REPORT ON THE SEMINAR ON

Testing Aviation English

Hosted by EUROCONTROL Institute of Air Navigation Services, in Luxembourg, on September 4th and 5th 2003

Contents

- 1. Letter from the President
- 2. English Language Testing in the context of Aviation English Professor Dan Douglas, Iowa State University
- 3. The PELA Test an Update Mr Adrian Enright, Training Development and Harmonisation, Eurocontrol
- 4. ICAO Language Proficiency Requirements: The Solution or a Problem Ms Elizabeth Mathews, ICAO PRICESG Linguistic Adviser
- 5. Ethics and quality control in language testing Mr Nick Saville, ALTE Code of practice Working Group
- 6. Pilot and Controller Skills and Competencies Dr Jeremy Mell, ENAC, Toulouse
- 7. Testing English in the workplace Mr Nick Saville
- 8. Future meetings and Directions for ICAEA discussion

Paris January 2005

Dear Forum Participants

The clock ticks implacably towards the 5th of March 2008, the date when the new ICAO standards of English for controllers and pilots become applicable. This concentrates the minds of all those responsible for quality standards in civil aviation organisations and airlines. It is to the language training departments, where they exist, that people turn for answers to the multiplicity of questions that are raised. What do these standards really mean? What is level four? How many of our controllers and pilots reach operational competence in English? How many do not? How do we find out? The Forum in September 2003, held at EUROCONTROL Institute of Air Navigation Services, Luxemburg was focussed on the last point "Testing English for Aviation".

In two short days, a wide range of aspects of the topic were covered, ranging from presentations about both the theory and practice of testing languages, through the specific situation in aviation to commentaries from those at the pointy end, controllers and pilots on the job.

This event was enhanced greatly by the venue itself which is beautifully equipped for conferences. We were well received by courteous and efficient staff who bring so much professionalism to their work. In the preparatory stages of registration and hotel reservation, the work of Rosaria Corral and her department was particularly appreciated. All this assistance and hospitality is thanks to Adrian Enright, our man in Luxemburg, and the EUROCONTROL Institute of Air Navigation Services.

We wish to thank all those who came to give of their time and expertise as speakers in the programme. We wish also to thank all the participants. The interest you display and readiness to interact on topics central to our professions are what makes such a meeting worthwhile.

To everyone involved in this huge effort to progress towards a safer aviation world through better communication in English, let me say don't give up. You can make it happen.

Fiona A Robertson President

PS.

This is a very "post" PS, and I apologise to those whose very valuable and apposite contributions have had to wait so very long to see the light of day here on the website. But better late than never, and what is said here is, if possible, even more current and appropriate than it was when it was presented.

Fiona,

Paris Winter 2007

St. Pete Tower - Go-o-od evening! English Language Testing in the Context of Aviation English

Dan Douglas, Iowa State University

Introduction

I feel fairly certain that everyone here today is aware of how important English language testing is in the context of the newly approved policies that pilots, navigators, air traffic controllers, and station operators "demonstrate the ability to speak and understand the language used for radiotelephony communications" and of the vital consequences the policies will have for applicants for licensure, for currently employed controllers and pilots, for English language instructors and program administrators, for policy makers at all levels, and most importantly for airline passengers and crews. I also suspect that nearly every participant at this seminar has also been dissatisfied at one time or another with the quality of the English language tests with which they are familiar: with the content and methods of the tests, with the apparent lack of relationship of the tests to "real life" language use, with the time and expense involved in testing, and with the clarity and completeness of the reports of results. Such dissatisfaction gives rise to an overall uncertainty about whether the tests are fair, accurate, and useful. Our purpose for the next 50 minutes or so is to take a thoughtful look at English language testing as it may be realized in international civil aviation, considering it in light of our best understanding of the nature of language knowledge and use, of principles of language assessment, and of the individual, pedagogical, and social impact of English language tests. We'll get down to more concrete specifics in the other presentations today from Adrian (PELA), Elizabeth (ICAO Requirements), and Nick (Quality control and Workplace testing).

Overview of English Language Testing

More than thirty years ago, John B. Carroll (1968) defined a test as "a procedure designed to elicit certain behavior from which one can make inferences about certain characteristics of an individual" (p. 46). We may say more specifically that language tests are procedures for eliciting language performances which allow us to make inferences about learners' language abilities. There are a couple of important points to note about this rather neat definition that complicate matters somewhat. First, much of what we think of as language ability – qualities such as knowledge of grammar, vocabulary, textual organization, and so on – is internal to the learner and thus not directly observable or measurable. Therefore, we observe language performances, whether these consist of participating in an oral interview, writing an essay, or marking answers on a multiple choice test, and then use this information to infer the state of the learner's language ability. Second, although we could measure certain aspects of a language performance directly, such as by counting the number of words correctly used, for example, or the number of phonemes accurately pronounced, such measurements would apply only to the test situation itself when we would really like to generalize to language use outside the test. We could, of course, observe learners while they are actually engaged in the language behavior we're interested in – requesting permission to change altitude while cruising at 11,000 meters or informing an aircraft on current weather conditions - and decide, on the basis of how successful they seem to be, whether their language abilities were adequate. Some people might consider this a direct measure of language ability; the problems with it are efficiency – it would take a lot of time and effort – and consistency – each test taker's situation would be different. We should note in passing that since language

ability is a cognitive attribute, it is not possible to observe or measure it directly, of course, and in this sense, there is no direct test of language ability. However, in another sense, we often distinguish among tests in terms of the directness of their relationship to authentic language use. Thus, a speaking test in which the test taker actually has a conversation or interview with another person is said to be a direct test of speaking; a speaking test in which the test taker talks to a tape-recorder or computer in response to written prompts is called a semi-direct test; and an estimate of speaking ability based on a test of listening, for example, is sometimes called an *indirect* test of speaking. This problem with the relationship between internalized language knowledge and external language performance is why there is such an emphasis in language testing on systematic procedures that are limited in scope and the same for everyone being tested. Language testers try to devise procedures that mirror as much as possible features of actual communicative language use, while at the same time making efficient use of test takers' and scorers' time and providing clear, useful information to those interested in the results, including the test takers themselves, of course. Because language testing plays such an important role in the lives of test takers, those who produce the tests, those who administer them, and those who use the results have an ethical responsibility to ensure that the entire enterprise is carried out as reliably and fairly as possible so that the interpretations we make about language abilities and the decisions we make as a consequence of those interpretations are appropriate for both the individual being tested and the intended purpose of the test.

This discussion of English language testing in the context of international civil aviation will be concerned with four general areas of interest: 1) needs analysis, 2) construct definition, 3) test usefulness, and 4) operationalization. Before beginning on these topics, however, I would like to review very briefly a perspective on test purposes which will no doubt play a role in our discussions of assessment for civil aviation: proficiency, placement, and achievement tests. I will also touch upon two approaches to assessment of particular interest for us, testing language for specific purposes, and performance testing.

Test Purposes: Inferences and Decisions

There are two main reasons why we give tests: we wish to make *inferences* about aspects of individuals' language ability, and/or we need to make decisions that will affect their lives in sometimes profound ways. For example, we make inferences about the level of a person's vocabulary knowledge, knowledge of cohesion, or language functions; we may decide, at least partly on the basis of test performance, to give someone a certificate, to place someone else in an intermediate class, or to reject an applicant for admission. Often these two reasons are subsumed under a single term purpose because we usually make decisions based on inferences. That is, we may decide to admit an applicant to a training program because we infer that she has the requisite level of language ability to succeed. Three test purposes that seem to me to be relevant for civil aviation are proficiency, placement, and achievement. The three are, of course, related, and there is a great amount of overlap in their influence on test development and administration, but they are different enough that it is seldom possible to use one test for more than a single purpose. Proficiency tests are intended to measure ability to use English in specific situations in which the test takers will find themselves in the future, regardless of the circumstances in which they acquired the language. Since they are often taken months before a test taker actually needs to use the language for the actual purpose, proficiency tests are therefore often used to look forward, aiming to help make predictions about the probability that a particular applicant will be able to cope with the demands for English language use in the context of, in our case, civil aviation. The likelihood that a

person with a given level of ability in English will be able to succeed will vary depending upon the specific requirements of the job and the person's own linguistic, educational, and cultural background. In a way somewhat similar to weather forecasting, proficiency testing involves a large number of factors interacting in complex ways, so that we must be very cautious in the interpretations and decisions we make on the basis of test results. As the name implies, *placement tests* are used to place candidates into instructional programs relevant to both their current level of proficiency and specific courses, materials, and methods offered in the program. In our case, placement tests might be tailored to match the instructional programs that many of you work in, providing English language instruction to ATCs, pilots, and other personnel. By their nature, placement tests are designed to look at current needs in light of currently available solutions, as opposed to looking forward as with proficiency tests. *Achievement tests*, also known as *progress tests*, are designed to look back at past performance, indicating how well learners have mastered the material in an instructional program. Achievement tests thus reflect a specific curriculum of study and the materials used in the classes and help define where a student should be after a given amount of time.

I've described these three purposes briefly because it may be that we will want to consider the need for all three with regard to the new civil aviation policies and the testing and teaching that will accompany them. As I indicated above, it is usually not advisable to try to use one test for all three, or even two, of the purposes. Proficiency tests, which will almost certainly be our first and major concern, are designed to distinguish among test takers with a wide range of ability, in reference to a standard of performance defining language use in a target situation. Proficiency tests cannot reflect any particular teaching approach or set of materials since the test takers come from a wide variety of backgrounds and programs. Placement tests, on the other hand, are designed to distinguish among test takers who have already taken the proficiency test and are therefore a much more restricted group in terms of their levels of ability. Since the purpose of the placement test is to place them into a particular program of study where they will receive instruction most relevant to their needs, the test should reflect the approach and materials the test takers will encounter in the program. Placement tests, which reflect the entire range of skills and the number of levels available in the program, are usually much more finely-tuned and detailed than are proficiency tests. The achievement test is designed to make only a two-way distinction, really, distinguishing those who have learned the material taught from those who haven't. Achievement tests thus reflect only the specific material the learners experienced and are usually given at the end of the course of study, or, when designated as progress tests, during it as a monitor on continuing development.

A Specific Purpose Approach to Assessment

Language tests may be developed with reference to situations of use defined very broadly or quite narrowly. For example, in many instructional circumstances we may not know with any precision what the learners will use the language for when they leave the program – tourism, immigration, diplomacy, business, education – and so the teaching and testing associated with such a program must necessarily be fairly general in attempting to cater to the variety of interests and needs of the learners. On the other hand, many times we do know precisely what the learners will be doing with the language when the leave our programs, and this happy situation in the context of civil aviation leads us to *specific purpose* English testing. I've defined specific purpose language testing as follows:

A specific purpose language test is one in which test content and methods are derived from an analysis of a specific purpose target language use situation, so

that test tasks and content are authentically representative of tasks in the target situation, allowing for an interaction between the test taker's language ability and specific purpose content knowledge, on the one hand, and the test tasks on the other. Such a test allows us to make inferences about a test taker's capacity to use language in the specific purpose domain.

Douglas 2000, p. 19

There are a number of ideas in this definition that we could discuss for more time than we have today – content and methods being derived from a situational analysis, authenticity, interaction between language ability, content knowledge, and test tasks, and making inferences about capacity to use a language in a specific domain of use – so rather than dealing with them independently, I will refer to each of them within the context of the rest of the discussion as I outlined at the beginning.

Just to get us in the mood, let's listen to a minute or so of Pilot-ATC radio traffic from St. Petersburgh airport in Florida:



As I think is clear from this example, prescribed language is the basis for communication in the context of civil aviation, but there is also quite a bit of non-prescribed usage as well. It is not enough to test for control of only the approved phraseology. We need to understand more about communicative language use during actual performance in the context of civil aviation.

Which brings me to my next topic - performance testing - in which test takers are asked to engage in production tasks that mirror features of "real life" language use. Certainly the testing of aviation English will involve to one degree or another "a performance by the

candidate which is observed and judged using an agreed judging process" (McNamara 1996, p. 10). Among the challenges of performance testing are the interaction between content knowledge and language knowledge discussed above, but also insuring uniformity across topics and disciplines, training raters to evaluate complex performances, and the administrative costs associated with most evaluation programs. Again, these are thorny topics that are not quickly resolved.

I'd like to move on now to the topics I announced at the beginning of my talk, four important aspects of test development: 1) needs analysis, 2) construct definition, 3) test usefulness, and 4) operationalization.

Needs Analysis

Needs analysis is a concept/procedure that has been around since the earliest days of English for specific purposes in the 1960s. It involves analyzing and describing the target language use situation – in our case, the context of civil aviation – as input to the development of the test or tests. The outcome of the analysis is an account of the purpose and process of communication in the target situation, including a description of the salient communicative tasks. A framework that I recommend for analyzing the TLU situation is SPPCTLNG!

Setting
Participants
Purpose
Content
Tone
Language

Norms of Interaction

Genre

W can measure only a limited number of aspects of this complex situation, and this brings us to the topic of construct definition.

Construct Definition

An important part of language test development is being very clear and precise about what it is we wish to measure. As I discussed earlier, language knowledge is an internal, cognitive concept, and our only understanding of it is inferential – we can only infer what's inside the black box by observing outward manifestations of it. Our various inferences are called *constructs* in the testing business and when we define what it is we wish to measure with our test, we are creating a construct definition. A construct definition may be theory-based or needs-based. Example of a theory-based definition from *Occupational English Test*:

OVERALL COMMUNICATIVE EFFECTIVENESS	
Near-native flexibility and range _ _	
Limited	

INTELLIGIBILITY Intelligible _ Unintelligible
FLUENCY Even _ Uneven
COMPREHENSION Complete _ Incomplete
APPROPRIATENESS OF LANGUAGE Appropriate _ _ _ Inappropriate
RESOURCES OF GRAMMAR AND EXPRESSION Rich, flexible _ _ Limited

Notice that the focus here is on the linguistic system, a theory of what it means to know a language. There is no reference to specific content or background knowledge. A needs based definition, on the other hand, does assume content knowledge:

Example of needs-based definition from *PELA* (1994):

THE CANDIDATE SHOULD BE ABLE TO UNDERSTAND COMMUNICATIONS BETWEEN PILOTS AND ATCOS, BETWEEN ATCOS AND ATCOS, IN WHICH OF THE PILOT OR ATCO:

- 4.1 READS BACK/ACKNOWLEDGES ATCO'S MESSAGE
- 4.2 REQUESTS APPROVAL OR CLEARANCE
- 4.3 GIVES INFORMATION
- 4.4 REQUESTS INFORMATION
- 4.5 GIVES/REQUESTS REASONS
- 4.6 REQUESTS OF THE ATCO TO DO SOMETHING
- 4.7 CHECKS, CONFIRMS AND CLARIFIES
- 4.8 GIVES/DENIES APPROVAL AND CLEARANCE (ATCO'S ONLY)

SHOULD BE ABLE TO RECOGNISE THE COMMUNICATIVE FUNCTION OF MESSAGES WITH AND WITHOUT EXPLICIT INDICATORS

Aspects of language ability that might appear in a needs based construct definition include the following:

Construct Definition

- Language Knowledge
 - Grammatical Knowledge
 - Textual Knowledge
 - Functional Knowledge
 - Sociolinguistic Knowledge
- Strategic Competence
 - Assessment

- Goal setting
- Planning
- Execution
- Specific Purpose Background Knowledge
 - Content schemata
 - Formal schemata

Not all of these features may appear in any given definition, for which ones we might wish to assess depends upon our test purpose. For example, consider the place of aviation background knowledge in the construct definition for a test we might consider designing. In the case of a proficiency test for novice air traffic controllers, for instance, with widely varied background knowledge, we may not wish to include BK in the construct definition and make inferences only about components of language ability. On the other hand, in a test for experienced candidates for license renewal, with fairly homogeneous levels of background knowledge, we may indeed include it in the construct definition, and the inferences we would make would be about the ability to process specific aviation information by means of English. You can see a potential problem here: if a candidate does poorly on the test, we will not easily know whether the performance was due to a deficiency in English ability or to a deficiency in specific aspects of aviation knowledge. As I said earlier, the relationship between language knowledge and background knowledge is complex and not well understood. My own view is that we are better off to include both in the construct definition, since both are always present in any case.

Usefulness

The notion of test usefulness is fundamental to the language testing enterprise. When a teacher asks another, "Do you know of a good test to give my ESL students on the first day of class?" they're really talking about test usefulness. When ICAO administrators ask us to develop "a good test that will demonstrate a candidate's ability to speak and understand the language used for radiotelephony communications", they're really talking about test usefulness. Clearly, the most important quality of a test is its usefulness for its intended purpose. But how can we develop a useful test? How can we tell if a test is useful before we give it? How can we tell whether it's useful *after* we give it? Bachman and Palmer (1996) have argued that a test's usefulness depends upon several qualities which complement each other and together add up to overall usefulness, and the concept provides a framework for providing evidence for making a reasonable case that the test provides a firm foundation for the *inferences* we wish to make about people, and the *decisions* we are called upon to make about their lives. We've already referred to some of the qualities of test usefulness, which, modified a bit from those discussed by Bachman and Palmer, include *construct validity*, *reliability, authenticity, impact, and practicality*.

Construct Validity

At the heart of English language testing practice lies the question of the extent to which the test results give us useful information with respect to the needs of the various stakeholders: pilots, navigators, air traffic control officers, administrative personnel, instructors and training program supervisors. *Test validity* involves evidence that the interpretations these various stakeholders make of test results are justified in terms of what we infer about a test taker's

knowledge of Englishⁱⁱ and what decisions we make on the basis of test results. This requires that the tests are shown to provide consistent and accurate results each time they are administered, and that performance on the test can be shown to be related to language use in situations of interest outside the test – engaging in actual radiotelephony communication. A key term here is *evidence*: test developers, whether they are large corporate entities like Educational Testing Service or teams of ESL instructors, need to, and indeed, will wish to demonstrate that their tests in fact measure what they are intended to measureⁱⁱⁱ and that the interpretations and decisions made on the basis of test performance are justified. The higher the stakes, the more serious the consequences of the test are for the test takers and the stronger such evidence should be, in fairness to the test takers, to provide maximum usefulness for training and certification purposes, and for public administrative accountability. It is an aphorism in language testing that just because a test *looks like* a good test of speaking ability, for example, it does not necessarily follow that it in fact *is* a good test. Evidence is required.

Reliability

Reliability is about providing evidence that our measurement of language ability is *consistent*, both across individuals being given the same test, and within individuals taking the test at different times. When we measure an object with a ruler, say, we expect to get the same measurement each time we measure the object, unless, of course, we have reason to believe its size has changed since the last measurement, and we expect that when we measure two different objects, any difference we get in the measurements is due to real differences in their size. Not so with language tests, unfortunately. Factors that can lead to inconsistency, low reliability, in language tests include physical and psychological characteristics of the test takers, the nature of the test tasks and their difficulty for the test takers, and the clarity and precision of the criteria by which the test performance is rated. Generally speaking, the more opportunities the test takers are given to demonstrate their abilities, and the more variety there is in the types of tasks so that the test takers can show their abilities in different ways, the more reliable the test will be.

Authenticity

Authenticity refers to the degree to which a test mirrors real-life language use. Note the word "degree" – test tasks vary in the degree to which they reflect actual communication. Also, we talk about authenticity with regard to entire communicative tasks, not merely the language aspects. Finally, it is important to understand that authenticity resides in the mind of language users, not in the tasks or instances of language themselves. Following Widdowson (1979) we distinguish between *genuine* and *authentic* tasks: the first refers to the task as carried out in its original context – real life language use – while the second refers to the mirror image – a simulation of real life language use. Genuiness is a dichotomous attribute of an instance of communication - all or nothing. For example, on 2 July, 2002, the pilot of a Russian aircraft heard the following communication from a controller:

"TU 154, descend flight level 350, expedite, I have crossing traffic."

Apparently, the pilot was receiving conflicting information – this message from the ATCO and information from the automatic Traffic Alert and Collision Avoidance System, which recommended maintaining level flight. The pilot was engaged in the genuine task of deciding which advice to follow. This situation resulted in tragedy, as you might recall. The event is genuine, as is the language accompanying it.

Now, this same situation, if replicated in a language test, no matter how technically realistic, would lose its genuiness but could be authentic to one degree or another. It would be relatively easy (though expensive!) for us to construct a test, using the appropriate technology, including a cockpit simulator, and a trained air traffic control officer to replicate many of the features of the original situation, thus producing a high level of authenticity; the task will never be genuine, of course, because, in the end, in a test, there is no airplane! We can only strive, by means of the simulation of as many features of the original communicative event as possible, to maximize the likelihood that the test takers will regard the test task as authentic.

How can we design authentic tests and tasks, or evaluate the degree of authenticity in a given test or task? I like to look at authenticity from two perspectives: situational authenticity and interactional authenticity. The first can be analyzed in terms of the SPPCTLNG characteristics – we use the framework to compare the features of the genuine, real life situation with those of the test task – the greater the congruence (while realizing that it can never be complete), the greater the degree of authenticity. The second aspect of authenticity, interactional authenticity, refers to the degree to which the language user's language and background knowledge are engaged in carrying out a task: it presupposes an appropriate area and level of *background knowledge*, a wide range of areas of *language knowledge*, language *functions* other than simple demonstration of language ability, opportunities for use of *communication strategies* (assessment, goal setting, planning), and *positive affect* with regard to the task.

Impact: The Ethics of English Language Testing:

All English language tests have consequences, or washback, for test takers, for their employers, for instructors, for program administrators, and ultimately, in the case of international civil aviation, for those who travel by air, the passengers. The tests that will be developed in the context of the ICAO policies will have enormous personal significance for the individuals who take them, since they will be instrumental in life-decisions about whether an applicant will be certified for employment or not, or whether an existing employee is able to continue in her or his job or is promoted. Such tests affect not only the test takers, of course, but also their supervisors who must ensure that qualified personnel are employed and assigned. The social consequences of language testing raise such important ethical issues as fairness to all test takers, equal access to tests and information about them, accountability for decisions taken based on tests, and the impact tests have on test takers, instructors, and training programs. Standards for testing practice, which have been codified by such organizations as the International Language Testing Association (ILTA), the American Psychological Association (APA), the American Educational Research Association (AERA), the National Council on Measurement in Education (NCME), and the Association of Language Testers in Europe (ALTE) can help ensure that test developers adhere to professional procedures in test production, but it is also important that the users of test results —instructors, program administrators, and policy-makers at all levels — are aware of the ethical aspects of English language testing as well.

Practicality

The final aspect of test usefulness, practicality, refers to the availability of the resources necessary to develop, administer, score, and report the results of the test. For example, as I mentioned earlier, we could design an extremely realistic test by using the latest flight and

ATC simulators, but does every site where the test is to be given have access to such equipment? Is it of standard quality and design? Are trained technicians available to operate and maintain the equipment? But even if we back away from this extreme example, with any test that we might design, we must take account of the practical aspects of human, financial, technological, and physical resources required. Rather than view issues of practicality as roadblocks to the realization of the wonderful test we've designed, it is better to treat practicality as just one among the five components of usefulness that need to be managed during the process of test development. If the resources required to develop and administer the test are greater than those available, the best-designed test in the world will never be constructed, given, or used. Which brings me to my last topic, operationalization — the actual carrying out of the steps in constructing and administering a test.

Operationalization

Operationalization is the process of using the information generated by the needs analysis, construct definition, and test usefulness analysis to produce specifications, or a blueprint, for the test, and actually constructing test tasks. In this section, we'll take a quick look at the components of specifications, also considering the use of computer technology in English language testing, since computer assisted language testing (CALT) has become such a force in the field.

Specifications

There are two levels of specifications, depending on whether we're designing the entire test, planning a subtest, or focusing on different task, or item, types that make up the test. The specifications document functions first as an *aide memoir* to help ensure that we've touched all the bases in designing our test, and second as a *record* of how each task type is to be constructed in specific enough terms so that someone not familiar with all the discussion and thought leading up to the construction of the test could reproduce the test from the blueprint.

Components of Specifications

- Describe the purpose(s) of the test
- Describe the TLU situation and list tasks
- Describe the characteristics of the test takers
- Define the construct to be measured
- Describe the content of the test
- Describe the test tasks
- Provide samples of tasks/items
- Develop a plan for test evaluation

An important aspect of the plan for constructing the test is how it is to be delivered to the test taker, how the test taker records her/his responses, and how the responses are to be scored. Increasingly, test developers are exploring the potential of computers to mediate in these areas, and this is my final topic for discussion this morning.

Computer-based Testing:

An area of interest in English language testing that deserves discussion here, given that test delivery is an important issue in an international testing enterprise, is computer-based assessment. A growing trend in language testing, as in many other aspects of our lives, is the use of computers in test construction, delivery, administration, and scoring. Computers have been a part of testing since 1935, in fact, with the introduction by IBM of an electronic test scoring machine, capable of scoring multiple choice tests 10 times faster than humans, and with greater accuracy (Fulcher 2000). This innovation made possible the rapid scoring of large numbers of answer sheets and formed the basis for multiple choice tests that are the mainstay of much large-scale English language testing today. More recently, it was a relatively easy step to transfer paper-and-pencil multiple choice tests to computer platforms, as for example, the TOEFL program did in the 1990s. Computer technology allows for many innovative means for test providers to deliver test content and for test takers to record their responses, including the use of audio and video, graphic images and animation, drag-and-drop and highlighting tasks, to name just a few. Computers can facilitate the administration of writing and speaking tests by making text and audio file transfer easy and fast. They can also facilitate the scoring of written essays (for example, e-Rater, Educational Testing Service, 2003c), and spoken responses (for example, the *PhonePass Spoken English Test*, Ordinate Corporation, 2002). Another innovation in testing made possible by computers is the development of *computer adaptive tests* (CATs), in which each test item is "tagged" with the probability that test takers of various levels of ability will answer it correctly, so that a test taker who answers a given item correctly will be presented next with a more difficult one, and when an item is missed, it will be followed by an easier one, until the computer program is able to estimate the test taker's ability level – the Cambridge Business Language Testing Service (BULATS) is an example.

If computers offer potential advantages for English language testing, including speed, accuracy, efficiency, and convenience, as well as multimedia capabilities that can enhance authenticity and interactiveness, they also carry with them the potential for creating various types of error, or bias, iv in the assessment of English language ability, known as the test method effect. The very medium of delivery, to the extent that it differs from much natural communicative language use, as well as whatever degree of computer skills may be necessary to manipulate the input and response technology, can detract from the measurement of English language ability: the score will to some extent become an indicator of computer familiarity, so that test takers who are entirely comfortable with computers may do better on the test irrespective of their English language ability than those who are less familiar with the technology. It may also be the case that test developers will be constrained in the types of tasks they include on the test to those the computer hardware and software can support, but which are unable to reflect the complexity of natural language use. Thus, test developers and researchers, as well as those who use test results, need to be aware of the limitations of computer-based tests as they interpret test performance as indications of learners' English abilities, and as they attempt to generalize the test results to language use outside the test situation.

Conclusion

Few of us doubt the importance and complexity of English language testing in the context of international civil aviation today. That importance and complexity will continue to increase

exponentially. I look forward to the discussions during this seminar and anticipate great progress in the future of the assessment of civil aviation English.

¹ Since language ability is a cognitive attribute, it is not possible to observe or measure it directly, of course, and in this sense, there is no *direct* test of language ability. However, in another sense, we often distinguish among tests in terms of the directness of their relationship to authentic language use. Thus, a speaking test in which the test taker actually has a conversation or interview with another person is said to be a direct test of speaking; a speaking test in which the test taker talks to a taperecorder or computer in response to written prompts is called a *semi-direct* test; and an estimate of speaking ability based on a test of listening, for example, is sometimes called an *indirect* test of speaking.

² Language knowledge is usually thought of in terms components: *grammatical knowledge* (e.g., sounds, vocabulary, sentence structure), *textual knowledge* (e.g., the structure and organization of written and spoken discourse), *functional knowledge* (e.g., how to convey information, to learn, to influence people and events with language), *sociolinguistic knowledge* (e.g., knowledge of dialects, styles, levels of formality, idioms), and *strategic competence* (knowledge of how to use language appropriately in specific situations) (Bachman & Palmer 1996).

³ Ideally, this is stated in a *construct definition*, a clear statement about what aspects of language ability the test was designed to measure.

⁴ Testers distinguish between *bias*, an undesirable property of a test that some test takers are unfairly advantaged or disadvantaged by certain characteristics of the test, and *discrimination*, a desirable property that the test distinguishes clearly those test takers who possess the aspects of language being measured from those who do not.

The PELA Test – an update

Adrian ENRIGHT, Eurocontrol

- Introduction
- Linguistic ability
- Test structure
- Rating scales
- Placing PELA
- Current development

In the late 1970s and early 1980s three highly publicised accidents in which language was a contributary factor, raised public concerns about safety. EUROCONTROL took the initiative and in June 1988 organised the first English language workshop held at the Institute of Air Navigation Services in Luxembourg.

The workshop concluded that air traffic controllers should demonstrate a predetermined level of knowledge and skills in the English language, especially in listening, pronunciation and comprehension, to enable them to carry out their tasks in such a manner as to contribute positively to the safety of air traffic.

Communication is a powerful tool but subject to breakdown when used inter-culturally. Adherence to procedures and rules can lessen, but not eliminate misunderstanding. Communication in aviation, based on a common language, should be our goal but political and economic forces prevail.

In the meantime, as we encourage others to improve their English let us not forget that the native speakers have an obligation to ensure that they also speak correctly and coherently. Safety has no place for jargon, delivered at a fast pace and disguised by an incomprehensible accent. Tests must be designed to encourage safe practice and to eliminate bad habits.

PELA is designed to reflect a range of tasks undertaken by air traffic controllers but with the specific focus on language rather than operational procedures. A level of proficiency in general English is necessary prior to commencing the specialised ATC English training.

The test has been specifically constructed to be administered to student controllers just prior to the completion of institutionalised training - a prerequisite if you like for licence or certification.

- Candidates are expected to adhere strictly to published ICAO radiotelephony phraseology, and where necessary,
- demonstrate an ability to communicate effectively in natural English in non-routine and emergency situations.

Candidates have to display the ability to produce messages in natural language in both usual and unusual situations that necessitate departure from ICAO phraseology.

The appropriate linguistic response must be made to a message transmitted by a pilot.

It is important to resolve misunderstandings in communication that may be due to limited language competency, noise and/or distortion, or to a stress-induced situation. To achieve effective communication a strict adherence to published ICAO radiotelephony phraseology is required

The test consists of three mandatory papers:

Paper 1 (Listening) comprises 8 sections (80 test items + the three first items which are not rated) in booklet form. The candidate listens to recorded pilot messages and writes the required information (short text responses) in the space provided. It is not a test of written English and the candidate is not penalised for grammatical errors. Some sections are played twice. The test lasts approximately 40 minutes.

Paper 2 (Oral responses) presents the candidate with a series of 32 charts (the first two are not rated) depicting aerodrome, approach radar or en-route radar environments. The positions of aircraft on each chart are indicated by appropriate symbols. The aircraft in communication with the controller (candidate) is highlighted. For each chart the candidate listens to a recorded pilot message and then makes an appropriate response which is recorded for later analysis.

A trained rater scores the candidate's responses for language performance AND for the appropriateness of the response. Appropriateness is safety-related. The recording is played once. The test lasts about 20 minutes.

Paper 3 (Oral interaction) is presented in two phases on a one-to-one basis between the candidate and a trained interlocutor. Communication is not visual, with the candidate being seated behind a screen, or ideally in another room.

In phase 1 the interlocutor plays the role of a pilot experiencing a developing unusual situation. The candidate is the controller who must ascertain and understand the problem being experienced by the pilot.

In phase 2 the candidate is required to brief the ATC supervisor (the interlocutor) about the unusual situation and may be asked to confirm and clarify information.

The test is currently being revised with a modified format, one book containing the texts and questions together, an improved balance between different types of item and the number of items being increased to eighty.

The revised rating scales applicable to Paper 2 and Paper 3 of PELA have evolved from a comparison of the currently valid PELA rating scales with the ICAO recommendation for an international Standard of language proficiency for pilots and controllers (level 4 on the ICAO rating scale).

The descriptors for features specific to PELA have been retained (phraseology, formulaic phraseology and the impact of appropriateness on safety. In PELA, pronunciation remains at a higher level than at ICAO level 4.

PELA is a test designed to satisfy:

ICAO SARPS

- ESARR 5
- Common European Personnel Licence for air traffic controllers
- ICAO language proficiency standard
- ICAO published phraseology

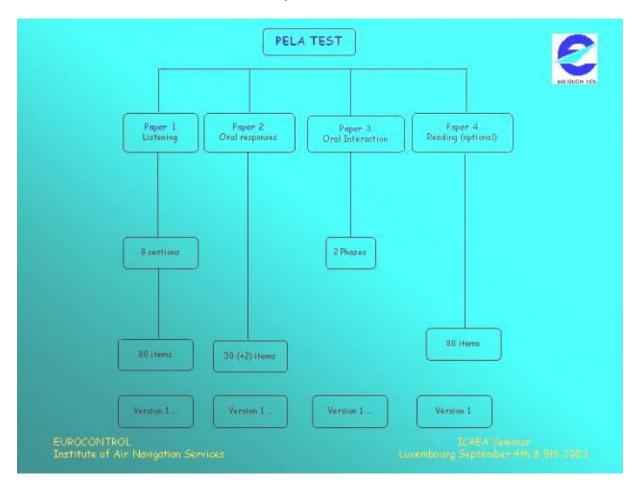
A web-based PELA test comprising:

- Paper 1 Listening
- Paper 2 Oral responses

has been developed. It renders administration more efficient and is more appealing to candidates.

The PELA Test has been evaluated and today, personnel in more than 20 European States are accredited to administer and rate the test.

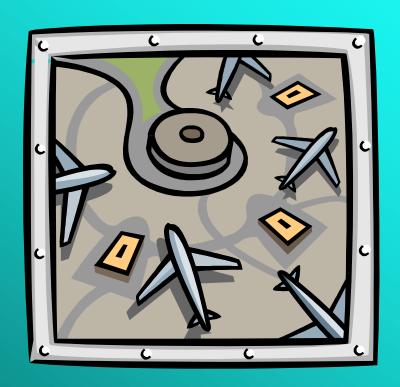
Belgium, Bulgaria, Czech Republic, Denmark, Estonia, Finland, France, FYROM, Hungary, Latvia, Lithuania, Luxembourg, Moldova, Poland, Portugal, Romania (ROMATSA & CAA), Slovakia, Sweden, Switzerland and Turkey.





PELA

"Putting safety into aeronautical communication"





PELA Proficiency in English language for Air Traffic Controllers

A test to meet international language proficiency standards

Adrian Enright
EUROCONTROL



PELA

- Introduction
- Linguistic ability
- Test structure
- Rating scales
- Placing PELA
- Current development



PELA

- Language is still a contributory factor in many aircraft accidents
- EUROCONTROL initiative June 1988
 - 1st English Language Workshop
- Workshop concluded that controllers
 - should demonstrate listening, pronunciation and comprehension skills in English



PELA - the design

- · Reflects a range of ATC tasks
- Focuses on language
- · Requires a level of general English
- · Is for student air traffic controllers



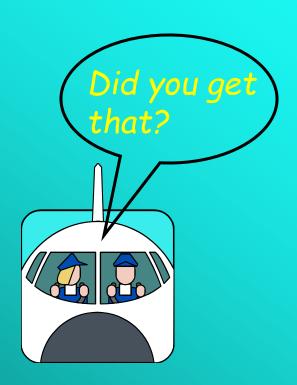
PELA

- Introduction
- Linguistic ability
- Test structure
- Rating scales
- Placing PELA
- · Current development



PELA - linguistic ability







PELA - linguistic ability

- to produce messages in natural language:
 - in both usual & unusual situations
- to give appropriate response to pilot messages

continued...



PELA - linguistic ability

- · resolve misunderstanding due to:
 - limited language ability
 - noise and/or distraction
 - stress
- maintain strict adherence to ICAO phraseology



PELA

- Introduction
- Linguistic ability
- Test structure
- Rating scales
- Placing PELA
- · Current development



PELA - Test Structure

- · General layout of the test
 - Paper 1 Listening
 - Paper 2 Oral Responses
 - Paper 3 Oral Interaction, and
 - Paper 4 Reading (optional)

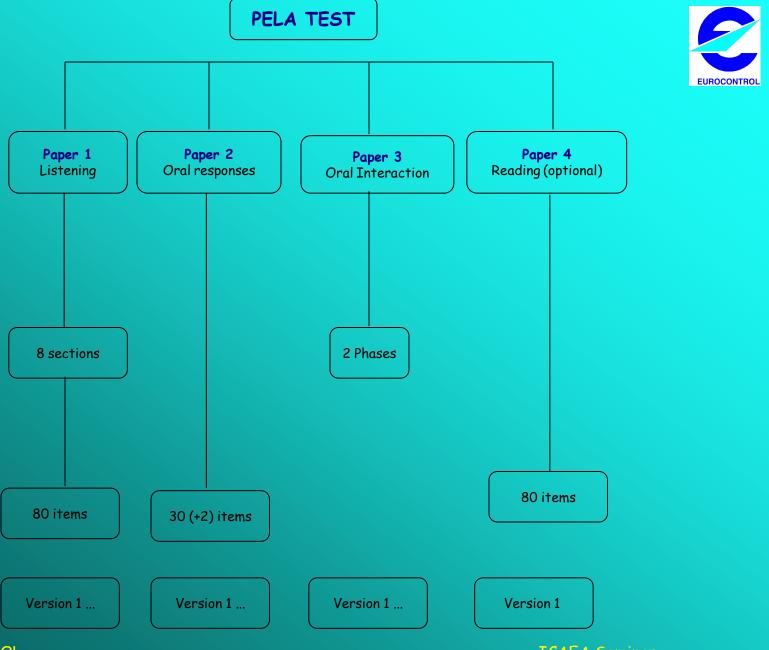


PELA Plain English + R/T

 Candidates are expected to adhere strictly to published ICAO radiotelephony phraseology,

and where necessary,

 demonstrate an ability to communicate effectively in natural English in nonroutine and emergency situations.





PELA - Paper 1 Listening

All versions have the same format and presentation

- Existing test items retained but number increased from 68 to 80
- Better balance of ATC functions (TWR, APP & ENR)
- Improved layout less distraction for candidates
- Simplified tables with easier to read instructions
- Pass mark increased from 50% to 70% (56/80)



PELA Paper 2 Oral Responses

- More versions of TWR, APP and ENR
- Revised charts
- Two practice items included
- Number of test items increased from 28 to 30
- Pass marks for language performance & appropriacy both increased to 70% (21/30)



PELA - Paper 3 Oral Interaction

- New scenarios developed for tower and approach functions
- Common PELA charts (same as for Paper 2) applied throughout



PELA - Paper 4 Reading

- Format revised
- One book includes texts and questions together
- Improved balance of item types
- · Test items increased to 80



PELA

- Introduction
- Linguistic ability
- Test structure
- Rating scales
- Placing PELA
- · Current development



PELA - Rating Scales

- Rating scales for Paper 2 and Paper 3 are aligned to the ICAO standard of language proficiency required for aeronautical communication (Level 4 operational)
- Terminology standardised

continued...



PELA - Rating Scales

- Descriptors for features specific to PELA retained - phraseology, formulaic phraseology and the impact of appropriacy on safety
- "Pronunciation" for PELA remains at a higher level than ICAO level 4



PELA

- Introduction
- Linguistic ability
- Test structure
- Rating scales
- Placing PELA
- · Current development



PELA - a test to satisfy

- · ICAO SARPS
- · ESARR 5
- Common European Personnel Licence for air traffic controllers
- · ICAO language proficiency standard
- · ICAO published phraseology



PELA

PELA is fully compliant with ICAO language proficiency standard (Level 4) for aeronautical communications

PELA = ICAO + RTF
 (for air traffic controllers, in English)



PELA

- Introduction
- Linguistic ability
- Test structure
- Rating scales
- Placing PELA
- Current development



Current Development

Web-based PELA

- Paper 1 Listening
- · Paper 2 Oral Responses
- · More efficient administration
- More appealing to candidates



PELA - Implementation

20 EUROCONTROL and ECAC States

- Belgium, Bulgaria, Czech Republic, Denmark, Estonia, Finland, France, FYROM, Hungary, Latvia, Lithuania, Luxembourg, Moldova, Poland, Portugal, Romania (ROMATSA & CAA), Slovakia, Sweden, Switzerland and Turkey.





EUROCONTROL
Institute of Air Navigation Services

ICAEA Seminar Luxembourg September 4th & 5th 2003



The PELA Test





TESTING ENGLISH IN THE WORKPLACE

Nick Saville Cambridge ESOL Director Research & Validation

The present paper will address the needs that exist to assess language in a working environment and in the case of specific purpose testing ask the question how specific the language should be in the assessment.

One of the effects of a global economy has been the growing number of transnational companies and companies with branches and subsidiaries across the globe. As a result languages other than the language of the country are required to communicate. Furthermore computer screen-based technologies and working habits and the intensive use of e-mail and internet have driven the need for communication skills and harmonised working practices.

Communicating in the workplace may take the following forms:

- establishing and maintaining business contacts,
- carrying out routine business transactions,
- making plans and arrangements,
- giving and obtaining information about work,
- conducting and attending meetings with all the attendant skills such as note taking, questioning, confirming, summarising, making presentations etc..

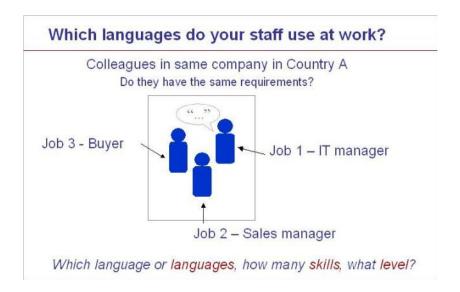
Companies are not specialists in language education. One may wonder whether schools are able to give priority in the curriculum to subjects that meet the language requirements of work forces. In Italy, Silvio Berlusconi has set three 'I' goals in education and training: 'inglese', 'informatica' and 'impresa'. But this will take time to have any far reaching effect.

Companies need to know how to:

- use an international standard to define language skills and the required level and
- recruit new staff and /or train existing staff to the appropriate level of language.

There are considerable risks involved if the assessment is unreliable.

Colleagues in Country A Colleagues in Country B Same company Clients in Country C



We will look at two case studies of European companies.

Example 1:

A company is setting up a new operation - there are a large number of job applicants whose language skills need to be tested. All of them for English, and some for French or German. They only require a brief summary of the applicants' overall ability, but they need it for all of them within 2 weeks. Only BULATS has both the flexibility to deliver the tests and the results within 2 weeks, and the range to do this in a standardised way across 3 languages.

In our first example we will take the aerospace industry in Europe with multinational collaboration between partners in France, Germany, Spain and the UK. There is a multicultural workforce incorporating 30 nationalities. Assessment is needed across 4 languages and for all levels of ability for recruitment, determining language training needs and for checking the effectiveness of the training given.

Some examples of job profiles:

- A Dutch contract administration manager based in Germany responsible for co-ordination between the contracts department, manufacturing, the banks and the customer. Fluent English is mandatory and a good knowledge of German and French is also required.
- A Spanish production manager based in Spain receives a lot of visits from colleagues outside Spain, e.g. people interested in seeing technical progress and how it is applied. He needs English for both technical aspects of his working the office and for social purposes.
- An English embedded software engineer based in France responsible for the new generation of on-board computers. He works with an international team and uses English for his work in the office. He needs French for everyday communication in the community and also for good will within the company.

Example 2:

An international food processing group, leading supplier of food processing equipment making pasta, chocolate, breakfast cereals, snack foods etc. Their headquarters are in Switzerland but they are present in over 100 countries with a staff of 6,500 employees.

Multilingual assessment is required to audit all the employees' language skills e.g. 200 tested in one day in one office. The following languages had to be assessed: English 84%, German 10%, Spanish 4%, French 2 % in order to test employees who will be posted abroad.

Every 3 months 20 staff at headquarters must be assessed in English. There are 5 of the staff to be selected with very high level skills e.g. a detailed report is needed of their speaking and writing ability.

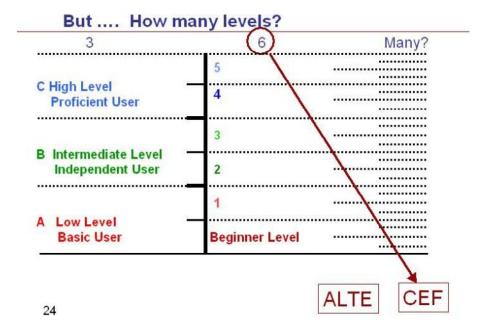
In this example, the company has a smaller number of personnel to be assessed four times a year. This again requires flexibility in administration, being able to run the test at times to suit the company. Of these people, 5 are at the top level in terms of their work and the linguistic demands on them - i.e. they're senior managers, legal advisers, etc, and the company needs to be absolutely sure of their ability to use the language at that level. In this case, the client needs a detailed report on all the candidates, particularly on the five top personnel. A suitable service from BULATS could be providing the Standard BULATS Test and the Writing Test for all staff, and then provide the BULATS Speaking Test for the five top personnel.

What kind of language tests do companies need?

- They must be linked to international standards so internationally recognised for mobility purposes.
- They must also have multilingual options for comparisons across languages and be flexible in administration with quick results and flexible options for testing skills including speaking and writing.
- They must be easy to interpret with meaningful results for non-specialists using Can Do descriptions.
- Finally, they must be supported by professional advice e.g. what level is adequate for a job or future training?

Language levels should be described in terms that are understandable by users who are not language specialists e.g. "Can understand with ease virtually everything heard or read. Can summarize information etc."

Language levels can be represented by ladders or steps but how many levels should be used? The Common European Framework (CEF) contains six.



Like other frameworks, the CEF uses two dimensions: a set of levels and descriptive categories like language skills.

The Common European Framework (CEF)

C Proficient User

C1 Effective
Operational Proficiency

B2 Vantage*

B1 Threshold*

A2 Waystage*

A Basic User

A1 Breakthrough

Using only two dimensions however is an oversimplification and other factors may be introduced like domains of use such as Work, Study or Tourism.

Three dimensions of CEF: Levels, Skills, Domains, e.g. for work or study purposes Reading, Writing, Listening, Speaking Work Study Levels

The time it takes a learner to go from one level to another is another consideration. It may take a beginner 375 hours to reach level 2.

Language skills might be classified as follows. In certain areas, particular skills may be more critical than others like interaction for pilots and air traffic controllers.

A Framework of Language Levels

Assessment grid - How many skills?

				E V	EL	- S	
		B/T	1	2	3	4	5
	Listening						
UNDERSTANDING	Reading						
ODEAKING	Spoken interaction						
SPEAKING	Spoken Production						
WRITING	Writing				1.2		

Here are examples of listening Can Do descriptors at the breakthrough level and level 5.

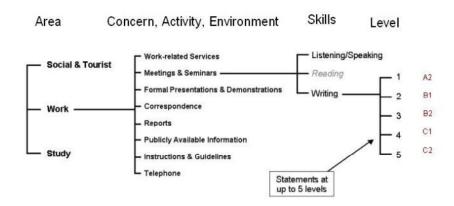
A Framework of Language Levels

Self-assessment

A Framework of Language Levels

Self-assessment

We can see in this representation how the CEF can be interpreted to help the test developer to define a specific language use domain.



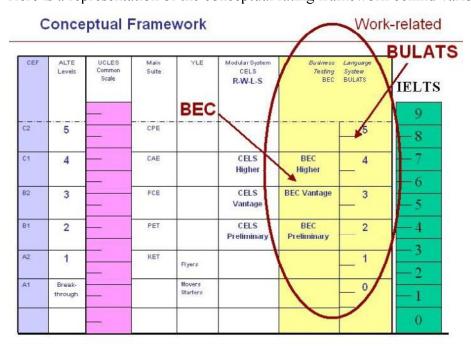
e.g. participating in meetings and seminars in the office

Can Do statements provide meaningful descriptions of the levels: "Can make outgoing calls, pass on simple messages e.g. 'Mr Flores' flight is late'. Can write reports of a factual nature of sets of instructions in own field, but work still requires checking. Mid Level Report Writing."

Assessment is needed in the workplace mainly to provide management information for:

- evaluating the existing level of staff within a company (language audits);
- evaluating the language level of new staff and job applicants (recruitment)
- placing staff on suitable courses for language training;
- evaluating language training courses;
- standardising levels across transnational companies in different locations and for different languages.

Here is a representation of the conceptual rating framework behind various tests.



Various specific test characteristics can be determined and questioned:

<u>Distinguishability:</u> is it possible to accurately describe a specific purpose language use domain, and if it is, can we demonstrate that this domain is distinguishable from any other domain?

Criticism: as all specific purpose contexts share language with other areas there are no clear borders so it is not really possible to tightly define what we mean by a specific language use domain.

<u>Authenticity:</u> is it possible to create a truly authentic test that is practical? Two perceptions of authenticity can be distinguished:

- Situational authenticity in which the task reflects the domain in terms of appearance and content;
- Interactional authenticity where the task reflects the domain in terms of the cognitive processing required.

Criticism – It is not possible. No problem to make such a test – e.g. we could ask candidates to wear a hidden microphone over a period of time while they work and then later listen to the tape and make a judgement. However, concerns such as time (how long to wear the apparatus, how long to award a score); ethics (could/should we only tape work-related material; if the job is high-stakes and language is critical, health and safety issues etc.);

<u>Replicability:</u> The likelihood is that all test events will be different and uncontrolled. However, this perspective only sees authenticity from the situational perspective. There are theoretical issues and criticisms of Language for Specific Purposes (LSP) tests:

Non-Language Factors: In a test of language in a specific domain is it possible to separate ability in the domain from language ability in the domain?

Criticism – No it is not, at least not definitely. There have been suggestions that we need an *indigenous* scale (created by linguists in association with LU domain specialists. Such a scale has been made (Abdul Raof 2001), but not trialled in a large scale test. Questions then are: Who should use the scale, language specialists or LU Domain specialists? What inferences can we draw from the test score?

Threshold Level:

Is it possible to describe the minimum language proficiency level required for specific purpose proficiency is attainable? Dan Douglas gives a working definition of specific purpose assessment: "One in which test content and methods are derived from an analysis of a specific purpose target language use situation (TLU) so that test tasks and content are authentically representative of tasks in the target situation, allowing for an interaction between the test taker's language ability and specific purpose content knowledge on one hand and the test tasks on the other."

In a recent review O'Sullivan characterised LSP tests:

- LSP tests tend to be "market driven",
- many tests are infrequently updated and do not change to reflect changes in the TLU domain or in language testing theory,

- few tests include skills options e.g. all four skills,
- low level of support is provided for most tests,
- there is a tendency against highly specific tests.

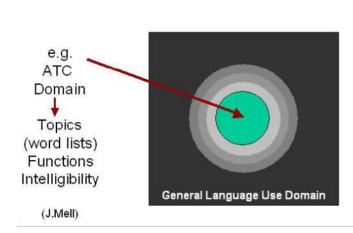
Some interesting points can be taken from this review:

- Large-scale tests tend to have originally been produced at the behest of government agencies (though the trend is that international tests are more and more being produced to meet either perceived or established market needs in other words the TLBPs are more and more market driven).
- As the markets (and the test taking population) change there is little sign that the tests have been revised to meet the change (the BEC revision seems to be the first such project, there has been no literature in the public domain on many of the tests reviewed so it has not been possible to make a judgement regarding them in this respect).
- There has not really been a tendency for changes in proficiency language testing practice to be reflected in TLBP practice (the exception among the tests reviewed seems to be the BEC).
- Few TLBPs include papers related to the four skills of Speaking, Writing, Reading and Listening (the principal exception again being the BEC suite).
- There appears to be a relatively low level of support material available, though all of the UK-based tests offer practice or past papers at no charge to test takers.
- There is a clear tendency against very highly specific tests, for example a test of language for chartered accountants. Instead, the tests on the market appear to be more general in nature, though set in the context of business or commerce.

All tests can be placed somewhere along a specificity continuum, from the non-specific or general to the highly specific.

The notion of the CORE language is represented here being distinct from general language use, but that in any LTU domain there will be not only a distinct core, but also an area in which language can be shared either with the general domain or with other specific domains.

The notion of 'Core' SP language



It is this notion of CORE that lies at the heart of LSP testing.

Both of these tests are from the domain of business language testing:

"General Business English" - Example Tests

Business language domain

Test 1

Items	Main Skill Focus	Focus	Format	Marks
1-40	Grammar	Incomplete Sentences	4 option MCQ	40
41-60	Grammar & Vocabulary	Error Recognition	4 option MCQ	20
61-100	Reading for specific information	Reading Comprehension	4 option MCQ	40
			Total Marks	100

Example Tests

Business language domain

Test 2
From
BEC Reading

Items	Main Skill Focus	Focus	Format	Marks
1-5	Reading & Vocabulary	Understanding intended meaning (short texts, e.g. signs)	3 option MCQ	5
6-10	Reading & Vocabulary	Understanding basic vocabulary (from short input – e.g. table of contents)	Matching (5 from 8)	5
11-15	Reading interpreting visual information	Interpreting information from charts	Matching statements to chart data	5
16-22	Reading Comprehension	Comprehension of report (approx 150-200 words)	T/F/not included	7
23-28	Reading Comprehension	Comprehension of written text (approx 300-400 words)	3 option MCQ	6
29-40	Grammar (in context of reading text)	Grammar use in context (Rational deletion cloze)	Cloze (3 option MCQ)	12
41-45	Reading & information transfer	Reading for specific detail from two written inputs	Form completion	5
			Total Marks	45

Comparing the two tests in terms of what they test and how they test it is quite interesting.

Comparing the Tests

	TEST 1	TEST 2
The construct	Unclear, general or business focus?	Business focus, though elements of general
The method	MCQ only	Varied
Skills coverage	Listening & Reading	4 Skills
Measurement qualities	High reliability (alpha)	Good reliability (alpha)
Degree of specificity	Not specific	Moderately specific
Authenticity	Not authentic	Quite authentic
Impact of non-language factors	Low	Low, but greater than Test 1
Reporting of test performance	Norm referenced — On a scale of 0-990	Criterion referenced - Pass/Fail grades linked to CEF

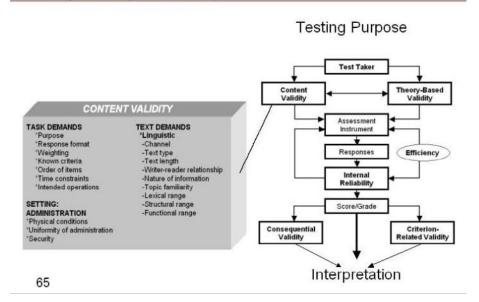
O'Sullivan (2003) argues that the notion of degree of specificity brings with it the related notions of **generalisability** and **situational authenticity**. Where a test is seen to be positioned towards the specific end of the continuum, the potential for generalisation for test performance beyond the specific situation is reduced – it is difficult to imagine a test that could be placed at the extreme end of the continuum as this would be focused only on a very limited 'core' language.

In an attempt to re-conceptualise the specificity argument he proposes looking at the continuum. O'Sullivan suggested that we see the LSP test as somehow focusing on the CORE and/or the GENERAL along a sliding scale.

As tests are more specific they become more situationally authentic (where the task reflects the domain in terms of appearance and content) though less generalisable beyond the context or domain being tested.

This diagram is limited by our ability to represent the notion of general and core in anything but a two dimensional diagram. In reality, once we move beyond the distinguishable core we are in the domain of general language use; the figure implies that only a part of this domain can be represented in the test sample. What is successfully represented in the figure is the idea that when a test is more 'specific' in its focus, the greater will be the importance of the core and when as the test is more 'general' in focus, the less part can be played by language from a specific core. The question we must now ask is how do we know that a test is either specific or general in focus?

In order to re-conceptualise specificity we need to go back to test validity and ensure that we have a clear idea of what we mean by content validity.



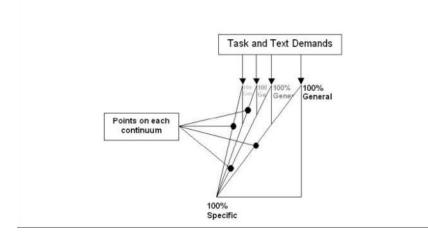
This figure (on the right) is a representation of a the basis for a new set of validation frameworks being developed by Cyril Weir and Barry O'Sullivan ant CRTEC, the University of Surrey Roehampton, UK.

Here we can see where the different elements of a test fit together in terms of validation. A closer look (on the left) at what we mean by content validity shows that it can be seen to have the following aspects:

- task demands
- text demands
- setting and administration demands

The former two of these demands have a very definite role to play in defining the specificity of a test. If we can demonstrate that an LSP test is different from a general purpose test in terms of these demands then the distinguishability argument can be rejected. The problem is how can we do it?

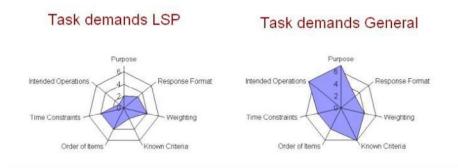
One way to do this is to visualise specificity not along a single continuum but as being related to the position of a test along a whole series of continua – each one related to an element of task and text demands.



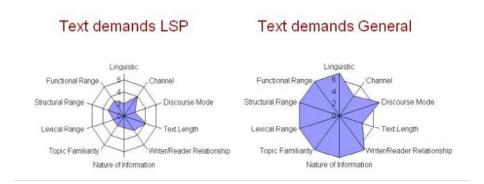
In order to put this idea to the test, lets now look at how Test 2 compares with a general purpose language test designed for a test population at the same level.

The two tests were looked at by a small group of testing specialists. They were asked to rate each test on each of the task criteria using a Likert scale (1 = highly specific, 7 = completely general). When the two tests were compared it was found that there were clear differences in terms of task demand (where an element was seen to be not applicable it was given a score of 4.)

Comparing Task Demands

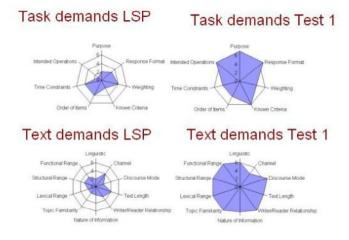


When the same was done with the text demands it became clear that the differences were even greater.



Looking back at the two tests we compared earlier, we can see that there is a clear difference between them (though the profile for Test 1 seems to be very close to that of the General test looked at earlier).

Comparing the two tests:



The relevance is important.

This demonstrated that we can observe differences between different specific purpose tests (though admittedly Test 2 is a poor example of this type of test – even though it is marketed as such!)

If this is applied to a highly specific test, such as a test for air-traffic controllers for example, the profile becomes even more acute – with many of the elements rated at 1 (highly specific)

Task demands Highly Specific Text demands Highly Specific



We have seen that situational authenticity is related to content validity – and that this is all seen to be a part of the notion of test specificity.

Authenticity Situational - related to content validity Interactional - related to cognitive processing THEORY-BASED VALIDITY INTERNAL PROCESS Goal setting and monitoring Visual recognition Pattern synthesizer Test Taker **EXECTUTIVE RESOURCES** *Linguistic knowledge -Grammatical (lexis, syntax) -Textual -Functional Theory-Based Validity Validity -Sociolinguistic 'Content knowledge - Internal - External Assessment

However, the other aspect of authenticity is interactional - where the task reflects the domain in terms of the cognitive processing required.

For our test to be truly interactionally authentic, it should elicit the same cognitive processing as a task in the actual TLU domain context.

We can only discover if a test is interactionally authentic by a post test analysis, and in order to do this we must go back to the candidates. Processes such as questionnaires, interviews, retrospective or 'real time' protocol analysis can be used to elicit the evidence needed to demonstrate this.

We have seen that situational authenticity is related to content validity – and that this is all seen to be a part of the notion of test specificity.

Overall conclusion

- Level Screening for level is important
- Skill Identification of skill for particular jobs is important
- Domain General or Specific Will a general purpose test do the trick?
- Practicality is a major concern in developing new and highly specific tests

In language testing, practicality is always going to be central to our ability to devise and administer tests; no theoretical rationale is of particular value if it cannot be operationalised. O'Sullivan comments that: "All tests can be seen as lying on a specificity continuum between the highly specific and the general purpose. This continuum is multi-componential and includes the twin aspects of authenticity – situational and interactional. A specific purpose test will be distinguishable from other tests (both specific and general purpose) in terms of the domain represented by the demands of its tasks and texts, and in terms of the cognitive processing it elicits." This definition is not meant to counter that offered at the beginning of this talk (Douglas 2000) but to offer a means of operationalising that definition.

If we return to the original criticisms, LSP tests are distinguishable from both general purpose tests and from other LSP tests and authenticity (both situatiuonal and interactional) are elements of specificity.

<u>Distinguishability:</u> Is it possible to accurately describe a specific purpose language use domain, and if it is, can we demonstrate that this domain is distinguishable from any other domain?

Answer: LSP tests are distinguishable from both general purpose tests and from other LSP tests.

<u>Authenticity</u>: Is it possible to create a truly authentic test that is practical?

Answer: It is possible to create tests which approach the TLU domain for authenticity, but it must be done for both interactional and situational authenticity. Authenticity (both situational and interactional) are elements of specificity

Non-Language Factors: In a test of language in a specific domain is it possible to separate ability in the domain from language ability in the domain?

Answer: Non-language factors have little impact in practice

To conclude we may say that screening for level is important just as identification of skills for particular jobs is important. Practicality is a major concern in developing new and highly specific tests.



Ethics and Quality Control in Language Testing

A Quality Management System for International Language Examination Providers

Nick Saville

ALTE Code of Practice Working Group

Outline of presentation

- Part 1 Introduction and background
- Part 2 The ALTE QM System
- Part 3 Self-evaluation and quality improvement
- Part 4 Setting standards issues of contextualisation
- Part 5 Conclusions and Discussion

Part 1 – Background

- The Association of Language Testers in Europe
- How long has ALTE existed?
- Members of ALTE
- Aims of ALTE

Aims of ALTE

- To establish common levels of proficiency in order to promote the transnational recognition of certification in Europe
- To establish common standards for all stages of the language-testing process
- To collaborate on joint projects and in the exchange of ideas and know-how

Background

Summary quote:

ALTE has developed, not only a Code of Practice, but a Framework of Levels which ... has contributed to raising international debate about levels and standards. ALTE holds very useful conferences, encourages exchange of expertise among its members and has certainly raised the profile of language testing in Europe.

Web-site: Alte.org

Developing the ALTE Code of Practice

... common standards for all stages of the language-testing process

1994 - ALTE published its first Code of Practice – SEE HANDOUT

Initially drafted and discussed by ALTE members in 1991-93

It drew on *The Code of Fair Testing Practices in Education* produced by the Washington D.C. Joint Committee on Testing Practices (1988)

It was a broad statement of what the *users* of the examinations should expect

Focused on the *roles and responsibilities* of stakeholders in *striving for fairness*

- SEE HANDOUT

Developing the ALTE Code of Practice

Striving for fairness for stakeholders of the examinations

The Code of Practice identifies the roles of three groups of stakeholder in the testing process:

- the examination developers e.g. members of ALTE
- the examination takers primary users who take the examinations by choice, direction or necessity
- the examination users secondary users who require the examination for some decision-making or other purpose

Developing the ALTE Code of Practice

The Code of Practice lays down four broad areas of responsibility:

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

Striving for fairness is a *shared responsibility* involving all stakeholders

Considerations for ALTE Members

In developing the Code of Practice ALTE Members were asking themselves the following questions:

Are we doing the right things?

Are we doing them right? How

Can we do things right every time? How well

 But the 18 statements in the COP provided little practical guidance on implementation and on setting standards

Web-site: Alte.org

Considerations for ALTE Members

Two aspects of quality needed to be considered:

- Better understanding of Principles of Good Practice theoretical concerns and knowledge
- The practical application of the principles within each ALTE organisation
 - developing better systems
 - managing innovation and change
 - monitoring standards

ALTE Principles of Good Practice

A supplementary document Principles of Good Practice for ALTE Examinations was co-ordinated by Milanovic and Saville - 1991 and 1993

Discussed at ALTE meetings between 1992 and 1994

Re-drafted by ALTE COP working group – 2001

This document sets out in more detail the *principles* which ALTE members should adopt in order to achieve high professional standards

Council of Europe publication: Language Examining and Test Development (Strasbourg, October 2002)

Web-site: Alte.org

The ALTE Principles of Good Practice

Based on VRIP features

- Validity
- . Reliability
- Impact
- . Practicality

Plus additional feature

. Quality of service

Note on Practicality

A practical examination is one that does not place an unreasonable demand on available resources (money, people, technology, organisational structures, time/space)

If available resources are exceeded, then the examination is not practical. In this case, the examination developer must:

- modify the design of the examination
- make a case for an increase or reallocation of resources

If the former, is it still "useful"?

Practicality concerns at least the following:

- Design features, e.g. related to format and content of the four skills approach
- Test production features, e.g. replicability of tasks, availability of pre-test candidates
- Availability of the examination (frequency, location etc.)
- Level of fees to be paid by test takers
- Central Costs in terms of production, marking, scoring, validation
- Local costs in relation to administration at centres
- Central Administration entry procedures, exchange of data with centres, collection of fees, despatch of materials, marking and grading procedures, issue of results
- Local Administration at Centres
- Security
- Special circumstances (arrangements for candidates with special needs)

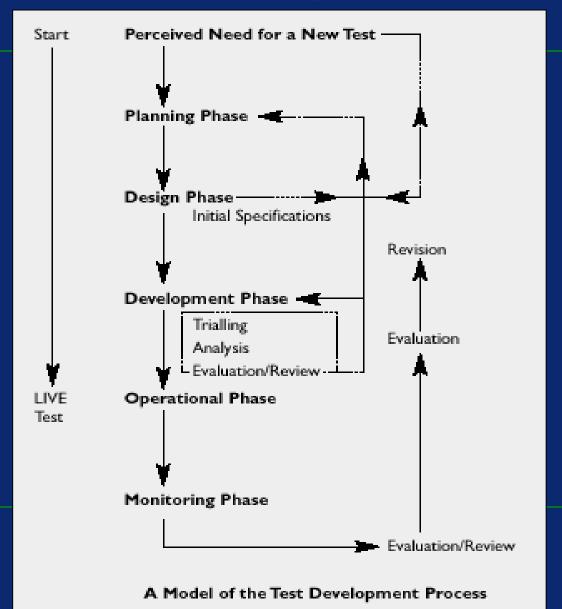
Practicality

Whenever a new or revised examination is to be developed, there should be procedures in place to address at least the following areas:

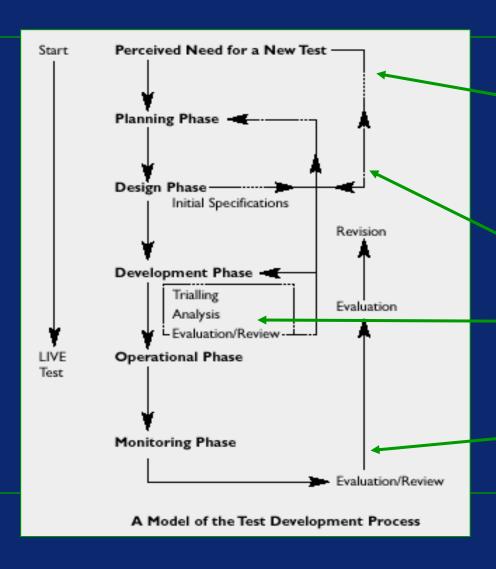
- the management structure for the development project
- a clear and integrated assignment of roles and responsibilities
- a means of monitoring progress in terms of development schedules and resources
- a methodology for managing the examination production process when the examination becomes operational:

item writing, vetting, moderation, pre-testing, item banking, question paper construction

The Test Development Process



Linking Quality and Test Development



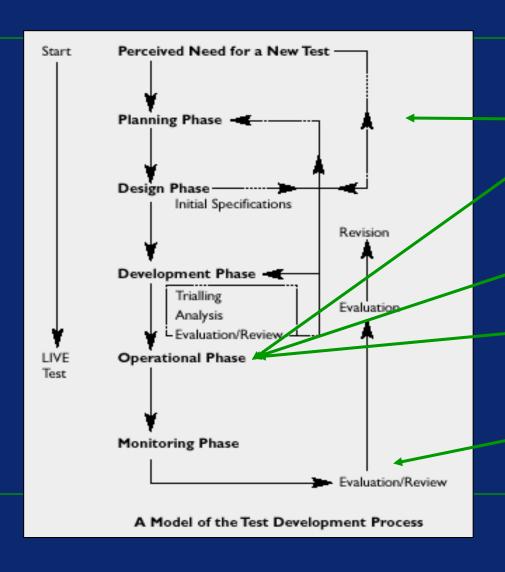
Identifying stakeholders and their needs

Linking these needs to the requirements of **test usefulness** (VRIP)

- theoretical
- practical

Long term, Iterative Processes - a key feature of validation

Linking Quality and On-going improvements



Test Development and routine production

Administration & Logistics

Marking Grading & Results

Test analysis and Postexamination Review

Part 2 - a Quality Management System

2000 onwards – recent attempts at "putting the principles into practice" (COP Working Group – 5 members)

2001

- Revised documents COP as self-evaluation checklist
- Consideration of International QM Systems

The European Foundation for Quality Management EFQM

The EAQUALS Inspection Scheme

Also many schemes within *national* education systems

The Aim of a QM System?



Aim = To improve customer satisfaction

In language testing, who are your customers?

Features of QM Systems

Effective initiatives for QMS usually involve:

A public statement – often in the form of a *Code of Practice* or Charter

A commitment to the *change process* typically involving the following steps:

- Define your mission = role of institution, future ambitions/aims
- Assess what you currently do
- Identify areas in need of improvement
- Decide on measures of improvement and the action plan
- Carry out action to bring about improvement
- Review progress and revise plan

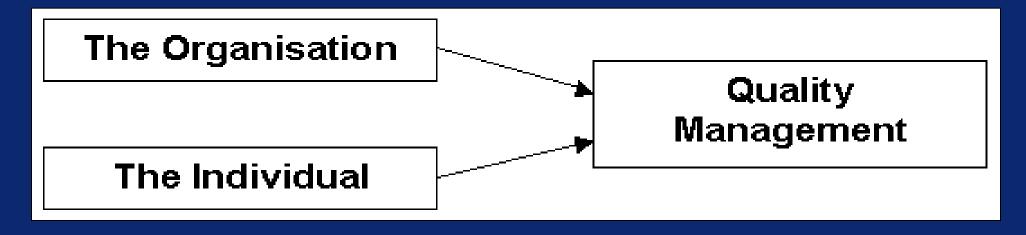
Applying an ALTE QMS

QMS approach:

Seek to understand the nature of your "organisation" and to involve all participants (stakeholders)

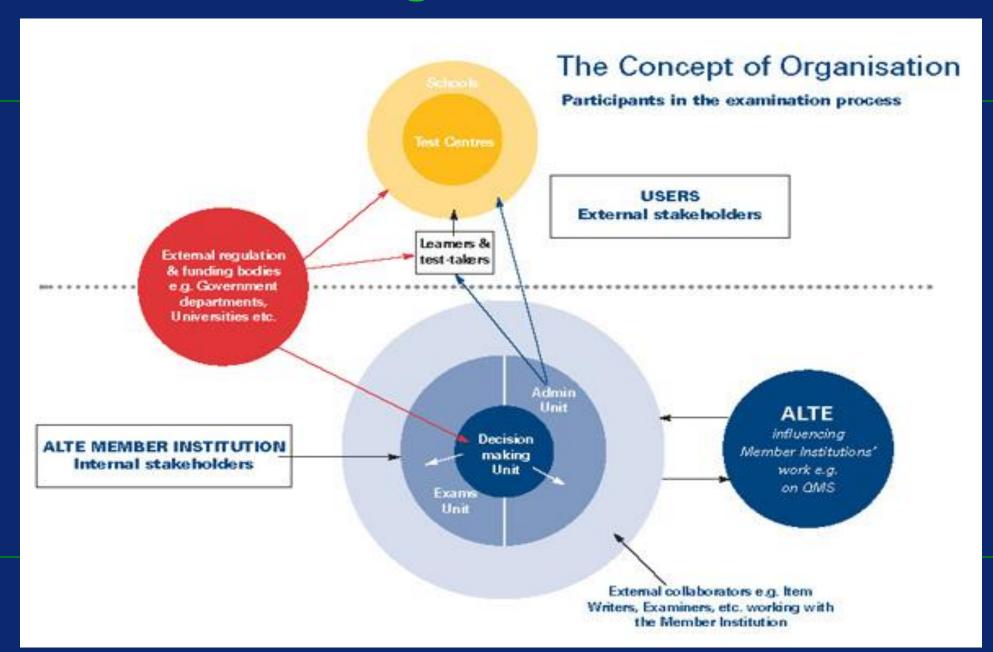
Applying an ALTE QMS

A successful QMS requires behavioural and cultural change - and three components of your systems must be aligned:



This relates to innovation and change management theories

BUT each ALTE organisation is different!



Applying an ALTE QMS

- Establish desired outcomes and impacts within each member organisation
 - Move towards good practice
 - Aim at *continuous* improvement
- Discuss and agree on minimum standards but establish "best practice" models as long-term target
- Monitor quality through self-assessment
- Seek confirmation that standards are being met
 - e.g. through peer review within ALTE

Current phase: ALTE QMS

Adopt the QMS approach and the revised Code of Practice checklists

Apply the checklists to the 4 aspects of the Test Development and Administration Cycle

Test Design and Construction

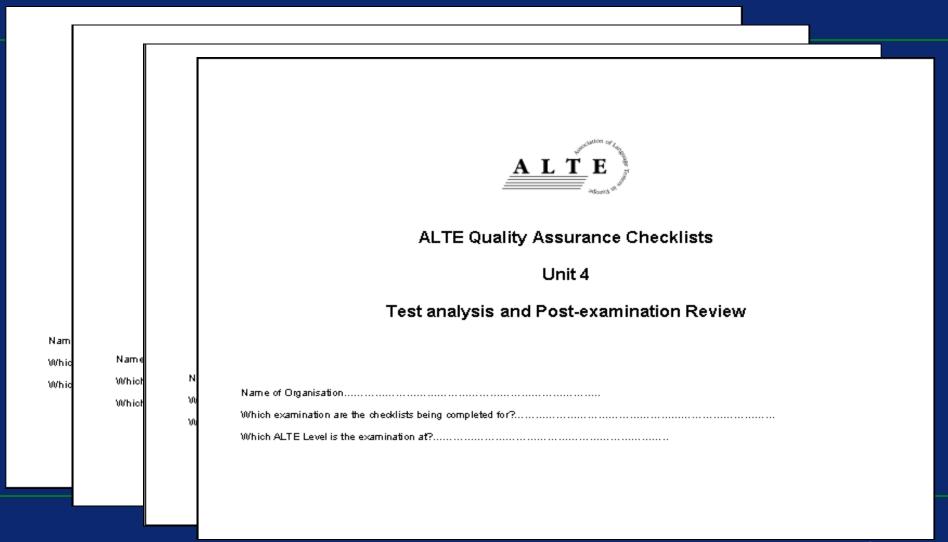
Administration

Processing - marking, grading, issue of results

Analysis and Review

(c.f. Model of Test Development noted above)

ALTE QA Checklists - Units 1 to 4



Unit 1 - Test Construction



A. Conceptual phase

- A1. Describe how and why the examination was developed.
- A2. Describe what the examination assesses.
- A3. Describe what the examination should be used for.
- A4. Describe the population for which the examination is appropriate.

Unit 1 - Test Construction



- A. Conceptual phase
- B. Test development, test construct & context

- B1. Explain how the content and skills to be tested are selected (for each component of the examination).
- B2. Describe the routine procedures used to ensure the appropriateness of the examination for the intended candidates.
- B3. Describe how examination tasks are reviewed and revised to avoid potentially insensitive content or language.
- B4. Describe procedures that help to ensure that differences in performance are related to the skills under assessment rather than irrelevant factors.
- B5. Describe the process of examination development.
- B6. Describe procedures to determine reliability. For example: by skill/paper, overall for the whole examination.
- B7. Describe procedures to determine validity. For example: Construct validity, Content validity, Criterion-related validity.

Unit 1 - Test Construction



- A. Conceptual phase
- B. Test development, test construct & context
- C. Communication with External Stakeholders

- C1. Describe how you provide examination users and takers with information to help them judge whether a particular examination should be taken.
- C2. Describe how you make such information equally available to all candidates.
- C3. Describe how you provide candidates with information they need in order to be familiar with coverage of the examination, the type of task formats, the rubrics and other instructions and appropriate examination-taking strategies.
- C4. Describe which samples or complete copies of relevant documents you provide.

Example from the Conceptual Phase



ALTE Quality Assurance Checklists

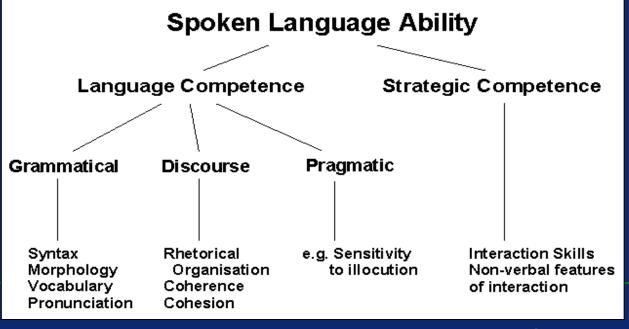
Unit 1

Test Construction

iii. How do you relate your examination to a model of communication and communicative language ability?

e.g. Saville & Hargreaves model of language ability which drive the UCLES Main Suite examinations

This question relates to the construct validity of the test



Example from the Development Phase

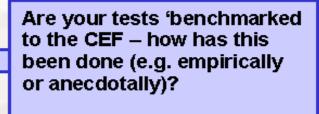


ALTE Quality Assurance Checklists

Unit 1

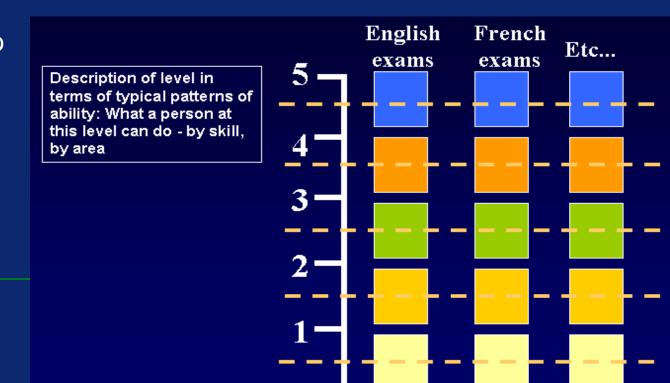
Test Construction

iii. What use, if any, is made of the Common European Framework of Reference?



This question also relates to various aspects of validity:

Construct-related Criterion-related Content-related



Example from the Communication Phase



ALTE Quality Assurance Checklists

Unit 1

Test Construction

i. How are these materials distributed to ensure adequate availability?



CPE

Business English

A variety of distribution channels would ensure that all test takers have ready access to all relevant documents – also a variety of distribution outlets (schools, guidance advisory services, etc.)

This question relates to "informing test takers" and thus to the impact of the test (consequential validity)

EXAMINATIONS EFL HOME PAGE NEWS SUPPORT CONTACTS CAMBRIDGE EFI ON-LINE Downloadable Documents General Documents Young Learners English Tests Regulations, Publishers List Handbooks and Examiner Reports for YLE Lower Main Suite Exams Handbooks for KET, PET Academic Purposes Handbooks for IELTS Upper Main Suite Exams Handbooks, Examiner Reports Other Exams

and Specifications for FCE, CAE,

Handbooks and Examiner Reports

and Specifications for BEC.

CELS

Recognition leaflets

Publishers List

by exam and by country

Information on exam recognition,

Publishers that prepare material for learning Cambridge EFL exams

Unit 2 - Administration/logistics



ALTE Quality Assurance Checklists

Unit 2

Administration & Logistics

i. What guidelines do you provide on:

Suitable venues - location, space etc.?

Seating plans for examination rooms?

Specifications for audio or video equipment?

Suitable rooms for oral assessments?



In how much detail are these requirements specified (and to what extent are their implementation monitored?)

This question relates to the important area of setting test performance conditions – as changes in conditions can affect performance this can be seen as a validity/reliability issue

Unit 3 - Marking, Grading, Results

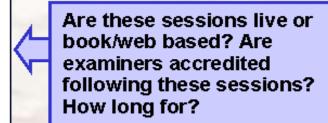


ALTE Quality Assurance Checklists

Unit 3

Marking Grading & Results

iv. Do examiners for speaking/writing attend regular co-ordination sessions? If so, is this obligatory and how often does it take place?



Examiner recruitment & training documents/regulations

This question relates to the establishment of procedures to ensure rater reliability

Unit 4 - Test analysis and post-exam review



ALTE Quality Assurance Checklists
Unit 4

Test analysis and Post-examination Review

i. Do you compare results of different groups in terms of psychometric characteristics in order to detect unwanted differences such as test or item bias?

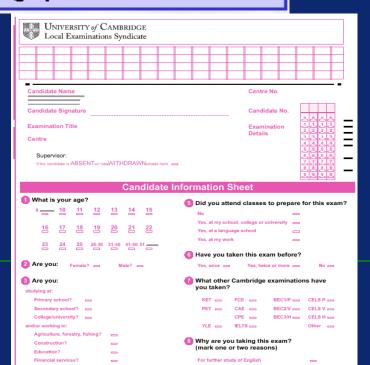


Before doing this you will need:

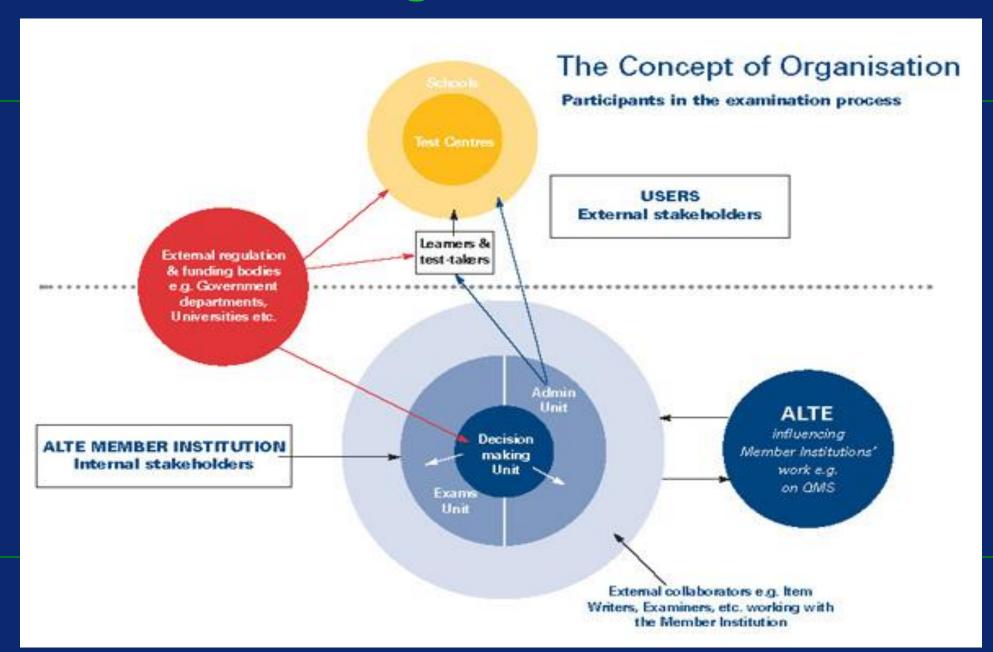
To identify the groups that may be 'at risk'

To have access to relevant demographic data

This question relates to the estimation of test bias – a validity issue



BUT each ALTE organisation is different!



Current phase: ALTE QMS

Identify within each organisation:

Current strengths

Areas in need of immediate improvement

Areas for long-term development

Jointly agree on activities to support ALTE Members in raising standards:

e.g. workshops, sub-groups projects etc.

Part 3

Self-evaluation and quality improvement

Timescale:

- 2000 develop QMS
- 2001-2 revise COP as checklists and conduct piloting
- 2002-3 use checklists for self-assessment, revise checklists, begin peer monitoring and peer support
- 2004 set up formal monitoring systems

Summary of progress so far

٠	2000/1	develop QMS, set up COP Working Group
٠	2001/2	revise COP as checklists and conduct piloting
•	2002	use checklists for self-assessment revise checklists — put into Excel format add glossary and examples focus on test analysis (item-based tests) present work at conferences
	2003	extend use of checklists, establish examples of good practice, extend peer monitoring and peer support
•	2004	set up formalised monitoring systems

Latest results of the exercise:

- Trend analysis
- Use of Code of Practice checklists for determining future activities
- Data gathering and analysis

Trends so far:

Test Design and Construction

Administration

Processing - marking, grading, issue of results

Analysis and Review

- Deciding on future activities:
 - Routine procedures for data gathering and analysis
 - Responsibilities of test providers for the social effects of their tests
 - Dealing with candidates under special circumstances

Part 4

QMS and the setting of minimum standards:

issues of contextualisation and variation between the testing bodies

QMS and the setting of minimum standards

- Power mechanisms in the setting of standards part one:
 imposing norms versus norm negotiation
- Power mechanisms in the setting of standards part two:
 the attraction of those who have the most (symbolic) power and the most prestige

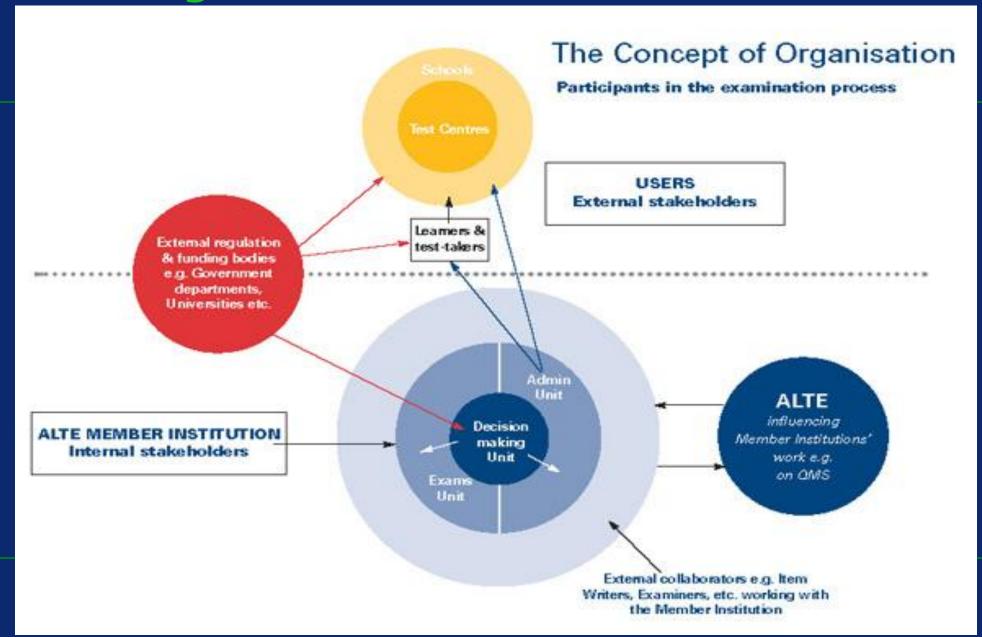
How to set ALTE Quality Standards?



QMS and the setting of minimum standards

- Quality management within ALTE: a step wise process to set MINIMUM standards
- Setting standards: taking the dynamics of cultural and social diversity and variation into consideration

Each organisation is different



Conclusions

The world of testing is more than concepts: it is above all about people

But, as testing is also about power and politics, we have to be aware of the power mechanisms that are in play ...

Contacts:

ALTE COP WG Members

Piet van Avermaet - CNaVT (KU Leuven, Belgium)

Henk Kuijper - CITO Groep (Netherlands)

Ramon Parrondo - Instituto Cervantes

Nick Saville - Cambridge ESOL Examinations

Peter Villads Vedel - Danish Ministry of Education

Acknowledgement: several slides were based on a presentation for ALTE members made by Dr Barry O'Sullivan, Reading University



Pilot and controller skills and competencies

Dr Jeremy Mell
ENAC, Toulouse
jeremy.mell@enac.fr



What we are looking for...

- For setting test objectives:
 - break up the complex system of language and skills of communication into manageable bits for learning and testing
- For writing test items/learning materials:
 - means to contextualise language in meaningful ways
- First steps:
 - Choose amidst the cacaphony of terms: ... requirements, competence, competencies, skills, functions, topics, roles, objectives, elements, conditions, tasks, content, knowledge...
 - Focus on practical abilities as opposed to theoretical knowledge



The bigger picture...

- Pilot/controller competencies
 - ICAO: annexe 1 (Personnel Licensing)
 - Eurocontrol: Common Core Content & Training Objectives (ATC)
 - Australian National Training Authority: Competency Standards (PIL)
 - **–** ...
- Human factors competencies
 - ICAO Human Factors Training Manual, (Doc 9683-AN/950)
 - SHEL model (F Hawkins)
 - **–** ...
- Language competencies for task of radiotelephony communication
 - language content
 - language skills



Language content: overview

- words and phrases
- grammatical stuctures
- prosodic patterns
- phonology



What ICAO tells us (annex 1)....

- Words and phrases
 - «common, concrete and work-related topics »
- Grammatical structures
 - «relevant grammatical structures and sentence patterns are determined by language functions appropriate to the task »
- Phonology
 - «accent which is intelligible to the aeronautical community »



Sept 2003

Content selection

- Language items are <u>exponents</u> of semantic and discourse values:
 - words/phrases are exponents of concepts within certain domains/topics
 - grammatical structures and prosodic patterns are exponents communicative functions related to a given task
- Exponents in aviation radiotelephony may be standard phraseology OR plain language Competencies (J Mell)



Topics: aviation

Domains

- **—** ...
- aerodrome
- aircraft
- approach
- ATC automation
- ATC staff
- cartography
- goods



Topics: non-aviation

Domains

- communication
- geography
- goods
- human behaviour
- laws
- malfunctions
- **–** ...

- Sub-domains (goods)
 - load fixing
 - load distribution
 - loading/unloading
 - packaging
 - substances
 - types of cargo



Language functions - overview

- speaker's intention in producing a given utterance (speech act)
- 4 major groups in aeronautical R/T:-
 - 1. Triggering actions
 - 2. Sharing information
 - 3. Managing the relationship
 - 4. Managing the dialogue



Language functions - groups 1-2

- 1. Triggering actions
 - orders
 - requests and offers to act
 - advice
 - permission/approval
 - undertakings
 - **—** ...

- 2. Sharing information
 - present states/actions/events
 - future states/actions/events
 - immediate/recent past states/actions/events
 - past states/actions/events
 - necessity
 - feasability/capacity
 - **—** ...



Language functions - orders

controller

- give an order
- give an amended order
- give a negative order
- give alternative orders
- give a sequence of orders
- cancel an order

pilot

- announce compliance
- announce non-compliance



Language functions - exponents

- phraseology (ICAO)
 - TURN LEFT
- plain language (eg. assistance to an inexperienced pilot)
 - I WANT YOU TO TURN LEFT BY...
 - NOW YOU MUST TURN LEFT...

—



Language functions - groups 3-4

- 3. Managing the relationship
 - greetings
 - thanks
 - satisfaction/complaint
 - reprimand
 - concern/reassurance
 - apologies
 - **—** ...

- 4. Managing the dialogue
 - opening/closing
 - (self-)correction
 - readback/acknowledge
 - checking
 - repetition
 - (dis-)confirmation
 - clarification
 - relaying
 - **...**



Language functions - marking

- politeness
- directness
- subjectivity
- probability
- certainty

- formality
- assertiveness
- insistence
- •



Phonology - the problem

- Phonology is a crucial parameter
 - sound channel imperfections + absence of visual channel
 - NNS greater reliance on acoustic input
 - least described + least susceptible to conscious control
- Impacts of <u>L1 phonological transfer</u> and <u>regional accents</u> on hearing and speaking (scale of difficulty)
 - NS NS
 - NNS NNS (same L1)
 - NS NNS
 - NNS NNS (different L1)



Phonology - some solutions

- follow ICAO recommendations
 - alphabet, numbers
 - use a dialect/accent which is intelligible to the aeronuatical community
- work towards intelligibility of plain English
 - renounce any NS target for pronunciations
 - define core phonology, eg. Lingua Franca Core (LFC), Jenkins, 2000
 - foster principle of accommodation
 - replacement of 'high-risk' L1 features to increase communicative efficiency
 - a natural inclination if intelligibility is a salient feature of the interaction



Language skills

- cogniton
 - short-term memorisation, rapid response, multi-tasking...
- reception (listening)
 - low acoustic quality, regional varieties, no visual channel, unexpectedness, «hearback »,...
- expression (speaking)
 - reconcile concision/non-ambiguity, clear enunciation, alternate use of phraseology and plain language, alternate use of mother tongue and English...
- relation
 - co-operation (pooling information, collaborative problem solving...)
 - conflict (negotiation, authority)



What more....where to find it?

- Inventories of topics/functions for pilots and controllers for each phase of flight, which are attested in corpora of appropriate authentic communication (needs analaysis)
 - existing inventories (see bibliography)
 - recordings transcribed and analysed
 - subject-matter experts
- Phonology of EIL
 - ongoing research (Jenkins,...)
 - SIGs, journals
- Skills
 - human factors studies
 - training manuals

Meeting about the Association

Subject: Future directions of the association

As background to the discussion we were able to view on the screen the newly installed version of the ICAEA web site, for which the Association is particularly indebted to Bozena Slawinska and PATA (the Polish Air Traffic Agency) for its production, maintenance and much appreciated layout. The various topics of the web site served as themes of the discussion as well as the practical problems of input, updating documents and managing the contents of the site. The management and actual running of ICAEA also led to much discussion during the meeting as well as ideas on how to envisage future structures and the need to call upon volunteers amongst members not only for organising future seminars but also to participate in the development of the web site.

History of the Association

Fiona Robertson and Philip Shawcross reminded the participants of the original aims of the "Aviation Forums", set up in 1984 and developing into the Association during the early 90's, as being a means of bringing people working in a similar professional field but from various skill areas together to exchange expertise, teaching and learning practices and current information concerning safety issues in the aviation world.

These forums rapidly expanded from one day events to three day forums for over one hundred people from around the world. A solid network of participants was thus established and clearly met a need amongst the world wide language training community.

The creation of the "Newsletters" and regular "Seminars" on specific themes, maintained further links by promoting the exchange of information and worked towards enhancing the status of the professionals engaged in language training for the aviation sector. (The contents of the newsletters and the themes of the Seminars can now be viewed on the web site)

The Association withdrew into a period of hibernation from 1996 to 2002. Fortunately due to the enthusiasm and dynamism of Elizabeth Mathews and Jeremy Mell as well as the hard work done by Fiona, Philip and Bozena to get the Association started again the 2002 seminar took place in Warsaw with more than 50 participants followed by the 2003 Luxembourg seminar with some 80 delegates.

The original aims of the "Aviation Forums" are still extremely pertinent as can be testified by the involvement and interest of the participants at both the Warsaw and Luxembourg seminars.

Managerial and Operational themes

The budget of the Luxembourg seminar was presented.

The registration fees from 71 paying delegates generated an income of 8,460 euros. The expenses amounted to approximately 5,530 euros including the travel expenses.

It was pointed out that these expenses were exceptionally low for this seminar and it would be unrealistic to project similar sums for future seminars. The budget was tight and it was felt that for future seminars a higher fee for participation might have to be envisaged.

The current savings accounts of the association allowed for a small financial safety margin.

The status of the Association

Governed by a 1901 French law which provides the framework for non profit making organisations, this legal frame provides sufficient liberty and form to justify its maintenance for immediate future needs. ALTE is organised under a European legal structure .Its organisation as a consortium, rather than an association, may be interesting for us to examine. Further discussion about this topic could be organised on the web site.

The makeup of the governing body

ICAEA is run by four volunteers, one President and three Vice presidents, during their spare time. As they all have full time jobs, help is urgently needed in the organisation of the seminars. Fiona Robertson felt she could not take on the organisation of another seminar before 2005.

Decisions had to be taken about ICAEA policy and regular secretarial work to relieve the burden in some of these areas on the current team. As the association grows, so do the expectations of participants for better services, more information concerning the ever changing field of aviation and the desire to be informed about current issues such as safety, new rules and regulations and operational procedures. A certain frustration was felt amongst the governing body that because of limits on their time they were unable to meet these growing needs. So a call for volunteers was raised in the following areas:

- developing the bibliography on the web site
- managing the web site
- being responsible for web site features such as research, book reviews, resources
- organising secretarial aid
- organising and running seminars
- participating in the maintenance of the list serve
- representation of the association at international conferences and events such as TESOL or IATEFL

Sponsorship

Everyone was aware that a certain amount of sponsorship would be necessary for ICAEA to be able to develop. Different views were aired on the nature and implications of such sponsorship and the profile of the sponsors.

Links with ICAO

Elizabeth Mathews expressed her desire to involve ICAEA in the ICAO symposium in Montreal in May 2004 and in the subsequent regional seminars designed to inform the civil aviation authorities and the aviation community at large on the content and repercussions of the latest ICAO Language Proficiency Requirements.

Teacher Workshops

There was a widespread need felt for teacher workshops to be held under the auspices of ICAEA. Several offers of hosting had been made and the team wished to materialise this. In point of fact, it has been necessary to wait for the September 2005 Forum in Besançon.

Extending ICAEA's activity in Asia and Latin America

As opportunities arose, ICAEA wished to establish nuclei in South East Asia and Latin America with the co-operation of local members with a view to reaching professionals in these regions.

PARTICPANTS

ACS NAGY	Ms	Maria	Language Teacher	Hungaro Control Hungarian	HUNGARY
ALDEA	Ms	Violeta	ATS Expert	ATS ROMATSA	ROUMANIA
ANDRADE	Ms	Iolanda	Teacher	NAV.EP, Centro de Formação	PORTUGAL
BAKER	Mr	Robin	Head of ATS Licensing	CAA	UK
BEHALKOVA	Ms	Denisa	English Language Instructor	Czech Airline Training Centre	CZECH REPUBLIC
BELLEC	Ms	Joan	Language Teacher	CLA	FRANCE
BLAHAKOVA	Dr	Radoslava	Head of Lang Dept.	ATC Admin of Czech Rep.	CZECH REPUBLIC
BOGOMAZYUK	Mr	Oleksandr	Senior Expert Training and	Ukraine State Air Traffic	UKRAINE
BOOOMAZIOK	IVII	Oleksalidi	testing	Enterprise	UKKAINE
BOIA	Ms	Maria	ODR Expert	ROMATSA	ROUMANIA
			En all an entre Survivited		
BRAYKOVA	Ms	Mariya	Eng Language Specialist	ATSA	BULGARIA
CALIMAN	Ms	Ioana	Aeronautical Inspector	Romanian Civil Aeronautic Authority	ROUMANIA
CARSON	Dr	Joan	Professor	Georgia State University	USA
CONNOLLY	Mr	Finian	Managing Director	Executive & Prof. Training Inst Ltd.	IRELAND
CORDEIRO	Ms	Isabel	Teacher	TAP Air Portugal	PORTUGAL
CORNELL	Ms	Sarah	Training Consultant	Télélangue	FRANCE
DANCER	Mr	John	Head, ATS Standards Dept.	CAA	UK
DAUGHERTY	Mr	Lawrence	Consultant, International Program Development	Embry-Riddle Aeronautical University	USA
DE JONG	Dr	John	Director	Language Testing Services	NETHERLANDS
DE JOING	D.	H.A.L.	Birector	Earliguage Testing Services	HEITIEREZHOS
DE MONTALTE	Capt.	Hugh		Air France	FRANCE
DEPERNET	Ms	Odile	English Language Trainer	Air France Language Academy	FRANCE
	Prof.	Tatiana	Head of Fac. of Foreign Langs.		RUSSIA
DOBRUNOVA				Civil Aviation Academy St Petersburg	
DOUBENOK	Mr	Dmytro	Inspector Controller	Ukraine State Air Traffic Enterprise	UKRAINE
DOUGLAS	Prof.	Dan	Professor, TESL/Applied Linguistics	Iowa State University	USA
DUFOUR	Ms	Francoise	Language Trainer	Forem Belgium	BELGIUM
DUGONJIC- HERCOG	Ms	Andjelka	Aviation English TEacher	Croatia Control Ltd.	CROATIA
DUNCIC	Mr	Gradimir	Head of Ground Studies	JAT Flight Academy	SERBIA and MONTENEGRO
ENRIGHT	Mr	Adrian	Training Development and	Eurocontrol Institute of Air	LUXEMBOURG
FERDINANDOVA	Ms	Marieta	Harmonisation Expert, Human Resources	Navigation Services Air Traffic Services Authority	BULGARIA
			Directorate	D. 677 D. 11 11 11 11	
FITZGERALD	Ms	Faith	General Manager	RMIT English Worldwide	AUSTRALIA
FLOOD	Ms	Michèle	Training Coordinator	SEFA	FRANCE
GERIGHTY	Dr	Terence	Director ELT Training	ELT Banbury	ENGLAND
GODMET	Ms	Carmel	English Language Training Coordinator	DGAC	FRANCE
HAILA	Mr	Jussi	Air Accident Investigator	AIB	FINLAND
HALL	Mr	Tony	Director, Embry Riddle	Embry-Riddle Aeronautical	USA
			Language Institute	University	
HAUSER	Mr	Christian	ATC Training and Documentation	Austro Control	AUSTRIA
HUANG	Ms	Min-Huey	Flight Operations Training Coordinator	China Airlines	TAIWAN
IHARA	Capt.	Hiroshi	Japan Airlines	Flight Personnel Qualification	JAPON
KANTOLA	Сарі. Mr	Erkki	Air Accident Investigator	AIB	FINLAND
KELLY	Mr	Michael	Japan Airlines	Pilot Trainee Ground Instructor	JAPON
KIM	Ms	Kun-Ok	Japan Airines	Chung-Ang University	KOREA
		Natalia	Course Manager		
KORABLOVA	Ms	Helena		ANS Training Centre	LATVIA
KUKOVEC	Ms		Lecturer of English	University of Ljubljana	SLOVENIA
KULTON	Ms	Florence	Radiotelephony Instructor	Air France	FRANCE
LEE	Mr	Youndae	Deputy Director	Korean Civil Aviation Safety Authority	REPUBLIC OF KOREA
MAIA	Ms	Isabel	Teacher	TAP Air Portugal	PORTUGAL
МАЛС	Ms	Jadranka	English Language Consultant	Croatia Control Ltd.	CROATIA
MAK	Ms	Karen P.Y.	Language and Communication Devepmt.	Cathay Pacific Airways	HONG KONG
MARINCIC	Prof	Mirna	Head of Staff Planning and Training	Croatia Control Ltd.	CROATIA
MATHEWS	Ms	Elizabeth	Linguistic Adviser	ICAO	CANADA
MAZAL	Ms	Linda	English Teacher	ENAC	FRANCE
MELL	Dr	Jeremy	Language Instructor	ENAC	FRANCE
MERTENS	Ms	Christine	Director	Technifutur	BELGIUM
MITRYUSHKINA	Ms	T.	Director	Volga-Dnepor Training Center	RUSSIA
WILLYLOSHKINA	1419	1.	DICCOL	vorga-Direpor Training Center	ROSSIA

NESKOVIC	Ms	Ljiljana	Senior Adviser in English Language	FATCA	SERBIA and MONTENEGRO
O'LOUGHLIN	Dr	Kieran	Academic Director	RMIT English Worldwide	AUSTRALIA
OSTAPETS	Ms	Tetyana	Teacher of English	Ukraine State Air Traffic Enterprise	UKRAINE
OZEGOVIC- KOSIER	Prof	Milena	ChiefInstructor	Croatia Control Ltd.	CROATIA
PARRIS	Ms	Christy	General Manager	Air France Language Academy	FRANCE
PATASIUTE	Dr	Juraté	RT phraseology instructor	S.e. "Air navigation"	LITHUANIA
REES	Ms	Lynette	Linguistic Coordinator	Strasbourg Entzheim Airport	FRANCE
ROBERTSON	Ms	Fiona A.	Teacher	CLA, Université de Franche Comté	FRANCE
ROMELE	Ms	Mara	Personnel Manager	Latvian Air Navigation Service Enterprise	LETTONIE
SAMPSON	Mr	Charly	Head of ATC	Eurocontrol	LUXEMBOURG
SANCHO	Mr	Patrick	Chief Ground Instructor	DGAC/SEFA	FRANCE
SAVILLE	Mr	Nicholas	ESOL Director	UCLES	UK
SHA	Capt	Chi-Ping	744 Check Pilot	China Airlines	TAIWAN
SHAWCROSS	Mr	Philip	Teacher	English for Aircraft	FRANCE
SIPOS	Mr	Gabor	Language Teacher	Hungarocontrol/ROK	HUNGARY
SLAWINSKA	Ms	Bozena	Head of EFL Training	PPL/PATA	POLAND
SMITH	Mr	Russell	Manager International Training	NAV Canada Training Institute	CANADA
STROTHER	Dr	Judith	Vice President	Virtual Languages Inc.	USA
STUMBERGER	Ms	Karmen	Head of Training	CAA	SLOVENIA
TROTT	Mr	Robert	IFATCA Representative	Guild of Air Traffic Control Officers	ENGLAND
WON	Mr	Sang Woong	Instructor	Flight Crew Training Centre, Korean Airline	KOREA
YOO	Mr	Seunghun	Assitant Director	Korean Civil Aviation Safety Authority	REPUBLIC OF KOREA
ZEPPIERI	Mr	Carlo	Coordinator	ENAV SPA	ITALY
ZIVOTIC	Prof.	Ana	Sen. Advisor Eng Lang and	Federal Air Traffic Control	SERBIA AND
. 2 2			Aviation Phraseo.	Authority of Serbia and Montenegro	MONTENEGRO