Apr 24th, 11:05 AM - 11:20 AM

Commercial aviation language testing: Barriers to success? A case study

Henry Emery
Latitude Aviation English Services, UK, henry@latitude-aes.aero

Commercial aviation English testing: barriers to success? A case study

Henry Emery

Latitude Aviation English Services, Plymouth, UK

Author note

Henry Emery is Managing Director of Latitude Aviation English Services Limited.

Correspondence concerning this article should be addressed to Henry Emery, Latitude Aviation English Services, 38 Edith Avenue, Plymouth, Devon, PL4 8TJ, United Kingdom.

Contact: henry@latitude-aes.aero

Abstract

In 2008, in response to the International Civil Aviation Organisation (ICAO) Language Proficiency Requirements (LPRs), the flight training organisation Oxford Aviation Academy and the aviation English service provider Emery-Roberts set out to jointly develop the English Test for Aviation (ETA), an English Language Proficiency test for commercial pilots. The ETA was launched in 2010 and was the first test to achieve conditional ICAO endorsement in 2012, though ceased operations in early 2013. Section 1 of this paper describes the development of the ETA and the business plan for test roll-out. Section 2 describes and discusses the regulatory, commercial and practical challenges that the owners of the ETA faced with the development, maintenance and distribution of the ETA which eventually led to the withdrawal of the ETA from the market in 2013. Then, taking the ETA as a case study, section 3 discusses the state-of-the-art of aviation language testing, and presents the case that, in the current climate, weak regulation allows poor testing to thrive and commercial interests to prevail over test quality with the result that validity and reliability remain worryingly absent in aviation language testing today.

Keywords: Aviation English, aviation language testing, ICAO Language Proficiency Requirements, test validity and reliability, aviation regulation, language assessment policy and politics

1 Test development and roll-out

1.1 Background

The flight training organisation Oxford Aviation Academy (OAA) and the aviation English service provider Emery-Roberts (ER) elected to joint develop an aviation English test for pilots for a number of reasons:

Firstly, OAA, a well-respected aviation training organisation, trained student airline pilots at its ab-base at London Oxford airport. OAA's student intake comprised both domestic and international students training under self-sponsored and airline-sponsored flight training programmes. Under United Kingdom Civil Aviation Authority (UK CAA) regulations, OAA's flight examiners were permitted to endorse a pilot's licence for language proficiency at level 6 (expert) according to the International Civil Aviation Organisation (ICAO) Rating Scale during flight examinations, with those students not achieving level 6 required to undergo formal language testing with an approved Language Assessment Body. As such, OAA had an internal requirement to provide language testing for its students as part of the pilot training and licensing process.

Secondly, OAA's base in Oxford was part of a network of flight training centres serving individuals and airline customers worldwide. With a strong reputation, global access to market and the infrastructure necessary to distribute a language test internationally, OAA felt that a language test capability would add value to its service portfolio.

Thirdly, with an existing partnership for the provision of aviation English training, OAA and ER felt they had the resource and combination of English language and aviation subject matter expertise necessary to develop a specific-purpose aviation English test for pilots.

Fourthly, in 2008, OAA and ER conducted market research which included an appraisal of the aviation English tests available at the time, and found that there were no tests available that offered

evidence of test quality commensurate with the high-stakes, safety-driven aviation English testing context[1]. Indeed, in a survey of aviation English tests in the same year, Alderson concluded that 'we can have little confidence in the meaningfulness, reliability, and validity of several of the aviation English tests currently available for licensure' (Alderson, 2008, p1).

OAA and ER began developing the English Test for Aviation (ETA)[2] in the autumn of 2008. At the time, ICAO guidance on aviation English testing was as follows:

> A statement of evidence for test validity and reliability should be accessible to all decision-makers, in plain, layperson language (ICAO, 2010, section 6.3.3.1); and
>
> A description of the development process … should be accessible to all decision-makers (ICAO, 2010, section 6.3.2.2).

Thus, OAA and ER began developing the ETA using standard procedures in test construction with a view to honouring ICAO guidance.

1.2 Test development

In late 2008, a review of the language testing literature was conducted with a particular focus on performance and task-based testing and the testing of speaking and listening. Then, with the ongoing support of language testing consultant, Dr Rita Green[3], the test and task specifications, and examiner and test taker instructions and rubrics were developed. A production team of aviation Subject Matter Experts (SMEs), English Language Experts (ELEs), audio engineers, voice actors, flight simulator engineers and graphic designers was formed, and prototype tasks and associated audio-visual content were developed and piloted on a small population of flight instructors and student pilots in the advanced stages flight training at OAA. Following test piloting, the test documentation was revised, and marking keys and

[1] For the purposes of this paper, test quality refers to validity and reliability. In essence, validity is the extent to which a test measures what it sets out to measure, and reliability is the consistency of measurement.
[2] The ETA had three parts: a radiotelephony role play task, an integrated listening and speaking task and a picture description task with a follow up discussion. The test took approximately 35 minutes, was administered with a single candidate by an English language expert and an aviation subject matter expert and was designed to measure at ICAO levels 2-5+.
[3] www.test-development-training-analysis.com

procedures for test administration and assessment created. In early 2009, a plan for test trials and subsequent analysis of test performance was developed and in mid-2009, OAA and ER began seeking candidates for test trials. Conducting test trials is a major hurdle for anyone developing a test and yet it is essential: 'However well designed an examination may be, and however carefully it has been edited, it is not possible to know how it will work until it has been tried out' (Alderson et. al., 1995, p73). As per standard practice in language testing, OAA and ER needed to trial the ETA on an adequate number of candidates who were representative of the target test population in terms of characteristics and language proficiency level. This meant finding:

1. Candidates who did not have English as a first language;
2. Candidates who presented a wide range of language ability from low level users (ICAO level 2) to high level users (ICAO level 6); and
3. Candidates who were:
    a. Student pilots at the late stages of advanced commercial pilot training (multi-engine, instrument rating); and
    b. Licensed commercial and airline pilots.

The challenges were twofold. On one hand, it was difficult to find enough pilots to participate in test trials. The stream of students completing airline pilot training at OAA's base in Oxford, and the stream of airline pilots passing through OAA's Type Rating Training Organisation (TRTO) at London Gatwick airport were steady but relatively small in volume. Even if one were able to persuade such individuals to give up their free time for a test trial, it would have taken many months and considerable expense to conduct the necessary number of trials on candidates from these streams. On the other hand, it was difficult to find pilots via OAA's network in the UK presenting the low and intermediate levels of language proficiency targeted by the ETA (ICAO levels 2, 3 and 4). The majority of the international students at Oxford presented strong language proficiency on account of the fact they had already spent approximately 18 months immersed in an English-speaking professional training environment. Similarly,

those airline pilots passing through the TRTO at Gatwick also presented strong language proficiency. Furthermore, even when candidates with lower-level language proficiency were identified, it was difficult to persuade them to participate in test trials. In a regulated environment where licences (and possibly training success or employment) are contingent on language proficiency at or above ICAO level 4, there is a lot at stake for candidates with lower levels and therefore little incentive to disclose their language proficiency in a test trial. Fortunately, after some months, OAA and ER found an airline partner in Russia with whom to cooperate for test trials. The airline kindly agreed to bring a number of line pilots to its base in Russia over a two-day period during which trials could be conducted. Trials were scheduled for March 2010. In late 2009, the full test blueprint - multiple versions of parts one, two and three of the test - was produced to specification by the production team.

In early 2010, the existing test documentation was compiled to create the ETA Test Manual which was harmonised with OAA's aviation training organisation Quality Manual and submitted to the UK CAA for inspection. In March, one SME and one ELE conducted 4 trial tests at OAA's base in Oxford before travelling to the Russian Federation to trial the complete test blueprint on the 27 pilots presented by the airline partner. Pre-and post-test questionnaires were administered to the trial candidates and a smaller number of post-test interviews with candidates were conducted. In the same month, the author delivered a peer-reviewed poster presentation at the International Language Testing Association Language Test Research Colloquium at the University of Cambridge on the work conducted on the ETA to date (Emery, 2010). In April, the UK CAA approved OAA as a Language Assessment Body.

In June 2010, a four-week standard-setting session was conducted with a team of five SMEs and five ELs. ICAO Rated Speech Samples (ICAO, 2006) were used for calibration before the team turned to table-top rating and discussion exercises using videos and audio recordings of trial test performances. The remaining trial test performances were rated by the team according to a pre-defined model, the data from which, along with candidate questionnaire responses, were analysed by Dr Ute Knoch at the University of Melbourne. Later in 2010, an expert judgement session was conducted in which 32 independent pilots, air

traffic controllers, language teachers and language testers watched videos of trial test performances and made judgements about the extent to which the test engaged the language abilities set out in the test specifications. The data from expert judgement were analysed and published, along with a history of test development and results of analysis of test trials and standard setting, in a report that was available for download from the ER website. It is believed that ER was one of two organisations in the world ever to have published information about the performance of their test in the public domain. In November, the author gave a talk about developments in testing language for specific purposes with reference to the ETA at the Language Testing Forum at Lancaster University, a talk which was eventually published as a paper in the peer-reviewed Language Assessment Quarterly (Emery, 2014).

1.3 Business plan

In late 2010, OAA and ER officially launched the ETA and began marketing and distributing the test. The business plan included OAA and ER staff providing direct testing services via OAA group training centres, as well as licensing the test to third party organisations such as regulators, airlines and aviation and language training organisations for a fee per candidate. All test activity was monitored from a central administration unit at OAA's base in Oxford. Test centres were set up and by 2011, there were 4 test centres running in the UK, Spain, Cameroon and Hong Kong.

The price of a commercial aviation English test at the time ranged from £80[4] to £160, so to compete effectively, the fee to take the test at OAA was set at £145 per candidate. The licence fee for third party use of the test was set at between £70 and £95 depending on candidate volume.

In the aviation English testing market, competitors were reporting accumulated candidature as >23,000 (Australia), >10,000 (Belgium) and >8,000 (UK). In order to recover development costs, the ETA needed to attract approximately 3,500 candidates in the five-year period to 2015.

---

[4] All financial figures are approximate and are quoted in British Pounds Sterling.

1.4 ICAO endorsement[5]

Earlier in 2010, ICAO acknowledged that 'disparities in the quality and appropriateness of language proficiency testing have been identified as obstacles to the effective implementation of ICAO's related safety objectives' (ICAO 2010b, p30) and announced a new collaborative endorsement process for aviation English tests. In November 2011, after three years of work and approximately £125,000 of investment, ER submitted the ETA to the newly-opened ICAO Aviation English Language Testing Service (AELTS) and in March 2012, the ETA became the first test to receive an endorsement from ICAO:

> ICAO would be pleased to identify your test as conditionally endorsed on its website and later as fully endorsed when changes are made. (ICAO test evaluation report on the ETA)

To summarise the conditions of endorsement, ICAO were concerned with the 'limited trialling used to confirm [the test] meets key criteria for validity and reliability' (ibid). Under the terms of conditional endorsement, OAA and ER were requested to 'prepare and submit a test improvement plan to ICAO and work with ICAO, as needed, to make the necessary changes … [addressing] all issues noted in the evaluation report and propose specific corrective actions leading to measureable outcomes' (ibid). OAA and ER agreed to the conditions offered by ICAO including a 12-month period in which to conduct this work. In May, with Dr Luke Harding of Lancaster University, OAA and ER drew up and submitted an improvement plan to the AELTS which was accepted in July 2012. OAA and ER also established a relationship with the university through which students on Lancaster's MA in Language Testing by Distance programme used ETA test data to research test quality. Khadijah Jilani looked at the relationship between task specification and observed candidate performance in part 3 of the ETA for her MA dissertation in 2012 (Jilani, 2012).

---

[5] Initially, the ICAO AELTS offered two levels of endorsement: 'conditional endorsement' and 'full endorsement'. ICAO later changed the wording to 'partial recognition' and 'full recognition' respectively. For the purposes of consistency, the word 'endorsement' is used in this paper to refer to what was 'endorsement' and is now 'recognition'.

The ETA's owners were committed to working with ICAO and conducting the necessary trial and revision work to achieve full ICAO endorsement. At the same time, the owners were mindful that for the test to succeed, it was crucial to develop business. But they were confident: they had a well-developed aviation English test with demonstrable evidence for quality, a test with competitive advantage as the first in the world to be endorsed by ICAO and one which, for some months, held the unique status of being the *only* test in the world to be endorsed by ICAO. The test had good access to market through OAA's strong flight training brand and international network of flight training centres. The owners felt that the test had pricing which was competitive and commercial targets that were achievable. And there was genuine interest from leading airlines. Given the test and the regulatory and commercial context in which the test was launched, one might have predicted commercial success. However, the ETA was beset by insurmountable regulatory, commercial and practical challenges to which this paper now turns.

2: Regulatory, commercial and practical challenges

For an aviation English test to be valid for personnel licensing in a given state, the Test Service Provider (TSP) needs approval from the National Aviation Authority (NAA) of that state. Without approval, test scores cannot be used for license issue or renewal. This was true in all 190 ICAO member states, regardless of whether a TSP had ICAO endorsement or not. In spite of the evidence for test quality provided by the test owners, and in spite of the ETA's ICAO endorsement, OAA and ER were unable to gain national regulatory approval in many key ICAO member states. This represented the single most significant barrier to market entry for the ETA. The reasons for this were many and varied.

Firstly, the time and costs associated with achieving regulatory approval in multiple states were significant. The nature of the approval process varied considerably from state to state. In some states, approval was a straightforward administrative procedure that was free-of-charge. In several cases, presentation of the ETA's UK CAA approval certificate alone was sufficient evidence for some NAAs to accept ETA test scores. As the UK is an English-speaking country with a generally well-respected

regulatory authority, a UK CAA-approved aviation English test was perceived as a mark of certain quality. In Europe, as the UK was a member of the Joint Aviation Authorities (JAA[6]), many (but not all) JAA member states accepted ETA test scores under the principle of mutual acceptance with no further requirement for NAA approval. However, in other states, approval was a lengthy, detailed and sometimes very expensive process. For example, a large national airline wished to use the ETA. In addition to submission of the ETA Test Manual and payment of approval fees, the NAA required a two-day site visit to OAA's base in Oxford for two flight standards and personnel licensing inspectors. Stipulated in the inspectors' travel costs were business class flights and 4-star hotel accommodation as well as time charged at a daily rate. The total cost of approval came to approximately £10,000. For both the potential airline customer and for OAA and ER, this cost was prohibitive.

Secondly, many regulators would not engage in discussions about approval at all, effectively blocking OAA and ER's access to market. NAAs in two key markets were unable to approve the ETA because they had existing contracts with TSPs which precluded approval of others, effectively creating monopolies in both states. The quality of the approved tests in use would not feature in discussions. In one large state, OAA and ER had a business relationship with a local partner with potential for a network of six test centres across the country, and two large airlines and the airline pilots' union showed serious interest in using the test. But in spite of the ETA's ICAO endorsement, the NAA simply would not engage with the owners of the ETA, and instead continued with the exclusive use an aviation English test that it had developed itself, a test that had no evidence for validity and reliability. While the NAA cited a national regulatory framework that would not accommodate approval of third party aviation English TSPs, the ETA's owners sensed that the NAA was protecting its own interest rather than pursuing the effective implementation of the ICAO LPRs. The experience was similar with NAAs in two other large ICAO member states. In both states, language proficiency was regularly featuring in the national media,

[6] The Joint Aviation Authorities was body representing the civil aviation regulatory authorities of a number of European states. The establishment of the European Aviation Safety Agency (EASA) created a Europe-wide regulatory authority which has since absorbed most functions of the JAA (in the EASA Members states).

in one case, newspapers reporting the pilot community's serious dissatisfaction with testing services provided by the NAA, and in the other, reports of concern about the language level of licensed pilots. In spite of this turmoil, and in spite of the fact that the tests used in these ICAO member states provided no evidence for test quality, neither NAA would consider using the ETA. Again, test quality would not feature in the discussion.

The generally low level of knowledge of aviation English testing, of test development and administration, and of test validity and reliability presented a significant challenge for the owners of the ETA. In an 'off-the-record' conversation, a personnel licensing inspector at one NAA admitted that they didn't have any of the expertise necessary to evaluate the quality of a language test. However, the same NAA continues to charge a fee of approximately £10,000 for the first three years of approval for a 'service' to industry which is little more than an administrative exercise which completely ignores – by the NAA's own admission – basic and fundamental evaluation of test quality. In the same 'off-the-record' conversation, the ETA test owners suggested that the NAA adopt ICAO endorsement as an approval standard to overcome the obvious knowledge gap. The response from the NAA was a firm 'no'. Essentially, the NAA feared pushback from their approved TSPs, some of whom were rumoured to have submitted their tests to AELTS and failed to achieve ICAO endorsement.

In 2010, ICAO acknowledged the problem of poor assessment literacy:

> Many States lack the expertise and resources to implement the ICAO guidance established to assist them to oversee, select or develop appropriate tests. (ICAO 2010b, p30)

This knowledge gap prompted ICAO set up the AELTS to support NAAs in selecting quality tests:

> The primary goal of the new testing endorsement initiative will be to provide a pool of testing systems of appropriate design and content, and which additionally meet well-defined standards of good practice from which States can then choose. (ICAO 2010b, p30)

However, for the owners of the ETA, ICAO endorsement failed to have any positive impact on discussions with NAAs about test approvals. The most favourable responses the test owners received from NAAs regarding the ETA's ICAO endorsement were platitudes about the general desirability of ICAO endorsed tests, and applause of the test owner's commitment to quality. Yet sadly, in no instance did ICAO endorsement actually facilitate the process of NAA approval. In fact, as alluded to above, the ETA owners suspected that reverse may have been true: some NAA's may have perceived an ICAO endorsed test provided by a third party as a potential threat to their sovereign testing and personnel licensing responsibilities.

In the process of achieving approvals, OAA and ER also encountered NAAs that were not even aware that ICAO had launched the AELTS to help states with test selection and approval. Of those that did, it was often the case that the NAA did not understand what 'conditional endorsement' and 'full endorsement' actually meant. When the ETA was submitted to one NAA, the head of personnel licensing responded with a terse email stating that 'ICAO does not endorse tests', and implied that the ETA was being falsely promoted. When the ETA was submitted to another, the personnel licensing inspector stated that because the ETA's ICAO endorsement was 'conditional', it was not of sufficient quality for NAA approval, quite regardless of the fact that none of the tests approved for use in the same state was endorsed by ICAO or provided evidence for test quality.

A desire for a practical and expedient 'solution' to the ICAO language proficiency requirements coupled with a lack of assessment literacy led some NAAs to discharge the responsibility for language testing to personnel such as flight and aeromedical examiners, with one NAA issuing a one-page guidance document as a substitute for the use of a valid and reliable test instrument and suitable personnel training and monitoring. In some states, NAAs have issued guidance on the structure and tasks of an aviation English test, but have delegated the responsibility of writing, administering and assessing tests to the various approved individuals and organisations within the state. Consequently, in these states, tests are developed and administered with little or no field-specific expertise or personnel training, and no

regulatory oversight of test activities. In one large ICAO member state, the NAA endorsed pilot licences for language proficiency at ICAO level 4 by post with no test event at all, effectively selling language proficiency endorsement in exchange for a fee equivalent to the price of a cappuccino. In Europe, the mutual acceptance of aviation English tests under EASA has meant that the poorest quality test approved in any EASA member state has become accepted for personnel licensing across the continent. This has reduced test quality to the lowest common denominator, with commercial success determined not by quality, but by price, convenience and glossy marketing campaigns.

The effect of such regulations on the ETA was disastrous. Due to (sometimes wilful) ignorance of test quality in the creation and implementation of national and international aviation regulations, the competitive playing field was and still is extremely uneven. Why would a company as cost-sensitive as an airline pay £145 for the ETA when a test with the same NAA approval and claims for fitness-for-purpose was available for less than half the price?

In addition to the challenges created by weak regulation, some potential customers were fearful of using an ICAO endorsed test. The quote below from correspondence with a potential customer illustrates the perception that quality testing would somehow be more 'difficult' and result in fewer test takers achieving ICAO level 4:

> [Organisation] will evaluate ICAO next April with old traditional method, face to face interview within 20 minutes. At this moment, they are afraid that if they apply new method, all will be failed and they scare. [Sic] (Personal correspondence)

Scoring ICAO level 3 or below results in suspension or loss of licence which is both potentially disastrous for a pilot's career and extremely costly and disruptive to airline operations. Why would an airline or an individual pilot take that risk?

In terms of test administration, the ETA was developed in strict accordance with ICAO guidance and therefore required, amongst other things, two examiners: an SME and an ELE, both of which played

the combined role of interlocutor and rater during a test with a single candidate. Competing tests operated differently, many ignoring best practice as outlined in ICAO document 9835: some operated with a single human interlocutor/rater, some were computer-based tests with a single rater, and one test was administered and assessed exclusively by a computer with no human-mediated or assessed component. Some potential licensees of the ETA were understandably sceptical about adopting a test that required both an SME and an ELE. Adherence to ICAO guidance was rarely a feature of the ETA that contributed to successful business outcomes.

Finally, OAA and ER faced various practical challenges associated with the development, administration and maintenance of the ETA which can be summarised as follows:

1. Executing the ICAO improvement plan: Given the challenges associated with conducting initial test trials, how could trial candidates best be found and trials paid for to meet the obligations for full ICAO endorsement? How could this work be completed while coping with existing and new business? Would it require additional human resource? If so, where would the budget be found?
2. Adapting the test to meet specific purposes: OAA and ER received enquiries from a range of potential test users including general aviation aeroclubs with members flying recreationally under the privileges of private pilot licences, and a helicopter manufacturer that was interested in integrating the ETA into its type rating training programme linked to aircraft sales. Given the principle that test validity is linked to the degree to which a test engages the characteristics and target language use domain of the test takers, how could the owners of the ETA address the requirements of both a mountain rescue helicopter pilot and a private pilot flying circuits for fun, in a single test designed for commercial multi-engine fixed-wing pilots?
3. Absorbing approval costs: Some potential airline customers understandably required the ETA to be approved for use before committing to purchasing. However, some NAAs required payment of approval fees and proof of test capability (personnel training, administrative systems etc.) and, in some cases, observation of a test in progress locally, before issuing approval. For the owners of the

ETA, it was common to encounter requirements for up-front set up, training and approval fees before securing new business. This had obvious implications for resource and cash flow alongside the inherent business risk. What if this work was conducted but approval not forthcoming? What if, subsequent to approval, the potential customer elected to use a different test?

Together, these regulatory, commercial and practical obstacles eventually led the owners of the ETA to conclude that, in the regulatory and commercial climate of the time, the ETA was not a commercially viable activity. The ETA ceased operations and was withdrawn from the market in the spring of 2013.

In late 2016, another ICAO endorsed test announced closure. This was a test that, in the same year, reported accumulated candidature of >28,000 and had achieved full ICAO endorsement for the testing of pilots (the first and currently the only test to have achieved this highest level of endorsement). No doubt the test owner's reasons for the closure were complex and varied, yet looking at the macro level, the closure of a second ICAO endorsed test is perhaps indicative of a trend in aviation language testing: Rather than witnessing the stabilisation and spread of quality aviation English tests, we are instead witnessing a decline. Is it that quality aviation English testing is simply unviable?

3 Commercial language testing: Barriers to success.

In the aviation industry, language testing has a direct impact on the careers of individuals, airline and air traffic control operations, aviation training and most importantly, on the safety of aeronautical communications. Consequently, test quality is paramount. As ICAO states:

> [Aviation] language proficiency testing … [is] a case of exceptionally high-stakes testing. Inadequate aviation English testing can result in either serious safety gaps or have highly negative social and economic consequences. (ICAO 2010, Section 6.2.2.1)

Regulatory guidance material produced by ICAO and, in Europe, EASA, clearly points to the need for TSPs to provide evidence for the quality of their test instruments in the process of NAA evaluation and approval of language tests:

> TSPs should supply documented evidence of the validity and reliability of their testing methods. (ICAO 2010, Section 6.3.3.1)

> The assessment documentation should include at least … documentation demonstrating the assessment validity, relevance and reliability. (EASA, 2011, Section N, 4, IV (P28))

If not legally binding, this guidance is clear. In the field of language testing, the International Language Testing Association (ILTA) advocates the communication of information about test quality both as good practice and as a professional responsibility:

> [TSPs] preparing and administering publicly available tests ... should publish validity and reliability estimates and bias reports for the test along with sufficient explanation to allow potential test takers and test users to decide if the test is suitable in their situation. (ILTA, 2007)

With the introduction of the ICAO language proficiency requirements in 2003, much aviation English testing has and continues to take place around the world. Dozens of tests have been developed by regulators, airlines, air navigation service providers, air traffic control agencies, language schools, and entities whose sole activity is aviation English testing. The exact number of tests in use around the world is unknown. In April 2017, the author conducted a two-hour Internet search and counted 33 active commercial aviation English TSPs. Some of these TSPs make claims about test quality. Below is selection of claims for test quality from the websites of commercial TSPs which were live at the time of writing (author's emphasis):

> [TEST] is supported by extensive experience and expertise (both linguistic and operational), by **a substantial programme of research**, and by **strict procedures in item design, test delivery, administration and security, examiner training and continuous monitoring of performance**.

> Data collection and analysis of test items and test tasks and the system of on-going, test-by-test monitoring of the [TEST] examiners and assessors continue to ensure the **integrity, validity and reliability** of the Testing System and its component parts.
>
> Test [SIC] is **reliable, valid** and practical.
>
> Language specialists and seasoned pilots have combined their expertise to design an exam that is **fair, reliable**, relevant and personal.
>
> It is a **valid and reliable** qualification proficiency test for pilots, controllers and aviation cadets in their last year of training.
>
> Developed by [TSP], the [test] provides both trainee and licensed pilots with an innovative and **highly effective** means of determining English language proficiency for international radiotelephony communications.

There are many methodologies that a TSP might use to validate a test, that is, to demonstrate that a test is fit for its intended purpose. As Xi notes, 'in the last several decades, language testing has evolved into an independent field that is characterised by well-articulated theories of validity and sophisticated validation methodologies' (Xi, 2014, p189). It is the job of the TSP to build an argument for test validity based on a clear rationale and/or empirical evidence in support of all validity claims (ICAO, 2012). One very important aspect of validity is reliability. As reliability is a necessary condition for test validity (i.e. a test cannot be valid if it is not reliable), an evaluation of evidence for reliability is arguably the first consideration in the evaluation of a language test. As Bachman notes, 'The need to estimate and report information about reliability and measurement error … is explicitly stated in the professional standards for language testers. The appropriate estimation of reliability is not only a matter of good testing practice; it is also a professional responsibility of all who use language tests' (Bachman, 2004, p154). In other language testing contexts that are considered high stakes, for example, testing English for Academic Purposes (EAP) for university admissions or the testing of English for medical purposes, TSPs publish information about the reliability and measurement error of their tests in the public domain as standard practice. (Indeed, many TSPs publish a substantial quantity of research on test validity each year. One leading provider of EAP testing states that its test 'is supported by more than 240 peer-reviewed research

reports, books, journal articles and book chapters regarding test design and validity' (ETS, 2017)). Yet in aviation English testing, at the time of writing, none of the 33 TSPs surveyed for the purposes of this paper provided any substantive evidence for the quality of their tests in the public domain.

Today, the state of aviation English testing is as follows:

- Of the four tests to have achieved ICAO endorsement to date, two TSPs have withdrawn their tests from the market (one TSP's conditional ICAO endorsement lapsed in 2014 though the test still appears to be in use);
- There is just one ICAO endorsed tests for air traffic controllers; and
- There are no ICAO endorsed tests for pilots at all (ICAO, 2017).

It is troubling that in such a safety critical industry which is otherwise tightly regulated and where the stakes in language testing are so high, quality aviation English testing is a rare exception rather than the norm. The absence of test quality is a serious threat both to the fairness of aviation English tests and to the effective implementation of the ICAO LPRs. One wonders how many pilots and controllers have, as a result of poor language testing, had their licenses taken away unfairly, or how many have scored ICAO level 4, 5 or even 6 who should have had their licenses removed because of inadequate language proficiency.

Taking the ETA as a case study, one can summarise three broad issues that result in what can only be described as the failure of language testing to meet the needs of the aviation industry: the regulatory issue, the cost issue and the population issue.

The regulatory issue

For reasons of poor assessment literacy coupled with commercial and political factors, the regulation of aviation English testing is weak. Consequently, a plethora of tests which do not maintain even minimum professional standards for quality thrives unchecked. Such tests are unlikely to contribute

meaningfully to aviation safety or offer meaningful information to the regulators, airlines, air navigation service providers or individuals that use them.

The cost issue

Developing, administering and maintaining a quality language test is expensive. It takes time, involves a large team of personnel and requires field-specific expertise. In commercial language testing, these costs are recovered through fees. Due to weak regulation and no doubt due to poor assessment literacy, many TSPs do not attend to quality in accordance with the recommendations of ICAO document 9835 and therefore do not carry the associated cost burden. Therefore, such tests are able to operate profitably, even with low prices. This makes it extremely difficult for a quality commercial aviation English test to compete effectively and attract the candidature necessary to recover costs, let alone make a return on investment.

The population issue

For the purposes of comparison, let's briefly examine language test populations in a context other than aviation. The testing of English for Academic Purposes (EAP) for college and university admission is a field which is dominated by a small number of key players. Like aviation English testing, EAP testing is considered 'high-stakes' insofar as test scores have a significant impact on those that use them. Therefore, the investigation and reporting of quality in EAP testing comes as standard practice. Some EAP TSPs report candidature in the millions. For example, more than 30 million candidates have taken the Test of English as a Foreign Language (ETS 2017), and 2.9 million candidates took the International English Language Testing System (IELTS, 2017) in 2016 alone. To compare EAP test populations with those of aviation English, in 2010, the global population of pilots and air traffic controllers was 530,410

(ICAO, 2010c). However, the test taker population is much smaller than the total population due to the adoption in the vast majority of states of the following ICAO recommended practices[7]:

> '... language proficiency [of personnel] who demonstrate proficiency below the Expert Level (Level 6) should be formally evaluated at intervals in accordance with an individual's demonstrated proficiency level, as follows:
>
> a) those demonstrating language proficiency at the Operational Level (Level 4) should be evaluated at least once every three years; and
>
> b) those demonstrating language proficiency at the Extended Level (Level 5) should be evaluated at least once every six years. (ICAO 2010, Section 1.2.9.7)

Therefore, personnel that have achieved ICAO level 6 are exempt from further testing and have been permanently removed from the test taker population. In addition, those that achieve ICAO levels 4 and 5 are required to take tests at intervals of three and six years respectively. This means that the total annual aviation English test taker population is a fraction of the total population pilots and air traffic controllers. In terms of standardised international high-stakes language testing, this population is small indeed. As discussed above, for a quality test to be commercially viable, a TSP needs to attract a sufficient volume of candidates to recover costs and generate profit. Given the disproportionately large number of TSPs competing for a share of a relatively small and poorly regulated market, the complete absence of evidence for the quality of aviation English tests should not come as a surprise.

14 years have passed since the introduction of the ICAO LPRs, and today, unfortunately, there is little evidence of improvement in the standard of aviation English testing. In fact, aviation English testing appears to be travelling in the wrong direction: since the introduction of the ICAO AELTS in 2012, the number of credible aviation English tests and the number of tests that publish evidence for quality has

[7] Note that for pilots in Europe, EASA extended the testing interval for those demonstrating ICAO level 4 from 3 to 4 years, and for those demonstrating ICAO level 6, implemented a 9 year testing interval rather than lifetime exemption.

actually declined. Given the significant issues outlined above, one might argue that in the current climate, quality aviation English testing is simply an elusive goal. As Alderson (2012, p402) noted:

> Meanwhile, testing continues, no doubt frequently using unreliable or at best unproven instruments … the implications of this for aviation safety are obvious.

REFERENCES

Alderson, J. C. (2008). *Final Report on a Survey of Aviation English Tests.* Unpublished manuscript, Lancaster University, Lancaster, United Kingdom. Retrieved on 20/06/2017 from http://www.ealta.eu.org/documents/archive/Alderson_2008.pdf.

Alderson, J. C. (2011) *The Politics of Aviation English Testing.* Language Assessment Quarterly, 8:4, 386-403

Alderson, J.C., Clapham, C. & Wall, D. (1995) *Language Test Construction and Evaluation* Cambridge University Press

Bachman, L. (2004) *Statistical Analyses for Language Assessment.* Cambridge University Press, Cambridge.

EASA (2011) *Acceptable Means of Compliance and Guidance Material to Part-FCL.* EASA: Cologne

ETS (2017) *About the TOEFL iBT Test* retrieved on 20/06/2017 from https://www.ets.org/toefl/ibt/about

ETS (2017) *Test and Score Data Summary for TOEFL iBT Tests January 2016– December 2016* retrieved on 20/06/2017 from https://www.ets.org/s/toefl/pdf/94227_unlweb.pdf

Emery, H. (2010) *Developing an English Language Proficiency Test for Pilots.* Poster presented at the 32nd ILTA Language Testing Research Colloquium, University of Cambridge

Emery, H. (2014) *Developments in LSP Testing 30 Years On? The Case of Aviation English.* Language Assessment Quarterly 11:2, 198-215

IELTS. (2017) retrieved on 20/06/2017 from https://www.ielts.org

ILTA. (2007) *Guidelines for Practice* retrieved on 20/06/2017 from http://www.iltaonline.com/page/ITLAGuidelinesforPra

Jilani, K. (2012) *Here Is The Picture. Describe Everything You Can See: A Comparison of Intended and Observed Production of Language Functions in Part Three of the English Test for Aviation.* Unpublished Masters Dissertation, Lancaster University

ICAO. (2006) *Rated Speech Samples.* Montreal: International Civil Aviation Organization.

ICAO. (2010a). *Manual on the implementation of ICAO Language Proficiency Requirements.* (2nd ed.). Doc 9835 AN/453. Montreal: International Civil Aviation Organization.

ICAO. (2010b) *ICAO to Endorse Testing for Language Proficiency.* ICAO Journal 65:4, 2010, 30-31

ICAO. (2010c) *Global Air Transport Outlook 2010* Circular 333. Montreal: International Civil Aviation Organization.

ICAO. (2012) A *Guide to Submitting Validity Evidence* Montreal: International Civil Aviation Organization.

ICAO. (2017) *ICAO Recognized Tests* retrieved on 22/06/2017 from https://www4.icao.int/aelts/Home/RecognizedTests

Xi, X. (2014) *Methods of test validation*. In Sohamy, E. & Hornberger, N. *Encyclopedia of Language and Education, 2nd Edition, Volume 7: Language Testing and Assessment* 177-196 Springer: New York