International Civil Aviation Organization

A39-WP/249
TE/97
23/8/16

WORKING PAPER

ASSEMBLY — 39TH SESSION

TECHNICAL COMMISSION

Agenda Item 35: Aviation safety and air navigation standardization

THE NEED FOR REVISION OF ANNEX 1 LANGUAGE PROFICIENCY SARPS FOR PILOTS AND AIR TRAFFIC CONTROLLERS

(Presented by Brazil)

EXECUTIVE SUMMARY

The adoption of the Annex 1 Language Proficiency Standard and Recommended Practices (SARPs) in 2003 represented a considerable progress towards safer radiotelephony communications. However, applied linguists, language testing scholars and experienced test developers and assessors have identified in the past thirteen years some serious deficiencies in the ICAO language proficiency policy, which may significantly impact safety. For this reason, experts have recommended that the policy should be reviewed. This paper provides a brief discussion of the main weaknesses identified in the ICAO policy related to the language proficiency requirements for pilots and air traffic controllers (ATCs). It highlights the importance of having a clear definition of the abilities that should be assessed as well as the necessity of native speakers also being formally evaluated on their abilities to use the language effectively and to manage communication successfully.

Action: The Assembly is invited to:

a) call on the Council to establish a study group to develop an ICAO rating scale and a set of holistic descriptors based on clear, explicit and relevant definitions of the specific language abilities which need to be assessed in this particular environment;

b) call on the group to consider reviewing the Annex 1 Language Proficiency SARPs as a matter of utmost importance taking into account the recommendations made by researchers in the fields of applied linguistics and language testing, considering the current policy poses significant threats to safety

Strategic Objectives:	This working paper relates to the Safety Strategic Objective.
Financial implications:	The activities referred to in this paper will be undertaken subject to the resources available in the 2017-2019 Regular Programme Budget and/or from extra budgetary contributions. Highly qualified experts from Civil Aviation Authorities, associations and non-profit organizations may advise ICAO without any financial implications.
References:	Annex 1 — *Personnel Licensing,* 1.2.9 Language Proficiency Doc 9835, *Manual on the Implementation of ICAO Language Proficiency Requirements*

1. INTRODUCTION

1.1 In March 2003 the Council adopted the Standards and Recommended Practices (SARPs) concerning the Language Proficiency Requirements (LPRs). The Proficiency Requirements in Common English Study Group (PRICESG) had taken on an enormous challenge and had produced fair materials. However, as discussed by Alderson, Clapham and Wall (1995), no matter how responsible test developers are in order to make sure their tests are valid and reliable, "problems with a test or associated procedures may only emerge once a test has been in operation for some time"(p. 218). This is the reason why, as they argued, test content, test administration and test marking need to be a monitored ongoing process, so that they "can be modified and improved in the light of their performance and of research and feedback" (p. 218). This applies to the ICAO policy for two reasons: a) the rating scale is a very important part of the marking process; b) the content of the ICAO language proficiency tests had to be based on the ICAO rating scale, as the tests needed to assess what the rating scale descriptors described. Unfortunately, since the publication of the ICAO LPRs, neither the policy nor the rating scale has changed. The SARPs have been implemented in at least 67.75% of States (according to the working paper A38-WP/37 presented by ICAO at the 38th Assembly), the rating scale has been used for more than a decade, extensive research has been carried out, many studies have put forward suggestions on how to improve the LPRs, but the policy still remains unchanged.

1.2 There has been a fair bit of criticism over the ICAO policy and the quality of its rating scale in the applied linguistics, language testing and aviation communities. Prestigious researchers around the world, such as Charles Alderson, from Lancaster University, UK, and Dan Douglas, from Iowa State University, USA, both Professor Emeritus, have called for a revision of the ICAO LPRs. Douglas (2004) argued there are areas of ambiguity and uncertainty that need clarification, whereas Alderson (2009) named critiquing and revising the ICAO rating scale as a particularly important research area, having, one year later, questioned: "are they" (the scales) "sufficiently explicit and relevant to guarantee that any test constructed on the basis of the ICAO scales will indeed be at the 'right' level or do the scales represent an uncertain and unstable foundation?"(Alderson, 2010, p. 64).

2. DISCUSSION

2.1 A broad range of issues have been discussed concerning the ICAO language proficiency policy. Most of them are derived from a foundational problem, which is the lack of a precise definition of the domain of English for radiotelephony communications (Douglas, 2004; Moder & Halleck, 2009; Emery, 2014). This lack of clarity in relation to the nature of the language that needs to be assessed has created some significant problems in the ICAO policy, including the rating scale. ICAO has been very clear in the DOC 9835 that the purpose of the test is to assess plain English, not phraseology, and that some phraseology can be included in the test as prompts or scene setters, as long as its incorrect usage is not rated (6.3.2.9). However, research (Davies, 2001; Ryan, 2007; Knoch, 2009) has indicated that it is difficult for subject matter experts to separate language ability from technical knowledge. In addition to that, it is neither possible nor desirable to separate one from the other (Knoch, 2009; Emery, 2014). As pointed out by Estival, Farris, & Molesworth (2016), "while ICAO's reason for not including the assessment of standard phraseology in assessments designed in relation to the LPRs is understandable, it also creates a tension between policy and reality that is difficult for test developers to reconcile" (p. 101).

2.2 A consequence of this obscurity in relation to the nature of the English used in radiotelephony communications between pilots and ATCs is that the construct of what needs to be tested (the definitions of the specific language abilities that need to be assessed) is also not clear. Emery (2014) believes that "the ICAO guidance (…) is of little practical use in the definition of the test construct and

the development of test specifications" (p. 206). In 2004, Douglas had already claimed it was necessary to undertake in-depth research to investigate the nature of both standardized phraseology and plain language, the relationship between them, and the conditions in which they were used. More information would have been crucial to define a clearer construct and to develop a rating scale that would effectively represent the target language. As pointed out by Bachman and Palmer (2010), rating scales must be defined from a clear definition of a test construct. Hence, the publication of unclear guidelines encountered resistance not only among scholars, but also among aviation professionals. Kim and Elder (2014) pointed out aviation experts became reluctant towards the ICAO policy because they believe some features of the construct are not a good reflection of the target language. Furthermore, Kim (2013) believes the non-native speakers' resistance towards the ICAO LPRs will only change after the policy and the construct underpinning it are revised. As Douglas (2004) pointed out, it is very important to have a clear picture of what is being assessed, as well as a "clear, complete and unambiguous definition of the construct to be measured in relation to the purposes for which the measurement is being made" (p. 250). Therefore, test developers have been having to deal with very difficult challenges (Moder & Hallack, 2009; Emery; 2014; Estival et al, 2016; Garcia, 2015).

2.3 Some descriptors which were included in the rating scale are considered to be irrelevant by scholars, experienced pilots and ATCs, whereas some important abilities were not taken into consideration (Kim & Elder, 2009; Read & Knoch, 2009; Scaramucci, 2011; Foy, 2012; Monteiro, 2012; Kim, 2013; Douglas, 2014; Prado, 2015). Kim and Elder argued that ICAO oversimplified the test construct by not including the assessment of important communicative abilities. Kim claimed that "strategic competence for accommodation, and shared responsibility for lack of success of communication by participants should be incorporated into the radiotelephony communication construct and any tests which are designed to reflect this" (p. 107). Communication failures do not happen only because of lack of proficiency in the language, but also because of lack of knowledge of how to use the language (lack of communicative abilities), which can and need to be trained. Unfortunately, as pointed out by Douglas (2014), interactional competence in the context of the ICAO LPRs is not assessed in any country.

2.4 It is not only important to assess non-native speakers' interactional competence, but also native speakers'. Native speakers may also interact poorly and may not communicate effectively. Unfortunately, in spite of the fact that the DOC 9835 supports the idea of understanding the English used in radiotelephony communications in the context of English for Lingua Franca (ELF), as stated in section 5.3 ("the burden of improved communication should not be seen as fallen solely on non-native speakers"), the current policy has put the burden of effective communication on non-native speakers, as native speakers do not need to be formally evaluated. As Estival et al (2016) argued, both native and non-native speakers are accountable for communication problems. Therefore, all pilots and ATCs, regardless of their first language, should be trained to communicate effectively in English. They expanded "pilots who are native English speakers commit, in some cases, as many communication errors as English as a second language pilots" (p. 199). Read and Knoch (2009) argued that the ICAO LPRs not only have placed the onus on non-native speakers to improve their proficiency, but have given "native-speaking aviation personnel no incentive to develop their communicative competence in ELF terms" (p. 21.7). As Douglas (2014) advocated, the assessment should test linguistic awareness and the ability to successfully manage communication, and also the "abilities to accommodate their use of English in the context of intercultural communication" (p. 2). Similarly, Monteiro (2012) emphasized the relevance of raising pilots and ATCs' "awareness of the linguistic, discursive-interactional and intercultural factors" (p. 64) in order to improve radiotelephony communications safety. Foy (2012), who is a native speaker pilot, also discussed the importance of training cultural sensitivity and pointed out that level 6 pilots tend to assume everybody understands them, therefore they do not employ strategies to make sure the messages are actually comprehended. Nevertheless, as Williams (2016) found out, most native speaker pilots and ATCs

think they should be required to take Aviation English tests, a total of 63.5% of the participants in his study. Moreover, 77.4% of the non-native speakers think alike.

3. CONCLUSION

3.1 Given that the importance of reviewing the Annex 1 language proficiency SARPs for pilots and ATCs has been demonstrated, it is proposed that they be amended, as follows:

3.2 After Paragraph 1.2.9.4, another Paragraph with a deadline for the implementation of the new requirement will have to be included.

3.3 After Paragraph 1.2.9.6, another Paragraph with a new deadline for the implementation of the new requirement will have to be included, as follows: "As of (to be determined by ICAO), the language proficiency, interactional competence and linguistic awareness of aeroplane, airship, helicopter and powered-lift pilots, air traffic controllers and aeronautical station operators shall be evaluated."

3.4 Paragraph 1.2.9.7 should include an interval of assessment for Level 6 individuals. The intervals must be borne out by evidence. They will need to be verified by post hoc research aiming at investigating if the intervals align with actual language decay and, in case they do not, they have to be re-established.

3.5 Note 1 in Section 1.2.9 should be deleted, whereas the note in Appendix 1 should be clarified.

3.6 It is also proposed that a study group be established to carry out a comprehensive review of the holistic descriptors and the ICAO rating scale. As a consequence of the promotion of the best practices in language testing, safety will be enhanced. The study group should be composed of both native and non-native speakers, experienced pilots and ATCs with different backgrounds, experienced aviation English testers, aviation English trainers, aviation English material developers, applied linguists, and language testing scholars.

3.7 The appointed study group will need to come to detailed and clear definitions of the specific language abilities which need to be assessed in this particular environment (the test construct) as well as review the holistic descriptors according to this construct, and develop, trial and validate a rating scale which reflects the defined construct. Post hoc validation studies must be conducted in order to improve the rating scale and the holistic descriptors.

— END —

Note.- Due to a limitation on the number of pages, the references list was not provided. The references list may be sent by e-mail upon request. Please send a request to angela.garcia@anac.gov.br