaders provided us with valuable feedback about all aspects of KET and PET: the content and format of each paper, the administration procedures and also the extent to which the tests reflect what happens in classrooms around the world.

We received encouraging responses from examination centres in that the tests are meeting the goals of learners around the world. The feeling from the majority of centres was that they are pleased with the current examinations, giving an average satisfaction rating of over 4 (where 1=not satisfied, 3=satisfied and 5=very satisfied). This feedback is summarised in the following table:



The links that we have with teachers, students and, more broadly, with examination centres, are extremely important to us. It is through the dialogue that we have with our customers that we can ensure that our tests reflect and reinforce the best in current classroom practice.

Another part of our review process included research and consultation with other EFL practitioners, including the Senior

Team Leaders (STLs) who are responsible for the training and monitoring of Oral Examiners. We asked them about their overall satisfaction with the Speaking tests; we also gave them the opportunity for more detailed comment on each part of the KET and PET test and the appropriateness of the materials for the age and cultural background of their candidates. Finally, they were asked to comment on the more specific aspects of assessment and examiner training. Their feedback on the Speaking Tests was very positive and results of the survey are reported below (values are on a rising scale of 1 to 5).

Summary of Senior Team Leader Feedback	KET	PET
How satisfied are you with the current Speaking Test?	4.1	4.1
How effective do you feel the interlocutor frame is in KET Part 1?	3.9	
How effective do you feel the personal tasks are in KET Part 2?	3.1	
How effective do you feel the non-personal tasks are in KET Part 2?	3.9	
How suitable do you feel the current format of PET Part 1 is?		3.5
How suitable do you feel the current format of PET Part 2 is?		4.2
How suitable do you feel the current format of PET Part 3 is?		4.6
How suitable do you feel the current format of PET Part 4 is?		4.3
How appropriate is the content and format of the materials for the typical age range of candidates in your country/region?	3.6	4.1
How suitable is the content and format of the materials for the cultural backgrounds of the candidates in your country/region?	3.9	4.4
How appropriate are the current assessment scales and procedures?	3.9	4.5
How satisfied are you with the current training materials?	4.1	4.4

In general, STL feedback on a range of issues relating to the Speaking tests was very positive but it did highlight a few areas for improvement. The changes that we are considering making to KET Part 2 (discontinuation of the personal question type) and to PET Part 1 (the introduction of a more interlocutor led frame) are direct responses to the information received.

All the stakeholders who were canvassed made a number of suggestions for changes to the tests to ensure that they continue to reflect current teaching practice, and this feedback has resulted in the consideration of some new item types. Here is a more comprehensive list of the areas for possible change that were highlighted in the consultation process:

- Introducing short personal messages (such as e-mails and 'post-it' messages) in Reading Part 1 of the PET test, to be used in addition to the public notices/signs/labels and thereby extending the range of genres.
- Constraining the number of words students are required to write in the PET Writing Part 1 transformations to focus the task onto the correct identification of the target structure.
- Replacing the current form-filling task in PET Writing Part 2 with a guided writing task with a strong communicative purpose. This would expand the range of text types that PET students produce in the writing component, in line with the feedback we received from schools on what actually happens in the classroom.
- Making the start of the PET Speaking test more interlocutor-led.
- Having only Non-Personal Question types in Part 2 of the KET Speaking test

- Introducing a vocabulary testing exercise that will expand on the current KET Part 2 Reading task.
- Broadening the range of input and output text types in the KET Part 7 Information Transfer task to include such texts as notices and diary entries.

You will see from the above list of possible changes that any modifications to the tests will be minor, and that some papers may remain completely unchanged.

All of the proposed changes were suggested by the people who teach and administer KET and PET. Before any of the changes can be implemented, however, new task types are being extensively trialled.

PET was last revised in 1994, while KET has only been in existence since November 1994. The review process is very carefully considered and we are very aware that any modifications to our tests have a wide impact – on teachers, candidates and publishers. With this in mind, we give test-users at least 2 years' notice of any changes to the test specification, and the first administration of updated papers for KET and PET is likely to be March 2004.

The UCLES EFL staff seminar programme

In November 2001 Dr Antony Kunnan of California State University, Los Angeles, ran a series of workshop sessions for EFL staff on Structural Equation Modelling. He also presented a seminar on Test Fairness where he explored the relation between test fairness and test development. Dr Kunnan argued that test fairness concerns should be examined at all stages of test development. Reviewing the traditional definitions for fairness, he reiterated the importance of four major fairness issues listed by APA Standards (1999): fairness as *lack of bias*; fairness as *equitable treatment in the testing process*; fairness as *equality in outcomes of testing*; and fairness as *opportunity to learn*. He touched upon the importance of absence of bias from test content as a criterion for quality fairness where the following should be controlled: bias due to inappropriate selection of test content; bias due to different focus in curriculum in achievement testing; and bias due to a lack of clarity in test instructions or from scoring rubrics that credit responses more typical of one group than another.

Dr Kunnan proposed a Fairness Model where test fairness qualities could be examined under three main headings: *main quality*, *focus points* and *evidence generation methods*. Main

quality concerns issues such as construct validity, access, administration, absence of bias and value, impact and responsibility. Focus points is related to a particular quality being examined, i.e., content-related validity. Evidence generation methods is concerned with the methods employed to study each point e.g. expert judgements, checklists, external comparison, statistical analysis, etc. The proposed Model offers an integration of traditional concepts of validity, reliability, and practicality with newer concepts such as absence of bias, fair access and fair administration. The Test Fairness Model also suggests evidence generation methods for the different test fairness qualities. The Model provides a framework for test developers to conduct a self-examination of their test development systems in terms of fairness issues.

Following his visit to Cambridge Dr Kunnan attended the ALTE Conference in Budapest, Hungary, where he presented his Fairness Model in a plenary address. In his talk, Dr Kunnan sought to relate his approach to the issue of fairness as it is dealt with in the ALTE Code of Practice (see also Nick Saville's article on page 12).

Report on the BAAL/CUP Seminar 'Young Language Learners: Towards a Research Agenda'

FIONA BALL, VALIDATION OFFICER, UCLES EFL

In June 2001 the Language and Literacy Studies Research and Teaching Group at Manchester University held its inaugural seminar. This event was organised by the British Association of Applied Linguistics and sponsored by Cambridge University Press. UCLES' interest in this seminar stemmed from our recent expansion into EFL tests for children and an ongoing interest in related teaching and learning issues. Five papers were presented, interspersed with discussion on the similarities and differences between foreign language teaching and the more marginalised teaching of 'community' languages in various countries and sectors. This article reports briefly on the content of the seminar and considers the contribution UCLES EFL has to make in this field.

1 What can YLL research show? What is the way forward?

Professor Katharine Perrera (vice-chancellor, University of Manchester) suggested that boundaries are changing for everyone involved in language teaching. The issues of globalisation and the requirement for an additional language were raised. Professor Perrera confirmed that the early years are the most important for language learning, an idea reiterated by other speakers.

2 Languages at primary school as a matter of national and international policy

Professor Dick Johnstone (director, Scottish Centre for Information on Language Teaching & Research) spoke on the key conditions for success in language learning in the primary context. Professor Johnstone suggested that the underlying language competence of children is encouraged through socialising activities rather than formal pedagogy. Worryingly, he found little evidence of primary-secondary development in foreign language ability, perhaps due to children not being aware of the language system or lacking creativity due to learning a foreign language in chunks. On a more positive note, Professor Johnstone noted that children develop language awareness through learning another language which sensitizes them to learn more about language in general. At secondary level attainment rather than proficiency is the current aim, taught by rote learning. Secondary children know that they are not proficient (e.g. cannot interact on school trips), meaning that there is little evidence for second language acquisition in secondary classrooms.

Professor Johnstone concluded that there are several important factors for teaching children languages:

- Starting young
- Time and intensity of teaching
- Use of computers
- Children's experience of languages at home, with relatives etc
- Encourage cultural exchanges between schools in different countries

3 Research on Hungarian young learners: What has been done and what needs to be explored

Dr Marianne Nikolov (University of Pecs, Hungary) explained that there was no formal evaluation of language teaching in Hungary where up to two modern foreign languages (MFLs) are taught in primary school (age 6–14). There are differences between national and regional curricula and no correlation between weekly hours and performance. In Dr Nikolov's experience, children become unwilling to use a second language at around 11 years old and the mother tongue is used to control children and to negotiate roles in activities. The problem of measuring outcomes was a principal concern in this educational system.

4 Teaching modern foreign languages in primary schools: A question of experience

Dr Patricia Driscoll (Canterbury Christ Church University College) described the realities of MFL teaching in primary schools. She found teacher experience to have the main effect on teaching style and outcomes in two very different Local Education Authorities. Using interviews and observation techniques, Dr Driscoll found that children wanted to learn more MFLs than were available to them and that teachers' 'mental scripts' of past lessons, subject knowledge and beliefs influenced their teaching style. Body language and context were also found to be important. Comments from children suggested that they had strong views about their language learning experiences, however had no forum in which to raise them. Dr Driscoll concluded that there is a need to embed languages in primary culture.

5 Schooling the mind: Children, Language and Learning

The highlight of the seminar was Professor Ruqaiya Hasan's presentation (Macquarie University, Australia). Professor Hasan suggested that babies from two weeks of age are designed to learn, so it is up to teachers to find out what students know and to be

responsive to them as language learners. Learning is the interaction of the learner's mind with problems. Language is both meaning potential (the experiential function of language) and relationships (interpersonal function) – the two functions must work together for development to occur. Children do not create ambiguities as they require themselves to actively participate with others in a context of situation. Professor Hasan concluded that the problem with current teaching practices is that programs designed to assess language (words and patterns) are not designed to find out what a child can do with language.

The ways forward: setting the research agenda

At the end of the seminar delegates revisited the themes raised during the day and suggested the following key issues for a research agenda:

- Learning a language must be viewed as appropriate and useful to learners and teachers;
- Educational policy must consider the context and conditions of individual regions and communities;
- Parental, children's and other views should be taken into account for both policy and practice;
- A bottom-up approach to policy is preferable to imposition by local or national agencies.

Conclusion

This event provided a forum for informed discussion of current educational issues focussing on the policy and practice of teaching and assessing young language learners. The general feeling was positive in terms of what needs to be achieved, combined with a sense of the enormity of the task ahead for policy makers, practitioners and researchers in this area. There is a lack of research, for example, on language awareness and children's experiences of language learning.

In relation to our own work, the seminar confirmed that we are continuing to develop the Cambridge YLE tests in line with current

thinking about how children learn a foreign language and that the tests provide appropriate support for children as they learn how to communicate effectively in another language from an early age. They aim to fulfil the 'key conditions for success' by including appropriate child-centred activities and topics and by focusing on meaning and social interaction. Furthermore, performance by young learners in our YLE tests has the potential to provide a rich source of data for investigating second language acquisition and learning among children which could significantly help to increase understanding in the field. Our knowledge and experience of language testing in general, and of young learner assessment in particular, mean that we have a valuable contribution to make in the ongoing debate about YLL.

References and further reading

- BAAL: <http://www.baal.org.uk>
- UCLES YLE: http://www.cambridge-efl.org/exam/young/bg_yle.htm
- Language & Literacy Studies in Education:
<http://www.education.man.ac.uk/rgs/>
- Nuffield Languages Inquiry:
<http://www.nuffieldfoundation.org/languages/inquiry/in0000000555.asp>
- Andrews, J and Chohnacka, E (1998): "Washback and wash-forward in language testing", in *Conference Papers: Innovations and Outcomes in English Language Teacher Education*, British Council, pp. 221–230
- Aubrey, C, David, T, Godfrey, R, and Thompson, L A (2000): *Early Childhood Educational Research: Issues in Methodology and Ethics*. London & New York: Routledge Falmer
- Barton, A (2000): "Gender differences in pupils' perceptions of modern foreign languages", in Chambers, G (ed) *Reflections on Motivation*, CILT
- Cook, G (2000): *Language Play, Language Learning*. Oxford: Oxford University Press
- Moon, J and Nikolov, M (eds) (2000): *Research into Teaching English to Young Learners, International Perspectives*, Pécs: University Press
- Thompson, L A (ed) (1997): *Children Talking: The Development of Pragmatic Competence*, Clevedon: Multilingual Matters

Studies in Language Testing

Qualitative approaches to language test validation are now making a much more significant impact on the field of language testing. We have tried to emphasise the important role such approaches can play by publishing volumes in the *Studies in Language Testing* series, e.g. 'Verbal protocol analysis in language testing research: a handbook' by Alison Green. The latest volume in this series is entitled 'A qualitative approach to the validation of oral language tests', and focuses on the area of speaking assessment.

We are especially pleased to be able to publish this volume by

Professor Anne Lazaraton, who has been working closely with staff at UCLES over the last ten years. Her contributions to our work have not only been stimulating in the academic sense but have also made a very valuable contribution in practical and extremely important ways. They have, for example, helped UCLES staff in the development and revision of speaking tests both in relation to content and also in relation to the procedures which are necessary to monitor and evaluate how oral assessments are carried out.

Direct oral assessment is one of the cornerstones of the UCLES

approach to language testing. However, it is well known that direct assessment is fraught with difficulties. At UCLES we believe it is important that we work towards a better understanding of these difficulties and seek to manage and control them in the most effective way. The Performance Testing Unit, part of the Research and Validation Group within the UCLES EFL Division, is specifically charged with conducting research and co-ordinating the research of others to further our capability to carry out direct assessment in speaking and writing as effectively as possible. The task is on-going but we can see clearly how the quality of our assessments has improved over the years and continues to do so.

Professor Lazaraton's research related to Cambridge EFL examinations has engaged with a number of assessments and has built on work conducted by the UCLES EFL Division. Between 1990 and 1992 she worked closely with the UCLES team on the Cambridge Assessment of Spoken English (CASE). This assessment was developed largely as a research vehicle and Professor Lazaraton's work focused on using a qualitative discourse analytic approach to further understanding of the speaking test process with particular reference to the role of the examiner. The work subsequently contributed significantly to the development of monitoring procedures for a wide range of Cambridge examinations.

The work on CASE was followed by work on the Certificate in Advanced English (CAE), situated at Level 4 in the Cambridge/ALTE level system. This research was intended to evaluate interlocutor adherence to the CAE *interlocutor frame* and to analyse

interlocutor speech behaviour, which led to the development of the CAE Examiner evaluation template. Professor Lazaraton then conducted similar work in relation to the Key English Test (KET) at Level 1 in the Cambridge/ALTE level system and comparative research across the two levels.

Professor Lazaraton also carried out a number of studies focusing on *candidate* behaviour, as opposed to examiner behaviour, in speaking tests. This work focused on CAE, the First Certificate in English (FCE) and the International English Language Testing System (IELTS). The work on candidate behaviour started with a CAE study and was followed by one on FCE, the most widely taken UCLES EFL examination. Professor Lazaraton investigated the relationship between the task features in the four parts of the FCE Speaking Test and candidate output in terms of speech production. In 1997 Professor Lazaraton was asked to work on IELTS again with particular reference to candidate language. This work made a valuable contribution to the revision of the IELTS Speaking Test, introduced in July 2001.

Anne Lazaraton has always understood the tensions that exist between researching issues in language testing and delivering reliable and valid language tests. While situated firmly at the research end of the language testing cline, her energy, enthusiasm and openness have meant that she has been able to share with us what she has to offer, which in our view is of enormous value. Her work emphasises the value of building research into the on-going validation and improvement of language testing tools and procedures, leading to assessments of ever-improving quality.

Announcement of the winner of the IELTS MA Dissertation Award 2001

As part of the tenth anniversary of IELTS in 1999, the IELTS partners – UCLES, The British Council, and IDP Education Australia – agreed to sponsor an annual award of £1000 for the MA dissertation in English which makes the most significant contribution to the field of language testing. In its inaugural year the award went to joint winners in Australia and Canada. In 2001 the award went to a Korean student studying at the University of California, Los Angeles (UCLA).

For the award in 2001, submissions were accepted for dissertations completed in 2000. The IELTS Research Committee, which comprises members from the three partner organisations, met in October 2001 to review the shortlisted submissions. The Committee was extremely impressed at the high standard of all the shortlisted dissertations which, it was believed, reflected well on their authors and on the universities where they studied. The research topics were varied and included the four macro-skills of listening, speaking, reading and writing, as well as interpretation of

test results. It was agreed that the candidates and their work are indicative of the considerable worldwide interest in language testing. All the dissertations merit a wider audience and the findings are publishable.

After careful consideration, the Committee decided to announce one winner : Sang Keun Shin – *An Exploratory Study of the Construct Validity of Timed Essay Tests*.

In addition, it was felt that two other dissertation authors should be mentioned for the quality of their contributions: Eleftheria Nteliou – *UCLES 'Main Suite' Speaking Tests: Describing the Test-takers' Language Output in terms of CALS Checklist of Operations at KET and FCE Levels* (Reading University, England) and Nick Boddy – *The Effect of Individual Interpretation of the Elicitation Phase of the IELTS Speaking Test on its Reliability* (Macquarie University, Australia).

The Committee considered the winning dissertation by Sang Keun Shin to be an excellent example of applied linguistics

research within the language testing domain. This dissertation reported on an exploratory study examining the construct validity of timed essay tests by comparing the composing processes of L2 writers in test and non-test situations. Following a case study approach, Sang Keun Shin controlled for cultural differences by limiting the study to Korean subjects. While the researcher investigated the writing process of the subjects in authentic, non-timed situations, the timed-test situations were simulated. The findings are important and have a direct bearing on test design.

The dissertation was extremely well-crafted in all respects. The research rationale was clearly stated and justified and the research questions, presented as objectives, were coherent, well organised and highly perceptive. The study was well-situated within the body of research literature, the review of which was both thorough and well-argued. The methodology was imaginative, appropriate and clearly explained. Moreover, the research design demonstrated a good control of variables and a clear understanding of design limitations. The findings and their interpretations were presented as thoughtful, practical and well-reasoned conclusions.

The Committee believed that the study constituted a significant contribution to an understanding of the differences and similarities between the processes L2 learners use when composing academic

course papers and writing for 'sit-down' timed writing essays; it also has relevance for an understanding of the key issues relating to construct validation.

The abstract from the award-winning dissertation is presented below :

Sang Keun Shin : An Exploratory Study of the Construct Validity of Timed Essay Tests

This exploratory study examined the construct validity of timed essay tests by comparing the composing processes of L2 writers in test and non-test situations. Five second language writers who were native speakers of Korean participated in the study. They retrospectively reported their drafted composing processes in which they engaged when producing their papers for their courses. They were also asked to write timed impromptu essays on an assigned topic and then to recall their composing processes they went through.

The results of this study indicated that the composing processes in test and non-test situations are different, thus supporting claims that essay tests may not elicit samples of writing which truly represent test takers' writing ability because writing done in such a controlled context implies a composing process that is radically different from the test takers' normal writing processes.

IELTS MA Dissertation Award 2002

For 2002, the entry procedures and timetable for the award are given below:

Submission and evaluation procedures

The full dissertation abstract, accompanied by both the Introduction and Method chapters together with a reference from your supervisor, should be submitted to:

Dr Lynda Taylor / Stuart Shaw

EFL Division

University of Cambridge Local Examinations Syndicate

1 Hills Road

Cambridge

CB1 2EU

United Kingdom

- The IELTS Research Committee, which comprises members of the three partner organisations, will review the submissions and shortlist potential award winners.
- For all shortlisted dissertations a full copy of the dissertation will be requested and a further reference may be sought.
- Shortlisted dissertations will be reviewed and evaluated by the IELTS Research Committee according to the following criteria:
 - Rationale for the research
 - Contextualisation within the literature

- Feasibility of outcomes
- Design of research question(s)
- Choice and use of methodology
- Interpretation and conclusions
- Quality of presentation
- Use of references
- Contribution to the field
- Potential for future publication
- The Committee's decision is final.

Timetable

The following timetable will apply in 2002 :

1 June Deadline for submission of dissertation extracts and references to UCLES

1 August Deadline for submission of full copies of shortlisted dissertations (and further references if required)

October/November Meeting of IELTS Research Committee

November/December Announcement of award

Details of the application process for the IELTS MA Dissertation Award 2002 can also be found on the IELTS website – www.ielts.org

Reader Questionnaire

Research Notes has reached its seventh edition since its first issue in March 2000. It has an increasing readership throughout the world of teachers, administrators and other language testing professionals.

UCLES EFL is interested in your views about the content and approach of *Research Notes*. Please take the time to complete and return this short questionnaire about your background, interest in and opinion of *Research Notes*. Your response will help to inform the future development of this publication and will provide UCLES EFL with a clearer picture of the needs and interests of its audience. This questionnaire can also be found on *Research Notes* website at: http://www.cambridge-eft.org/rs_notes/

Please send your response to UCLES EFL by the end of May 2002. A summary of responses to this questionnaire will be included in Issue 9 of *Research Notes*, to be published in August 2002.

Fax the completed questionnaire to +44 1223 460278 or send the completed questionnaire to:

RN7 Questionnaire
EFL Information
University of Cambridge Local Examinations Syndicate
1 Hills Road
Cambridge
CB1 2EU
United Kingdom

What is your main occupation?

- Lecturer Researcher Teacher Administrator
 Other – please specify

.....
.....

How many issues of RN have you read?

- this is my first one 1–2 3–4 5–6

How did you find out about RN?

- Direct mailing to a centre/school/institution/other
 Website
 Word-of-mouth
 Conference/promotional presentation
 Other – please specify

.....
.....

Have colleagues been interested in RN?

- Yes / No

How many people have shared this copy?

- none 1 2–5 6 or more

In what ways have you found *Research Notes* interesting or informative?

.....
.....
.....

How do you find the spread of topics?

- Inadequate Adequate More than adequate

What topics would you like to see covered in future issues?

.....
.....

Do you prefer a mixture of topics or a theme for each issue?

- mixture themed issue

Have you accessed further information from UCLES after reading RN? Yes No

How did you obtain further information?

- website writing phoning email
 other – please specify

.....
.....
.....

Did you receive sufficient information to answer your query?

- Yes No

Any other comments on *Research Notes*.

.....
.....
.....

Thank you for your time.



For further information:

EFL Information
University of Cambridge
Local Examinations Syndicate
1 Hills Road
Cambridge CB1 2EU
United Kingdom
Tel: +44 1223 553355
Fax: +44 1223 460278
e-mail: efl@ucles.org.uk
www.cambridge-efl.org

