
Research Notes

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



UNIVERSITY *of* CAMBRIDGE
Local Examinations Syndicate

ResearchNotes

Contents

Introduction	1
EFL Research at UCLES	2
The EFL Local Item Banking System	5
Developing Language Learning Questionnaires (LLOs)	6
Issues in Speaking Assessment Research	8
The UCLES/CUP Learner Corpus	10
Studies in Language Testing	11
Further Information	12
The UCLES EFL Test Development and Validation Group	

Introduction

Welcome to *Research Notes*, the newsletter on current developments in the research, validation and test development work carried out by the EFL Division at UCLES. This re-launched edition of *Research Notes*, is the successor to the publication of the same name that UCLES published in the early 1990s.

In this issue, we give an overview of the work of UCLES' EFL Test Development and Validation Group, as well as looking in more detail at a number of other topics, including:

- the EFL Local Item Banking System
- Developing Language Learning Questionnaires
- Issues in Speaking Assessment Research
- the UCLES/CUP Learner Corpus
- the Studies in Language Testing series

In future issues, we will aim to provide information about the full range of our research activities and to report progress on them as they develop and results become available.

Research Notes is intended to reach a wide audience of those involved in the Cambridge examinations around the world and more generally those people who are interested in the theoretical and practical issues related to language assessment.

Research Notes is being distributed to all UCLES EFL centres and other key contacts. If you would like to receive additional copies or if you would like a personal subscription to the newsletter, please complete and return the form on page 12. We would also be interested in receiving your feedback on *Research Notes* – whether you find it interesting and useful, how appropriate you find the level of presentation and what else you would like us to cover.

EFL Research at UCLES

UCLES established a dedicated EFL Evaluation Unit in the Summer of 1989. This unit, headed by Dr Michael Milanovic, working with Nick Saville, was given the role of establishing a validation programme and research agenda specifically focusing on the EFL examinations. It was the first unit of its kind within a UK EFL examinations board.

In the first year or so of the Unit's existence, the work of Lyle Bachman and colleagues on the Cambridge TOEFL Comparability Study influenced a number of the activities of the Evaluation Unit. This study, which was carried out between 1987 and 1989 is described in full in the first volume of the UCLES/CUP series in Language Testing (see page 11).

There were, however, many new projects which emerged in the early 90's as the result of work being carried out at that time within the EFL Division itself. In his series editor's note in volume one of the series (1996), Michael Milanovic was able to describe seven key areas in which major advances had been made over a period of 5 or 6 years. These are as follows:

- major revisions of existing exams and development projects for new exams
- increased pretesting of materials and the setting up of an electronic item banking system
- the rationalisation of data capture to allow for routine analysis of examinations
- research on the triangulation of test content, candidate background and test performance
- the rationalisation of systems to support oral examiners throughout the world
- research projects into the direct assessment of spoken and written skills
- item writer training programmes and increased investment in the development of key personnel

Since the publication of the first volume in 1995, seven more volumes in the *Studies in Language Testing* series have appeared in print and a further four are now nearing completion for publication in 2000. The full list is given on page 11 of this newsletter and this provides an idea of the scope of work which the validation team is involved in these days.

Summer 1999 marked the 10th anniversary of the setting up of the Test Development and Validation Group, as it is now known, which expanded to 25 staff. The following list of projects gives an idea of the range of activities undertaken by members of the Group:

The Framework Project

A key objective was to establish a framework of five levels against which our examinations could be placed. This has led to the Cambridge 5-level system of examinations – KET, PET, FCE, CAE and CPE. In order to ensure that each level was suitably differentiated, a

system for describing, calibrating and storing test materials and information about test takers was required. This work covers the following areas:

UCLES Common Scale Project (5 level system)

Calibration of all item-based materials and equating of tests (e.g. by investigating candidates doing two exams, and the use of specially prepared anchor tests)

Item banking

Development of the Local Item Banking System (LIBS)

Development of a descriptive system for item based tests

Investigation of candidate background in relation to performance on tests

The Work of ALTE and links with other international organisations (e.g. Council of Europe)

In 1990 UCLES began collaborating with other institutional providers of language examinations within the context of the newly formed ALTE – the Association of Language Testers in Europe. This work has coincided with UCLES' interest in a framework of language proficiency (as described above) and has led to collaboration on a range of related projects within the ALTE group and with other international organisations, such as the Council of Europe and EAQUALS. The following list includes some of the key areas of activity which have been carried out:

ALTE and Common European Framework

ALTE *CAN DO* project

Development of *CAN DO* scales

Validation of the scales

Linking learner-responses to their performance on examinations

Linking ALTE Can Do Statements to the Common European Framework

Production of Multilingual Glossary of Testing Terms in 10 languages

Production of guidelines for training item writers, including the Council of Europe *Users Guide for Examiners* as supplement to the Common European Framework

Development of Content analysis checklists for analysing and comparing examinations

An evaluation of the Council of Europe's Vantage Level

Validation of Item-based tests – Reading, Listening and Language Systems

UCLES EFL examinations include many item-based tests to assess reading, listening and language systems (grammar and vocabulary). Over the past 10 years, systems have been developed which allow routine analysis of all materials used in the examinations to take place at the level of individual items. In addition, systems to collect and store data relating to the candidates who have taken the examinations have been put in place. This has allowed for the following activities to take place:

Development of improved systems for data collection and management (i.e. score data and demographic data)

Routine analysis and evaluation of test materials and candidate performance including: routine Item analysis, reliability studies including use of generalizability theory

Investigations into the factors affecting the performance of candidates taking Cambridge exams:

Candidate background and test performance

Development and validation of specialised questionnaires focusing on

Candidate background information

Attitudes of stakeholders

Language Learner Questionnaires (see page 5)

Investigations using advanced analysis techniques

Item level factor analysis

Structural Equation Modelling (EQS)

Qualitative methods of validation making use of protocol-based studies to investigate examiner and candidate behaviour (e.g. a pilot project focusing on CAE Reading – Paper 1)

Research into Performance Tests (Speaking and Writing)

A key feature of the Cambridge examinations is the assessment of speaking and writing ability and therefore research and development in these areas have been a major focus. Some of the issues relating to research in the area of speaking are described in more detail by Lynda Taylor on page 8.

Focus on Speaking

The Cambridge Approach to the Assessment of Speaking (e.g. article by Saville and Hargreaves 1999)

The recording/archiving of Speaking Tests including the setting up of a digitised corpus of Cambridge speaking tests

The analysis and transcription of Speaking Tests Research reported in article on FCE/CPE speaking tests by Young and Milanovic (1992)

Studies using Conversation Analysis of speaking test data by Anne Lazaraton on CASE, CAE, KET, FCE, IELTS (*Studies in Language Testing* volume under preparation, see page 11)

Discourse and Grammatical analysis of speaking test data (e.g. working with Alan Tonkyn of Reading University)

The setting up of the professional support system for Oral Examiners

Improved systems for monitoring examiners e.g. using data collected from mark sheets and from examiner monitoring checklists

The development of a Common Scale for Speaking Tests

Specific rating scale development and validation projects include the CASE project 1991-94), the FCE revision (1994-6), the IELTS and CPE revisions (1999 ongoing), use of Rasch-based analysis – partial credit, FACETS).

Focus on Writing

Development of a Common scale for writing

Development of the Learner Corpus with CUP (see page 10)

Marker strategies projects (e.g. see *Studies in Language Testing* 3 Milanovic, Saville and Shen 1996)

Investigation into the computer-marking of compositions (ongoing 2000)

Consequential Validity (e.g. Impact Studies)

Tests provided by major examination boards have an impact on educational processes and on society in general. While the effect of tests on language teaching and learning has received some limited attention in recent years, the more general concept of impact and how it may be investigated has not been systematically addressed in this field. The concepts of consequential validity and test impact are being addressed by UCLES EFL within a long-term research programme. In particular, we have set out to investigate the impact of IELTS which is currently taken by candidates seeking admission to higher education in the UK, Australia, Canada and the USA. This forms part of the IELTS Research Programme.

In order to understand the impact of an international language test or examinations like IELTS, instruments and procedures need to be developed which focus on areas which are crucial in considering the impact of language tests: the content and nature of classroom activity in test-related classes; the content and nature of teaching materials based on the test; the views and attitudes of user groups towards the test; the test-taking population and the use of test results. The first two of these areas concern the effects of the test on other systems, such as the administrative and academic contexts where the test results are

used, and on the attitudes and behaviour of stakeholders.

In relation to the attitudes and behaviour of test takers, the development of Language Learning Questionnaires has been an important project in recent years.

Projects in this area include:

IELTS Impact Studies – 4 sub-projects 1996-99

Language Learning Questionnaires (working with UCLA and Jim Purpura - see page 6 and *Studies in Language Testing* Volume 8)

Technology in Language Assessment

This area includes Computer-based testing – both development and validation. Projects include Linguaskill, CommunicAT and BULATS Computer Test. Other areas of investigation include the automated marking of writing and administration of speaking tests via technology rather than in a face-to-face context with an examiner.

The IELTS research programme

IELTS, (the successor to ELTS) celebrated its 10 anniversary in 1999. IELTS is jointly owned by UCLES EFL, the British Council and IDP Education Australia and as part of the collaboration between these three parties, further developments have been introduced in the 1990s (e.g. IELTS 1995) and a research programme related to IELTS has been established. Some of the activities which have been carried out include:

Revision of IELTS, leading to IELTS 95

Work of Caroline Clapham on testing academic reading (as reported in *Studies in Language Testing* Volume 4)

Research into the use of the General Training Module (1999 ongoing)

Establishment of an IELTS Research Committee to handle the publication of IELTS Research Reports and the research programme funded by the IELTS partners (including the enhanced distribution of the call for research proposals from 1999)

Principles of test development

UCLES also carries out research into theoretical and methodological issues underlying the development and construction of language testing instruments. Examples of this work include:

Development of a principled approach to test development and validation – e.g. the model-based approach

Development of Item writer Guidelines, especially in collaboration with ALTE partners

Applying the principles of test development to consultancy projects, e.g. the PETS Project in China, Projects in the Baltic States, Mexico SEP Project, etc

Reporting on research and validation

A final, vital element in the research programme is the publication of research outcomes. This includes a large number of papers in various journals and conference proceedings, as well as the series *Studies in Language Testing* (see page 12).

References:

Saville, N. and Hargreaves, P. 'Assessing speaking in the revised FCE' in *ELT Journal* Volume 53/1, Oxford 1999

Young, R. and Milanovic, M. Discourse variation in Oral Proficiency Interviews', in *Studies in Second Language Acquisition* Vol. 14 No.1, Cambridge 1992

The EFL Local Item Banking System

Simon Beeston, EFL Pretesting Manager, UCLES

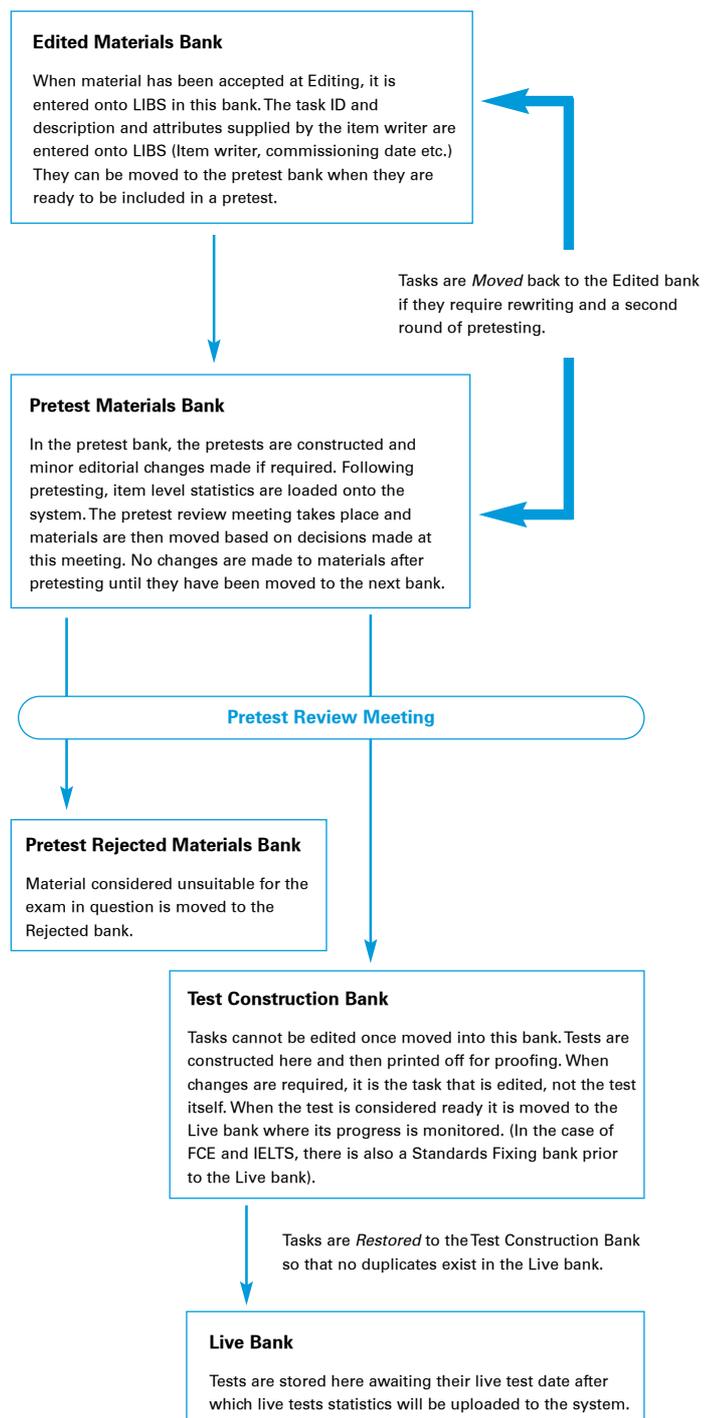
The development of Item Banking is one of the most important developments in recent years. It gives test developers immediate access to a large supply of pretested test items, the characteristics and performance of which has already been established. This allows for more efficient test development and provides a highly significant resource for research projects.

The Local Item Banking System (LIBS) is a software application that has been designed specifically for the UCLES EFL examinations. The system is designed to provide a comprehensive database of all current EFL material whilst supporting all of the activities typically associated with test development such as editing, pretest review and test construction.

Since August 1998, the EFL Test Development and Validation Group have been entering exam material onto the Item Bank and training staff in order to support the introduction of LIBS, which will house all of the material used for the Cambridge EFL Main Suite examinations, as well as IELTS, the Business English Certificates and a number of other examinations.

LIBS has a number of important benefits for UCLES, allowing the rapid construction of examination papers and providing an enormous resource for research projects. Amongst the most important benefits are:

- All EFL examinations can be stored and constructed using a standard procedure;
- Copies of tasks are stored for each stage in the test development process so that, for example, the pretest version can be compared with the live test version if required;
- Test statistics are automatically loaded onto the system and then checked by Validation staff;
- Tests can be created already formatted with item and section numbering automatically added to the test;
- Reports can be routinely generated to provide, for example, test construction reports, answer keys, pregrading estimates, a comparison of pretest and live test statistics, etc.
- Each task has a set of customised attributes which are automatically generated for the task itself and for its items;
- Question paper preparation schedules are automatically calculated on the basis of the required date with e-mails automatically sent to designated users to warn them of late papers;
- Each task has a documented history showing which banks and which tests it has been in;
- All test material is stored on a secure server with only authorised staff having access to the specific banks on which they work.



Developing Language Learning Questionnaires (LLQs)

Nick Saville, Manager, EFL Test Development and Validation Group, UCLES

For several years, UCLES has been working with researchers in the UK and North America to develop research instruments that will enable us to develop a better understanding of the effects of the psychological, social and cultural contexts in which assessment takes place. This is an important and under-researched subject, and one which is particularly important in the case of the Cambridge EFL examinations which are designed to assess communication skills that are of direct relevance to the test takers' educational, professional and social lives.

This article looks at one of the key elements in this research programme – the development of a bank of Language Learning Questionnaires (LLQs) – and in particular the collaborative work between UCLES EFL and Dr Jim Purpura from Teachers College, Columbia University (New York). This work will become an increasingly important part of UCLES' research and development agenda and of projects such as the revision of the Certificate of Proficiency in English.

As part of this effort, UCLES has been developing and validating a bank of questionnaires to investigate the background characteristics of the UCLES candidature in relation to learning strategies and styles. These questionnaires will ultimately be used alongside examinations and tests in order to examine the relationships between test taker strategies and styles and their performance on language tests and on self assessment instruments (what learners think they can do in their target language).

In 1991 UCLES EFL together with the Language Assessment Working Group at UCLA (the University of California at Los Angeles) began the development of a bank of questionnaires which were designed specifically to investigate the background characteristics of Cambridge EFL Examination candidates, and to examine how these characteristics relate to performance on the Cambridge exams.

Purpura, together with Lyle Bachman and Sarah Cushing Weigle at UCLA identified two sets of background factors that were hypothesised to affect second language acquisition – and by extension, second language test performance. The two factors are grouped as Strategic and Socio-psychological and can be further divided as follows:

a. Strategic factors

- (i) Cognitive strategies
- (ii) Metacognitive strategies
- (iii) Communication strategies

b. Socio-psychological factors

- (i) Attitudes
- (ii) Anxiety
- (iii) Motivation
- (iv) Effort

At that stage of the research, *personality factors* (e.g. self-esteem, extroversion) and *cognitive style factors* (e.g., field dependence/independence, visual/aural, analytic/gestalt) were not included in the design for reasons of feasibility. However, these factors have drawn a considerable amount of research attention in the past few years and might be a possible direction for further questionnaire development. In addition, demographic factors were excluded since UCLES had already implemented a questionnaire – the Candidate Information Sheet – to collect these data on a routine basis.

The UCLA team proposed taxonomies for each of the seven questionnaire item banks, developed pilot versions of the questionnaires, administered the pilot questionnaires to students, and analysed the initial questionnaire data by means of reliability analyses and produced a revised pilot version of the questionnaires.

The process of developing these questionnaire item banks was documented in an internal report produced by Bachman, Purpura and Weigle (1992) and reported in *Research Notes* in September 1992. UCLES then revised these questionnaires (especially in relation to British English usage) and administered them to a number of candidate groups in order to collect more data from the target population. Since then Jim Purpura has continued the work on validating the questionnaires.

Strategic Factors

In the first instance Purpura concentrated on the strategic factors and analysed the data on *cognitive* and *metacognitive strategy* use using exploratory factor analysis (EFA). In this way he was able to reduce the number of questions to include only those items that best measured the underlying traits. The results of these analyses were reported at the 1995 Language Testing Research Colloquium and a paper on this research is in the 1995 LTRC Proceedings, *Fairness and Validation in Language Assessment* edited by Antony Kunnan (1998) and published as part of the series *Studies in Language Testing* (see p. 11).

Purpura then used the results of these analyses to produce a new version of the questionnaire. With this he collected questionnaire data together with candidates' responses to an FCE Anchor Test from three countries. Using these data he carried out a construct validation study in which he looked at the relationships between strategy use and performance on second language tests. The entire study is consolidated in full detail in the forthcoming Volume 8 in the series *Studies in Language Testing*.

This study documents the design, development and construct validation of a cognitive strategy use questionnaire and a metacognitive use questionnaire with relation to a theory of human information processing. It then relates these test taker characteristics to test performance. The study shows that cognitive and metacognitive strategy use must be used in concert with one another to produce best results in the tests.

The study also modelled strategy use and test performance with high and low ability candidates, finding that although high and low ability test takers utilise many of the same strategies on language tests, the ways in which these strategies are used by each group can produce significantly different end results. Finally, Purpura shows how the findings of his study can be implemented in the language classroom to provide students and their teachers with concrete information on their strategy use to be used for training purposes.

In 1994 UCLES collected data relating to the Communication Strategies but there were some serious questions raised about the traits in this questionnaire, and given the recent research in this area, it was decided to suspend the collection of more data until further attention could be given to the construct.

Socio-Psychological Factors

Most recently (1997 to 1999) the work has focused on validation of the questionnaires relating to the *socio-psychological factors*: attitudes, anxiety, motivation and effort.

The *Attitudes Questionnaire* was designed to measure the following four underlying factors:

- (i) Attitudes towards Learning English
- (ii) Attitudes towards English speakers
- (iii) Interest in Foreign Languages
- (iv) Perception of Task Difficulty

The *Motivation Questionnaire* was designed to measure the following three underlying factors:

- (i) Integrative Motivation
- (ii) Instrumental Motivation
- (iii) Achievement Motivation

The *Anxiety Questionnaire* was designed to measure the following three underlying factors:

- (i) Class Anxiety
- (ii) Language Anxiety
- (iii) Test Anxiety

Data have been collected and analyses have been carried out on all three questionnaires in order to provide reduced versions that best measure the underlying traits. The following kinds of analysis have been used in all cases: reliability analyses, exploratory factor analyses and item-level Structural Equation Modelling (SEM).

The *Effort Questionnaire* was designed to measure one underlying factor; reliability analyses, EFAs and item-level SEM were carried out and the data supported a one-factor solution.

One aspect of the work which is particularly noteworthy is the use of item-level SEM to validate questionnaires as this technique had not previously been used in applied linguistics or language testing

research. This is now being written up by Purpura and will be submitted for publication.

Current developments and future research

We are now at a stage where the bank of questionnaires has been validated and will soon be available to be used for a number of purposes. The following uses are anticipated for a range of user groups:

- to provide *test takers* with information on their language learning behaviours;
- to provide *teachers* with a profile of their students' language learning behaviours;
- to provide *curriculum planners and administrators* with a profile of the language learning behaviours in their schools;
- to provide *UCLES* with data to continue examining the relationships between test taker background characteristics and performance on language tests.

For UCLES, the opportunity to collect data on a wide scale suggests an agenda for future research in, for example, the following three areas:

- a series of cross-validation studies could be carried out to provide further evidence of the validity of the questionnaires and their underlying traits;
- a series of test validation studies looking at the relationships between *individual socio-psychological* factors and test performance (e.g. to investigate the relationship between anxiety and test performance);
- a large-scale study looking at the relationship between *several* background factors and test performance in order to determine the differential effects of multiple background factors on test performance.

Computerisation of the questionnaires

The next phase of the development work for the questionnaires will be to produce a user-interface on computer so that the questionnaires can be administered effectively to large numbers of people. This is being undertaken as part of the UCLES developments in the field of Computer-based Testing.

It is anticipated that users will be able to assess their language ability using computer-adaptive tests (e.g. *CommuniCAT*) and also to evaluate their approach to learning using the language learning questionnaires. In this way learners and their teachers will be able to target the learners' strengths and weaknesses more effectively and thus be able to tailor the learning process to the individual's specific requirements.

We will report on this continuing research, and on the use that can be made of the data from the questionnaires in future issues of Research Notes.

Issues in Speaking Assessment Research

Lynda Taylor, Performance Testing Unit Co-ordinator, UCLES

Research into issues surrounding the testing of spoken language dates back over 20 years. During the 1980s a substantial amount of work was carried out on speaking tests produced and used in North America, focusing primarily on issues of content validity, construct validity, concurrent validity, and reliability and rating procedures. Relatively little attention, however, was paid to other important speaking assessment issues such as:

- the design of the speaking test tasks;
- the nature of the candidate's spoken discourse or language;
- the language and behaviour of the oral examiner;
- the nature of the rating instruments being used.

Towards the late 1980s language testers began to recognise that we can consider the testing of spoken (and written) language in terms of a number of different variables or 'facets'; furthermore, these facets can interact in complex and sometimes unpredictable ways. During the early 1990s the challenge for language testers was to identify these various facets and to develop ways of understanding, explaining and accounting for the interactions between them.

In order to do this, it proved helpful to develop a conceptual framework or model. Language test specialists at UCLES presented just such a model at the LTRC conference in 1993 (*Studies in Language Testing*, Volume 3, page 3). This framework has proved helpful in identifying the various different avenues of research which are possible. It suggests that there are potentially complex interactions between many different variables in the activity of speaking assessment and that, as far as possible, they need to be anticipated by the test designer. In many cases, however, the effect of these interactions on the process of assessment and the resulting score are only now beginning to be better understood.

The 1990s in particular have seen an increasing understanding of some particular facets of speaking assessment, in terms of both process and product.

UCLES has recently been engaged in various strands of research to investigate three specific research issues:

- the background characteristics of the candidate or test taker;
- the language produced by the oral examiner;
- the nature of rating scales.

Investigating background characteristics of the test taker

In relation to background characteristics, each candidate brings a complex set of variables to the testing situation. These variables relate to:

- who the candidate is: age, gender, etc.;
- what the candidate knows in general terms: knowledge schemata;
- how the candidate behaves and what s/he feels: affective schemata;
- how the candidate thinks: cognitive and metacognitive skills;
- the sample of language elicited.

In order to understand and explain candidate performance fully, all these features need to be investigated. This type of work is especially relevant in the context of Cambridge's large-scale assessment of speaking where the size and distribution of the candidature make the range of background variables considerable.

Candidate Information Sheets are now routinely administered to over 600,000 EFL candidates annually enabling UCLES to gather a large amount of demographic data such as age, gender, nationality, first language, etc. for research purposes. When analysed and compared with candidates' performance data on the speaking tests, the results enable us to investigate possible areas of bias.

This information is complemented by the research described in Nick Saville's article on page 6 on language learner questionnaires for investigating test takers' affective schemata and their cognitive and metacognitive skills.

Investigating the language produced by the examiner

Since the early 1990s the UCLES EFL Division has been systematically examining the instructional language and verbal prompts of oral examiners for a number of the Cambridge EFL exams. A series of projects has been carried out in collaboration with Professor Anne Lazaraton at George Mason University, who is the author of the forthcoming Volume 14 in the *Studies in Language Testing series*. Her studies have made use of conversational analytic techniques to highlight the problems of variation in examiner talk across different candidates and the extent to which this can affect the opportunity candidates are given to perform, the language sample they produce and the score they receive.

The results of these studies have confirmed the value of using a highly specified interlocutor frame in speaking tests which acts as a guide to assessors and provides candidates with the same amount of input and support; the interlocutor frame is now an integral design feature of the Cambridge approach to speaking assessment.

In addition, it has led to the development of an approach for assessing oral examiner performance based upon an Oral Examiner Monitoring Checklist. This checklist helps to monitor systematically how well oral examiners are performing during live examining which is especially important for the purposes of standardisation in large scale speaking assessment. Analyses of the data it gathers can be used to provide feedback to examiners and trainers on how oral interview techniques might be improved.

Investigating the nature of rating scales

Considerable attention has also been paid in recent years to the nature of the rating instruments used in speaking assessment, i.e. the number of scales, the nature of scales and the length of scales.

Traditionally, opinions as to the number of scales to use in speaking assessment have varied. Some tests have adopted a holistic approach in which a single scale covers all the components of communicative competence, perhaps implying that skill development has a flat profile. However, it is also conceivable that an individual speaker may not in fact develop different facets of communicative competence at the same rate. For example, a speaker's socio-linguistic and strategic competence might develop substantially over a period of several years, but their grammatical competence may remain relatively underdeveloped. Clearly there is a question in oral assessment about whether to adopt a *holistic* or an analytic approach, although in the UCLES EFL Speaking Tests we believe we have learned to exploit the strengths of both approaches.

Investigation into rating scale development at UCLES has grown steadily in recent years. Rating scales for the Cambridge EFL Speaking Tests are developed simultaneously with the design and development of the test format and in the light of the model of communicative competence being adopted. Scales are often broken down into several main areas or criteria: grammatical, discourse, pragmatic and strategic competence. The grammatical scale may be further broken down into grammar, vocabulary and pronunciation. The discourse competence scale relates to organisational features of the spoken language, such as coherence and cohesion, while the pragmatic and strategic competence scales relate to communications skill and strategies. Sometimes these criteria may be combined into a holistic or global scale.

Once a scale or series of scales has been devised, however, the question of how long they should be needs to be addressed. Traditionally, relatively little time has been devoted by language testers to the empirical validation of rating scales prior to their tests being widely used; in general, they have addressed the question of scale length intuitively, often without a clear rationale for the number of points used. In recent years, however, language testers have begun to address the issue of rating scale development from an empirical rather than an intuitive perspective, using discourse analytic techniques combined with statistical procedures, such as Rasch partial credit

analysis, for their construction and validation. The Cambridge Assessment of Spoken English (CASE) is a good example of a Cambridge EFL Speaking Test which was developed using a combination of the new qualitative and quantitative techniques that have become more readily available during the 1990s.

Summary

The last few years have seen a much clearer identification and understanding of the many different research avenues open to us in the field of speaking assessment. The development of new and sophisticated qualitative and quantitative methods to facilitate our investigative and analytical research work has been equally important over this period; advances in discourse analysis and in multi-faceted Rasch measurement are just two examples. At Cambridge we are committed to the large-scale face-to-face assessment of speaking skills and we are therefore taking full advantage of these new tools to help us with our programme of ongoing test design and validation.

The UCLES/CUP Learner Corpus

Andrew Boyle, EFL Computer-Based Testing Officer, UCLES

David Booth, EFL Test Development Officer, UCLES

Computer-based corpora have an increasingly important place in Applied Linguistics research and it is also felt that Corpus Linguistics may have a role to play in the development and validation of language examinations. In light of these advances UCLES is making increasing use of techniques linked to Corpus Linguistics in various aspects of its work.

A Learner Corpus (see Leech, 1998) is a database of language produced by foreign language learners. UCLES has been working with Cambridge University Press (CUP) over the past few years to develop a Learner Corpus based on responses by candidates in the Writing components of EFL examinations (e.g. Paper 2 for FCE, CAE, CPE).

The Learner Corpus currently contains about 10 million words and is made up of the responses to examination questions by candidates from a wide range of backgrounds. This corpus is obtained by keying in the exact responses made by candidates, including all the errors. It also contains very comprehensive data relating to the candidates themselves; this information is routinely gathered from the Candidate Information Sheets which are filled in by all candidates taking the examinations and from the scores awarded to them on the other components and for the examination as a whole. It is important to point out, however, that no information is stored about individually named candidates as their names are removed when the scripts are keyed in. All information is therefore anonymous.

To date, scripts have been selected from the Upper Main Suite examinations (FCE, CAE and CPE) with the selection being determined according to the first language group of the candidates. The corpus is currently being expanded to include scripts from the Lower Main Suite exams (KET and PET), from the Business English Certificates (BEC), and from a wider range of candidate backgrounds.

The users of the corpus work with sophisticated Windows-based software, which allows them to carry out a wide range of searches and concordances. CUP have so far used the corpus to inform their development of learners' dictionaries and other publications.

One of the distinguishing features of the corpus which CUP have developed is that it can be tagged for learner errors. So far about a quarter of the corpus has been annotated in this way. This allows the user to search on particular types of error, as well as on specific words or first language groups. Sub-corpora can be set up (for example of responses to a particular examination) to further refine searches.

From the point of view of UCLES, the development of this corpus has a number of benefits. Firstly, it allows us to develop an archive of

candidates' responses to the examinations which can be made readily available to staff in an electronic form. Secondly the scripts together with the performance and demographic data provide a valuable resource for test development and validation purposes.

UCLES-EFL is developing approaches in several areas. A project is underway to grade candidates' essays using sophisticated Natural Language Processing (NLP) technology. In this project it is hoped that NLP technology can be used to grade essays into broad bands (for example ALTE levels). The Learner Corpus will be used as a resource and a model for the NLP software, which will in effect compare the language in a new essay with those already graded in the corpus. By doing this it can provide an appropriate grading of the new essay. Questions with shorter text answers (for example comprehension or Use of English items) are being developed for inclusion in future computer-based examinations using a more computationally lightweight approach. This approach may well be informed by analysis of corpus data.

Exploratory work is being done by querying the corpus to find out collocational information on words. New item types being explored for the revised version of CPE focus on collocation as a feature of advanced language knowledge. Item designers have access to the Learner Corpus to inform item construction and to validate their intuitions about language features and frequencies associated with different levels of examination in the learner corpus.

UCLES hopes to improve its approach to the specification of word lists at different levels and develop a comprehensive lexicon for use in exams. This lexicon could be validated with reference to the Learner Corpus and other publicly available corpora. This innovation would have particular relevance for the suite of Business English Certificates.

References:

Alderson, J.C. (1996): 'Do corpora have a role in language assessment?' in Thomas, J. & Short, M. (eds). Using Corpora for Language Research London: Longman.

Hargreaves, P. of the UCLES EFL Division, 'How Important Is Collocation In Testing the Learner's, Language Proficiency?' to appear in a forthcoming LTP collection.

Leech, G. (1998): 'Learner Corpora: what they are and what can be done with them', preface to Grainger, S. Learner English on Computer London: Longman

Tono's website <http://www.lancs.ac.uk/postgrad/tono/> provides useful links to many publicly available learner corpora.

Studies in Language Testing

Already published:

1. **Lyle F. Bachman, F. Davidson, K. Ryan, I-C. Choi** *An investigation into the comparability of two tests of English as a Foreign Language: The Cambridge—TOEFL comparability study*, Cambridge, 1995 (ISBN 0-521-48467-7)

This book documents a major study, which compares the Cambridge First Certificate in English (FCE) with the Test of English as a Foreign Language (TOEFL) and investigates similarities in test content, candidature and use.

2. **Antony John Kunnan** *Test taker characteristics and performance: A structural modelling approach*, Cambridge, 1995 (ISBN 0-521-48466-9)

This book investigates the influence of test taker characteristics on test performance in tests of English as a foreign language by exploring the relationships between these two groups of variables. Data from a test taker questionnaire and performance on several tests including the FCE and the TOEFL were used for the study.

3. **Michael Milanovic, Nick Saville** *Performance Testing, Cognition and Assessment: Selected papers from the 15th Language Testing Research Colloquium*, Cambridge and Arnhem, Cambridge, 1996 (ISBN 0-521-48465-0)

This book contains a selection of research papers presented at the 15th Annual Language Testing Research Colloquium (LRTC). The Colloquium was jointly hosted by the University of Cambridge Local Examinations Syndicate (UCLES) in Cambridge and CITO in Arnhem, in the Netherlands. At the Cambridge venue, the papers were presented on the themes of performance testing, and at Arnhem they covered aspects of communication in relation to cognition and assessment. A selection of papers has been made in order to achieve a balanced coverage of these themes.

4. **Caroline M. Clapham** *The development of IELTS: A study of the effect of background knowledge on reading comprehension*, Cambridge, 1996 (ISBN 0-521-56708-4)

This book investigates the ESP claim that tertiary level ESL students should be given reading proficiency tests in their own academic subject areas, and studies the effect of background knowledge on reading comprehension. It is set against a background of recent research into reading in a first and second language, and emphasises the impact schema theory has had on this. The book is a useful resource for those involved with IELTS and others interested in the testing of English for academic purposes.

5. **Alison Green** *Verbal protocol analysis in language testing research: A handbook*, Cambridge, 1998 (ISBN 0-521-58635-6)

Verbal protocol analysis (VPA) is a methodology that is being used extensively by researchers. Recently, individuals working in the area of testing, and in language testing in particular, have begun to appreciate the roles VPA might play in the development and evaluation of assessment instruments. This book aims to provide potential practitioners

of VPA with the background to the technique and a good understanding of what is entailed in using VPA in the context of language testing and assessment. Tutorial exercises are presented which enable the reader to try out each of the different steps involved in VPA.

6. **Multilingual Glossary of Language Testing Terms**, Cambridge, 1998 (ISBN 0-521-65877-2)-also available on CD ROM

A multilingual glossary has a particular significant role to play in encouraging the development of language testing in less widely taught languages by establishing terms which may be new alongside their well-known equivalents in the commonly used languages.

The glossary contains entries in ten languages: Catalan, Danish, Dutch, English, French, German, Irish, Italian, Portuguese and Spanish.

7. **Davies et al.** *Language Testing Dictionary*, Cambridge, 1999 (ISBN 0-521-658764)

It contains some 600 entries, each listed under a headword with extensive cross-referencing. The selection of headwords is based on advice from people in many countries who are involved in language testing, scanning of current textbooks in this field and of dictionaries and encyclopedias in adjacent fields (eg psychometrics, applied linguistics, statistics).

8. **James E. Purpura** *Learner strategy use and performance on language tests*, Cambridge, 1999 (ISBN 0-521-658748)

This book investigates the relationship between learner strategy use and performance on second language tests, by examining the construct validity of two questionnaires designed within a model of information processing that measures test takers' self-reported cognitive and metacognitive strategy use. This book investigates how learner strategy use influences test performance, and how high performers use strategies differently from low

Forthcoming:

9. **Anthony Kunnan** *Fairness and validation in language assessment*
10. **Micheline de Chalhaub-Deville** *Issues in Computer-adaptive testing of reading proficiency*
11. **Catherine Elder ed** *Experimenting with uncertainty*
12. **Cyril Weir, Yang Huizhong, Jin Yan** *An empirical investigation of the compotentiality of L2 reading in English for academic purposes*

In preparation:

Kieran O'Loughlin *An investigatory study of the equivalence of direct and semi-direct speaking tests*

Anne Lazaraton *A quantitative approach to the validation of oral language tests*

Further Information

UCLES provides extensive information on the examinations and assessment services referred to in this newsletter. For further information, visit the UCLES EFL website

www.cambridge-efl.org.uk

or contact [performers](mailto:performers@ucles.org.uk).
EFL Information
University of Cambridge Local Examinations Syndicate
1 Hills Road
Cambridge CB1 2EU
UK

Tel. +44 1223 553822
Fax. +44 1223 553068
e-mail harding.a@ucles.org.uk

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
UK

Tel. +44 1223 553355
Fax. +44 1223 460278
e-mail alte@ucles.org.uk

If you would like further copies of this issue of Research Notes, or if you would like to be added to our mailing list (all registered UCLES centres receive copies automatically), please complete and return this form to EFL Information at UCLES. Please photocopy the form if you wish.

Please send me ... extra copies of this issue of Research Notes.

Please add me to the Research Notes mailing list.

Name

Job Title

Institution

Address

.....

.....

Country

Please let us have any comments about Research Notes, and what you would like to see in further issues:

The UCLES EFL Test Development and Validation Group

Group Manager	Nick Saville
PA to Group Manager	Susan Chapman
ALTE Assistant and European Projects Co-ordinator	Marianne Hirtzel

■ Performance Testing Unit

Performance Testing Co-ordinator	Lynda Taylor
Performance Testing Officers	Janet Bojan Val Sismey
Performance Testing Administrator	Rowena Akinyemi

■ Centre Inspections

Centre Inspections Officer	Morag Bean
-----------------------------------	------------

■ Analysis and Interpretation of Results

Grading Co-ordinator	Neil Jones
EFL Analysis and Evaluation Co-ordinator	Chris Banks
Assistant Validation Officers	Kerri Gardner Helen Marshall
Validation Clerks	Rouman Marinov Kevin Jones

■ Commissioning, Pretesting, Item Banking and Question Paper Preparation

EFL Pre-Testing Manager	Simon Beeston
EFL Question Paper Preparation Administrator	Mike Northfield
EFL Materials Administrator	Sue Ellis
EFL QPP Clerk	Gill White
Pre-testing Administrator	Kate Talbot
Pre-testing Clerks	Tracey MacArthur Gordon Craft

■ Test Development and Institutional Testing

Assistant Test Development Officer	David Booth
Assistant Test Development Officer	Ed Hackett
Validation Administrator Institutional Testing	Mary Russell