

Standardization in Language Rating: an Association of Language Raters

Henry Emery
emery-roberts
Plymouth, United Kingdom

In any test used to measure language proficiency in aviation, language raters are required to assess the speech of pilots and air traffic controllers. As such, language raters play a pivotal role in the implementation of the ICAO Language Proficiency Requirements. This paper highlights the need for regulation in language rating, and proposes a mechanism for establishing a standard interpretation of the *ICAO Rating Scale and Holistic Descriptors*.

The need for regulation in language rating

In response to the ICAO language proficiency requirements, English language teaching for aeronautical communication is witnessing an unprecedented growth in approaches, methodologies and training materials. With the implementation date of the language proficiency requirements now on the horizon, airlines, air navigation service providers and training organizations around the world are developing and delivering English language training and testing programmes so that operations personnel may meet the ICAO requirements.

Naturally, syllabus designers, language teachers and test developers seek to deliver training programmes which cater to the specific needs of language learners. As a result, language training and testing programmes reflect the wide variety in learning needs and contexts of operations personnel throughout the world.

In this diverse language training environment, what every pilot and air traffic controller shares in common is the criteria by which their language proficiency is assessed. Every testing programme and test instrument developed to measure the language proficiency of aviation operations personnel will employ the *Rating Scale and Holistic Descriptors* in each of the 190 ICAO member states.

In any test used to measure language proficiency in aviation, language raters are required to assess the speech of pilots and air traffic controllers. As such, the language rater plays a pivotal role in the language training and testing process. It is the language rating community that carries the ultimate responsibility for the effective implementation of the ICAO Language Proficiency Requirements.

Language testing in aviation is extremely ‘high-stakes’. The professional lives of pilots and air traffic controllers at stake, and the operational capability of airlines and air traffic control centers and future budgets for language training and testing may be greatly affected by the results awarded by language raters. However, above all, it is ensuring a standard of safety in aeronautical communication worldwide that rests with the competence, skills and good judgment of the language rater.

In such a ‘high-stakes’ environment, language testing needs to be accountable to the stakeholders in the aviation industry. State regulators, managers of airline and air traffic management service providers, trainers, pilots and controllers, and ultimately, the flying public, need to be able to trust global language assessments and to have confidence that licensed operations personnel are competent communicators in the English language. As those who are obliged to learn and use English on the frequency, pilots and controllers deserve to have their language proficiency assessed fairly, and to know that their counterparts around the world have been assessed according to the same standard.

Despite the ‘high-stakes’ nature of aviation language testing, language rating is, at present, an unregulated area of activity.

The Language Rater

In order to inspire the trust and confidence of the stakeholders in aviation language testing, language testers around the world must provide test results which are valid, consistent and reliable. In order to achieve this, it is essential that a global standard in the application of the assessment criteria is established. The levels of proficiency described in the *Rating Scale* and *Holistic Descriptors* need to be applied to the speech of aviation operations personnel in the same way in order to ensure a standard for clear and safe communication in the skies. Uniformity in what is considered ‘Operational’ features of pronunciation, structure, vocabulary, fluency, comprehension and interactions needs to be established according to a standardized interpretation of the *Rating Scale* and *Holistic Descriptors*, regardless of what test instruments are employed and in what member state language testing is conducted.

As with the nature of the use of descriptors in the assessment of spoken language, the *Rating Scale* and *Holistic Descriptors* are open to interpretation. In listening to the same speech sample, what one rater may see as ‘creative use of basic grammatical structure’ (level 4), another may see as poor control of grammatical form (Level 3). One rater may feel that a pilot’s speech is so heavily influenced by first language as to make it frequently unintelligible, awarding a Level 3. Conversely, another rater may be quite familiar with the pilot’s first language, and therefore find that the same speaker’s speech ‘rarely interferes with ease of understanding’, and award a Level 5. As language raters come from a wide range of professional, cultural and linguistic backgrounds, nowhere is the issue of interpretation of assessment criteria more acute than in language rating for aviation.

Language raters are inevitably influenced by the following factors:

The rater’s first language

The level of English language proficiency (if the rater is a non-native speaker)

The degree of familiarity with aviation operations and aeronautical communication (if the rater has a professional background in language and linguistics)

The degree of familiarity with language and linguistics (if the rater has a professional background in aviation operations)

The degree of experience in language assessment and using language descriptors

The degree of training in the application of the rating scale

The extent and frequency of exposure to international accents

The extent and frequency of exposure to a particular accent

Given the diversity of background of the people involved in language rating in aviation, there is an obvious need to establish a standard interpretation of the *Rating Scale* and *Holistic Descriptors*, and to ‘calibrate’ the international rating community.

Intra-rater reliability

As it is necessary to calibrate the instruments on a flight deck in order to gain accurate and reliable readings, so it is essential that a human rater is calibrated to produce reliable test results. Studies show that individual raters are unlikely to produce consistent results over time, and that without calibration, the longer a rater works, or the higher the number of candidates rated, the more inconsistent the results become. To illustrate, a rater listens to a speech sample of a controller and awards a Level 5. Six months later, the same rater listens to the same speech sample, and awards a Level 4. These inconsistencies in an individual rater’s performance threaten intra-rater reliability.

Inter-rater reliability

As an individual rater may fail to produce consistent results over time, then a group of raters may fail to produce comparable and consistent results. This can occur at an organizational level, for example, within a team of six active raters in one test organization.

The same issue also occurs at a national level. In a member state where there is a national aviation training academy, an air traffic service provider, a number of airlines and several private training institutions, each with their own test instrument and group of raters, the reliability of test results from that state alone may be questionable.

When we examine the issue of inter-rater reliability on an international level, and consider the international context of language rating in aviation and the diversity of background of language raters worldwide, then we see how the reliability of test results may be thrown into serious doubt.

Many different tests have been developed to assess the language proficiency of pilots and air traffic controllers with huge variations in quality and integrity and approaches to rater training. Nonetheless, it is essential that test results are not only reliable, but comparable. If the objective of the ICAO language proficiency requirements is to achieve a standard level of language proficiency among operations personnel for safe and effective communication, then a system for ensuring a standard interpretation of the assessment criteria and the reliability of test results around the world is fundamental in achieving this end.

An Association of Language Raters

It is the aim of this paper to propose a mechanism to set and maintain a global standard for the interpretation and application of the *ICAO Rating Scale* and *Holistic Descriptors*. Through international co-operation and self-regulation, an Association of Language Raters, an international panel of raters from operational and linguistic backgrounds, would work together to minimize variation and discrepancy in language rating. Such an Association would help language raters around the world meet a global standard, and to produce test results which are consistent and reliable.

In order to establish of an Association of Language Raters, this paper proposes the following phases.

Phase 1: Consultation and funding

An Association would need to:

1. Seek endorsement from ICAO.
2. Seek endorsement and funding from state regulators and aviation organizations.
3. Seek support from academic organizations, such as ICAEA, ALTE and ILTA.
4. Invite raters from operational and linguistic backgrounds to act as associates.
5. Formalise the structure and procedures of the Association.

Phase 2: Establishing a language rater accreditation service

An Association would need to:

1. Gather samples of speech of pilots and air traffic controllers from organizations around the world which are involved with testing English language proficiency for aviation.
2. Convene to:
 - Rate the gathered speech samples.

- Produce guidance material for language raters on the standard procedure for the application of the Rating Scale and Holistic Descriptors.
- Organize the rated speech samples and guidance material into ‘assessment packs’.

Phase 3: Implementing a language rater accreditation service

The proposed procedure for rater accreditation is as follows:

1. A potential language rater applies to the Association of Language Raters for rater accreditation.
2. The Association provides an ‘assessment pack’ to the potential rater.
3. The rater rates the samples and returns the ratings to the Association for assessment.
4. If the rater has adequately demonstrated an ability to rate speech samples according to the standard set by the Association, an official certificate of accreditation is sent to the language rater.
5. A certificate of rater accreditation is valid for two years, after which the rater must re-apply for accreditation.

Principles of an Association of Language Raters

An Association of Language Raters would

- set a global standard in the application of the Rating Scale
- accredit language raters to rate for licensing purposes
- regulate, monitor and support the work of language raters
- serve the interests of the test taking community
- be answerable to the aviation industry
- be transparent
- be representative of the international community
- be democratic in its organization
- have associate raters with a limited tenure
- be made up of linguists and operations personnel
- be made up of native and non-native speakers
- be non-profit making

Pilots and air-traffic controllers deserve to be assessed by expert language raters who are sufficiently trained and qualified to perform the task. It is not enough to claim to be able to assess language using the *Rating Scale* and *Holistic Descriptors*; raters must demonstrate their ability to rate language according to a global standard.

Language raters need to inspire the trust and confidence of the test-taking community, and be accountable to the stakeholders in the aviation industry. An Association of Language Raters would work with the objective of establishing and maintaining standards in language assessment, pursuing professionalism, transparency and fair play in aviation language testing.

— — — — —