

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

Saturday 10th May 2008

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 is the work of a language rater 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 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 practise independently of a testing system. A parallel was drawn with the legal profession in the USA whereby practitioners of law a 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 practise 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, of 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 have 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