9th ICAEA Forum

ICAEA

Testing for ICAO compliance:
Best practice in aviation English proficiency assessment

Event Report

Over 100 delegates from 33 countries attended the 9th ICAEA Forum 'Testing for ICAO compliance: Best practice in aviation English proficiency assessment' which began with the association's Annual General Meeting on the afternoon of 7 May and ended at lunch after two accreditation meetings and a tour of Warsaw on Saturday 10 May.

The Forum lived up to its name; in the fullest sense it was a true forum or exchange of views and information. It enabled those who attended to consider testing from every angle: from the regulatory point of view with Nicole Barrette-Sabourin (ICAO), from theoretical and academic standpoints of best practice with Magdalena Vecerova (Eurocontrol consultant), Angela ffrench and David Booth (Cambridge ESOL), methodologically by seeing how four different test providers saw the design and administration of their tests (ELPAC, RELTA, TEA and VAET) and through the eyes of four different test users. ICAEA was extremely fortunate to have so much expertise and experience under one roof. Finally, everyone was also able to familiarize themselves with the latest aviation English publications and services in the exhibition area.

The ensuing discussions were of a remarkably high quality: lively and well-informed. The warmth, friendliness and enthusiasm of all those present reflected the generosity of our hosts, the Polish Air Navigation Services Agency, and demonstrated perfectly what ICAEA is all about.

You will find the full proceedings of the forum below.

Note:

In organising its events and publishing their proceedings ICAEA is offering a forum for those members of the aviation community involved in the use, training and testing of English to express and exchange views in an atmosphere of tolerance and mutual respect in the interests of the community as a whole. The views expressed and the materials presented during the forum and reported in the proceedings are the authors' own and are in no way endorsed by ICAEA or by any other third party.

9th ICAEA Forum

Testing for ICAO compliance:
Best practice in aviation English proficiency assessment

Programme

Wednesday 7th May   **ICAEA 2008 Annual General Meeting**

Thursday 8th May   **OPENING WELCOME ADDRESS:**
PANSA
ICAEA President

"What do we want to test?"

PLENARY SESSION 1:
Nicole Barrette-Sabourin
Training Officer, Flight Safety Section, ICAO
'Implementation of ICAO language provisions: Issues and Challenges'

PLENARY SESSION 2:
Magdalena Večeřová
Language testing consultant and expert to Eurocontrol
'Test evaluation: Do our tests fulfil the ICAO requirements?'

Exhibition time

BREAKOUT SESSIONS 1A, B & C:
Test design and construct rationale

BREAKOUT SESSIONS 2A, B AND C:
Test assessment and implementation rationale

Ben Rimron (Mayflower)
Michael Kay (RMIT)
Larry Rothenberg (Pearson-Ordinate)
Adrian Enright & Magdalena Vecerova (Eurocontrol)

PLENARY SESSION 3:
Summary of breakout sessions by four moderators

Friday 9th May   **'How do we test it?'**

PLENARY SESSION 4:

Angela ffrench & David Booth

Cambridge ESOL

'Delivering Validity and Reliability - practical concerns in delivering assessments'

PLENARY SESSION 5:

Marjo Mitsutomi, PhD, and **Larry Platt,** PhD

University of Redlands

'Accreditation and Aviation English: Why bother?

Henry Emery

Emery-Roberts Consultants

'A test rater accreditation project'

PLENARY SESSION 6:

Four presentations of feedback on test implementation experience

Bozena Slawinska, PANSA, on using ELPAC in Poland

Milina Kolomparova, EPST, Slovakia, on using TEA in Slovakia

Krzysztof Sysio, LOT, on using RELTA with pilots in Poland

Christel Schipper, LVNL, on using RELTA with ATCOs in Netherlands

Saturday 10th May   **'Where do we go from here?'**

MEETINGS ON EXPLORING IMPLEMENTATION:

Test rater accreditation project

Chaired by **Henry Emery**

Emery-Roberts Consultants

Testing accreditation processes,

Chaired by **Dr. Marjo Mitsutomi & Dr. Larry Platt**

University of Redlands

INTERNATIONAL CIVIL AVIATION ENGLISH ASSOCIATION

Warsaw, 7 May 2008

Welcome by ICAEA President

Dear Members,

We are delighted to welcome you to Warsaw for the third time in our history. You have joined us from over thirty countries in all five continents to attend ICAEA's 9th forum on the theme: '**Testing for ICAO compliance: Best practice in aviation English proficiency testing**'.

Warsaw occupies a very special place in the history of the association and without the Polish Air Navigation Services Agency's generous hosting and support this event would not have been possible.

It is only eight months since we last met in Cambridge and yet much has happened in that time in the environment in which we exercise our professions as well as inside ICAEA under an enlarged and strengthened Board.

ICAEA has always offered a forum for the meeting and exchange of ideas. This is a role which the association fully intends to pursue and develop. However, in the rapidly changing world we all know, the time has perhaps come for ICAEA to think about formalising and harnessing the expertise and experience built up over the years among its members for the benefit of the aviation community as a whole, i.e. to look outwards in a more proactive way. This is a question I invite you all to reflect on.

To return to matters at hand, we are extremely fortunate that such a group of experts in the field of language assessment has been willing to address us at this 9th forum. I think you will agree that the programme will keep us busy over the next three days!

Thank you for being here, sharing this event with us and contributing to its success.

Wishing you all a very enriching and enjoyable forum,

Philip

Philip Shawcross
ICAEA President

**POLISH AIR NAVIGATION
SERVICES AGENCY**

Warsaw, 9 May 2008

Address by the acting President of PANSA

Ladies and gentlemen,

PANSA has been involved with the International Civil Aviation English Association for several years. I am sure that Mr. Rodak conveyed to you our warm wishes of success for your forum when he spoke to you yesterday.

We are indeed very happy to host this event in Warsaw because we at PANSA consider English language proficiency to be a safety-critical issue and are convinced that your association is committed to raising standards in the industry. Recognizing this fact and our shared objectives, we have tried to support your work in different ways.

When I was technical director I gave permission for your website to be hosted on our server as I recognized its potential for information exchange and I am very glad to see how it has prospered and developed.

As an independent company whose objective is to provide safe air traffic flow to the international community, we are very much in favour of promoting your activities as we believe we have common concerns. This is the reason we have hosted your events and sponsored your website and your Board meetings.

I see that ICAEA has a great potential for promoting a better understanding of the role of proficiency testing in air traffic control and I agree with my Rodak that the association's future lies in a closer co-operation with bodies such as IFATCA and CANSO.

I hope you will enjoy this forum and your time in Warsaw. It is a pleasure to see people from so many different countries coming together to work towards a single goal.

I wish you every success.

Kizysztof Banaszek

**Krzysztof Banaszek
Acting PANSA President**

PLENARY SESSION 1

Nicole Barrette-Sabourin, Training Officer, Flight Safety Section, ICAO

BIO DATA

Mrs. Nicole Barrette-Sabourin is a training officer and the Project Manager for Language Proficiency Requirements issues in the Flight Safety Section of the Air Navigation Bureau of ICAO.

She has been a staff member of ICAO since 1983, worked in the TRAINAIR Central Unit since the inception of the programme in 1988 until 2005 and now belongs to the Aviation Training Policy and Standards Unit. She holds a Master's Degree in Educational Technology from Concordia University in Montreal, Canada.

She is presently involved in the ICAO endorsed Government Safety Inspector Programme, the ICAO/ACI Airport Management Programme and the IATA Training and Qualification Initiatives. In April 2006, she has served as the Secretary of the Proficiency Requirements in Common English Study Group (PRICE SG) during its fifth meeting, organized the technical work programme of the Second ICAO Aviation Language Symposium conducted in May 2007 and developed and conducted language proficiency implementation plan workshops in the ICAO regions.

"Implementation of ICAO language provisions: Issues and Challenges"

SUMMARY

In September 2007, the ICAO Assembly adopted resolution A36-11 - Proficiency in the English language used in radiotelephony communications. The resolution urges States that will not comply with language proficiency requirements to provide ICAO with their implementation plans and directs the Council to support Contracting States in their implementation of the language proficiency requirements by establishing globally harmonized language testing criteria. The presentation will provide a status report on the development of implementation plans. A description of the work carried out to date on the development of language testing criteria will also be provided.

Implementation of ICAO language provisions:
Issues and Challenges

Nicole Barrette-Sabourin
Training Officer and LPR Project Manager
Flight Safety Section, Air Navigation Bureau, ICAO

1.   Introduction

1.1   The lack of English language proficiency has played a key contributing role to several accidents in civil aviation. The mid-air collision that occurred over New Delhi in 1996 brought this safety concern to the forefront when 349 passengers and flight crew lost their lives.

1.2   Resolution A32-16 was adopted during the 32nd session of the Assembly of ICAO. This resolution urged ICAO to consider the matter with a high level of priority and to "…take steps to ensure that air traffic controllers and flight crews involved in flight operations in airspace where the use of the English language is required, are proficient in conducting and comprehending radiotelephony communications in the English language"

2.   Chronology of events

2.1   Since the adoption of this resolution in 1998, several actions were taken.

- In 2000, the Proficiency Requirements in Common English Study Group (the PRICE SG) met for the first time.
- During the 33rd session of the Assembly Contracting States requested that the language proficiency requirements being developed should not be limited to English.
- In March 2003, the language provisions were formally adopted to become applicable in March 2008.
- At the time of adoption, the Council requested that a mid-point review be conducted on the status of implementation of the requirements.
- This mid-point review was followed by another in April 2007.
- In October 2008, the Assembly held its 36th session and adopted resolution A36-11.
- The language proficiency requirements became applicable on 5 March 2008.

3.   Overview of the ICAO language provisions

3.1   The language provisions can be found in ICAO Annexes 1, 6, 10 and 11. It is in Annex 10 that the use of the English language is specified. Paragraph 5.2.1.2.1 states that "The air-ground radiotelephony communications shall be conducted in the language normally used by the station on the ground or in the English language." Paragraph 5.2.1.2.2 states that "The English language shall be available, on request from any aircraft station, at all stations on the ground serving designated airports and routes used by international air services."

3.2   Annex 1 specifies to whom the language proficiency requirements apply and the level of proficiency for the language used in radiotelephony communications

- The requirements apply to pilots, air traffic controllers and aeronautical station operators involved in international operations

- These personnel should speak and understand the language used for radiotelephony communications (Paragraphs 1.2.9.1 and 1.2.9.2)
- They should demonstrate this ability to speak and understand the language used in radiotelephony communication to the ICAO Operational level 4 (Paragraph 1.2.9.4)
- And personnel who are proficient below level 6 (Levels 5 and 4) should be formally evaluated at regular intervals. (Paragraph 1.2.9.6) with a recommended practice of every 3 years for level 4 and every 6 years for level 5.

3.3   Annex 6 Part I and III specify the role of operators in ensuring that flight crew demonstrate the ability to speak and understand the languages used in radiotelephony communication to the level specified in Annex 1. Annex 11 makes a similar statement for air traffic service providers.

4.   ICAO Deliberations

4.1   When the Council adopted the language provisions in March 2003, it had requested that their implementation be monitored. The understanding was that if States encountered major difficulties in implementing the language proficiency Standards in paragraph 1.2.9 of Annex 1, it would be possible for the Council to consider reviewing the applicability date. Therefore in November 2005 a survey was sent to all ICAO regions. The results of this survey were considered in June 2006. 36 States and two international organizations responded. It was noted that 10 out of the 36 responses were from States that have English as a first language. As the results of the survey were considered inconclusive, it was agreed to retain the applicability date of March 2008. Contracting States of ICAO were informed accordingly and reminded of the need to take the necessary measures to implement the language proficiency requirements in a timely manner.

4.2   A second survey was conducted in October 2006 and its results were reviewed by the ICAO Air Navigation Commission in April 2007. 59 States completed in part, or in full, the questionnaire and three States had responded to the State letter without completing the questionnaire. The table below summarizes this survey results. The majority of respondents indicated that language provisions would be implemented by 5 march 2008.

	No. of States who have replied indicating:		
	Full implementation	Partial implementation	N/A
Will all pilots with air transport pilot's licence (ATPL) and commercial pilot's licence (CPL) be at least at Level 4?	48	18	
Will air traffic controllers be at least at Level 4?	41	21	
Will aeronautical station operators be at least at Level 4?	22	5	30
Will pilots with private pilot's licence (PPL) be at least at Level 4?	42	12	

4.3   In June 2007, the Council considered a report on the status of implementation of the language proficiency provisions as well as options concerning the applicability date. The

Council decided to retain the applicability date of 5 March 2008 as it established a milestone and maintained the focus on implementation of these standards. It was also recognized that some States would not meet the applicability date. Therefore a resolution was developed and adopted during the last Assembly.

5.   Addressing Non-compliance

5.1   There are four articles of the Convention on International Civil Aviation (the Chicago Convention) that impact non-compliance with the language proficiency requirements. These are articles 33, 38, 39 and 40, which will be briefly described below.

5.2   Article 33- Recognition of certificates and licenses

This article stipulates that if Contracting States meet the minimum requirements specified for licenses and airworthiness certificates, then these licenses and certificates will be automatically recognized by other Contracting States. This means that non compliance with the ICAO language proficiency requirements could invalidate recognition of licenses.

5.3   *Article 38 – Departures from international standards and procedures*

This article addresses differences with the international standards. Contracting States that do not comply with international standards or that adopt regulations and practices that differ from the international standards must notify ICAO of these differences. Once ICAO is notified, it advises all other Contracting States of the differences.

5.4   *Article 39 – Endorsement of certificates and licences*

The endorsement of licenses for language proficiency is addressed through article 39 of the convention and also in Annex 1, paragraph 5.1.1.2 XIII. Article 39 stipulates that when an international standard is not met, the licence should be endorsed accordingly. Annex 1 stipulates that from 5 March 2008 an endorsement on language proficiency should be included on the licence. Whether a licence holder involved in international operations meets or does not meet the language proficiency requirements, his or her licence should be endorsed.

5.5   *Article 40 - Validity of endorsed certificates and licenses*

Article 40 stipulates that no pilot with a license endorsed under article 39 shall participate in international navigation, except with the permission of the State or States whose territory is entered. This means that a pilot that does not meet the minimum language proficiency requirements and has a licence so endorsed may continue to operate internationally, provided that the States over which and to which the flight is conducted are in agreement. This could involve multiple bilateral agreements for these licence holders.

6.   English language proficiency in international operations

6.1   In some regions of the world, there is an agreement that a language other than English may be used in radiotelephony communications. During these international flight operations, pilots and controllers conduct radiotelephony communications in Russian or Spanish, for example. In these cases where air traffic services are provided in Russian or Spanish, article 40 would not apply. However, the lack of language proficiency in English would limit the operational area in which pilots could fly. In any case, the controllers involved in international

operations are still required to have at least ICAO Operational Level 4 in English as stipulated in Annex 10 (see paragraph 3.1 above).

6.2 The use of a language other than English in international operations endorsement of licences has implications on the endorsement of licences. This can be better understood through examples of flight operations in such a region.

Example 1

Consider the flight from Ashkhabad, Turkmenistan
 to Minsk, Byelarus illustrated in the map below:
Air traffic services during this flight are provided in Russian. Presuming that the flight crew are proficient in Russian at ICAO Operational Level 4 and that their licence are endorsed in accordance with the requirements of Annex 1, article 40 would not apply and bilateral agreements with Russia, Ukraine and Byelarus would not be needed. However, English language air traffic services would still be required in accordance with Annex 10.

Example 2

Consider the flight from Ashkhabad, Turkmenistan to Stockholm, Sweden illustrated in the map below:

Air traffic services during this flight are provided in Russian and in English. Presuming that the flight crew are proficient in Russian and English at ICAO Operational Level 4, and that their licences are endorsed for these two languages in accordance with the requirements of Annex 1, article 40 would not apply. Once again, English language air traffic services would still be required through the whole flight in accordance with Annex 10.

If the flight crew is not proficient to ICAO Operational Level 4 in English, their licences should be endorsed in accordance with article 39. In this case, the permission of States that do not provide air traffic services in Russian would need to be obtained in accordance with article 40.

7.   Assembly Resolution A36 -11

7.1   One of the consequences of example 2 above is that States need to enter into numerous bilateral agreements. Another consequence and far worse in terms of safety is that States could ignore the standard. In its consideration of the issue, the Council of ICAO decided that further actions were necessary to a) mitigate the risks of non-compliance with the language proficiency requirements and b) strengthen the implementation of the language proficiency requirements in a way that could not be ignored. To achieve this, the Council of ICAO proposed and the Assembly adopted Resolution A36-11.

7.2   Through A36-11, States have an additional three years (or until March 2011) to comply with the language proficiency requirements. However, this is conditional on States providing their implementation plan as soon as practicable but no later than 5 March 2008. The implementation plans should also include risk mitigating measures as appropriate.

7.3   The resolution urges States to waive the permission requirements under Article 40 for pilots from another State provided that the implementation plan is acceptable to them. The resolution also urges States not to restrict their operations into other States provided that their implementation plan for air traffic controllers and radio station operators is acceptable. It is important to note that the implementation plans provided by States are not reviewed or approved by ICAO.

7.4   Further to the adoption of the resolution, States were advised of the adoption of this resolution and requested to provide ICAO with their implementation plans or their notification of compliance, as appropriate. The implementation plan should include information concerning:

- Notification of compliance/differences
- Focal Point contact details
- National Regulatory Framework
- Estimate of the national level of implementation for pilots and controllers (this item should be updated on a regular basis)
- Information concerning training and testing programmes
- Interim measures

7.5   A series of workshops in each of the ICAO region was conducted for the development of implementation plans and to provide detailed guidelines on the development of these plans. Eight workshops were conducted from December 2007 to early March involving 488 participants from 102 States. To date 137 States and territories have provided either a notification of compliance (46 States), an implementation plan (88 States), or contact details for their focal points (3 States). This information can be found on the ICAO Flight Safety Information Exchange website at http://www.icao.int/fsix/lp.cfm*

8.   Globally Harmonized Testing Criteria

8.1   Resolution A36-11 also includes a resolving clause that "Directs the Council to support Contracting States in their implementation of the language proficiency requirements by establishing globally harmonized language testing criteria". ICAO is presently developing guidelines on these criteria which will be initially issued as an ICAO Circular. Eventually these guidelines will be included in the second edition of Document 9835 – *Manual on the Implementation of ICAO Language Proficiency Requirements*. Its purpose is to provide practical guidance to States' to assess testing services, to support the language testing industry's self-regulation and to facilitate positive washback. An additional challenge in developing these criteria is to balance best practices with practicality, as well as ensuring that they are applicable in 190 Contracting States. The primary target audience for the document are civil aviation authority personnel that may not necessarily have English as a first language and that are not language testing experts.

8.2   The criteria will address the following issues:

- Test design and construct
  Including but not limited to purpose, rationale, compliance with holistic descriptors and rating scale
- Test validity and reliability
  Including but not limited to documented evidence and what it means in layperson terms
- Rating

Including but not limited to number of raters, initial and recurrent rater training
- Test administration and security
  Including but not limited to interlocutor and candidate instructions, evidence to be collected by raters, equipment requirements, record-keeping
- Organizational structure
  Including but not limited to separation between training and testing activities and adaptation to local requirements
- Qualifications
  For test development and design teams, test administration teams including interlocutors, and rating teams

8.3   In addition, phraseology in language testing will be addressed, bearing in mind the intent of the language provisions, the existing assessment of phraseology within the operational environment, the need to use correct standardized phraseology and the need to elicit plain language speech samples. Finally, the assessment of Level 6 proficiency will also be addressed within the context of Annex 1 standards.

9.   Implementation Support Activities

9.1   Since the adoption of the language provisions, ICAO has undertaken several activities to support their implementation. These are briefly described below:

- Manual on the Implementation of LPRs (2004)
- Rated Speech Samples (2005/2007)
  1. This CD was developed with the assistance of the linguistic sub-group of the PRICE SG.
  2. It provides speech samples of level 3, 4 and 5 speakers as well as a detailed rating for each.
- 11 Regional Seminars (2004-2007)
  1. These seminars were aimed at understanding the requirements towards facilitating their implementation.
- PRICE SG/05 (2006)
  1. Conducted to develop additional guidance based on practical experience in language testing.
- ICAO Aviation Language Symposia (2004/2007)
  1. The first one conducted shortly after the adoption of the provisions
  2. The second to further assist States with implementation
- Implementation Plan Workshops (2007/2008)
- Implementation plan guidelines (FSIX)
- Posting of State implementation plans (FSIX)

10.   Conclusion

10.1   ICAO's Safety Management Course describes the management dilemma that aviation organization face in a very useful manner. In order to achieve its production objectives, the management of any aviation organization requires the management of many business processes. Safety management is a core business function just as financial management, Human Resource management, etc. This brings about a potential dilemma for management and that is to ensure that safety is managed while also ensuring the financial viability of the organization. Investing

heavily in safety management (protection) will impact the financial health of an organization (production), while not investing in safety management could have catastrophic consequences.

10.2   Aviation organizations, including aviation language testing organizations have a responsibility to balance protection and production. It is critical that the language testing industry navigate within this safety space. To do so it needs to balance best practices with available resources without compromising safety and without overtaxing the system.

PLENARY SESSION 2

Magdalena Večeřová, Language testing consultant and expert to Eurocontrol

BIO DATA

After receiving her master's degree (MA) in English literature and philology Magdalena started her career as a language teacher in the Czech Armed Forces. She also studied at the Canadian Language Training Faculty in Ottawa where she specialised in ESP (English for Specific Purposes). Soon after her return from Canada she became a member of a new language testing team for the Czech military and was sent to the University of Lancaster in Great Britain where she completed another MA in language testing. She also co-operated with the Peacekeeping English Project of the British Council and provided training and workshops in language testing for participants from more than 20 countries (both in the Czech Republic and abroad).

In 2005 she started her co-operation with EUROCONTROL Luxembourg where she was involved in the development of the ELPAC test. Magdalena currently works as an independent language testing consultant and is involved in both the ELPAC test and a new test for student controllers which is being developed by EUROCONTROL.

"Test evaluation: Do our tests fulfil the ICAO requirements?" with Q & A

SUMMARY

Testing language proficiency in aviation is of very high stakes. Unreliable and invalid tests compromise safety and thus service providers, operators and regulators should make every effort to ensure that a valid and reliable test is being used and standards are being kept. Best practice has to be applied not only in the test development process but also in test implementation.

There are currently many language tests claiming that they meet the ICAO requirements and assess language proficiency according to the ICAO scale. This is, unfortunately, not always the case. This paper and the accompanying presentation provide practical advice on how to evaluate language tests and what to do when the test being used does not meet the ICAO requirements.

1. *Introduction*

As of March 5th 2008 air traffic controllers and pilots operating in internationally designated airspace and on international air routes have to demonstrate their proficiency in the language(s) they use for aeronautical communication. Level 4 of ICAO's language proficiency requirements is set as the operational standard which air traffic controllers and pilots must meet.

Air Navigation Service Providers, Aircraft Operators and National Regulatory Authorities should have tested all their personnel and taken appropriate measures where level 4 has not been achieved, or have in place an implementation plan to be published and completed by March 5th 2011 (ICAO Resolution A36-11, October 2007).
ICAO Doc 9835 (Manual on the Implementation of ICAO Language Proficiency Requirements), chapter 6 states: "*Reliable, effective, legitimate testing systems are required to*

*ensure that pilots and controllers have adequate levels of English language
proficiency. ...language testing for licensing purposes needs to be of highest calibre...*".

The information published on the ICAO website (http://www.icao.int/fsix) *shows a very
different picture from what was originally expected. Many States have not even started testing
and many are using unreliable and invalid tests. All examples used in this paper are based on
information from State Letters published on the ICAO website.

(*Most of the information previously found on FSIX has now transitioned
into the **Integrated Safety Trend Analysis and Reporting System (iSTARS)**. **)**

2. *Test evaluation*
There are currently many codes of practice providing guidance on how to develop, implement
and evaluate language tests. It is beyond the scope of this paper to discuss the test evaluation in
detail and thus only the main principles will be explained. More information can be obtained
from the links provided at the end of this paper.

2.1.   **Test purpose**
The ICAO Language Proficiency Requirements (LPR) require the use of specific purpose
language proficiency tests. The test content must be based on a needs analysis and test tasks
have to be representative of the target language situations.
The data published on the ICAO website are incomplete as many States have not yet submitted
the required information. Yet there are several instances where tests being used do not meet the
requirements.

Example 1:
*OPI (Oral Proficiency Interview) is used to assess the language proficiency of pilots /
controllers.*
Problem: A perfect example of a test misuse. OPI was not designed to test the language
proficiency of pilots and controllers and was validated for a different purpose. There is no
aviation element. Results from this test bear no reference to the ICAO scale and are thus
meaningless as they cannot be translated into ICAO levels.
Solution: A new test has to be used. It must trialled on a representative sample of the target
population and test tasks have to be representative of the target language situations.
*(Editor's note: This topic has been expanded upon in several exchanges by the author and
other contributors on the icaea_world forum in May 2008)*

Example 2:
Diagnostic instead of proficiency test being used.
Problem: No further information provided and thus it is difficult to comment on the test. In
general, diagnostic tests are used to identify areas where a test taker needs help.
Solution: Results from this test could be used by teachers to prepare remedial language
programmes. A specific purpose language proficiency test has to be used to evaluate the
language proficiency.

Example 3:
Placement tests and informal interviews being used.
Problem: Placement tests assess a student's level in order to place him or her in an appropriate
language course. Informal interviews might give an indication of someone's level, nothing
more. Test versions should be parallel and this cannot be the case with informal interviews.
Thus each candidate receives a different test.

<u>Solution:</u> Similar as above. Use the information obtained from the two tests to place test takers into various language classes. For licensing purposes use a specific purpose language proficiency test that meets the ICAO LPR.

2.2. **Reliability**

Reliability refers to the degree to which test scores are free from different types of chance effects. No test will achieve a perfect reliability (1.0) but in high stakes testing tests with reliability coefficients as close to 1.0 as possible should be used. In speaking tests intra- and inter-rater reliability should be continually checked. When inconsistencies appear the raters should undergo refresher training and be re-accredited upon successful completion of the training. Examples of problematic tests are presented below:

<u>Example 1:</u>
The test developer stated that tests were trialled and the results showed that the test was reliable.
<u>Problem:</u> If no further information is provided then the quality of the test is in doubt. Each statement has to be supported by evidence. Also, it is not enough to say that the trialled test version was reliable. The reliability of the test must be monitored during all test administrations.
<u>Solution:</u> There may be nothing wrong with the test itself. The problem might be on the side of the test provider who fails to communicate the necessary information. Request evidence to support the statements about the test reliability and information on how the reliability was and is computed.

2.3. **Validity**

Validity is a rather complex test characteristic and the scope of this paper does not enable detailed discussion. In general, a test is valid when it measures what it is supposed to measure. Validity pertains to the correctness of the inferences or decisions made on the basis of test scores. Therefore each test version should be trialled on a representative sample of the target population. Test items/tasks must represent the language skills needed in the specific content domain. Examples of several problematic tests currently being used are presented below:

<u>Example 1:</u>
Controllers are assessed in listening, speaking and reading.
<u>Problem:</u> ICAO rating scale includes listening comprehension and speaking. Testing reading does not seem to represent the language skills needed by controllers in their work. Levels assigned cannot be ICAO levels as there is no scale for reading comprehension. A perfect example of construct irrelevance.
<u>Solution:</u> If the listening and speaking tests follow all ICAO requirements on language proficiency testing they should be retained and used. The reading comprehension test should not be used for licensing purposes.

<u>Example 2:</u>
A speaking test (face to face) assessing all 6 ICAO levels. The test lasts approximately 20 minutes and there is one interlocutor and one rater.
<u>Problems:</u> Listening comprehension should be "*…assessed separately from speaking through an objective assessment tool. This is necessary to reduce the impact of variables associated with oral proficiency interfering with assessment of proficiency in listening comprehension*" (Michael Kay: Test Evaluation Criteria).
Also it would be impossible to assess if the test taker has the ability "*…to comprehend a range of speech varieties (dialect and/or accents) or registers.*" (Comprehension: Level 5 on the

ICAO rating scale). Level 6 goes far beyond the scope of aeronautical communications (vocabulary is idiomatic, nuanced and sensitive to register, level 6 speaker varies speech flow for stylistic effect, demonstrates comprehension of linguistic and cultural subtleties, is sensitive to verbal and non-verbal cues and responds to them appropriately, etc.). A 20-minute test cannot include tasks which would test language at all 6 ICAO levels in a valid and reliable way. This test seems to be a perfect example of construct under-representation.

Solution: Evaluate the speaking test thoroughly to see to what extent it is capable of eliciting a rateable sample of language. Review all test items / tasks and prepare new ones. Test tasks have to be representative of the target language situations. A separate listening comprehension test has to be developed.

Example 3:
Informal listening and speaking tests being used.
Problem: There is no more information provided but the word "informal" seems to suggest that the tests have not been validated and might be just general English language tests. They are probably unstructured and versions are not parallel. If this is the case then the results cannot be translated into ICAO levels.

Solution: Evaluate the test using the guidelines provided by ICAO and if the test does not conform to the guidelines do not use it for licensing purposes.

Example 4:
Oral interaction consisting of introduction, communication on common topics, auditing an aviation theme, auditing a real radio communication, testing task simulation.
Problem: The description of the test is not very clear. No sample test was provided and thus it is very difficult to see what the test takers actually have to do. Again, no listening comprehension test is included and thus it is another example of construct under-representation.

Solution: Evaluate the test using the guidelines provided by ICAO and if the test does not conform to the guidelines do not use it for licensing purposes.

Example 5:
Language proficiency test in plain English.
Problem: Radiotelephony is not plain English only. Phraseology is an integral part of pilot/controller communication. Such a test cannot assess language proficiency according to the ICAO scale.

Solution: This test cannot be used for licensing purposes.

2.4. **Fairness**

Each test should be fair to all candidates, regardless of their gender, age, language background, race, ATC rating or the aircraft type rating a pilot has. Each test version must be trialled on a representative sample of the target population and bias analyses have to be conducted to make sure that the test does not disadvantage a particular group of candidates. An example of a problematic test is presented below:

Example:
Test being used is a "sample interview and estimated grading".
Problem: It seems that the test being used has not been trialled and does not use the ICAO scale. A test should measure, not estimate. Test versions are probably not parallel and thus it is impossible to see whether and to what extent could some candidates be disadvantaged.

Solution: Use the results from the test for placement purposes and develop (or buy) a new test to use for licensing purposes.

3. **General considerations**

There are many States who until now (May 2008) have not provided any information on their compliance. Some, on the other hand, use "grandfather rights" to declare that all pilots or controllers are at level 4. Some state that as they have had no major incidents where lack of language skills would be a contributory factor, everybody must be at level 4 or above. This is an extremely dangerous approach. Nothing should be taken for granted and awarding levels of language proficiency must be based on measurement, not on assumptions.

4. **Basic guidelines for evaluation of language tests to meet the ICAO LPR**
➤ There should be separate tests of listening comprehension and speaking.
➤ Specific purpose language proficiency tests that meet ICAO LPR should be used (not diagnostic, placement, progress or plain language tests).
➤ Needs analysis has to be conducted.
➤ All test versions have to be trialled on a representative sample of the target population.
➤ Test items / tasks have to be representative of language skills needed in the content domain (radiotelephony). Different tests should be used for pilots and controllers.
➤ ICAO rating scale has to be used for assessment.
➤ Tests have to be sufficiently reliable to permit stable measurement of test takers' abilities.
➤ Tests are being validated and periodically reviewed.
➤ Tests are being sustained: new test versions are developed and trialled, testing standards (administration, interlocuting, assessing, …) are monitored and maintained.
➤ There exist comprehensive and clear guidelines for everybody involved in test administration.

Detailed guidelines for test design and evaluation can be found in the following links:

General Guidelines:
➤ ALTE Principles of Good Practice: http://www.alte.org/quality_assurance/index.php
➤ Code of Ethics for ILTA: http://www.iltaonline.com
  http://www.iltaonline.com/index.php/enUS/
➤ ILTA Code of Practice: http://www.iltaonline.com/index.php/en/resources/ilta-guidelines-for-practice http://www.iltaonline.com/ILTA-COP-ver3-21Jun2006.pdf
➤ The Standards for Educational and Psychological Testing:
  http://www.apa.org/science/standards.html
➤ Code of Fair Testing Practices in Education: http://www.apa.org/science/FinalCode.pdf
➤ EALTA Guidelines for Good Practice in Language Testing and Assessment
  http://www.ealta.eu.org/documents/archive/guidelines/English.pdf

Specific Guidelines
➤ http://www.icao.int/fsix/lp.cfm
➤ ICAO Doc 9835 (Manual on the Implementation of ICAO Language Proficiency Requirements),
➤ EANPG 48_Report_Appendix J: Recommended qualifications for raters of tests to meet ICAO LPR
➤ EANPG 48_Report_Appendix K: Recommended qualifications for interlocutors of tests to meet ICAO LPR
➤ Test Evaluation Criteria prepared by Michael Kay (ICAEA Board member)
➤ Selecting a language proficiency test to meet ICAO LPR (presented by Magdalena Vecerova at IALS/2)

Conclusion

Language testing in aviation is of very high stakes but not all service providers, operators and regulators seem to be aware of this fact. It is important to make every effort to ensure that valid and reliable tests are used. Unreliable and invalid tests compromise safety in aviation and this is what we all should want to avoid.

ELPAC Test
Test Design and Construct Rationale

Adrian Enright
EUROCONTROL

BIO DATA

Adrian qualified as an air traffic controller in the United Kingdom with all ratings for tower, approach and en route and has also worked as a controller at Maastricht UAC. He is currently working at the Eurocontrol Institute of Air Navigation Services in Luxembourg, in the Training Development and Harmonisation unit.

Adrian has been involved in developing English language proficiency tests since 1988. First as Project Leader for Eurocontrol's PELA Test for student air traffic controllers and for the last three years as Project Leader for ELPAC – a test of English Language Proficiency for Aeronautical Communication.

He is also a founder member of the ICAO PRICE Study Group (Proficiency Requirements for Common English), that strengthened the provisions of SARPs relating to aeronautical communications and that developed the ICAO language proficiency requirements. Adrian works closely with ICAO's European and North Atlantic Office in Paris on language proficiency matters as a member of the EANPG/COG ATM Training Task Force.

Currently, Adrian is working to assist states in implementing the ELPAC test. Development has started on a version of ELPAC to enable student air traffic controllers to meet the European Commission's language proficiency requirements for the student ATC Licence.

SUMMARY

The successful development of a language test can be accomplished by having a clear aim as to what the test should accomplish. Being guided by best practice in language test design at every stage of development and ensuring that the trialling population is truly representative of the target population is essential. An external audit of test development and design helps stakeholders meet their expectations.

1. **Rationale for the ELPAC test**
   Amendment 164 to ICAO Annex 1 (Personnel Licensing) which strengthened requirements for language proficiency came about as the direct result of the work of the ICAO PRICE (Proficiency Requirements in Common English) Study Group, of which the author is a member. PRICE also developed language proficiency requirements (LPR) in the form of holistic descriptors and the ICAO rating scale.

The European Union through its Directive on a Community ATC Licence has adopted the ICAO LPR. The net result being that pilots and air traffic controllers now require a language endorsement to their licence.

With extensive experience in developing the PELA test for student controllers, Eurocontrol was directed by its Member States to develop a new language test.

2. **Test objectives**

The primary objective was to design a specific purpose language proficiency test, in English, for air traffic controllers that would meet ICAO and EU language proficiency requirements. Test specifications were developed from a Needs Analysis. A thorough review of the needs analysis conducted for the PELA test showed that pilot / controller communications had not significantly changed and were still valid for aeronautical communication today.

3. **Development Team**

Eurocontrol's Member States contributed English language and ATC experts to the development team. The project is led by Adrian Enright (Eurocontrol) assisted by language testing expert Magdalena Vecerova of the Czech Republic. The team was guided during development by Dr. Rita Green, a consultant in language test design. Software development was (and remains) the responsibility of ENOVATE from Bergen, Norway.

4. **Test design**

The test examines aeronautical communication (phraseology and plain English) in conformity with the ICAO language proficiency requirements and holistic descriptors. The test ensures that all 6 ICAO language criteria at levels 4 and 5 are assessed. Language tasks are defined and items constructed so that suitable samples of language can be elicited during testing.

Based on experience with the PELA test it was decided at a very early stage that ELPAC would comprise two test papers. Paper 1 (Listening Comprehension) is web based – for test security and for economy of administration. Paper 2 is an Oral Interaction paper with visual and non-visual tasks.

5. **Test development – trialling**

From the beginning the team agreed that successful trialling will only be accomplished if the sample population accurately reflects the target population – operational air traffic controllers. Great care is taken to select candidates on the basis of gender, age, work, experience, ATC function, language background etc. ensuring they are proportionally representative of the true ATC population in Europe.

More than 700 controllers from 20 countries took part in the trialling exercises during the development of the first test versions which were released in July 2007. Each trial generated approximately 100,000 items for analysis. Test versions were trialled, analysed, evaluated, modified, and retrialled in a continuous cycle. Once test design is decided and items accepted, the test is subject to standard setting by a selected team of experts who do not have close knowledge of the ELPAC test. Finally, test versions are constructed to produce parallel versions. Meanwhile, software development proceeds hand-in-hand with trialling, and often in advance as feedback comes in from test takers, examiners and administrators.

Test development did not stop in July 2007. New versions are being developed, trialled, reviewed, re-trialled, etc.

6. **Test evaluation**

Together with data from the statistical analyses, the development team is receiving feedback from all participants (test-takers, administrators, examiners, markers), bio data on test takers, reports from administrators and test results and Paper 2 recordings.. All of this information helps the team to check how tasks and items are performing. Thus we are able to determine the reliability of items and evaluate the "fairness" of the test.

7. **Validation**

Throughout development the team periodically conducts an internal validation - to ascertain that the test is meeting its objectives and the stated language proficiency requirements.

Additionally, Eurocontrol contracted the Linguistics and English Language Department of Lancaster University under Professor Charles Alderson to carry out an external, independent validation of the ELPAC test. This external audit demonstrated to stakeholders that the ELPAC test meets their expectations in having a reliable tool to measure the language proficiency of their controllers.

8. **Dissemination**

Dedicated websites were set up to provide all stakeholders and interested persons with information about the ELPAC test. Sample tests are available for test format familiarisation. Presentations are made on a regular basis to international fora e.g. Eurocontrol groups, ICAO, IFATCA, ICAEA and individual controller associations. ELPAC test administrators are kept updated on development through reports on trialling and workshops.

9. **Conclusion**

Delivery of the ELPAC test is achieved by applying best practice in language test design during development; by ensuring that language skills and not ATC skills are assessed and by developing a specific purpose language proficiency test.

Test Assessment and Implementation Rationale

Adrian Enright
EUROCONTROL

SUMMARY

The need for a high level of security for such a high-stakes test was one of the factors that directed ELPAC development towards an internet based application. Language skills representative of aeronautical communication require both listening comprehension and oral interaction test papers which must reflect the linguistic actualities of the controllers' job. Maintaining testing standards of all involved in testing is critical to ensuring valid and reliable tests. A transparent approach to keeping stakeholders informed of developments is essential in maintaining a healthy and robust test.

10. **Rationale for assessment and implementation of the ELPAC test**

From the beginning, it was recognised that the ELPAC test had to achieve a very high level of security given the high stakes involved. Experience with developing a PC-based version of the PELA test stemmed from the desire of administrators to reduce the

amount of paper involved and the consequent time spent on test administration. This experience and the need for a very high level of security confirmed the decision to develop ELPAC as an internet based test, allowing that the Paper 2, Oral Interaction, has to have a direct human interaction (although all materials are stored in the database). The test has also to be as objective as possible in assessment.

11. **Paper 1 – Listening Comprehension**

A test of listening comprehension with routine and non-routine situations which are generic to all ATC functions. All items are based on authentic materials which are reviewed and if necessary rewritten in line with test specifications. Audio files are recorded using a large variety of nationalities relevant to European airspace. Items in Paper 1 are constructed and standard set to ICAO levels 3, 4 and 5. An algorithm calculates a candidate's level once a trained and accredited marker has completed marking the paper. Candidates provide their responses to items through a keyboard being required to enter numbers or words (maximum three or four). The duration of Paper 1 is approximately 40 minutes.

2.1 Using the keyboard to input data

The **ELPAC test does not require candidates to be computer literate**. Knowledge of computer programs is not expected. The only thing **candidates need to know** is **the keyboard layout**, or in other words, navigation functionalities – **how to delete something, how to make a space between words, etc**. The keyboard layout may vary from a country to country but the basic keys (delete, enter, space etc.) are the same. All test items are spaced in such a way that when a candidate has a sufficient knowledge of English and types slowly using only one finger, s/he will still have enough time to input all the answers. When examining all items in all ELPAC Paper 1 live test versions, one can see that the fastest speed required is 10 words per minute (short words such as articles or prepositions are each counted as one word). If a candidate does not understand the recording, then the problem is not in the keyboard input but in an insufficient knowledge of English.

A special procedure is in place to cater for candidates who are unable to use the keyboard to answer the test items. Such a candidate must state (in writing and prior the test) that s/he is unable to use one or more of the following keyboard functions:

a. Type with a speed of at least 10 words per minute (articles and prepositions each counted as one word).
b. Delete a letter.
c. Click on a box (using a mouse).
d. Make space between words.

This statement is submitted to the ELPAC National Administrator who co-ordinates further actions with Eurocontrol.

12. **Paper 2 – Oral interaction**

Two trained and accredited examiners assess candidate performance in visual and non-visual interactions. Candidates may select Tower, Approach or En route scenarios but these only serve to put the interaction into a familiar context – it is always the language proficiency being assessed not ATC procedures. Candidate performance is assessed at level 4 or level 5. Specific tasks assess the candidate's ability to give an opinion, put forward an argument, evaluate, speculate or hypothesise. The candidate is encouraged

to produce a rateable sample of extended speech. Interlocutors adhere strictly to a framework to ensure that each candidate receives the same test under the same testing conditions. The duration of Paper 2 is approximately 20 minutes.

13. **Maintaining testing standards**

All test administrators, Paper 1 markers and Paper 2 examiners are trained by Eurocontrol. Accreditation is by the National Licensing Authority on the recommendation of Eurocontrol. A user group of administrators meets at least once, if not twice, a year to exchange experiences and put forward suggestions for test improvement. Regular refresher workshops take place for markers and examiners. The performance of markers and examiners is monitored by reviewing marked Paper 1s and recordings of Paper 2 performances. Candidate anonymity is guaranteed.
New items for Paper 1 are trialled with each test version. These are then statistically analysed, evaluated and standard set before going forward to create new test versions. New Paper 2 tasks are trialled through familiarisation courses and training workshops.

14. **Updating software**

Continuous development takes place through feedback from all testing personnel and candidates. New software upgrades are produced on a regular basis in an evolutionary manner. A helpline (telephone and/or email) is available for administrators. Regular meetings, emails and conference calls ensure that test and software developers keep abreast of user requirements.

15. **Dissemination**

The ELPAC project disseminates information through a dedicated website (www.elpac.info) and makes available sample tests for familiarisation on www.elpacsample.info. Development, training and testing is carried out on separate, isolated servers for maximum security. New websites have been established for the version of ELPAC being developed for the student controllers. The project is subject to scrutiny by Eurocontrol Member States through such fora as the Training Focus group (TFG) and the Human Resources Team (HRT). Close liaison on language proficiency matters is maintained with ICAO through the PRICE Study Group and the COG ATM Task Force.

16. **Conclusion**

The ELPAC test is designed to assess English language performance at ICAO levels 4 and 5 for air traffic controllers. In meeting the need for security and economy of administration for such a high stakes test ELPAC is internet based. Sample tests are available for familiarisation and stakeholders are regularly informed of progress. Dedicated web sites allow easy access for information and a transparent approach to test design.

RELTA (RMIT English Language Test for Aviation)

Michael Kay
(RMIT)

Test Design and Construct Rationale

Test Assessment and Implementation Rationale

TEA (Test of English for Aviation)

Ben Rimron
(Mayflower College)

BIO DATA

Ben is an Aviation English course leader, teacher and TEA Examiner working at Mayflower College in Plymouth. He has a Bachelor's Degree in English Language, is DELTA-qualified and is an IELTS Speaking Examiner. He is concerned about the differences in perception between Education and Training in the world of language teaching and testing.

TEST OF ENGLISH FOR AVIATION (T.E.A.)

Approved English testing for aviation

The Test of English for Aviation is managed and administered by Mayflower College and:meets the ICAO English Language Proficiency Requirements (LPR's) for licensing purposesis approved by the UK Civil Aviation Authority (certificate here) and many other national authoritiesassesses spoken and listening ability according to the ICAO Rating Scale

http://www.maycoll.co.uk/aviation-english/test_of_english_for_aviation.htm

BREAKOUT SESSIONS 1 & 2

VAET (Versant Aviation English Test)

Larry Rothenberg
(Pearson Ordinate)

Versant Aviation English Test

The Versant™ Aviation English Test was created to help aviation employers meet the new requirements from the International Civil Aviation Organization (ICAO). The test was developed under a co-operative research and development agreement with the U.S. Federal Aviation Administration (FAA) and is compliant with the standards specified by ICAO.

https://www.versanttest.com/products/aviationEnglish.jsp

PLENARY SESSION 3

Summary of breakout sessions by four moderators

1. The Eurocontrol English Language Proficiency for Aeronautical Communication (ELPAC) test

Sessions presented by **Adrian Enright** & **Magdalena Vecerova**

Sessions reported **by Henry Emery**

Summary

ELPAC (English Language Proficiency for Aeronautical Communication) is an English language proficiency test developed by EUROCONTROL for Air Traffic Controllers to meet language proficiency requirements for both ICAO and the European Commission's Directive on a Community Air Traffic Control Licence. For further information, visit www.elpac.info.

Adrian Enright (ELPAC Project Manager, EUROCONTROL), accompanied by Magdalena Vercerova (Language Testing Consultant to EUROCONTROL) led breakout sessions on Thursday 8th and Friday 9th May on the design, development and administration of the ELPAC Test.

The presentations on the two days covered the following areas:Thursday 8th May

Rationale
Test Development
Design
Trialling
Evaluation
Validation
Dissemination
Friday 9th May

Rationale
Paper 1: Listening comprehension
Paper 2: Oral interaction
Maintaining Standards
Updating Software
Dissemination

What follows is a selection of the key questions posed by the session participants and the responses from Adrian and Magdalena, the breakout session leaders, over the two days. The questions are grouped according to the following three themes:

1. Test content and specification
2. Paper 1 (Listening Comprehension)
3. Paper 2 (Oral Interaction) delivery and rating

Questions and answers

1. Test content and specification

Q: How did the ELPAC development team specify the content of the test?

A: Language functions were selected on the basis of frequency and importance in aeronautical communications. Although the ELPAC development team benefited from previous experience with the PELA test, the ELPAC test specifications went through a number of drafts and revisions as a result of debate and discussion between subject matter experts and linguists before final test specifications, including content, were produced.

Q: Does ELPAC test phraseology?

A: ELPAC doesn't test phraseology per se. Phraseology is required as receptive and productive language in order to fully test operational language proficiency in the context of radio-telephony communications. Phraseology is taken as an integral part of aeronautical communication and thus in Paper 1 (Listening Comprehension) candidates listen to recordings of routine and non-routine situations. These recordings contain both phraseology and plain language. In Task 1a of Paper 2 (Oral interaction), candidates have to demonstrate proficiency in the appropriate use of standard ICAO phraseology, switching between standard ICAO phraseology and plain English, making an appropriate response to a pilot message, resolving misunderstandings,dealing effectively with the pilot/controller relationship, negotiating a developing unusual situation, making a verbal report in plain English A candidate who fails to comprehend recordings in Paper 1 or demonstrate proficiency in Task 1a of Paper 2 will fail to be awarded a level 4 overall, however strong the performance elsewhere in the test.

Q: Why does ELPAC not test beyond level 5?

A: The focus of ELPAC focus is on levels 4 and 5 for a number of reasons. Firstly, level 4 is considered to be a strong level of language proficiency which is sufficient for effective aeronautical communication; language proficiency above level 4 is an added benefit. Secondly, testing at levels 4 and 5 is an extremely complex task in itself. Thirdly, to test for level 6 proficiency would mean to move far beyond the realm of radiotelephony communications to test discourse and language use which is simply absent from aeronautical communication (Level 6 goes far beyond the scope of aeronautical communications (vocabulary is idiomatic, nuanced and sensitive to register, level 6 speaker varies speech flow for stylistic effect, demonstrates comprehension of linguistic and cultural subtleties, is sensitive to verbal and non-verbal cues and responds to them appropriately, etc.)..

2. Paper 1 (Listening Comprehension)

Q: How did the ELPAC development team produce the radiotelephony content for Paper 1?

A: Due to issues of human sensitivity and security, and the length of a typical R/T dialogue, it is problematic to use authentic radio-telephony scripts. The scripts for Paper 1 of ELPAC are based on transcriptions of authentic air traffic control recordings. Scripts are standardised for length, audio and linguistic quality and sanitised to remove sensitive references found in real communications, although the scripts do refer to real places,

airspace and operators (airports, FIRs and airlines). The scripts are re-recorded by EUROCONTROL personnel.

The authenticity of the recordings is maintained in that grammatical errors occur in the scripts which are representative of the use of English by the international community. Furthermore, the voices found in the re-recorded scripts represent a variety of international accents which are considered acceptable to the international aviation community, and typical of those found within the test population and to which the test population are exposed.

Q: How did the development team determine the difficulty of the items in Paper 1?

A: In the trialling process, the development team used two methods of determining the difficulty of items and thus helping to ensure test equivalence and stability. Firstly, standard setting sessions were conducted whereby pre-trained personnel were invited to sit Paper 1 and were asked to decide the level of difficulty of each item presented. Secondly, items were subject to statistical analyses which disclose levels of difficulty and determine the usefulness and effectiveness of test items. The results of the qualitative and quantitative data then informed the production of final versions of the test. Item-writing and trialling is an ongoing process in the maintenance of ELPAC, with new items being trialled as a part of live test administration. More information can be found on www.elpac.info under "Final test construction" and "ELPAC validation".

Q:What are the implications of testing listening comprehension in isolation as a separate skill? In particular, how far does responding to Paper 1 (listening Comprehension) of ELPAC rely on a candidate's secondary skills or knowledge which are not assessed, such as reading, writing and spelling?

A: Item responses in Paper 1 are limited to alpha-numeric data or very short answers, thus reading and writing skills do not interfere with item response. Paper 1 is marked in part automatically, in part by trained human clerical markers. Human markers of Paper 1 are assisted by marking sheets to maintain objectivity; spelling errors are acceptable so long as the meaning of responses is clear.

Q: Are candidates disadvantaged in that typing skills are necessary to respond to tasks in the computer mediated Paper 1?

A: The typing speed required to respond effectively in Paper 1 is maximum10 words per minute. Compared to the 27 words per minute of an average 'slow' typist, the ELPAC development team felt that slow typists are not disadvantaged in any way. Test versions were trialled on candidates who were instructed to type their answers using only one finger.

Q: Did the ELPAC development team consider presenting the rubrics, instructions and task items in the candidate's native language rather than in English?

A: Research suggests that candidate's are more comfortable and prefer to be tested in the target language of the test. Furthermore, to deliver the test in the many first languages of the test population would incur considerable time and expense.

Q: How is time managed during Paper 1 of the test and what time is given to candidates to respond to test tasks? Are candidates disadvantaged in any way by the timing of test delivery?

A: ELPAC candidates are given adequate time to read and listen to rubrics, instructions and test items, to listen to the test scripts, to respond to the items and finally to review the responses before the test automatically moves on to the next task. Air traffic control is a time driven profession with procedures and communications logically ordered and sequenced in time. To reflect this, candidates always know what information they are listening for before the information is presented, and responses are given in the order that information appears in the scripts. ELPAC sample test on www.elpacsample.info contains the same pauses and preparation times as real test versions and candidates are encouraged to take the sample test in order to become familiar with the test format.

Paper 2 (Oral Interaction) delivery and rating

Q: Does ELPAC use the ICAO rating scale?

A: Essentially, yes. However, for language rating in Paper 2, the ELPAC team standardised the meaning of the adverbs of frequency (sometimes, often, frequently etc) found in the criteria of the scales. This was done by quantifying the occurrence of a behaviour for a particular adverbial descriptor so that 'sometimes' becomes 25% of the time, 'rarely' 10% of the time, for example. The rating scale with the definitions of adverbs of frequency was trialled several times before consensus was reached. The team then produced DVDs of candidates performing in Paper 2 and demonstrating these adverbial descriptors for rater training purposes.

Q: How are the various components of the ELPAC test rated to produce an overall score?

A: Paper 1 is a discrete item listening test, and the candidate is awarded a level for comprehension. In Paper 2, performance is rated as a process through Tasks 1a, 1b and 2, and a level is awarded in each of the six language profiles of the rating scale, including comprehension.

In Paper 2 Task 1a, in which the candidate takes the role of an ATCO in radiotelephony role-play, the maximum rating that a candidate can be awarded is level 4 or 5 except for fluency (level 5 fluency cannot be awarded as the task does not elicit language at level 5 (speaks **at length** with relative ease on familiar topics). In Task 1b the candidate remains in the role of the ATCO and de-briefs the "watch supervisor" based on the events that took place in Task 1a. The candidate is rated across all six profiles, this time at below level 4, level 4 or level 5. In Paper 2 Task 2, the candidate describes and discusses an aviation-related picture. In this task the candidate is given further opportunity to demonstrate language proficiency at the extended level though producing stretches of fluent speech.

Paper 2 is rated as a 'process' in that an overall rating is given for the <u>whole</u> oral test event rather than ratings for each part (although if the candidate fails to perform at the Operational level with standard phraseology in Task 1a, s/he cannot be awarded an Operational level overall). It is often found that a candidate's performance varies through the different Tasks of Paper 2, as illustrated in Figure 1 below.
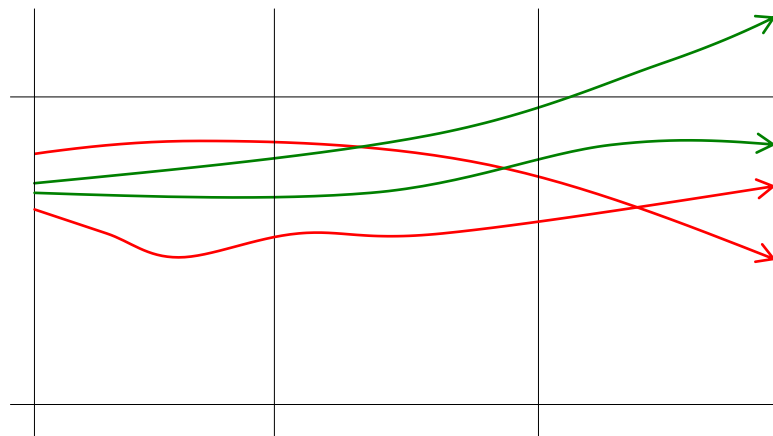
Q: To what extent can an interlocutor in ELPAC Paper 2 support a candidate who is struggling to comprehend the oral prompts?

A: The interlocutor has a series of scripted prompts to guide interaction in Paper 2. If a candidate fails to comprehend a prompt, the interlocutor may repeat the prompt once. If the candidate fails to comprehend the prompt for a second time, the interlocutor may read a pre-scripted paraphrase of the original prompt once. No further assistance to the candidate is given, and if comprehension is unsuccessful, the interlocutor moves on to the next item in the test process.

Q: How are the roles of rater and interlocutor managed during Paper 2?

A: In conducting Paper 2 tests, there is always a team of two test personnel, a subject matter expert (SME) and an English language expert (ELE). In Tasks 1a and 1b, one member of the team takes the role of interlocutor (nearly always the ATCO) while the other rates. In Task 2, the roles are switched.

Q: If each Task of Paper 2 (1a, 1b and 2) of the test is rated by the team of two and the two members of the team switch roles during the test, how is workload managed given that rating and interlocution are difficult roles to perform simultaneously?

A: It is recognised by the ELPAC development team that interlocution and rating at the same time is problematic. Therefore, one member of the team takes a passive role and rates the candidate's performance using the rating scale while the member of the team acting as interlocutor works with an overall scale, a simplified version of the scale which is easy to refer to and serves as an *aide memoire*. In Task 1a the interlocutor also uses a 'traffic light system' which makes it easier to evaluate the candidate's responses The examiner who is interlocuting and thus using the "overall scale" may refer to the analytic scale (with 6 criteria). Both examiners complete assessor sheets where they have to record the whole rating process.

Q: *When is a third rater used?*

A: A third rater is used when there is a discrepancy in the rating between the two members of the team which conducted the test. For objectivity, the third rater rates the sample 'blind' (s/he does not know the initial ratings) and is not made aware why there was a discrepancy between the raters.

Q: *What happens in the rating of Paper 2 if a candidate isn't particularly talkative by nature?*

A: It is essential that a full speech sample is elicited in Paper 2 that is demonstrative of the candidate's oral proficiency. In preparing to take ELPAC, candidates are encouraged to produce as much language as possible and it is made clear that it is the candidate's responsibility to 'perform'. Candidates are reminded that in Tasks 1b and 2, it's not the content of language that's being assessed, but rather the quality of language used to express themselves.

Q: *If a candidate fails to demonstrate level 4 in ELPAC, what is the recommended period before re-testing?*

A: The ELPAC team recommend a minimum period of 3 months during which a candidate may wish to undergo diagnostic testing and language training before returning for retesting with ELPAC.

2. The Royal Melbourne Institute of Technology (RMIT) English Language Test for Aviation: RELTA

Sessions presented by **Michael Kay** of RMIT

Sessions reported by **Marjo Mitsutomi**

The purpose of the aviation English proficiency test developed by RMIT is to determine whether aviators have attained the required ICAO English proficiency level in listening and speaking. Pilots and controllers are examined via different test forms, meaning that the two groups have separate tests although they are very similar in structure.

The focus of the test content is on radiotelephony (RT), phraseology and plain language. Knowledge of aviation content is required although not assessed. The test is advertised as being useful for licensing purposes at which stage the candidates are already skilled at the craft they are pursuing. It must, then, be noted that the RELTA is inappropriate for ab initio trainees whose skill base in the profession is inadequate to respond appropriately and correctly to the presented scenarios. Most native speakers without aviation content training would perform poorly in this test. The RELTA is a good example of a test of English for specific vocational purposes (ESP).

The test measures all six strands of the ICAO standard. Both forms of the tests consist of listening and speaking sections. The listening test is computer-delivered to multiple candidates at the same time. The areas tested are RT phraseology in routine situations, non-routine RT communication mixed with plain English, and lastly an interview in plain English. The speaking test has three parts: (1) routine phraseology, (2) non-routine phraseology mixed with plain English, and (3) a face-to-face interview in plain English with an interlocutor present (a live examiner, in other words). The last area requires the test takers to produce linguistically more complex language than the other two sections. Air traffic controller candidates have the additional component in their speaking test that they are required to be able to use appropriate language in three specifically distinct context, namely enroute, tower and approach.

The test design rationale, and correctly so, was based on the necessity to assess proficiency in English language usage in aviation specific situations. The ICAO standard and the scale were carefully analyzed at the onset in order to identify the language skills, both functional and notional, to guide the test writers in developing appropriate linguistic and applicable work-related prompts for the test takers. The focus of the test developers was on creating meaningful and authentic tasks which maximize the opportunity for engagement in the English language. The test development team decided that that the work-related language challenges with which pilots and controllers are faced are distinct enough to warrant a separate test for each group.

The RELTA currently has several different forms of the test for both pilots and controllers. It took the team at least two years to develop the first version, beginning with the thorough study of the ICAO specifications, followed by the prototype and its refinements, numerous triallings, creation of the test bank subject to constant review and monitoring of the items, until its first publication a few years ago. The test was developed with the assistance of target users, subjects matter experts, and other test development experts so that RMIT states with confidence today that the test is valid and reliable.

Mr. Kay's assessment of the RELTA is that it meets the standards for best practice in aviation English testing. The reasons are that is has been shown to be valid and reliable, it is authentic and interactive, and it is fair and practical. If RELTA is chosen as the preferred instrument of testing aviation English proficiency, the company highlights the following on its website:

RELTA is designed in response to the ICAO Language Proficiency Requirements
RELTA is delivered locally by accredited examiners
RELTA is assessed by accredited raters
RELTA assesses proficiency according to the ICAO Language proficiency Rating Scale
Validity and reliability of RELTA is rigorously checked
RMIT trains and accredits raters and examiners

It was very evident to Mr. Kay's audience that test development is serious business; it takes the input of several authorities in various areas of expertise, and it is very time-consuming, requiring the full-time energy and knowledge base of several individuals whose primary job description is nothing but test development and validation.

3. The Test of English for Aviation (the TEA test)

Sessions presented by **Ben Rimron** (Mayflower College)

Sessions reported by **Fiona Robertson**

The Test of English for Aviation (the TEA test)

This is a proficiency test of communicative competence of plain English in an aviation context for trained pilots and controllers. It tests the skills of speaking and listening. It takes place in the form of a one-to-one interview with two interactive listening comprehension sections.

There is a warm up work-related interview (Part 1) lasting 3 or 4 minutes, followed by the comprehension sections (Parts 2A and B) using CD recordings or non-routine situations with different accents.

The first section covers 20 situations and the candidate's task is to explain after each recording, what is happening.

The second section consists of 8 recordings concerning different, short, unusual and problematic situations. The candidate must ask for clarification or more information about each situation and after the reply, make a suggestion of help to find a solution.
Part 3 is a discussion using 2 aviation related pictures. The candidate describes the pictures and the examiner asks more probing questions designed to allow the candidate to produce longer statements and to shine, if possible.
The test lasts 20 minutes.

TEA elicits enough language to be assessable with the ICAO rating scale. A study has been made to check that the test elicits the functions itemised in the ICAO documentation. For this study, the 116 functions were grouped under 25 headings (e.g. orders) and using checklists a chart was made of the occurrence of these functions. It was thus shown, for example, that "clarifying" was used more at lower levels while "summarising" was more frequent at the higher levels.

Attention was also paid by the test developers to eliciting language that falls within the ICAO defined priority lexical domains.

This is a test of plain English in an aviation context, not of radiotelephony phraseology. It tests language, not operational procedures.

The standardisation and therefore the reliability of the interview is enhanced by having a standardised script and rubric which the examiners follow. There are multiple test versions. Each test is digitally recorded and the tests are regularly monitored.

There is rigorous examiner selection – examiners must have a minimum of IELTS level 7. They have intensive training and on-going monitoring, with re-certification every 2 years. They respect the ILTA code of ethics.

Intra- and inter-rater reliability studies are conducted, and a concurrent validity test was performed when the school asked the teachers of future candidates to place their students on

the ICAO rating scale. Comparison with the exam results subsequently yielded a very high correlation coefficient.

The assessment is by profile marking using the 6 scales. But for listening part 2A (the 20 short recordings) a quantifiable mark is given and a listening band-scale ceiling thus established. Scores are logged into a secure data-base by an administrator. The test recordings are archived. A photo of each candidate is taken immediately before the examination begins and this photo then appears on the certificate (which has numerous security features to reduce the risk of counterfeit).

For further information about TEA click here:

http://www.maycoll.co.uk/aviation-english/tea.html

4. The Ordinates Versant Aviation English Test (VAET)

Sessions presented by **Larry Rothenberg** (Ordinates)

Sessions reported by **Angela ffrench**

Thursday 8 May 2008

Overview

The Versant Aviation English Test comes in two forms: The Versant Aviation English Certification Test, and the Versant Aviation English Practice Placer Test.

The Versant Aviation English Certification Test lasts approximately 25-30 minutes. Each test event has to be taken through an authorized test partner, which is responsible for making appropriate test-taker identification checks and guaranteeing test security. Test takers are provided with anonymous test ID numbers to ensure data privacy.

The test-taker is provided with instructions and a test paper, and either makes a phone call or utilizes a PC to interact with the Ordinate testing system. Each test event is proctored.

The test has 8 sections, comprising 78 questions:
- Aviation Reading
- Common English Reading
- Repeat
- Short Answer Questions
- Readback
- Corrections and Confirmations
- Story Retelling
- Open Questions

Aviation and general English Language skills are tested. All instructions provided are in English.

Test-taker responses are captured and stored securely. Each response provides data which is used for the final calculation. Evaluation is made at the end of the test and the responses made by the test taker are machine scored. If for any reason a test is interrupted, it is discarded and the test taker has to start again. Each test is unique having been randomly compiled from a battery of test items.

A numeric score on a 10-to-70 point scale is posted on the Ordinate website approximately 5 minutes after the test has been taken. The test taker is given a breakdown of performance in each of the sub-skills, together with an explanation of the sub-skill scores.

The Versant Aviation English Practice Placer Test is a low stakes version of the test and is used:
- to assess overall ICAO level and place individuals into the right level for training
- to allow test-takers to practice for the Certification test
- to use as an instructional activity within training.

As for the certification test, test takers are provided with anonymous test ID numbers to ensure data privacy, and their responses are captured and stored securely.

The test lasts approximately 15 to 20 minutes and tests aviation and general English.

The test is a fixed format with 7 sections, comprising 54 questions:
- Aviation Reading
- Common English Reading
- Repeat
- Short Answer Questions
- Readback
- Corrections and Confirmations
- Story Retelling

Test takers are provided with an estimate of the current ICAO language level only.

For both versions, sample tests are available.

Development

The Versant Aviation English Tests contain two types of item: aviation specific, and general English. The aviation specific items are newly written and use is made of Oklahoma' State University's aviation corpus, radio telephony corpora, aviation standards publications and general aviation publications. The general English items are taken from the bank of materials and responses from other general English tests in the Versant™ suite. Materials produced by authors from one English-speaking state are reviewed by personnel in other English-speaking states to ensure they are authentic and that they comply with international standards. Pilots and ATCs are used for the recordings.

During the trial period, materials were tested on students from 90 countries representing 50 languages, providing over 100,000 responses.

Items for inclusion in live tests are pretested on over 500 students. These responses are recorded, transcribed and mapped onto a response network: tasks being mapped to criteria. The performance of every response is re-evaluated by the internal team to see if it is acceptable and every possible valid response is included in the marking database.

Issues arising from the break-out session

1. Concern was expressed about the inability to test interactions. Ordinate drew the groups attention to the 'Corrections and Confirmations' task. In this task tests takers are presented with a radiotelephony exchange between two speakers. After they have heard the exchange, the test takers have to imagine that they are one of the speakers and continue the dialogue. If they hear information which is incorrect, they have to correct the information; if the other Speaker includes a question or a request, they have to respond appropriately.

   The assessment model looks at other task types and cross refers to the assessment scales to see what predictive measurements can be derived regarding interactions. Ordinate suggests that from the thousands of measurements which are collected this is possible. However, it

was acknowledged that the test design did not allow a test taker to show competence in interacting by, for example, requesting clarification.

2. The group asked how a test taker could demonstrate control of grammatical structure and be able to form associated meanings, a requirement of Level 4. Ordinate drew attention to the 'Repeat' task type: this task starts with short sentences and then the sentences get progressively longer. Research has shown that there is a limit to the amount of information a second language learner can repeat without needing to call on the skill of transformation. The strings of questions presented are too long for an individual to be able to memorize in one hearing, and in attempting to repeat the information test takers have to transform the information, so demonstrating comprehension and control of grammatical structures.

3. The group asked how longer stretches of speech were tested and marked. Ordinate drew attention to the 'Story telling' task; in this task test takers listen to an input of about 30 seconds and are required to retell the story in up to 30 seconds. For the time being, this task is being developed for training purposes only and though included in the test is not included in the assessment.

    The 'Open questions' provide an opportunity for test takers to give extended responses, but these are currently marked by human raters.

4. ICAO Doc 9835 (2004) questions the appropriateness of computer marking. Ordinate suggested that this position may need to be reviewed in light of advances in technology.

Friday 9 May 2008

1. **Information**
    Information on test development, the development team, the process of development, methods for producing data, and validation reports are all available on the Ordinate website: www.ordinate.com

2. **Test availability**
    The Versant Aviation English Certification Test can be taken 24 hours a day through an authorized test partner. Tests are scheduled to take place every hour to take account of the time needed to read the instructions, ensure the test taker has taken the practice test etc.

    The Versant Aviation English Practice Placer Test is available to anyone.

3. **Control of the environment**
    In order to ensure each test taker has the same experience, the certification test is administered through an authorized test partner. Responsibilities therefore rest with both Ordinate and the authorized test partner.

    Ordinate:
    - provides the certification test and the practice test
    - monitors test statistics
    - maintains test data
    - identifies non-performing items
    - maintains the item pool by adding or deleting items
    - maintains the test development system

Authorized test partner:
- provides testing facility
- ensures item security
- screens test takers
- checks test taker's ID
- proctors the test event
- maintains records
- provides practice tests
- provides training (optional)

4. **Test administration**

   Tests are mostly administered through Jepperson Solutions, though other organisations which have recognised experience of going through licensing procedures are also used.

   The test is taken either by phone or via a PC. Cell phones and VoIP systems are not acceptable. Ordinate requires the centre to provide 3 submissions for quality checking purposes. Should 'call drop' occur, or if for any reason the test taker chooses to abort the session, the test has to be re-started with a new test configuration.

   All responses are sent to a central server which handles the assessment process and the result which is generated is posted to the Ordinate website.

5. **Test Security**

   Each test is unique and is generated randomly from an item pool. Responses are monitored in relation to the test taker's overall score and if 'item drift' is identified, i.e. an item is found to be consistently proving to be too easy (with security possibly having been breached) the item is removed from the item pool.

6. **Raters**

   Currently, the validation of test items is carried out by the internal development team, which comprises experts in the areas of linguistics, psychometrics and test development. It is anticipated that as entry numbers increase, external raters will have to be taken on or there may be negotiations with organisations such as ACTFEL.

7. **Reporting of information**

   A numeric score on a 10-to-70 point scale is posted on the Ordinate website approximately 5 minutes after the test has been taken. The test taker is given a breakdown of performance in each of the sub-skills, together with an explanation of the sub-skill scores. The reporting illustrates where there are strengths as well as weaknesses.

8. **Review process**

   A review process is available. All the tests are recorded and the system can identify strings of incorrect responses. These can then be listened to by human raters to see if there was a reason for the poor performance, e.g. extraneous noise which affected the test taker's ability to hear clearly.

PLENARY SESSION 4

Angela ffrench & David Booth, Cambridge ESOL

BIO DATA

Angela ffrench holds an Med in Language Testing from the University of Bristol and an MA from the University of Cambridge. Having been involved in teaching and testing English for 20 years, both in State and private sectors, Angela joined Cambridge ESOL in 1991 as an Examinations |Officers, with special responsibility for Speaking.

Since 2001 she has been responsible for the management of a suite of examinations ranging from upper intermediate to advanced levels. In this capacity she has been involved in all aspects of developing and administering tests and assessment criteria, in the training of item writers, in pretesting and analysing test material and in the grading of examinations.

David Booth has worked at Cambridge ESOL since 1998. He has worked across the business including roles in Research and Validation, Assessment and Operations and Business Management. He is currently the manager of the Business Management Unit which co-ordinates a number of marketing and business functions.

David has worked extensively overseas with the British Council as a teacher, teacher trainer and Director of Studies in Hong Kong, Malaysia and South Korea . David returned to the UK in 1998 to take up a post at Cambridge ESOL.

David has extensive teaching, training and assessment experience and has delivered seminars and presentations to a wide range of audiences around the world.

'Delivering Validity and Reliability - practical concerns in delivering assessments'

Angela ffrench and David Booth, Cambridge ESOL, Cambridge, UK

Abstract

Issues of validity, reliability, impact and practicality underpin the development and delivery of language proficiency assessments. Attention to these must be taken at all stages of the process of producing and delivering such tests. Cambridge ESOL, as a major provider of English language assessments, has developed detailed processes to ensure that this is the case.

This paper looks in detail at the process, sub-processes and process outputs and discusses the procedures, documentation and training required to implement them. The discussion will address issues such as test administration including security checks, examiner and administrator training, inspection criteria, policing irregular conduct, issuing results and awarding procedures.

Background

This paper is based on the presentation given to the International Civil Aviation English Association conference 'Testing for ICAO compliance: Best Practice in Aviation English Proficiency assessment'. The conference, sponsored by Polish Air Navigation Services Agency, was held in Warsaw Poland between the 7th and 10th of May 2008. The purpose of the presentation was to provide participants with an overview of the issues related to test administration to enable them to critically evaluate Tests of Aviation English, including those presented at the conference. As part of the presentation a checklist was presented to participants to focus their attention on key areas of concern. Extracts from the checklist are included in this paper.

The five phases of the Core Process of Test Delivery

Critical to ensuring the reliability and validity of any test is the requirement to ensure that tests are delivered in a secure and appropriate environment and that effective post test procedures are in place to guarantee test results. This paper outlines some of the main process elements which need to be put in place in order for this to be the case. The paper concentrates on three main phases of the test delivery process, Examination Administration, Post Examination Processing and Exam Review and Evaluation.

In order to understand the phases of test delivery we need a model of the process which defines the relationship between the different phases. This process in turn is part of more complex organisational interactions which positions the process in the context of the examination provider. The diagram below is a representation of a test delivery process. It was developed by Cambridge ESOL as part of its Quality Assurance procedures to help staff conceptualise the work of the organisation and understand the fundamental relationships between the core phases of the test delivery function.

Figure 1: The Five Phases of the Core Process
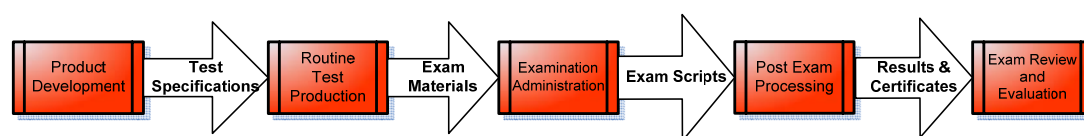
Approaches to the core phases of Product Development and Routine Test Production are discussed in other conference papers but it is worthwhile noting that the phases all relate to each other. Each phase has a process output which then feeds into the next phase. The quality of each of those outputs is critical in ensuring the validity and reliability of the test. Also worth noting are the sub-processes which are associated with the phases of the core process. In the case of routine test production, phases include commissioning (getting experts to write test material and test items), pre-editing and editing (checking the material is appropriate and will elicit appropriate responses, checking answers are unambiguous; in short using expert judgements), trialling and pre-testing (ascertaining the difficulty of the material), where necessary standards fixing (ensuring the measurement characteristics of test material, especially objectively scored test items, combine to give a score which can be related to a reporting scale such as the ICAO language proficiency scales), and test construction (ensuring that the different forms of the test are balanced and equivalent in terms of content and measurement characteristics). All these elements ensure that test materials, be they examiner prompt material, listening tasks and items or computer delivered prompts are of the highest quality to feed into the test administration phase.

Validity, Reliability Impact and Practicality issues in Test Delivery

The purpose of defining the phases of test delivery and listing the sub-processes which from part of those phases is to ensure that attention is paid to detailed requirements of Examination Administration, Post Examination Processing and Exam Review and Evaluation to support the validity argument for a particular assessment. Test providers need to provide evidence to the stakeholder community that their test is fit for purpose. The criteria by which it is decided that a test is fit for purpose is not only defined by test developers and test experts. The views and expertise of a much wider range of involved individuals and groups need to be taken into account. The role of the test developer is to provide evidence and explanation for the way in which the test is presented and to commit to gathering evidence relating to the performance of the test.

Reliability, in the simplest sense the consistency of measurement, is also a concern of the test developer. In relation to tests which report to the ICAO rating scale, issues of rater standardisation and training are of paramount importance. Rater standardisation is an involved process and Cambridge ESOL's approach to it, based on its significant experience of delivering face to face speaking tests on an international scale, is referred to later in this paper. Impact and practicality are also a concern for test developers and it is hoped that test providers will further explore the impact of their tests on the aviation community, including pilots and air traffic controllers, and work with national agencies and certified training agencies to ensure test administration is efficient and cost effective.

Examination Administration

The Examination Administration phase puts the candidate at the centre of the test process. In international testing the aim is to ensure that each candidate, regardless of where they are taking the test, can be confident that as far as possible the test will be

taken under the same conditions as all other candidates. In order to do this it is the responsibility of the test provider to ensure that its test venues are carrying out test administration to the highest standards and that test administrators are appropriately trained. The focus of this phase is to ensure that the exam scripts, which are the output from this stage, as far as possible represent the best samples of candidate language for assessment purposes. Scripts, in this context, is meant to include any output which is to be assessed be that answers on paper, spoken discourse or digital recordings or records of performance.

The questions below focus on the arrangements test providers make to accredit centres, venues or administrators, to provide training, and to audit and inspect them.

Figure 2: Checklist Exam Administration

1. Exam Administration		
1.1	What is the approval process for centres?	
1.2	How often is centre approval required?	
1.3	What requirements are there in respect of:	
	– seating arrangements	
	– acoustics	
	– security of materials	
	– information to candidates	
	– training and monitoring of centre staff	
1.4	How often are centres inspected?	
	– by whom?	

Procedures are invalidated if they are not monitored and policed. It is crucial that the test provider can control the environment in which the test is taken if they are to maintain the validity and reliability of the test.

In addition to this the quality and comprehensiveness of documentation is fundamental to ensuring that all participants in the Examination Administration phase understand their roles and responsibilities. Documentation can include:

- Handbooks for centres, giving precise information on how to receive, store and return materials, and how to administer the tests
- Centre information, so candidates know where and when it is possible to take the tests
- Regulations, so candidates know precisely what they must do
- Handbooks for teachers, which include: specifications for the overall test and for tasks within it; sample materials; advice on how to prepare for te test
- Publishers' lists of course books which are appropriate for test preparation.

Issues such as support and complaints procedures also need attention as is illustrated below:

Figure 3: Checklist Centre Support

2. Centre Support		
2.1	What documentation is provided?	
	- for centres	
	- for teachers	
	- for candidates	
2.2	How often is it updated?	
2.3	Are translations of information available?	
2.4	How are enquiries and complaints managed?	
2.5	How quickly are they responded to?	
2.6	Is there a list of FAQs?	

In delivering the assessment attention needs to be paid to the exam process and procedures for setting test dates and the secure dispatch and storage of test materials. Some key issues are illustrated below:

Figure 4: Checklist Exam Process

3. Exam Process		
3.1	How often can the exam be taken?	
3.2	What is the process for enrolling?	
3.3	How often can a candidate re-take?	
3.4	How much notice is given of exam dates?	
3.5	Do the exam dates take account of cultural/ religious festivals?	
3.6	How is the material despatched?	
3.7	How is the material received and stored?	

Oral Examiner Management

Much attention has been paid both in the literature on language testing and in the development of assessments of spoken language on the role of the examiner. Issues of inter-rater and intra-rater reliability are central concerns of assessing oral language proficiency. Alongside this requirement, as has been already stated, it is important to ensure that test takers have a similar test experience across sessions and test locations.

The central place of the examiner in delivering the test and in most cases assessing the candidates' responses has been a central concern of examination providers. The roles of interlocutor and assessor are sometimes separated and sometimes combined. When combined it is important for examination providers to ensure that the examiner has the resources and time to both manage the interaction and to assess the candidates' proficiency. In many cases, however, the roles are split, either with the use of two examiners or in the recording of answers. In some cases marks are weighted so that one examiner's score contributes more to the overall score. In others, there are multiple ratings including a 'blind' rating using a recording to adjudicate between differences in examiners' scores. The method by which scores are collected does have a direct impact on the validity and reliability of the test and test providers need to provide clear justification for the model they have chosen.

The process of examiner management involves a number of separate but interrelated stages. The stages outlines below are part of the examiner management system which Cambridge ESOL uses across the majority of its tests in test centres around the world. The stages are usefully referred to by the acronym RITCME, which stands for; Recruitment, Induction, Training, Co-ordination, Monitoring and Evaluation.

The recruitment stage is designed to identify suitable professionals who have the appropriate qualifications, experience and skills to examine Cambridge ESOL exams. Once identified the potential examiner attends an induction session. The induction phase allows the potential examiner to become familiar with Cambridge ESOL's credentials and profile, understand more about its assessments in general and about how they fit into the assessment process. The induction phase allows the potential examiners to be sure that they are willing to make the commitment necessary to be Cambridge ESOL examiners, and for Cambridge ESOL to be sure it wants to invest resource into the potential examiners.

The training and co-ordination processes are examination specific. The training includes detailed procedures for administering the test, in particular the role of the interlocutor, as well as advice on dealing with particular circumstances which may arise during the test and help on managing test materials. Following training, examiners can assess candidates for the examination for which they are trained. The co-ordination phase occurs at regular intervals after initial training and focuses on the assessments that the examiner gives. Examiners are standardised with examples of candidate performance at different levels and explanations of the characteristics of each level or performance band. Samples of ratings are taken from examiners, evaluated and recorded. If significant issues are found at this stage then examiners will not be allowed to continue to assess.

Regular monitoring of examiner performance takes place during live examining, through the examiner management system. A checklist is used to assess the examiner on their performance. Examiners can be rated as 'satisfactory', 'in need of further training or support', or 'unacceptable' in a number of areas. The scores are evaluated and decisions regarding the status of the examiner made.

The RITCME processes are implemented through a cascade system. The system is centrally managed and records kept on a central database. In this way examiner eligibility and performance can be monitored. The delivery of the training, standardisation and monitoring is local with senior team leaders and team leaders managing groups of examiners. This is illustrated below:

Figure 5: RITCME processes cascade system
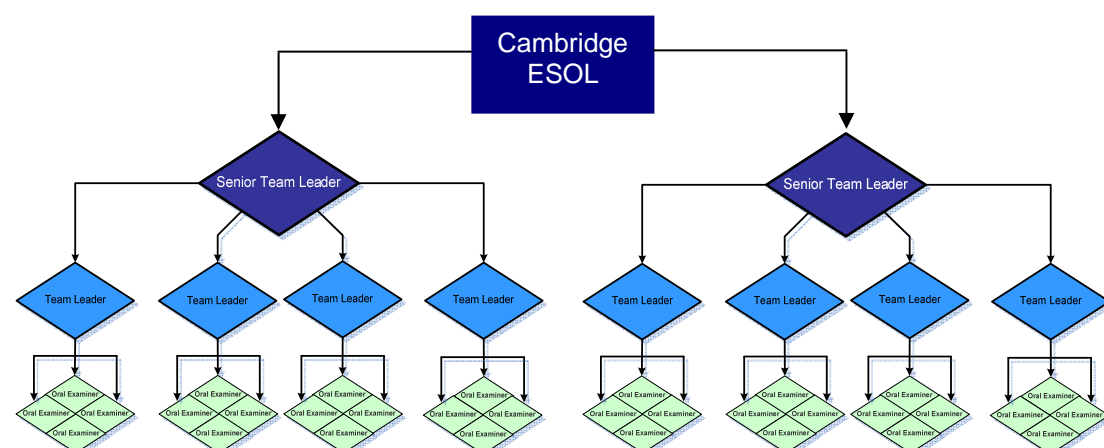
Security

Security of the delivery and storage of test material reduces the risk of tasks and items becoming known to candidates as does the use of multiple test forms. Multiple test forms also reduce the possibility of collusion where candidates co-operate to memorise test material. Internet forums for example may be used to exchange information about tests including their content. For computer-based tests alternative forms or an item pool approach are common ways to reduce the likelihood of tasks becoming known. Item pools also allow for the exclusion of items which behave in an unexpected way over time.

In addition to test material security, it is also essential that the identity of candidates is verified at the different stages of the test administration. This is likely to include registration, attendance at test administration session(s), results issuing and results verification by authorities. Picture identity is the minimum standard which needs to be implemented. If test reports are issued, security features need to be incorporated into the report or certificate. In addition to this, other biometric data such as fingerprints can be used to prevent identity fraud and impersonation.

Figure 6: Checklist Candidates

5. Candidates		
5.1	How is a candidate's identity checked?	
5.2	How do you ensure a candidate is not known to an examiner?	

Post Exam Processing

The output from the Examination Administration phase of exam scripts feeds into the Post Exam Processing phase. Sub-processes in this phase include; scanning of mark sheets, secure handling of scripts, marking, grading, the review of grades particularly at grade boundaries, issuing results and certificates and dealing with malpractice enquiries on results and remarking.

Figure 7: Core Process stages 3 & 4
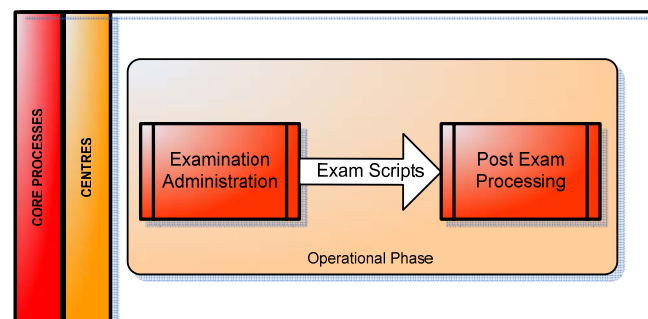
As has been discussed above, procedures need to be in place for ensuring examiners are rating candidates in a systematic and standardised format to ensure the delivery of the test is comparable for all candidates. Similarly, in the case of Aviation English tests the marking criteria, in particular for assessments of oral proficiency by interview, are guided by the level descriptors which have been determined and published by ICAO.

For listening tests, and for tests which use scoring algorithms there must also be a way of relating those scores to the aforementioned level descriptors. Pre-grading and standards fixing based on pre-testing or trialling is one common way of relating raw scores on a test to a standardised score. Test developers and test providers need to be clear on how cut scores are arrived at and how they relate to the scales. Issues related to item difficulty and item weighting need to be addressed. Do test providers claim, for example, that certain levels of proficiency are better tested by different parts of the test? How are these scores aggregated and reported?

Figure 8: Checklist Marking and Scoring

6. Marking and Scoring (where listening is assessed separately from speaking)		
6.1	How is the difficulty of items ascertained?	
6.2	Are certain task types weighted differently?	
6.3	How are the Listening items marked?	

Issues of test administration must also deal with reporting scores to the candidates and licensing authorities. What is the time frame for the delivery of results and how are they delivered? What are the procedures to challenge the result? These questions are reflected in the extract from the checklist below:

Figure 9: Checklist Processing Results

7. Processing Results		
7.1	How are results issued?	
7.2	How quickly are results issued?	
7.3	Is there an enquiries/remarking/appeals process?	

Exam Review and Evaluation

The diagram below shows the relationship between Post Exam Processing and Exam Review and Evaluation; the key outputs are results and certificates. The diagram also shows the feedback loop from review into improvements to the test delivery process.

Figure 10: Core Process stages 4 & 5
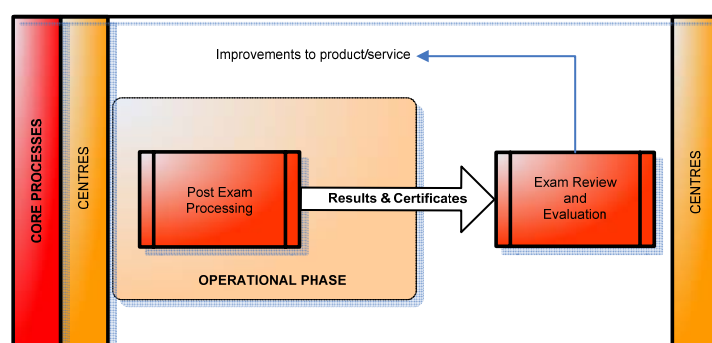
The key requirement of test providers is to ensure that the reporting relating to the evaluation of their test procedures is transparent. This can be affected by an open and consultative approach to test development and test implementation which is a characteristic of the test providers at the conference who were willing to discuss the development and delivery of their tests. Further evidence of this was provided by end users who provided case studies of implementation for the tests which revealed strengths and weaknesses as well as the desire for improvement and development of the testing organisations profiled.

Exam review and evaluation can take a wide range of forms but it should include the analysis of test sessions and the publication of data including grading data by region and by grade. Crucial to maintaining the trust of exam users is openness and honesty regarding the test process and the active involvement of stakeholders in improvements to the test delivery system.

Figure 11: Checklist Exam review and evaluation

8. Exam review and evaluation		
8.1	What analysis is carried out after a live administration?	
8.2	How are results publicised?	

Summary

When developing tests to meet particular purposes, the practical aspects of implementing test delivery and post examination processing should not be underestimated. These phases of the test delivery processes are critical to ensuring that tests remain valid and reliable measures of candidates' ability. The focus in this paper has been on ensuring the validity and reliability of the assessment of aviation personnel who are carrying out complex and involved processes. Indeed most often the requirement for them to use 'plain' English is in the context of the unexpected when standard phraseology is unable to effectively communicate the situation.

The test providers who attended the conference provided a great deal of evidence as to the measures they are taking to ensure effective delivery of their assessments. Similar scrutiny needs to be made of other tests which claim to be assessing English language ability as described in the ICAO scales.

University of Cambridge ESOL examinations

Cambridge ESOL examinations are the world's leading range of certificated exams for learners of English. Each year they are taken by over 2 million people, in 130 countries around the world. They include qualifications in general English, business and academic English and English for young learners. Cambridge ESOL also provides a suite of qualifications for language teachers which are widely accepted by employers in the English language teaching sector.

Cambridge ESOL is a founder member of the Association of Language Testers in Europe (ALTE), an international association with members representing 26 languages. ALTE is a (NGO) with Special Consultative Status with the Economic and Social Council of the United Nations and an (INGO) with participatory status in the Council of Europe.

Cambridge ESOL examinations are aligned to the Common European Framework of Reference (CEFR). The CEFR plays a central part in language and education policy within Europe, and increasingly in the wider world. Within Europe it serves policy goals of fostering linguistic diversity, transparency of qualifications, mobility of labour, and lifelong language learning. Beyond Europe it is also being adopted as a model for defining and assessing language proficiency levels.

The contributors

Angela ffrench

Angela holds an MEd in Language Testing from the University of Bristol, an MA from the University of Cambridge, and a Dip. TEFL.

Having been involved in teaching and testing English for 20 years, both in the state and private sectors, Angela joined Cambridge ESOL in 1991, with special responsibility for Speaking. Since 2001 she has been responsible for the management of a suite of examinations ranging from upper intermediate to advanced levels. In this capacity she has been involved in all aspects of developing and administering tests and assessment criteria, in the training of item writers, in pre-testing and analysing test material, and in the grading of the examinations.

Angela was a member of the PRICESG linguistic sub group which met in Montreal in 2005 to rate speech samples, and has recently become a member of the ICAEA board.

David Booth

David holds an MA in Education from the University of London and is an Associate Member of the Chartered Institute of Marketing.

David returned to the UK in 1998 to take up a post at Cambridge ESOL. He has worked across the business including roles in Research and Validation, Assessment and Operations and Business Management. He is currently the manager of the Business Management Unit which co-ordinates a number of marketing and business functions.

David has worked extensively overseas with the British Council as a teacher, teacher trainer and Director of Studies in Hong Kong, Malaysia and South Korea. David has extensive teaching, training and assessment experience and has delivered seminars and presentations to a wide range of audiences around the world.

Marjo Mitsutomi, PhD, and **Larry Platt,** PhD, University of Redlands (California)

BIO DATA

Marjo Mitsutomi has been affiliated with the University of Redlands for the past 10 years and currently serves as an Associate Professor in the School of Education. Her responsibilities include teaching graduate courses in second language acquisition theory and methods, pluralism in education as well as assessment and evaluation in higher education. Besides serving on various governance committees, she is part of the university-wide assessment team, the purpose of which is on-going assessment on campus in preparation for regular accreditation reports and visits.

Marjo has her Ph.D. in Applied Linguistics with a cognate in TESOL from Ball State University. A native of Finland, she is multilingual in Finnish, English and Japanese with a working knowledge of several other languages. Marjo began her language teaching career as a licensed secondary school instructor in English and German, but migrated to teaching English to non-native speakers of the language early on in her career. She spent six years at a Japanese university training teachers for the local junior high and high schools. For 10 years Marjo was involved in a FAA-certified flight school in Redlands, CA, as the Vice President of the business. She also earned a private pilot's license there specifically in order to understand aviation communication issues from the pilot's point of view. Together with her partner Kathleen O'Brien, FAA Safety Program Manager, the two received a national award from the FAA for pioneering "the English Project" as a safety initiative in aviation communication.

Marjo has served as the FAA linguistic consultant on various occasions and because of her aviation English expertise was appointed to the ICAO Price Study Group. Marjo's contribution there was to be a member of the team of specialists who wrote the ICAO English Proficiency Standard, ratified by the General Assembly in March of 2003. Currently, Marjo is interested in issues dealing with assessment and accreditation. She is a panel member for the bilingual teacher certification at the Commission on Teacher Credentialing, an organization which accredits all teacher training

----

Jerry Platt is a professor at the University of Redlands, in California. His areas of specialization include finance, statistics, and more recently geographical information systems. His most recent research focuses on collaborative work with Marjo Mitsutomi defining and analyzing the distribution of linguistically-isolated households in the U.S. Jerry also has been certified as a trained accreditation officer for both US and international agencies.Jerry came to Redlands as Dean of the School of Business, and as inaugural recipient of the Senecal endowed chair. He previously served as Dean at San Francisco State University, one of the largest and most diverse business schools in the world. He has been a senior executive at a Fortune 100 oil and transportation corporation and at a health services corporation, a consultant to numerous public and private organizations, and served as economic expert witness in more than 100 legal proceedings. Jerry was the first recipient of a doctorate in Public Administration at The Ohio State University, and earned a post-doctoral degree in Statistical Computing at Stanford University.

Aviation English Accreditation: *Why Bother?*

SUMMARY

In 2003 the International Civil Aviation Organization (ICAO), passed a resolution that designated English as the global standard for communication in international air travel, prescribed the level of mastery necessary to satisfy the standard, and specified the target population to which the standard would apply. The resolution delayed implementation and enforcement until 2008, providing time and reason for a plethora of education training programs and schools to emerge. We provide a context for understanding the potential value-added of developing an associated education accreditation process.

Appendix B provides internet-accessible references that provide additional support for the concept.

1. ## Introduction to Global Standards

> In my own profession, that of a seaman, the embarrassment arising from the many prime meridians now in use is very conspicuous, and in the valuable interchange of longitudes by passing ships at sea, often difficult and hurried, sometimes only possible by figures written on a black-board, much confusion arises, and at times grave danger. In the use of charts, too, this trouble is also annoying, and to us who live upon the sea a common prime meridian will be a great advantage.

Source: http://www.scribd.com/doc/2399636/International-Conference-Held-at-Washington-for-the-Purpose-of-Fixing-a-Prime-Meridian-and-a-Universal-Day-October-1884Protocols-of-the-Proceedings

The above passage by Admiral C.R.P. Rogers in 1884 nicely summarizes the need for a global standard for communication at sea, and nicely presages today's need for a global standard for communication by air. In air travel, there is a clear need for pilots and controllers to communicate with one another. The local language can suffice for a domestic flight, but a global standard is needed for international travel.

2. **Global Standards for Aviation English**

As a specialized agency of the United Nations, ICAO recognized the need for a global standard. The standard that emerged requires pilots to be proficient either in English or in each of the languages of countries over which they fly, whereas controllers must be proficient both in English and the local language. These rules assure continuous communication in a common language, and establish English as the global standard. The ICAO 2003 PRICE Study Group established benchmarks for proficiency.

3. **Current Status**

Scott Barrett suggests in his recent book (_Why Cooperate? The Incentive to Supply Global Public Goods_, Oxford University Press, 2007; p. 164):

> "Standards emerge whenever countries have strong incentives to coordinate.
> Global standards are much easier to supply than other global public goods. They
> do not need to be paid for, and they are easy to sustain once established. The
> problem with standards is _choosing_ them."

In the aviation sector, the choice has been made: English. However, unlike global agreement on a common time, agreement on a common language does require significant and varying costs, especially in countries where few natives speak English and the native language is not of Indo-European heritage.

So, we have an agreed global standard, a myriad of training programs to prepare pilots and controllers for a myriad of tests that purport to assess the ICAO benchmarks for proficiency in Aviation English, and a mandated timeline for implementation. We lack significant global enforcement mechanisms beyond the rather obvious risks from miscommunication in the air, resulting in meager levels of compliance.
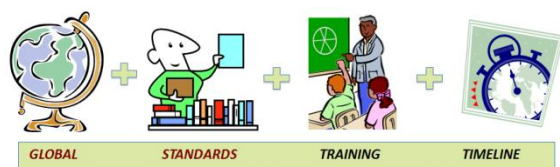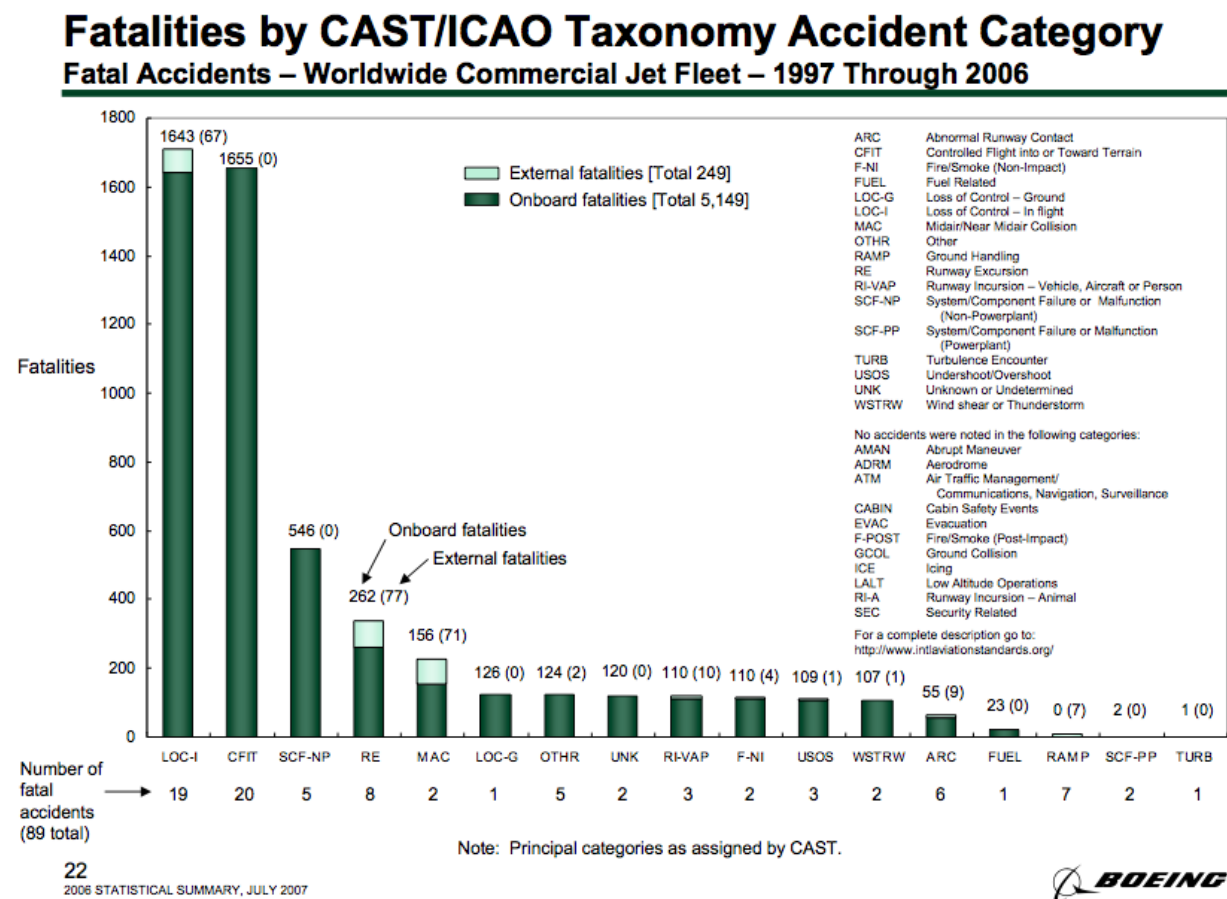

4. **The Value Proposition**

A common base for effective communication in global airspace is an important safety issue today, and will become even more important tomorrow. This statement implies that **data** supports the claim of an important safety issue today, and that **growth** patterns provide concrete evidence to support the claim it will become even more important tomorrow. This value proposition, once established, is compromised by the chaotic nature of today's training

environment for Aviation English. There is unlikely to emerge any coalescence around a single test instrument, and it is generally agreed that local variation in training programs is appropriate to accommodate the rich variety in cultures, native languages, learning styles, etc. If the value proposition is accurate, it best will be realized through a systematic accreditation process that aligns training programs with the ICAO proficiency standards.

5.   **Data**

There is substantial anecdotal evidence to suggest that language miscommunication has been a contributing factor in several airplane crashes, and in even more near-misses. Unfortunately, the coding mechanisms for recording cause of failure often obscure the role of language. Worse, there is evidence of deliberate withholding of such data, presumably to avoid increasing fears among an already skittish flying public. The result is surprisingly scant hard evidence to systematically support any claim that language communication "in global airspace is an important safety issue today". For example, consider the example below, reproduced from the Boeing web site:


Fatalities by CAST/ICAO Taxonomy Accident Category
Fatal Accidents – Worldwide Commercial Jet Fleet – 1997 Through 2006

Examining all fatal accidents during the past decade involving worldwide commercial jets, the report claims "no accidents were noted in … air traffic management, communications, navigation, (or) surveillance". The industry simple must become more pro-active in insisting upon honest and accurate categorization and reporting of all contributing factors to traffic mishaps.

6.   **Growth**

By contrast, there is substantial evidence to support the claim that "effective communication in global airspace … will become even more important tomorrow". In projecting air traffic for the year 2026, the same Boeing web site indicates that traffic within the Asia-Pacific region, with its great variety of native languages, will exceed air traffic within North America – the historically-dominant region that is comprised of but three native languages.
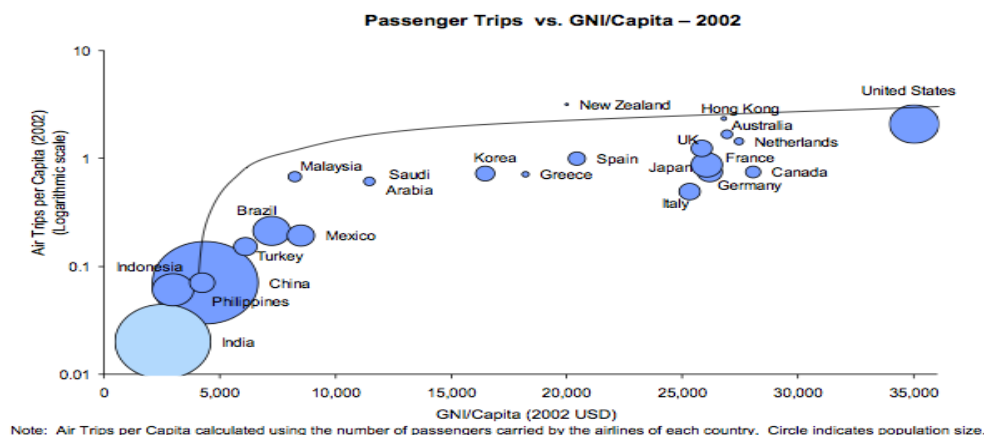

Air Travel and World Fleet
2007 to 2026
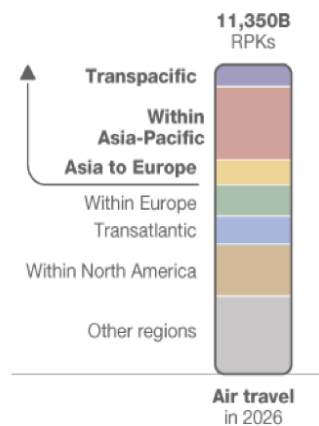
Additionally, there are clear indicators that per capita air travel traffic, which increases with wealth, is about to undergo dramatic shifts in demography. As the graph below shows, India and China are two very large countries (large circles) with relatively low per capita air travel (low circles); however, most prognosticators believe these two countries are experiencing large, sustained growth in wealth.

Air travel, which is correlated with wealth, has increased over the years with a rise in global GNI

Passenger Trips vs. GNI/Capita – 2002

Note: Air Trips per Capita calculated using the number of passengers carried by the airlines of each country. Circle indicates population size.

**Growth of Chinese Passenger Traffic ,
(2000 – 2005)**
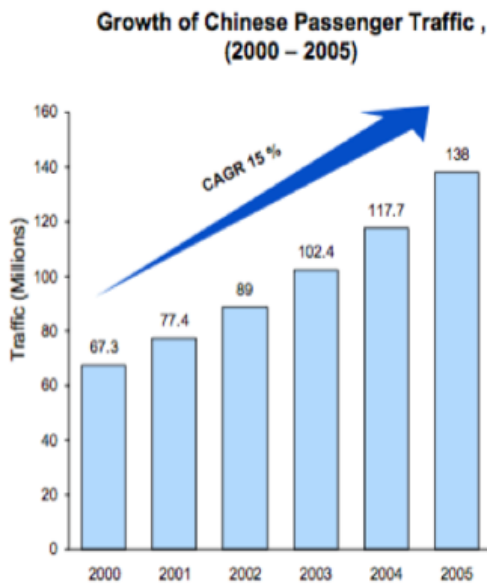
Growth Analysis

- China will be the fourth largest international market for leisure travelers by 2020, with about 100 million outbound Chinese tourists

- As per IATA, the region's current 23 percent share of the world's air traffic will grow to 50 percent by 2006

- *Route Analysis*
  – Total routes served by Chinese airline companies -1,200
    - Domestic – 975 (reaching 136 cities)
    - International – 225 (reaching 38 countries)

7.   **Accreditation**

The goal of accreditation is to ensure that the education provided by schools, institutions and programs meets the minimum acceptable levels of quality. The central theme of this paper is that the value proposition of Aviation English is best advanced through an accreditation process for its training programs. Just what is accreditation? Why does it matter? How can it help?

Ac-cred-it (trans. verb)   \ə-ˌkre-dət

Merriam-Webster:
   Etymology: Latin *accreditus,* past participle of *accredere* to give credence to, from *ad-* + *credere* to believe — more at creed Date: 1535

- **1:** to give official authorization to or approval of: a: to provide with credentials; *especially* : to send (an envoy) with letters of authorization b: to recognize or vouch for as conforming with a standard c: to recognize (an educational institution) as maintaining standards that qualify the graduates for admission to higher or more specialized institutions or for professional practice

- **2:** to consider or recognize as outstanding

- **3:** attribute, credit

The goal of accreditation is to ensure that the education provided by schools, institutions and programs meets the minimum acceptable levels of quality. Accreditation is "the stamp of

approval" for schools and/or programs in a particular discipline or industry. Institutions that seek accreditation recognize its importance by agreeing to a set of commonly adopted industry standards for quality assurance. The accrediting process is done by one's peers. Accrediting agencies are private educational associations of regional, national or international scope. The agencies develop evaluation criteria and conduct peer evaluations to assess whether or not those criteria are met. In terms of the so-called Level-4 proficiency benchmark, accreditation can address both general and specific questions of common interest, such as:

- *Who Decides One is at Level 4?*

- *On What Basis Does One Decide?*

- *Which Paths to Level 4 are Effective?*

- *Whom Can the Consuming Public Trust?*

- *How Can We All Continuously Improve?*

8. **The Education Specialty Analogy**

Accreditation is a common element of education systems all over the world. Some accreditation agencies are **general**, and assess overall performance, while others are **specific** to a particular task:

Institutional

- Institutional accreditation normally applies to an entire institution, indicating that each of an institution's parts is contributing to the achievement of the institution's objectives, although not necessarily all at the same level of quality.

Programmatic

- *Specialized or programmatic accreditation normally applies to programs, departments, or schools that are parts of an institution.*

Our interest is in a specific, specialized, programmatic accreditation. It would not assess the overall effectiveness of a flight school, an air traffic controller program, or an English language school. Instead, it would focus only on the Aviation English component of that school or
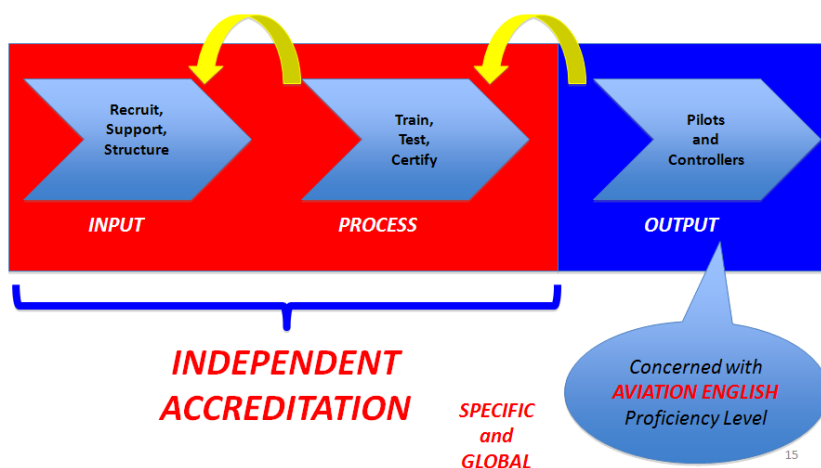
program, and assess it only with respect to the stated ICAO standards and benchmarks for Aviation English. An analogy may help.

There are thousands of universities around the world, and most countries have national or regional accreditation bodies to evaluate the overall effectiveness of the universities in meeting their educational goals. Many of these universities have schools of business as one patch of its educational fabric. There are only two global bodies for the accreditation of business programs within universities – EQUIS and AACSB. These two bodies accredit the business training subset of an institution, according to quite specific global standards that have been set by peers at business schools worldwide.

Negotiate:
✓Unifying Language of **Business**
✓Criteria for Recognizing Success
✓Reasonable Paths to Achievement

ACCREDITATION of Business Units within Colleges and Universities

Negotiate:
✓Unifying Language of **Aviation**
❖Criteria for Recognizing Success
❖Reasonable Paths to Achievement

ACCREDITATION of English Units within Schools Related to Aviation

9.   **Accreditation for Aviation English**

It is proposed that an international body be established to accredit schools and programs that purport to train and prepare pilots and controllers to be proficient in Aviation English by ICAO standards.

INPUT — Recruit, Support, Structure
PROCESS — Train, Test, Certify
OUTPUT — Pilots and Controllers

INDEPENDENT ACCREDITATION   SPECIFIC and GLOBAL

Concerned with *AVIATION ENGLISH* Proficiency Level

Necessary and/or Desirable Ingredients

The accreditation process should examine not only the measurement instrument(s) and output from the testing process for demonstrating proficiency in Aviation English, but also the inputs (recruitment; student support; organization structure; financial resources; personnel) and the processes (training; testing; certification). It is essential that accreditation be established as a holistic enterprise.

While it is both premature and presumptuous to specify standards today that should be set be agreement among peer schools and programs at some future date, there are certain ingredients that transcend specific criteria and can help guide in the formulation and establishment of the accreditation unit. It should be an independent, not-for-profit federation of schools and programs, with broad global representation that reflects the rich diversity in native languages and cultures across member states. Accreditation must be limited to Aviation English, and must be based upon a peer-review process that emphasizes absolute standards for performance while also recognizing relative advancement toward the absolute standard, thereby acting as a supportive organization that actively helps all programs improve.

- **INDEPENDENT, NOT-FOR-PROFIT Federation**
  - NOT ICAO, NOT ICAEA, NOT …
- **Accreditation at SCHOOL Level**
  - NOT at State level, Test level, or Individual level
- **Accredit Only the AVIATION ENGLISH PROGRAM**
  - NOT Overall Aviation or General Language Expertise
- **PEER-REVIEW Analysis**
  - Colleagues, NOT Bureaucrats or Regulators
- **IMPROVEMENT + OUTCOMES**
  - Both Relative and Absolute Assessments
- **FEEDBACK LOOP**
  - Supportive, NOT Punitive

Idealized Structure

A functional accreditation organization for Aviation English will require a small core of administrative leadership, preferably including expertise in both Aviation English and global business practice. There need to be both a Technical Panel of experts in Aviation English and a Stakeholder Panel to represent the interests of ICAO, pilots, controllers, trainers, test developers, manufacturers and other interests.
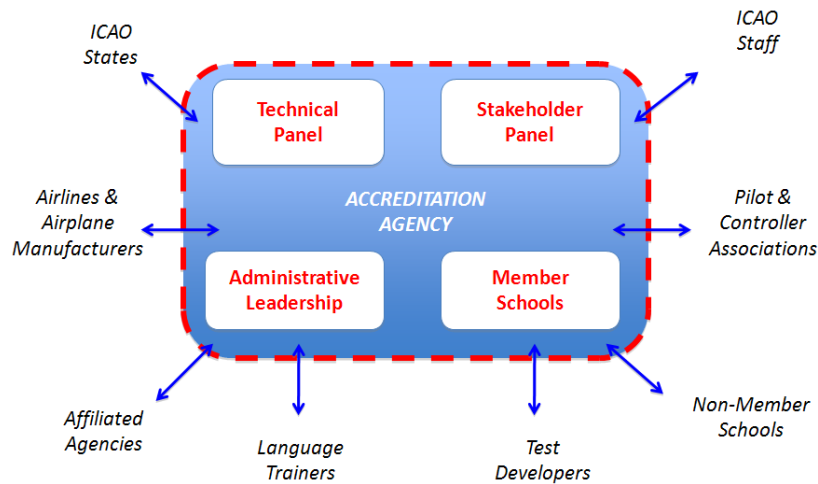
The actual accreditation process would proceed in two stages:

Stage I: A small set of Charter Schools are accredited by acclimation for three years

Stage II: These Charter Schools form a nucleus to begin the on-going accreditation process for programs or schools seeking:

- Initial Accreditation

- Re-accreditation.

Consistent with peer review, a condition for continuing active status is that each accredited program or school volunteer to participate in the peer review procedure for applicant programs or schools. To assure that the applicant has every opportunity to present their case, and to minimize costs, there will be heavy reliance upon written documentation supplied directly by the applicant, there will be use of videoconferencing procedures as appropriate, and there will be site visits as necessary.   Newly-accredited schools and programs contribute to restocking the peer review pool. Whatever the outcome, applicant schools and the peer-review team will be provided opportunities to provide feedback that can be used by the accreditation organization to continuously improve its processes and procedures.
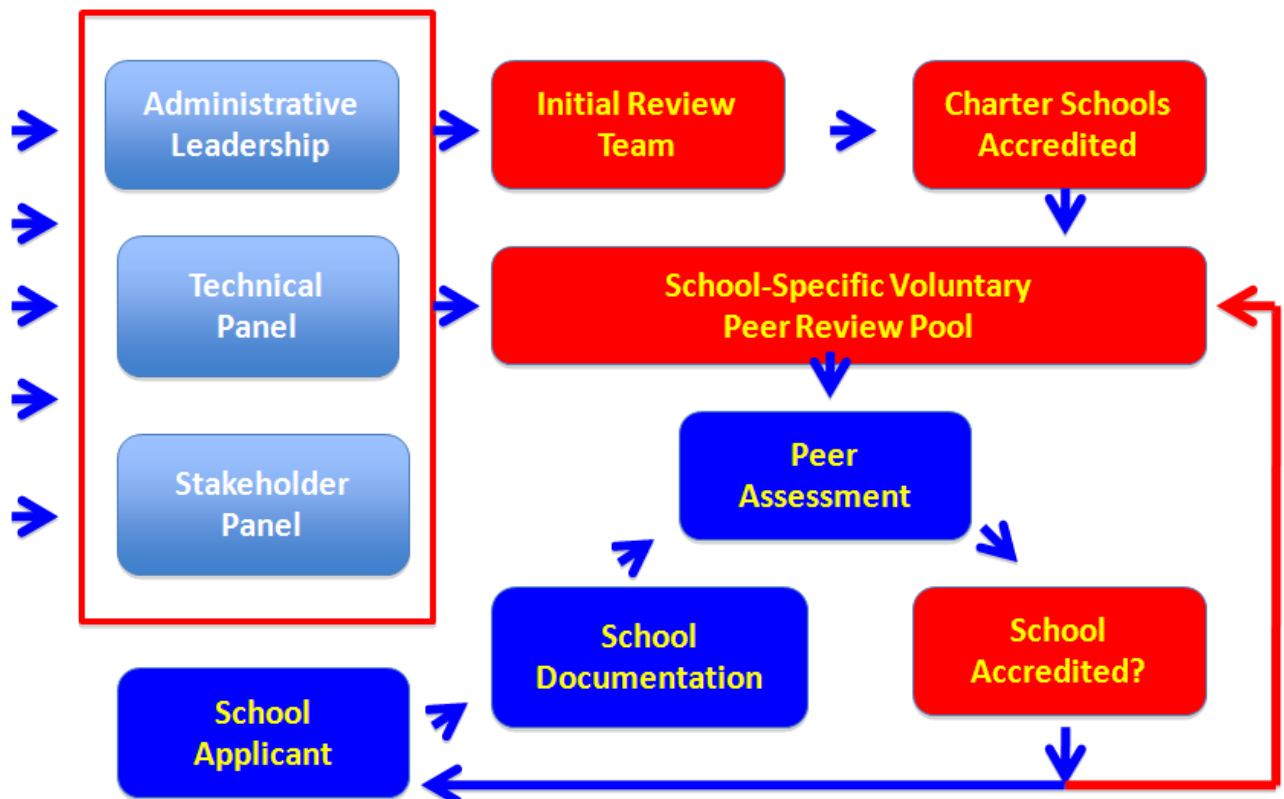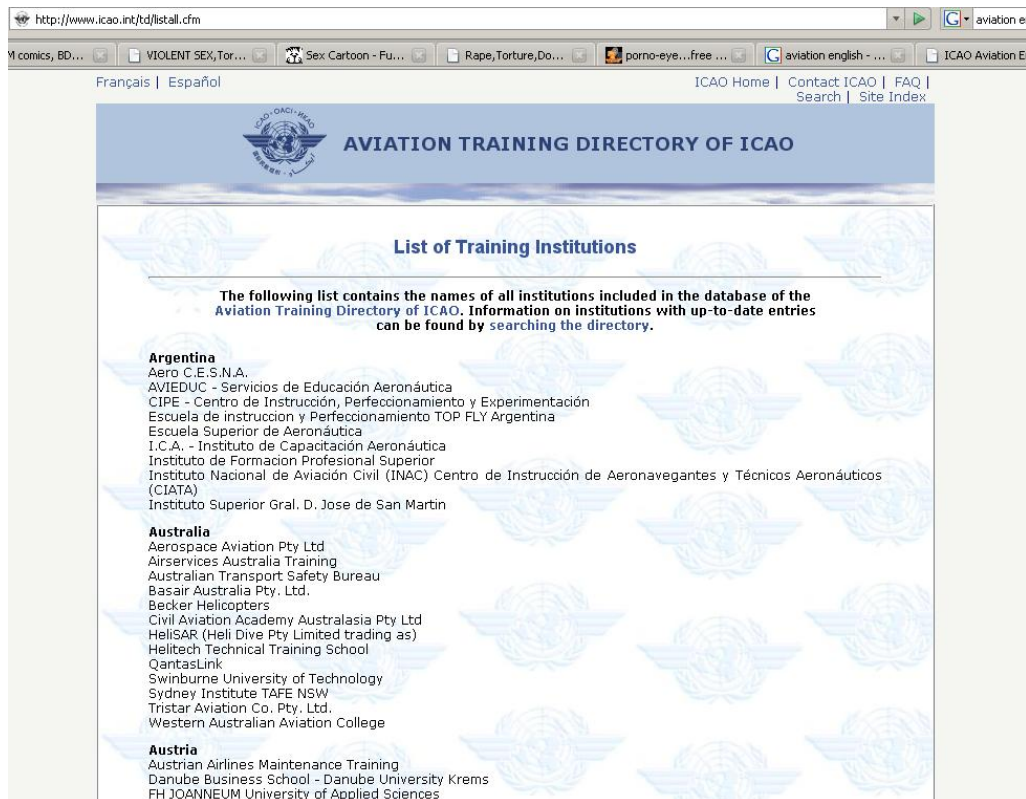
10. **Conclusion: Direction Matters**

There has been substantial progress over the past decade in developing and implementing a global standard for air traffic communication. However, the current state of affairs can be characterized as chaotic, inconsistent, and somewhat removed from the actual ICAO benchmarks for proficiency. Given nearly 200 member states, with very different needs, resources, and levels of preparedness, it is appropriate that there be multiple training paradigms and options, and that there be multiple testing instruments. What is missing is an accreditation organization that can help level the playing field, can keep all parties honest, and can protect the consuming public from the current glut of misinformation.

APPENDIX A: Consumer Beware?

Example 1

Example 2

Example 3

APPENDIX B: Internet-Accessible References

WHY BOTHER?

American Psychology Association -- Why Accreditation Matters:
http://gradpsych.apags.org/apr04/accreditation.cfm

WASC -- Why Accreditation?
http://www.acswasc.org/about_why.htm

Accreditation Commission for Health Care -- Why Accreditation is Important
http://www.achc.org/accreditation_importance.php

National Assessment and Accreditation Council -- Why Accreditation
http://naacindia.org/assessment.asp

UCEA -- Accreditation: A Detailed Explanation of Why it Matters
http://www.uceadirectory.org/accreditation.html

http://ezinearticles.com/?Do-You-Know-Why-Accreditation-Is-Important-In-E-Learning?&id=62976
Ezine -- Do You Know Why Accreditation Is Important In E-Learning?

http://www.insidehighered.com/news/2007/03/29/accredit
Inside Higher Ed -- Explaining the Accreditation Debate

South African National Accreditation System -- Why Accreditation
http://www.sanas.co.za/accreditation_why.php

ACCREDITATION AND SAFETY

USDA -- Rural Electric Safety Accreditation Program
http://www.usda.gov/rus/electric/engineering/2002/resap.htm

Washington Post -- Risk to Patient Safety Threatens Accreditation
http://www.washingtonpost.com/wp-dyn/content/article/2007/10/18/AR2007101801356.html

Psychiatric News -- Accreditation Group Raises Patient-Safety Consciousness
http://pn.psychiatryonline.org/cgi/content/full/39/20/12

National Rail Safety Accreditation -- SEE ATTACHMENT
www.transport.tas.gov.au/safety/rail_safety/rail_safety_-_accreditation

Language Differences as a Barrier to Quality and Safety in Health Care
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2078554

FINALLY,

I also think the AACSB standards may trigger some good thinking:
http://www.aacsb.edu/accreditation/standards.asp

Henry Emery, ICAEA Board

BIO DATA

Henry is a teacher, teacher-trainer and examiner of plain English for aeronautical communication. He is co-director of a language consultancy (Emery-Roberts Consulting), assisting organisations around the world with the implementation of the ICAO language proficiency requirements. Henry has just written an English language course book for pilots and air traffic controllers published by Macmillan.

In 2005, he sat on the ICAO PRICESG linguistic sub-group tasked with the development of the ICAO Speech Sample Training Aid. Henry is a vice-president of ICAEA and has particular interest in issues pertaining to the use of the ICAO Rating Scale for language proficiency assessment of flight crews and air traffic controllers.

A test rater accreditation project

SUMMARY

With few exceptions, language raters are required to assess the speech of pilots and air traffic controllers in tests used to measure language proficiency in aviation. However, raters around the world vary greatly in their interpretation of the ICAO Rating Scale and Holistic Descriptors. If the aim is for a standard level of language proficiency among air traffic controllers and flight crew, then a mechanism is required to standardise our understanding of Operational language proficiency.
Following on from previous papers and presentations on the subject of standardisation in language rating, the speaker gave an overview of progress to date with the Association of Aviation Language Raters project, described details of an IT platform for its administration and a method for standardising the rating of spoken language. He then outlined the challenges posed by implementing a system of rater certification and invited questions and comments from the audience in relation to these challenges.

https://www.polarisoffice.com/d/2RPZEqYr

PLENARY SESSION 6: **Presentations of feedback on test implementation experience**

1. Bozena Slawinska, PANSA, on using ELPAC in Poland

BIO DATA

Bozena Slawinska has worked in PANSA for almost 30 years. She has been teaching English to Polish student and operational ATCOs and other staff in her company. She is also involved in testing. She is a PELA test administrator and a certified RELTA and ELPAC rater. Bozena is a vice-president and the general secretary of ICAEA.

SUMMARY

My task here today is to talk about ELPAC test implementation in Poland. The ELPAC Test (English Language Proficiency for Aeronautical Communication) has been developed by EUROCONTROL (the European Organisation for the Safety of Air Navigation) to meet ICAO and European Commission language proficiency requirements in English for operational air traffic controllers.

Why ELPAC?
Before I start talking about ELPAC experience as such, I would like to answer the question some of you may want to ask, and that is why ELPAC and not a different test. The answer is very simple: because it was there.
Poland has been cooperating with Eurocontrol in the subject of testing since the early '90s. The first ideas about common tests for European controller were born in Warsaw during a visit of two gentlemen Wolfgang Leistico & John Rose who were then responsible for the PHARE training project for ATCOs from Eastern Europe. As a result the EPT – EUROCONTROL English Placement Test was created and is still used today for recruitment purposes. Then the idea of PELA was conceived developed and delivered. Poland, participated in the trialing of PELA and then in updating and trialing a new version. We never really managed to insert the whole of the PELA test into the training system but the listening part was incorporated into an ab initio training and was always seen as something very important, needed and relevant. With such a tradition we had to participate to ELPAC trialing.
Apart from wanting to participate in an international project we thought that it was a smooth way, of introducing a concept of a new test to the community of Polish ATCOs, well…. we were wrong.
Anyway, Taking the above experiences into account it was only logical that PANSA decided to use this particular test as a way to measure the proficiency of our ATCOs.
ELPAC was ready in summer 2007 and Poland signed the agreement with Eurocontrol in June. The training of markers and examiners started in the autumn. Poland trained 6 ATCO's to become markers and examiners in Paper 2 (interview) and raters in Paper1 (listening comprehension) and 3 language teachers (all with substantial experience in aviation English training) to become Examiners of Paper 2. All language teachers also became administrator of the test.

Implementing ELPAC
As soon as It was possible – that is when all our staff was trained by EUROCONTROL and accredited by Polish CAO, the testing in PANSA started on the 5th of November.
We have a self access room as a part of our training facility. There are 8 computers there, so it was logical to use this room for Paper 1 purpose.

At the peak time we examined 12 people a day 8 in the morning session and 4 in the afternoon. We started with the listening and continued with Paper 2. Then the computer room changed into the examination room.

One additional room was used so that oral exams could be carried out in two parallel sessions. All Paper 2 exams were recorded using small digital recording devices, which served us quite well.

Testing was done centrally. Smaller centers do not have facilities for testing. Not this time round anyway. Besides Polish controllers are used to this system because all training except OJT practice is done at Warsaw's ATC Training Centre.

At the moment we have just over 400 air traffic controllers. We tested 393 ATCO's   and 279 of them passed. That is just over 70,% of the test takers. 16 percent reached level 5, over 54% reached level 4 and 29% stayed below the minimum 4.

Test administration

The administration of the test was at times quite difficult because of the slowness of the internet connection. A lot of time was wasted because of that. Inserting candidate profiles, preparing or closing of the testing sessions, generally using the system is a bit complicated especially when testing is so intensive.

 A few times the system crashed during the testing session and on one occasion several times during one examination. But to restore it was fairly easy and not too disturbing for the test takers .They were warned about the possibility prior to each session.

We also had some other small technical problems but if in need we could always call Adrian Enright or technical support of Enovate in Norway.

Due to a shortage of staff, work load and some expected and unexpected absences just prior to the implementation, we were unable to run any familiarization courses. However the controllers were informed about the structure of the test, they were encouraged to access web site prepared by the ELPAC team containing the Sample test and the video of the Paper 2 exam.

The lack of regular familiarization training of ELPAC might be blamed for some percent of the level 3 results as 62 out of 114 failed only one paper and it was mostly Paper 1.

Feedback 1

A lot of students, especially those with the higher level of proficiency, were surprised to find their test easier than they expected. A lot of them thought the test was relevant to their job. As examiner I felt really well prepared for the job and the support we got from Eurocontrol and Enovate was really good. And one more thing. I've worked with Polish ATCO's for more years than anyone can remember. I taught all of them and I know the level they are at. In my opinion ELPAC is a very fair test, especially Paper 2

Feedback 2

The controllers generally felt it was a test relevant to their job but complained about the lack of a version specific to TWR/APP/ACC and just one sample version. A lot also considered Paper 1, listening comprehension, to be too difficult as it required good knowledge of the computer key board. Some of them criticized the lack of possibility to listen to the message again which they are so used to do in their job. They also felt there were too many questions in insufficient time.

Where are we today.

Poland filed a difference with ICAO, mainly because of the situation in the pilot's world, but also because Polish controllers, despite our efforts, were not fully compliant.

This put a totally new perspective on our efforts. But the work on improving the level of English of those who are below level 4 started in February when the first group of controllers went to England for a 3 week language course – first time in history.

We also bought the Skyguide program originally designed as a preparation for the PELA test and about a month ago people started to use it on a self access basis.

Any controller who feels ready for a resit - and also young people who are attempting to obtain the license - can take ELPAC during regularly organized sessions. Until the end of May these sessions are held every week.

As soon as this conference is over we will start to prepare the remedial in-house courses for those who need them.

Overall impression

The ELPAC project is an open project . Enovate, the software provider, launched an improved version of the ELPAC software system in January in which some of the administration tasks were improved. They also offer their help on the phone any time we need them. The same goes for Eurocontrol. New version of software is being prepared an will be available this year.

We have just returned from the 1st ELPAC User Group Meeting where we had a chance to discuss all aspects of the ELPAC testing, get rid of our frustrations and also discuss improvements that we would like to see in the test and in the system as a whole. Regular re-calibration courses for the examiners and markers are also planned by Eurocontrol so we get really strong support and supervision.

We do not have much time today but if you want to learn more about the test you can visit the ELPAC web site as well as contact me or any other ELPAC user.

PLENARY SESSION 6: **Presentations of feedback on test implementation experience**

2. Milina Kolomparova, EPST, Slovakia, on using TEA in Slovakia

BIO DATA

Milina Kolomparova has an MA in Psychology and English language and literature. She has worked as a psychologist for the Slovak Armed Forces for 5 years and as a part time teacher of English. She has been trained as a rater of aviation English and has worked as Aviation English Program manager for EPST Slovakia. She obtained accreditation for EPST Slovakia as TEA testing centre. She has received accreditation for testing of aviation English from the Czech, Slovak and Dutch CAAs.

PLENARY SESSION 6: **Presentations of feedback on test implementation experience**

3. Krzysztof Sysio, LOT, on using RELTA with pilots in Poland

BIO DATA

Captain Krzysztof Sysio is a Boeing 767 pilot flying for LOT Polish Airlines. He has general aviation experience as IR and chief instructor. He has been flying as an airline pilot for 20 years on Antonov 24, ATR 72 and 42 and Boeing 737 and 767. He has produced technical and aviation documents and worked as an official translator.

PLENARY SESSION 6: **Presentations of feedback on test implementation experience**

4. Christel Schipper, LVNL, on using RELTA with controllers in the Netherlands

BIO DATA

Christel has been working as a vice chairwoman of the exam committee for air traffic controllers for about 10 years.

In addition she is involved in different projects, like the implementation of the ICAO language requirements and since 2007. She is an Aviation English instructor and one of the 6 RELTA examiners at the LVNL (Dutch ATC).

SUMMARY

1. History
   In January 2006 a project team was formed to implement the ICAO requirements regarding the language proficiency.

   Team members were found within several segments of the training department and recruitment & selection.
   Our regulator was informed right from the start about all we did and therefore was able to approve the test we chose without any delay or discussion.

   The work was divided into 4 tasks:
   1 - a selection instrument to test plain English proficiency during the selection of ab initio's.
   2 - an Aviation English training program Initial Training.
   3 - an English language test for air traffic controllers according the ICAO requirements.
   4 - a training program for those controllers who did not meet the ICAO requirements after testing.

   Informing the controllers and gaining their support was one of the most important challenges for the project team.

2. Target group
   All the controllers and flight information service officers (FISO) from atc the Netherlands and two small regional airports, not operated by atc the Netherlands but for which we are responsible regarding training and competence. In total 300 people.
   All ab initio's who qualify for a student controller license. An average of 15 persons a year.

3. The test
   We did a thorough research for available tests meeting <u>all</u> the ICAO requirements.

Eurocontrol was one of the first that came into view. On the internet we found some tests and test developers approached us to sell their products.

Two serious candidates emerged from this lot.

The ELPAC test from Eurocontrol. We had the opportunity to participate in two trial sessions which were very useful and from which we learned a lot about testing. It is a high quality test with a solid scientific back ground.

The RELTA test from RMIT. One of the team members went to Melbourne and came back with a lot of information. This too was rated by the project team as a very good test.

Finally we presented a number of tests with the advantages, disadvantages and costs to the Operations management team.

The Management decided to go for the RELTA test because it caused the least strain on the duty rosters and could be administered by non controllers.

We then started the process of purchasing the test and the necessary equipment and the training of the examiners.

The rating is not performed by ourselves but by RMIT. We deliberately have the rating done by RMIT to exclude any discussion with the controllers about the results. We explained to them how the rating is done: by a team of professional raters.

4. The actual testing

On July 30, we started out on one of our regional airports, Eelde. Two examiners travelled to the northern part of the Netherlands and in a couple of days tested all controllers. In our personnel magazine "Behind the Screens" an article was published which served as an advertisement of how it was to take the test. During these first sessions we became more and more familiar with the test and the equipment.

In September through December the Enroute controllers had their turn at the RELTA test. The controllers at Beek and the FISO's at Budel passed the test in the last quarter of 2007. In 2008 all the Schiphol based tower-approach and ground controllers and FISO's will do the test. In March the FISO's from Lelystad airport took the test. We expect a small overrun into 2009.

5. Some numbers and results

6 certified examiners and more to be trained in the near future.

1 Program Manager with a back up.

Amsterdam En- route controllers: 79

Amsterdam FISO's : 22

Eelde Tower/Approach 13 and FISO: 6

Beek Tower/Approach: 16

Budel FISO's: 4

Lelystad FISO's: 6

Schiphol Tower Approach controllers: 18 (and many more to go)

Schiphol Ground controllers and FISO: 1 (and many more to go)

Ab Initio's: 5 (a continuing process all through the year)

Level 3: 1
Level 4: 11
Level 5: 92
Level 6: 96

Scoring range of the listening test: the majority has a score between 24 and 29.

6. What we think of the test and the work involved

We all believe it is a robust test and the results we've got so far are fair and correct. For the examiners it is fun to do, they get to know their colleagues from a completely different side then in a normal working situation. You create a totally new bond with a candidate, some are nervous and need to be calmed down a bit and especially in section 3, the interview, it is a challenge to make them talk about the aviation issues. You hear some really nice or interesting stories!

The work involved is not difficult but there are a lot of things that must be taken care of. The updating of the dongles, creating classes, enlisting students, planning the test and finding available examiners, uploading sound files and entering the listening test results in a database, distributing and logging the results. Etc.

The RMIT manuals are very useful and we've created an examiners checklist.

Another important thing that must be well organized is whether the test equipment is complete and working properly when we go to another test location. For instance we once forgot the adapter for the mixing panel, fortunately the technical department of the airfield had an adapter that fitted. Otherwise we could have gone home and messed up a duty roster for a whole day. That would have been a costly error.

Originally the speaking test is done with one laptop with the test and the recording software. We use two laptops, one with the test and the other one for the recording. We think it is more comfortable to keep an eye on the recording during the speaking test instead of switching from the test to the recording between the different sections. Sometimes the recording doesn't work that well but you'll find out after a whole section when you use only one laptop. With two you'll notice it immediately when there is something wrong and can take action to adjust the recordings.

7. Customer satisfaction

Our customers, the controllers, are quite happy with the RELTA test. Once they have taken the test their initial reluctance is gone and say they actually liked it. Especially the speaking test section 3. They sometimes grumble about the listening test but their overall performance is very well. (see 5) When they hear their marks for the listening test the grumbling ceases immediately!

We expect when people take the test for a second time and be more familiar with the test their hesitance or even anxiety will disappear completely.

All together I can state that the RELTA test is a good test, the results so far are just and fair, people like the test and feel comfortable when taking the test. Until now we received no complaints and nobody felt 'mistreated'.

Saturday 10 May 2008
MEETINGS ON EXPLORING IMPLEMENTATION:

1. Report on the ICAEA Meeting - A test rater standardisation and certification project (An Association of Aviation Language Raters)

Introduction

Henry Emery gave a presentation in plenary on the Friday afternoon entitled '*A test rater accreditation project*'. In his presentation, Henry outlined the factors which influence the work of language raters and highlighted the need for a standardised and harmonised interpretation of the ICAO Rating Scale and Holistic Descriptors. He moved on to propose a mechanism for international rater standardisation and certification and an on line system for its administration. (See http://www.icao.int/icao/en/jr/2008/6301_en.pdf for further information). He then touched upon some of the challenges that such a system could face, such as support and finance. The meeting on Saturday, chaired by Henry, was an open forum for discussion and debate following on from the presentation. What follows is a brief summary of the proceedings and outcomes of the meeting.

The meeting was attended by approximately 25 participants representing a wide variety of nationalities, cultures and professional backgrounds, including ICAO, ICAEA, former PRICE SG members, test developers, directors of training institutions, language teachers, language raters, academics, researchers, language consultants and pilots.

Discussion

The meeting was launched in earnest by referring back to the practical challenges relating to the implementation of a system of rater accreditation raised in the presentation on Friday. However, quickly coming to occupy centre stage on the agenda were the fundamental issues relating to the application of rating scale, the cultural, political and economic contexts in which raters work, and the feasibility of implementing a global system of rater accreditation.

The universality of the rating scales was the first major issue to be tackled. To what extent the work of a language rater is independent of a given testing system? One participant argued that rating is inextricably linked to the test instrument from which a speech sample is generated, and that it is therefore quite impossible for a rater to move easily from one test format to another. Conversely, it was argued that the rating scale is universal, and operates at a level above that of the test instrument. It was further maintained that while familiarity with a test format is essential to the work of a rater, the assessment of speech is governed by a common set of scales and criteria which can and should be applied equally to speech, and thus the skill of rating is transferable from one context to another. It was argued that deviations in the way the rating scale is applied resulting from peculiarities of a testing system will militate against universally accepted and understood levels of language proficiency.

One member of the group spoke of the experience of working simultaneously as a rater for two leading aviation language testing systems. In her experience, there was little difference in the way the rating scale was applied despite the differing nature of the testing systems, and it was suggested that the two tests produced similar and comparable test scores.

Following on from this was the issue of rater certification, and the question of whether raters can and should be certified to practice independent of a testing system. A parallel was drawn with the legal profession in the USA whereby practitioners of law are first required to pass a universal 'bar' examination before moving on to further professional examinations within the legal framework of an individual state. It was suggested that a system of rater certification could work in a similar way, with certification at a supra-national scale being the first step in the process of rater training, further training and familiarisation with a particular test instrument used in a region or within an organisation.

At this point in the meeting, a point of caution was raised. Given the current state of affairs in the implementation of the language proficiency requirements world wide, it was recommended that a system of rater certification at a supra-national scale be treated very carefully indeed. Many civil aviation authorities could see a system of rater certification as a 'fix' in the implementation of the language standards, and might gladly turn to an Association if it was felt that an Association would relieve the burden of rater standardisation and monitoring. This shift in the responsibility of implementation and oversight from governments to an Association is an undertaking which could have serious consequences for all those involved. It was argued that as implementation is the responsibility of the state, nothing should be seen to interfere with that sovereign responsibility whatever good intentions may be the driving force.

Where consensus was reached was on the view that aviation language testing is still in its infancy, and that it is uncommon in today's climate to find 'freelance' language raters who practice language rating from a number of testing systems as an independent professional activity in its own right. This view was supported by the evidence put forward in the meeting that many raters work for governments, public sector testing systems or for testing systems which operate within a fixed environment, and that raters are unlikely to migrate from one organisation or testing system to another or from one regulatory jurisdiction to another in significant numbers. It was felt that there is not yet the movement of skilled labour to create a demand for a top-down system of regulation and an internationally recognised rater qualification.

Discussions then moved on to the question of who a system of rater certification should target. It was suggested that rather than targeting individual raters, a system of rater certification should operate at the level of the testing system. It was proposed that test providers (test managers, senior rating team and rater trainers) ought to be the focus of rater standardisation and certification as a part of a broader test endorsement mechanism. That way, international standardisation would trickle down through the test management system to the raters operating within it. Furthermore, it was argued that an independent system of rater certification may undermine the work of established testing systems which may have in place effective systems of rater training, standardisation and monitoring.

On the other hand, it was questioned whether established test providers would benefit from a system of rater certification when they already have in place a management structure for rater training and monitoring. Furthermore, if the efforts of an Association of Aviation

Language Raters concentrated exclusively on working with established testing systems, then the Association would fail to reach the many raters working in isolation in different regions of the world without formal rater training and who operate outside of a carefully managed testing system.

Summary and conclusion

It became clear that to implement a formal, top-down system of rater certification on an international scale would be an extremely complex, politically sensitive and highly ambitious project requiring considerable resources of time, expertise and finance. Furthermore, it would require critical support and buy-in from regulators, governments, operators, academics and end-users in order to become a trusted, respected and therefore authoritative and effective system, buy-in that would take many years to secure and an authority that would take many years to develop. It was felt among the participants of the meeting that the initial framework proposed by the chairperson in the presentation on the previous day was perhaps premature, impracticable in the current climate, and too great an undertaking for ICAEA to pursue at this time. It was agreed that while it could be seen as a long-term vision, it was probably best left 'on ice', perhaps to be revisited in years to come.

Accepting, then, that the need for regulation was not yet driven by market forces (and, indeed, may never be), and that a top-down system of rater certification is impracticable at this stage, what guides the notion of international harmonisation in language rating? In whose interest would the work of an Association of Aviation Language Raters serve?

It was clear from the lively debate and passionate contributions to the meeting that in the broadest sense, ICAEA forum participants believe in the principle of self regulation, and that there would be great value in working together in the area of language rating in the near future. It was agreed that there ought to be *some* form of active international collaboration in the area of language rating, that activity should be at a relatively modest scale, that invitation to participate should be open, that participation should be voluntary, and that ICAEA has an important role to play in facilitating the exchange of expertise and understanding of language rating in the aviation context.

In conclusion to the meeting, the following ideas as areas of activity that might be pursued by ICAEA in the future were summarised:

The development of a profile of language raters around the world through gathering data such as biographical information, professional background, rater training, qualifications and experience, first language, employer, test instrument used and target test population. This would help ICAEA and the international community understand who raters are and the context within which they work, and would inform an approach to future rater standardisation projects.

Conducting a series of regional ICAEA rating workshops where test developers, raters, pilots, air traffic controllers and representatives of civil aviation authorities could be invited to share their rating expertise and experience on a voluntary basis. Participants could be invited to bring speech samples* to be rated by workshop participants for the purposes of standardisation. With agreement from the participants, data on the application of the ICAO rating scale, the performance of language raters and the speech samples themselves could be gathered and analysed with a view to producing future training and standardisation material.

For and on behalf of ICAEA, the chairperson would like to thank all the participants of the meeting for their time and enthusiastic contributions. We very much look forward to working together in this area in the future. If you have any questions comments or ideas on how to proceed, please contact henry.emery@icaea.net.

*speech samples which fit a pre-defined length and format

Saturday 10 May 2008
MEETINGS ON EXPLORING IMPLEMENTATION:

2. Training accreditation

Background and rationale for session:

The session was designed and led by Dr. **Mitsutomi** and Dr. **Platt** in order to expand on ideas and issues related to accreditation in the aviation English context. The discussion began with a summary of the main points of the conference presentation entitled "Aviation English Accreditation: Why Bother" and then proceeded to an exploration on existing vehicles and the potential development of new ones that might be of assistance in establishing such a quality control agency.

"Aviation English Accreditation: Why Bother" makes the case that all stakeholders in the growing aviation English industry would benefit from an accrediting agency. As ICAO member states continue to contemplate how they might meet the ICAO English proficiency standards in their own context, the question of how to judge the validity and integrity of an aviation English training program, whether private or public, becomes an imminent issue. Currently, there is no organization which is international, politically neutral and non-profit with the required qualifications and credentials to make such determinations.

The purpose of this session was to facilitate a discussion on the potential benefits of establishing an accrediting agency to oversee aviation English training programs. The facilitators sought both advice and information on the topic from the attendees in order to determine whether such an undertaking would be a desired and feasible outcome of the deliberations.

Discussion highlights:

The question of whether to accredit tests before training programs resulted in a spirited debate of the pros and cons of each.   After considering all sides of the issue, the majority of the participants leaned on the tentative conclusion that testing is still such a divisive topic that the accreditation process should begin elsewhere. Also, it was determined that sole focus on testing may encourage the broader issues of English education programs to be overlooked, which in the end contribute to candidates' success in passing tests, whichever ones they may be.

It was collectively agreed that the single most important reason for an accrediting agency is for the public at large to have the means to learn about aviation English education programs. Students, for example, need to know the qualifications of programs from among which they may choose their own. Civil aviation authorities benefit from publicized information on the credentials of various programs. Similarly, employers need to know whether an institution, school or program delivers the quality it promises in its advertisements. Although many programs claim to help candidates meet the ICAO proficiency standard by either providing the necessary training for it or by demonstrating their candidates' proficiency level by their test scores, there is no mechanism in place to investigate such claims or to inform the interested parties of the validity of such claims.

Some of the components of an aviation English education program to be considered for evaluation are as follows:

- Vision

- Financial resources

- Student pool/recruitment practices

- Student support

- Facilities

- Faculty qualifications

- Curriculum

- Outcome measures

- Job placement in the industry

The above list is far from complete but serves as an example of the factors about which an accrediting body would collect information to create a comprehensive narrative about the qualifications of a training program.

Potential outline for the accrediting process:

The conceivable process for creating an accrediting agency was determined to be a rather complex process which would require seed money as well as expert non-partisan leadership. Assuming that the two fundamental pieces were to be accomplished, the process for creating the agency becomes a concept that can be both realized and implemented.

The steps for implementation begin with assembling a governing board, which should consist of representatives of the various stakeholders as well as a host of experts from various fields, not excluding members of other pertinent aviation associations and organizations. One of the first functions of the board would be to ratify a set of standards to serve as the framework for aviation English education and institutions seeking accreditation. As accreditation is always a voluntary process, a representative sample of existing schools would be requested to volunteer as the first entities seeking accreditation. The process over time would contribute personnel to a peer review pool. Over time also, a critical mass of participating schools will be established, at which point the credible benefits of being an accredited entity become visible to the public at large.

Preliminary findings:

The exploratory session on establishing an aviation English accrediting agency was affirming of the concept in theory. Although the facilitators and participants shared many of the same questions, the overall sentiment was to begin to take steps in that direction. Mitsutomi and Platt offered their services in spearheading the exploratory phase of the project, which was deemed to be worthwhile and timely in the current political climate which surrounds aviation English testing and training programs world-wide.

EXHIBITORS

1. **Anglo-Continental**

2. **Edgewater**

3. **MacMillan Education**

4. **OISE**

5. **Oxford University Press**

6. **Pearson Ordinate**