



Civil Aviation Safety Authority
of Papua New Guinea

Advisory Circular

AC61-16

Guidelines for Testing and Approval of English Language Proficiency

Issue 1
31 October 2022

GENERAL

Civil Aviation Safety Authority Advisory Circulars (AC) contain information about standards, practices and procedures that the Director has found to be an Acceptable Means of Compliance (AMC) with the associated rule.

An AMC is not intended to be the only means of compliance with a rule, and consideration will be given to other methods of compliance that may be presented to the Director. When new standards, practices or procedures are found to be acceptable, they will be added to the appropriate Advisory Circular.

An Advisory Circular may also include **Guidance Material (GM)** to facilitate compliance with the rule requirements. Guidance material must not be regarded as an acceptable means of compliance.

PURPOSE

This Advisory Circular provides an acceptable means of compliance and guidance for the approval of English Language Proficiency for Pilots.

RELATED CAR

This Advisory Circular relates specifically to the Civil Aviation Rules listed in Section 4.

CHANGE NOTICE

This AC replaces Initial Issue dated 1 July 2017.

APPROVAL

This AC has been approved for publication by the Director of Civil Aviation

TABLE OF CONTENTS

1.	Introduction	3
2.	Scope	3
3.	Applicability	3
4.	Related Rules	3
5.	References	4
6.	Application and Qualification	4
6.1	English Language Proficiency Requirements for New Applicants	4
6.2	Existing Licence Holders	5
6.3	Language Proficiency Endorsement	5
6.4	Language Proficiency Assessment	5
7.	Means of Assessment of Language Proficiency	6
7.1	At the Radiotelephony Test	6
7.2	At a Flight Test	6
8.	ICAO Language Proficiency Rating Scale	7
Appendix 1 ICAO English Language Proficiency Rating Scale		8
Appendix 2 Guidance for ELP Assessors / Examiners		9
2.1	General	9
2.2	Assessment	9
2.3	Basic Assessment Requirements	9
2.4	Assessors	11
2.5	Criteria for the Acceptability of Language Assessment Bodies	11
Appendix 3 Criteria for Aviation Language Testing Providers		13
3.1	General	13
3.2	Test Design and Construct	13
3.3	Test Validity and Reliability	17
3.4	Rating	18
3.5	Test Administration and Security	19
3.6	Record-Keeping	23
3.7	Organizational Information and Infrastructure	24
3.8	Testing-Team Qualifications	26

1. Introduction

The International Civil Aviation Organisation (ICAO) of which Papua New Guinea is a Contracting State establishes international standards and recommended practices (SARPS) for, among other things, the licencing of aviation personnel. These SARPS which ICAO Member States are obliged to implement are published in Annex 1 – Personnel Licencing to the Convention on International Civil Aviation (ICAO Annex 1).

Applicable from 5 March 2008, ICAO introduced into ICAO Annex I a requirement for all pilots of aeroplane, helicopter, airship and powered lift aircraft, all flight navigators who use radiotelephony equipment in aircraft and all air traffic controllers, to be proficient in their command of the languages that they use for radio communication. A proficiency scale of 1 to 6 is specified, with Level 6 being the standard of an expert speaker of the language. ICAO Annex 1 specifies the minimum standard for the holder of a licence to be Level 4.

It also specifies that licence holders assessed as Level 4 or 5 shall be re-tested periodically but that a person assessed as Level 6 need not be re-tested. The maximum periods between tests stated in the ICAO documents for Levels 4 and 5 are recommendations only. If the language proficiency of an individual is assessed as being below Level 4 (i.e. 1, 2 or 3) that individual may not hold a licence that includes radiotelephony privileges. The ICAO standards apply to the language “used for radio communication”; ICAO permits languages other than English in circumstances where flight crew and air traffic controllers share another common language.

Section 75 of the Civil Aviation Act 2000 (as amended) states that PNG Rules shall be consistent with the standards of ICAO.

2. Scope

This Advisory Circular provides guidance for testing and approval of English Language Proficiency for applicants of a licence under CAR Part 61

3. Applicability

This Advisory Circular is applicable to all new applicants and existing foreign licence holders using or intending to use the privileges of a Part 61 licence. English Language Proficiency is not required to hold a validation however it is a requirement for the purpose of issuance of a PNG licence under the CAR 61. A foreign licence having an acceptable level of ELP may be transferable.

4. Related Rules

The civil aviation rules relating to the requirement of ELP are listed below; each of these rules has a requirement for documentation that could be met through ELP assessment by an authorised person /organisation and submission of Form CA 61/03 - ELP Assessment Notification.

61.9 (a) (3) Validation Permit for a Foreign Pilots Licence: Defines the requirement to have the ability acceptable to the Director in reading, speaking, understanding and communicating in the English language to enable the applicant to adequately carry out the responsibilities of a pilot exercising the privileges of a validation permit.

61.11 (b) Application and Qualification: Requires an applicant for a pilot licence or rating to have the ability acceptable to the Director in reading, speaking, understanding and communicating in the English language.

61.17(a) Examination Prerequisites: Requires an applicant to provide proof of identification as acceptable to the director.

61.43 (c) (d) Examination of Proficiency: Requires the holder of a licence or rating issued Part 61 to undertake and examination or test if the Director believes that the privileges of the licence or rating are being carried out in an incompetent manner.

61.103 (d) Student Pilot Licence: Defines the requirement to demonstrate an appropriate level of proficiency in English Language at ELP Level 4 or above and pass a Radio Telephony Practical Test.

61.155 (b) (2) Private Pilots Licence: Defines the requirement to demonstrate an appropriate level of proficiency in English Language at ELP Level 4 or above

61.205 (c) (2) Commercial Pilots Licence: Defines the requirement to demonstrate an appropriate level of proficiency in English Language at ELP Level 4 or above.

61.255 (b) (2) Airline Transport Pilots Licence: Defines the requirement to demonstrate an appropriate level of proficiency in English Language at ELP Level 4 or above.

The rules immediately above all relate to ELP as a licencing requirement.

5. References

The following documents are referenced in this Advisory Circular:

- **Civil Aviation Safety Authority PNG:**
- CAR 61 Pilot licences and Ratings
- **ICAO:**
- Annex 1; and
- Doc 9835 Manual on the implementation of ICAO Language Proficiency Requirements

6. Application and Qualification

6.1 English Language Proficiency Requirements for New Applicants

Rule 61.11(b) requires an applicant for a pilot licence to have sufficient ability in reading, speaking, understanding and communicating in the English language to enable the applicant to adequately exercise the privileges of that licence.

The acceptable means of compliance with rule 61.11(b) by all applicants for the issue of an aeroplane or helicopter private pilot, commercial pilot or air transport pilot licence is by demonstrating proficiency to at least Level 4 (Operational) of the ICAO Language Proficiency Rating Scale and the following tasks -

- communicate effectively in voice-only (radiotelephone) communications
- communicate on common, work-related topics with accuracy and clarity
- use appropriate communication to exchange messages and to recognize and resolve misunderstandings in a general or work-related context
- handle successfully, and with relative ease, the linguistic challenges presented by a complication or unexpected turn of events that occurs within the context of a routine work situation or communicative task with which they are otherwise familiar
- use a dialect or accent which is intelligible to the aeronautical community.

The foregoing requirement also applies to applicants for a higher aeroplane or helicopter pilot licence holder applying for a private pilot licence (PPL) or PPL holder applying for a commercial pilot licence (CPL) whose language proficiency has not been assessed.

6.2 Existing Licence Holders

As a result of their participation in the PNG civil aviation system, pilots who have been issued an aeroplane or helicopter licence under Part 61 before 31st May 2008, are considered to have demonstrated sufficient English language ability to adequately continue to exercise the privileges of that licence within the Port Moresby Flight Information Region. These persons do not need to be assessed unless they wish to have a language proficiency level endorsed on their licence, particularly, for pilots who fly internationally.

If the Director believes on reasonable grounds that an aeroplane or a helicopter pilot licence holder's English language proficiency is inadequate to safely exercise the privileges of that licence, the Director may, by notice in writing, require the holder to undertake a language proficiency assessment in accordance with CAR 61.43(c)(1) & (2)

6.3 Language Proficiency Endorsement

Applicants for an aeroplane or helicopter private pilot, commercial pilot or air transport pilot licence or other ICAO recognized licence holders who wish to have a language proficiency level endorsed on their licence, must have demonstrated English language proficiency to at least Level 4 (Operational) in all language categories specified in the ICAO Language Proficiency Rating Scale. Satisfactory evidence of such a demonstration is an assessment credit issued by a CASA delegate an Authorized Flight Examiner (AFE) or Aviation Training Organisation (ATO).

Applicants for the initial issue of a pilot licence will have their demonstrated language proficiency level endorsed on the new licence at no added charge. Existing licence holders who demonstrate language proficiency may, upon application to the Director and payment of the applicable licence amendment fee, have their language proficiency level endorsed on their licence.

Licences are endorsed as language proficiency level 4, 5 or 6 in accordance with the respective assessment credit. All English Language Proficiency demonstration currency periods are 3 years -

- Level 6 (Expert) no need for future re-assessment.
- Level 5 (Extended) valid for three years from the date of assessment
- Level 4 (Operational) valid for three years from the date of assessment

A current language proficiency endorsement held by a person applies to all pilot licences held by that person. A person who holds a current language proficiency endorsement and who applies for a higher or different type of ICAO licence will have that language proficiency endorsed on the new licence for the remaining currency period.

Aeroplane or helicopter pilot licence holders who do not have evidence of having their language assessed may not have their licences accepted by other ICAO Contracting States. Those who intend to operate in foreign airspace are strongly recommended to obtain a language proficiency endorsement.

6.4 Language Proficiency Assessment

The language proficiency assessment is a demonstration of the ability to communicate in an aviation context. Therefore, all language proficiency candidates are expected to have a basic aviation awareness of at least the subject matter contained in the Private Pilot Licence theory syllabus specified in AC 61-5, Appendix III.

Thus, as a prerequisite to undertake a language proficiency assessment, a person is to hold at least:

- written examination credits for all PPL or CPL theory;
- a Papua New Guinea aeroplane or helicopter pilot licence; or
- a valid armed forces flight experience and qualifications assessment indicating that the person meets the criteria for issue of a Papua New Guinea pilot licence; or
- a current foreign aeroplane or helicopter pilot licence.

To adequately cater for a wide variation in language proficiency levels, two forms of English language proficiency assessments are available and persons may undertake the assessment they consider most suitable-

a) Level 6 Proficiency Demonstration

The Level 6 Proficiency Demonstration is designed to confirm that native or very proficient non-native English speakers can clearly meet ICAO Level 6 language criteria. It is a direct assessment to confirm that the speaker can communicate at level 6 for pronunciation, structure, vocabulary, and fluency. This assessment is suitable for pilots who are confident that they are able to communicate at Level 6 in all respects. The only outcomes are “Level 6” or, if a candidate does not clearly demonstrate Level 6 proficiency, “not determined”.

The acceptable means of compliance with rule 61.11(b) for an applicant who does not clearly meet “Level 6” criteria on the first attempt is to complete a Formal Aviation Language Test.

b) Formal Aviation Language Test Provider

CASA will only assess candidates at ELP level 6. Candidates demonstrating ELP below level 6 will be directed to an organisation recognised by CASA as having the capability to provide Aviation Language Test services.

An organisation wishing to be recognised as a Formal Language Test provider must demonstrate the criteria for aviation language testing required by CASA and which is detailed in Appendix 3.

A candidate who does not attain an ELP of Level 6 (Expert), must complete a formal Aviation Language Test to demonstrate an ELP of Level 4 or 5, for the endorsement of their licence.

All tests are directly supervised by Testing Officers under the control of a delegated service provider and candidates are to provide photo evidence of identity as specified by rule 61.17(a) for written examinations.

Language proficiency candidates who wish to dispute an overall rating may apply to the Director, for a review of their language proficiency assessment.

7. Means of Assessment of Language Proficiency

7.1 At the Radiotelephony Test

During the practical test for the PNG FRTO RATING, a CASA authorised RTF Examiner who has Level 6 proficiency in English, will assess the applicant's proficiency in English. Where the examiner assesses the candidate as being Expert Level 6, he/she may certify to that effect by submitting a Form CA 61/03 to CASA.

If the Examiner considers that the applicant is not at Level 4, 5 or 6, the Examiner must not give a proficiency endorsement. The applicant should then seek an expert assessment, such as through a CASA approved language school as under (c) below. A FRTO RATING will not be issued unless and until the applicant has passed a practical radiotelephony test and has a valid endorsement of language proficiency at Level 4, 5 or 6.

7.2 At a Flight Test

Authorised Flight Examiners (AFEs) or CASA Flight Operations Inspectors or Authorised Personnel Assessors, who have themselves been assessed as proficient at Level 6 in English and are authorised by CASA, may include assessment of the language proficiency for new or existing holders of licences issued by CASA. This may be done as part of the skill test, proficiency check or assessment of competence that is conducted for the issue, revalidation or renewal of a licence, rating or certificate.

Examiners can only assess candidates as being Level 4, 5 or 6.

Where the assessment was conducted in association with a Skill Test, Proficiency Check, Assessment of Competence then the Examiner will complete the appropriate Examiners Report on form CA 61/03 confirming that the candidate has demonstrated an acceptable level of proficiency.

Where the examiner cannot be satisfied that the applicant is at the minimum acceptable level or better the examiner must not give a proficiency endorsement. In such cases the alternative means of assessment must be used as detailed below.

a) Through a Language School

Applicants choosing to be tested by a language school should verify that the school is approved by CASA for the purpose of language assessment.

b) At an Aviation Training Organisation

There are ATOs specifically approved by CASA to conduct language training modules that meet the requirements of Part 61, as part of an individual's overall training package.

c) Other Acceptable Means

Language proficiency may also be assessed by other means acceptable to CASA. Such means of assessment may be determined by an operator or organisation to make efficient use of their own resources, but in any case must be approved by CASA and meet these requirements before being put into effect.

One such acceptable means for Level 6 'expert speakers' only, is for the holder of a CASA examiner authorisation (as listed in (a) and (b) above) to conduct a language assessment as a face to face aviation-related conversation that is not associated with a Skill Test, Proficiency Check, Assessment of Competence or Flight Radio Telephony practical test.

The face to face language assessment will take the form of an aviation related conversation initially established through a general conversation, then radio phraseology and pronunciation. Discussion topics could include the planning of a navigation exercise; discussing en route navigational issues; NOTAMs and weather; conducting the pre-flight inspection of the aircraft or similar, so that the examiner can assess the candidate's language proficiency in accordance with Appendix 1 to this AC.

The student and the examiner must be together at the same location when the language assessment is to take place, so that the discussion is effectively face-to-face. The examiner must be satisfied that the person presenting themselves for assessment is who they claim to be. This may be achieved by means of photographic identification; (e.g. passport or driving licence with photograph).

8. ICAO Language Proficiency Rating Scale

The Language Proficiency Rating Scale in Appendix 1 to this advisory circular is extracted from Annex 1 to the Convention on International Civil Aviation. The overall proficiency rating is determined by the lowest rating level assigned in any particular category.

The Level 4 (Operational) descriptors are the safest minimum proficiency skill level determined necessary for aeronautical radiotelephony communications and represent the minimum required for a language proficiency level to be endorsed on a licence.

Appendix 1 ICAO English Language Proficiency Rating Scale

Level	Pronunciation	Structure	Vocabulary	Fluency	Comprehension	Interactions
Expert (Level 6)	Pronunciation, stress, rhythm and intonation, though possibly influenced by the first language or regional variation, almost never interfere with ease of understanding.	Both basic and complex grammatical structures and sentence patterns are consistently well controlled.	Vocabulary range and accuracy are sufficient to communicate effectively on a wide variety of familiar and unfamiliar topics. Vocabulary is idiomatic, nuanced, and sensitive to register.	Able to speak at length with a natural, effortless flow. Varies speech flow for stylistic effect, e.g. to emphasize a point. Uses appropriate discourse markers and connectors spontaneously.	Comprehension is consistently accurate in nearly all contexts and includes comprehension of linguistic and cultural subtleties.	Interacts with ease in nearly all situations. Is sensitive to verbal and non-verbal cues and responds to them appropriately.
Extended (Level 5)	Pronunciation, stress, rhythm, and intonation, though influenced by the first language or regional variation, rarely interfere with ease of understanding.	Basic grammatical structures and sentence patterns are consistently well controlled. Complex structures are attempted but with errors which sometimes interfere with meaning.	Vocabulary range and accuracy are sufficient to communicate effectively on common, concrete, and work-related topics. Paraphrases consistently and successfully. Vocabulary is sometimes idiomatic.	Able to speak at length with relative ease on familiar topics but may not vary speech flow as a stylistic device. Can make use of appropriate discourse markers or connectors.	Comprehension is accurate on common, concrete, and work-related topics and mostly accurate when the speaker is confronted with a linguistic or situational complication or an unexpected turn of events. Is able to comprehend a range of speech varieties (dialect and/or accent) or registers.	Responses are immediate, appropriate, and informative. Manages the speaker/listener relationship effectively.
Operational (Level 4)	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding.	Basic grammatical structures and sentence patterns are used creatively and are usually well controlled. Errors may occur, particularly in unusual or unexpected circumstances, but rarely interfere with meaning.	Vocabulary range and accuracy are usually sufficient to communicate effectively on common, concrete, and work-related topics. Can often paraphrase successfully when lacking vocabulary in unusual or unexpected circumstances.	Produces stretches of language at an appropriate tempo. There may be occasional loss of fluency on transition from rehearsed or formulaic speech to spontaneous interaction, but this does not prevent effective communication. Can make limited use of discourse markers or connectors. Fillers are not distracting.	Comprehension is mostly accurate on common, concrete, and work-related topics when the accent or variety used is sufficiently intelligible for an international community of users. When the speaker is confronted with a linguistic or situational complication or an unexpected turn of events, comprehension may be slower or require clarification strategies.	Responses are usually immediate, appropriate, and informative. Initiates and maintains exchanges even when dealing with an unexpected turn of events. Deals adequately with apparent misunderstandings by checking, confirming, or clarifying.

Appendix 2 Guidance for ELP Assessors / Examiners

2.1 General

- a) The language proficiency assessments should be designed to reflect a range of tasks undertaken by pilots/personnel but with specific focus on language rather than operational procedures.
- b) The assessment should determine the applicant's ability to:
 - i) Communicate effectively using standard R/T phraseology;
 - ii) Deliver and understand messages in plain language in both usual and unusual situations that necessitate departure from standard R/T phraseology.

Note: refer to the Manual on the Implementation of ICAO Language Proficiency Requirements (ICAO Doc 9835), Appendix A Part III and Appendix B for further guidance.

2.2 Assessment

- a) The assessment may be subdivided into three elements, as follows:
 - i) Listening: assessment of comprehension;
 - ii) Speaking: assessment of pronunciation, fluency, structure and vocabulary;
 - iii) Interaction.
- b) The three elements mentioned above may be combined and can be covered by using a wide variety of means or technologies.
- c) Where appropriate, some or all of these elements may be achieved through the use of the R/T testing arrangements.
- d) When the elements of the testing are assessed separately the final assessment should be consolidated in the language proficiency endorsement issued by the competent authority.
- e) The assessment may be conducted during one of the several existing checking or training activities, such as licence issue or rating issue and revalidation, line training, operator line checks or proficiency checks.
- f) The competent authority may use its own resources in developing or conducting the language proficiency assessment, or may delegate this task to language assessment bodies.
- g) The competent authority should establish an appeal procedure for applicants.
- h) The holder of a licence should receive a statement containing the level and validity of the language endorsements.
- i) Where the assessment method for the English language established by the competent authority is equivalent to that established for the assessment of use of the English language in accordance with AC61-16, the same assessment may be used for both purposes.

2.3 Basic Assessment Requirements

- a) The aim of the assessment is to determine the ability of an applicant for a pilot licence or a licence holder to speak and understand the language used for R/T communications.
- b) The assessment should determine the ability of the applicant to use both:
 - i) standard R/T phraseology;

- ii) plain language, in situations when standardised phraseology cannot serve an intended transmission.
- c) The assessment should include:
 - i) voice-only or face-to-face situations;
 - ii) common, concrete and work-related topics for pilots.
- d) The applicants should demonstrate their linguistic ability in dealing with an unexpected turn of events, and in solving apparent misunderstandings.
- e) The assessment should determine the applicant's speaking and listening abilities. Indirect assessments, of grammatical knowledge, reading and writing, are not appropriate.
- f) The assessment should determine the language skills of the applicant in the following areas:
 - i) pronunciation:
 - (1) the extent to which the pronunciation, stress, rhythm and intonation are influenced by the applicant's first language or national variations;
 - (2) how much they interfere with ease of understanding.
 - ii) structure:
 - (1) the ability of the applicant to use both basic and complex grammatical structures;
 - (2) the extent to which the applicant's errors interfere with the meaning.
 - iii) vocabulary:
 - (1) the range and accuracy of the vocabulary used;
 - (2) the ability of the applicant to paraphrase successfully when lacking vocabulary.
 - iv) fluency:
 - (1) tempo;
 - (2) hesitancy;
 - (3) rehearsed versus spontaneous speech;
 - (4) use of discourse markers and connectors.
 - v) comprehension:
 - (1) on common, concrete and work-related topics;
 - (2) when confronted with a linguistic or situational complication or an unexpected turn of events.

Note: the accent or variety of accents used in the test material should be sufficiently intelligible for an international community of users.

- vi) interactions:
 - (1) quality of response (immediate, appropriate, and informative);
 - (2) the ability to initiate and maintain exchanges:
 - (a) on common, concrete and work-related topics;
 - (b) when dealing with an unexpected turn of events;
 - (c) the ability to deal with apparent misunderstandings by checking, confirming or clarifying.

Note: the assessment of the language skills in the areas mentioned above is conducted using the rating scale in AC61-16 Appendix 1.

When the assessment is not conducted in a face-to-face situation, it should use appropriate technologies for the assessment of the applicant's abilities in listening and speaking, and for enabling interactions (for example: simulated pilot or controller communication).

2.4 Assessors

- a) It is essential that the persons responsible for language proficiency assessment ('assessors') are suitably trained and qualified. They should be either aviation specialists (for example current or former flight crew members or air traffic controllers), or language specialists with additional aviation-related training. An alternative approach would be to form an assessment team consisting of an operational expert and a language expert.
 - i) The assessors should be trained on the specific requirements of the assessment.
 - ii) The assessors should not test applicants to whom they have given language training.

2.5 Criteria for the Acceptability of Language Assessment Bodies

- a) To ensure an impartial assessment process, the language assessment should be independent of the language training.
 - i) To be accepted, the language assessment bodies should demonstrate:
 - (1) appropriate management and staffing;
 - (2) quality system established and maintained to ensure compliance with, and adequacy of, assessment requirements, standards and procedures.
 - ii) The quality system established by a language assessment body should address the following:
 - (1) management;
 - (2) policy and strategy;
 - (3) processes;
 - (4) the relevant provisions of CASA PNG or ICAO standards and assessment procedures;
 - (5) organisational structure;
 - (6) responsibility for the development, establishment and management of the quality system;
 - (7) documentation;
 - (8) quality assurance programme;
 - (9) human resources and training (initial and recurrent);
 - (10) assessment requirements;
 - (11) customer satisfaction.
 - iii) The assessment documentation and records should be kept for a period of time determined by the competent authority and made available to this competent authority, on request.
 - iv) The assessment documentation should include at least the following:
 - (1) assessment objectives;
 - (2) assessment layout, time scale, technologies used, assessment samples, voice samples;
 - (3) assessment criteria and standards (at least for the levels 4, 5 and 6 of the rating scale mentioned in AC 61-16 Appendix 1;
 - (4) documentation demonstrating the assessment validity, relevance and reliability;
 - (5) assessment procedures and responsibilities:
 - (a) preparation of individual assessment;
 - (b) administration: location(s), identity check and invigilation, assessment discipline, confidentiality or security;

- (c) reporting and documentation provided to the competent authority or to the applicant, including sample certificate;
- (d) retention of documents and records.

Note: refer to the 'Manual on the Implementation of ICAO Language Proficiency Requirements' (ICAO Doc 9835) for further guidance.

Appendix 3 Criteria for Aviation Language Testing Providers

3.1 General

The criteria listed below are formulated as self-contained statements. However, for personnel unfamiliar with the concepts of language testing, they may not all be self-explanatory. Several of the criteria require documented evidence to demonstrate that they have been met. Supplementary information has been provided in order to facilitate the implementation of these criteria as described below.

— **What it means.** Where it is necessary that testing organizations provide documented evidence to demonstrate that a criterion has been met. This paragraph describes the type of information required to complete an informed assessment.

— **Why it is important.** While for language testing experts the significance of the self-contained criterion statement may be obvious, it may not be so for personnel unfamiliar with this discipline. This paragraph justifies why a particular criterion is an essential element of testing best practices.

— **Additional information.** For several criteria, readers may feel they require more information. To assist them, this paragraph provides more explanation or links to references which may be useful.

3.2 Test Design and Construct

3.2.1 The test should be designed to assess speaking and listening proficiency in accordance with each component of the ICAO Language Proficiency Rating Scale and the holistic descriptors in Annex 1.

— **What it means.** Language tests for flight crews and air traffic controllers should specifically address the language skills of the ICAO Rating Scale as well as the holistic descriptors specified in Annex 1.

Testing service providers (TSPs) should be able to explain and justify their methods and approaches to testing with evidence that all components of the ICAO Rating Scale are addressed.

— **Why it is important.** The language proficiency requirements in Annex 1 specify that speaking and listening should be evaluated in the context of operational aviation communications. The holistic descriptors and Rating Scale were developed to address the specific requirements of radiotelephony communications. Each component of the Rating Scale is as important as any other. Tests developed for other purposes may not address the specific and unique requirements of aviation language testing.

— **Additional information.** The SARPs in Annex 1, Chapter 1, 1.2.9, require that the speaking and understanding proficiency of pilots and air traffic controllers be evaluated. Appendix 1 provides a Rating Scale that describes the range of language proficiency levels. Testing speaking and listening proficiency requires procedures that are different from the procedures that are used to test reading, writing or grammar. Testing reading ability, knowledge about English grammar or vocabulary items in isolation from their context is not consistent with the ICAO requirements.

3.2.2 A definition of test purpose that describes the aims of the test and the target population should be accessible to all decision-makers.

— **What it means.** Different tests have different purposes, and different target populations. If an existing test is being considered, it is important that the organization offering the test clearly describes the purpose of the test and the population of test-takers for whom the test was developed.

— **Why it is important.** A clear definition of test purpose and target population is a necessary starting point for evaluating the appropriateness of a test. The purpose and target population of a planned test influence the process of test development and test administration. For example, a test designed to evaluate the proficiency of *ab initio* pilots may be very different from a test developed for experienced or professional pilots; likewise, a test designed to measure pilots' or controllers' progress during a training programme may be inappropriate as a proficiency test for licensing purposes.

3.2.3 A description and rationale for test construct and how it corresponds to the ICAO language proficiency requirements should be accessible to all decision-makers in plain, layperson language.

— **What it means.** There are different approaches to proficiency testing for speaking and listening. Test developers should document the reasons for their particular approach to testing, in language that is comprehensible to people who are not experts in language test design.

— **Why it is important.** A description of the test structure and an easy-to-understand explanation of reasons for the test structure is one form of evidence that it is an appropriate tool for evaluating language proficiency for the ICAO requirements for a given context.

— **Additional information.** Nil.

3.2.4 The test should comply with principles of good practice and a code of ethics as described in Chapter 6 of ICAO Doc 9835.

— **What it means.** It is important for test developers to comply with a recognized code of good practice and ethics.

— **Why it is important.** Aviation language testing is an unregulated industry and has very high stakes. A documented code of good practice and ethics, along with evidence that the organization is adhering to that code, serves as an important stop gap in an unregulated system.

— **Additional information.** The Association of Language Testers of Europe publishes Principles of Good Practice for ALTE Examinations, available at www.ALTE.org.

3.2.5 The test should not focus on discrete-point items, on grammar explicitly or on discrete vocabulary items.

— **What it means.** Discrete-point items are individual test questions which are presented out of context. Examples are a vocabulary test in which test-takers are asked to provide definitions for a list of words, and a grammar test in which test-takers are asked to provide the past-tense forms of a list of irregular verbs. Discrete-point tests, also referred to as indirect tests, do not test language skills directly. Instead, they test individual, specific features of the language thought to underlie language skills. That is, they test knowledge about grammar, vocabulary, pronunciation, etc. This type of test is not appropriate for assessing aviation language proficiency.

— **Why it is important.** The ICAO language provisions focus on the ability to use the language. Discrete point tests do not evaluate a person's ability to use the language. Furthermore, test-takers who perform well on such tests often perform poorly on tests in which they actually have to use the language.

— **Additional information.** There are a number of different ways knowledge about language is tested, for example:

- a) multiple-choice questions in a series of unrelated sentences;
- b) identification of an error in a sentence; or
- c) written translation exercises.

For many people such tests have the advantage of being objective because they give a numerical score. However, the supposed objectivity of multiple-choice tests must be questioned in consideration of the choice of the particular items and questions selected for the test. It may be asked, why were they selected from the infinite number of potential items available? In other words, why were test takers asked to define certain words, or why were they tested on the use of a particular tense but not on their ability to ask clarifying questions? Speaking and listening tests, on the other hand, refer to a scale of proficiency rather than a numerical score.

The rating scale describes levels of proficiency which a panel of trained raters can use to assign the test-taker a level on a rating scale. The more directly a test performance is related to target performance, the more a test can be considered a proficiency test. For example, test administrators interested in an individual's speaking skills should arrange for an assessment of that individual's performance on a speaking task. Using this approach, speaking skills may be directly assessed during an interview or conversation or role-play, or are based on a recorded sample of actual speech. The goal of a proficiency test is to assess the appropriateness and effectiveness of communication rather than grammatical accuracy.

Grammatical accuracy should be considered only so far as it has an impact on effective communication, but evaluating an individual's grammatical knowledge should not be the objective of the test.

3.2.6 If comprehension is assessed through a specific listening section with individual items, it should not be done to the detriment of assessing interaction.

— **What it means.** Some language tests evaluate listening during an oral interaction such as a conversation, interview or role-play. Other language tests evaluate listening separately, in some cases via a series of individual listening items. An example of an individual listening item, in the aviation language context, might require a test-taker to listen to a pre-recorded conversation between ATC and a flight crew to identify relevant pieces of information.

— **Why it is important.** A separate listening test can provide information about comprehension independent of a person's ability to interact. In such tests, the communication is one-way, and the test taker does not have to participate in the way that is required by a conversation, role-play or other interaction.

— **Additional information.** It is important for the TSP to validate the method it uses to evaluate comprehension.

3.2.7 Proficiency tests that are administered directly may use face-to-face communication in some phases of the delivery but should include a component devoting time to voice-only interaction.

— **What it means.** Voice-only interaction is an important characteristic of aeronautical radiotelephony communications; when a pilot and a controller interact, they cannot see each other. Directly administered proficiency tests should simulate this condition of "voice only" in at least a portion of the test.

— **Why it is important.** When two people interact face-to-face, they use non-verbal cues (information other than words) to help them understand each other's messages. People's facial expressions, their body language and the gestures they make with their hands often communicate important information. Aeronautical radiotelephony communications do not benefit from such non-verbal cues; all radiotelephony communications are conveyed through words alone, which can be more difficult to interpret than face-to-face communication.

— **Additional information.** In a test that is administered directly, voice-only interaction can be facilitated by means of a telephone or headset via which the interlocutor and test-taker communicate while positioned in such a way that they cannot see each other.

An appropriate strategy may be to incorporate both direct and semi-direct methods in a single testing system. In any case, the method and approach taken should be clearly justified, with evidence for the rationale of that approach provided.

3.2.8 The test should be specific to aviation operations.

— **What it means.** Tests should provide test-takers with opportunities to use plain language in contexts that are work-related for pilots and air traffic controllers in order to demonstrate their ability with respect to each descriptor in the Rating Scale and the holistic descriptors.

— **Why it is important.** The ICAO Language Proficiency Requirements (LPRs) refer to the ability to speak and understand the language used for radiotelephony communications. It is important that flight crews and air traffic controllers be proficient in the use of plain language used within the context of radiotelephony communications in order to communicate safely on any operational issue that may arise.

— **Additional information.** ICAO language provisions require proficiency in the use of standardized phraseology and in the use of plain language. The assessment of standardized phraseology is an operational activity, not a language proficiency assessment activity.

While an aviation language test may include phraseology to introduce a discussion topic or make interaction meaningful to the testtaker, it is important that tests elicit a broad range of plain language and not be limited to tasks that require standardized phraseology.

The focus of a language proficiency test for compliance with ICAO requirements should be on plain language. The idea of a work-related context can be interpreted in different ways. The narrow view would seek to replicate

radiotelephony communications including both phraseology and plain language, as closely as possible. The broad view would elicit samples of interaction and comprehension on those topics occurring in radiotelephony communications without resorting to replicating radiotelephony communications.

These could be of a general piloting and controlling nature and involve question and answer routines, short reports or problem-solving exchanges, or briefings and reports.

A further step toward providing test-takers with a familiar aviation-related context would be to customize the tests for controllers or pilots. Thus, controllers would have the possibility of taking tests using or referring to a tower, approach or en-route environment; similarly, pilots would be able to take tests using or referring to an approach procedure. These should be seen as adaptations in the interest of the comfort of the test-taker, not as specialized tests of distinct varieties of language proficiency.

3.2.9 It is acceptable that a test contains a scripted task in which phraseology is included in a prompt, but the test should not be designed to assess phraseology.

— **What it means.** An aviation language proficiency test has different aims than a phraseology test. While an aviation language test can include some phraseology as prompts or scene setters, the purpose of the test is to assess plain language proficiency in an operational aviation context.

— **Why it is important.** First, tests of phraseology alone are not suitable for demonstrating compliance with ICAO language proficiency requirements. Second, using phraseology accurately is an operational skill which is very dependent on the operational context; and incorrect usage by a test-taker of a specific phraseology may be an operational error, rather than a language error. Phraseology must be taught and tested by qualified operational personnel.

— **Additional information.** Responses containing elements of ICAO phraseology should not be rated with regard to their procedural appropriateness or technical correctness during language proficiency testing. This practice could introduce confusion between the test-taker's operational knowledge and his/her language proficiency. It could also introduce contradictions between the regulators' established system of operational training/testing and language testing. Because of these contradictions, this practice could result in diminished, rather than enhanced, safety.

If phraseology is included in a test prompt, care should be taken that it is used appropriately and that it is consistent with ICAO standardized phraseology.

3.2.10 The test should not be designed to evaluate the technical knowledge of operations.

— **What it means.** Language tests should not assess either operational skills or the specific technical knowledge of operations. A language test is not an operational or technical knowledge test. For example, a language test item may prompt the test-taker to describe an operational procedure that involves a number of steps. A test-taker may provide a very clear description of that procedure but omit one of the steps. In such a case the rater may not recognize that the omission of that one step was an operational error and penalize the test-taker for that error. In responding to that same test item, another test-taker may correctly identify all the steps of the process (achieving technical accuracy), but do so with problems in pronunciation and fluency based on the ICAO Rating Scale. In this case, because of the test-taker's technical knowledge the rater may, perhaps unconsciously, assign a higher level of language proficiency than the test-taker should receive.

— **Why it is important.** If the distinction between language proficiency and technical knowledge is not very clear to the interlocutor and rater of an aviation language test, it may be easy to confuse one with the other. Such confusion may lead to test-takers getting penalized unfairly for technical errors; or to other test-takers getting rewarded, also unfairly, for their technical expertise. Another potential problem if very specific technical items are included in a language proficiency test is that they may require technical knowledge beyond that of a test-taker; for example, answers to questions concerning ground control procedures may not be known to en-route controllers. As a result, the test-taker may be unable to respond effectively, due to a lack of technical expertise rather than a lack of language proficiency.

— **Additional information.** Based on the above information, a prompt such as “What are the separation minima for aircraft being vectored for an ILS approach?” or “Describe the different flight modes of the A320 flight control system” are therefore not appropriate.

3.2.11 The final score for each test-taker should not be the average or aggregate of the ratings in each of the six ICAO language proficiency skills but the lowest of these six ratings.

— **What it means.** For each test-taker, scores should be reported for pronunciation, vocabulary, structure, fluency, comprehension, and interactions in accordance with the Rating Scale. In cases in which a test-taker is given different ratings for different skill areas — for example, 3 for pronunciation, 4 for vocabulary and structure, and 5 for fluency, comprehension and interactions — the overall score for that test-taker should be the lowest of these scores; in the above example, the test-taker’s overall score would be 3.

— **Why it is important.** This practice is critical because the Operational Level 4 descriptors are developed as the safest minimum proficiency skill level determined necessary for aeronautical radiotelephony communications. A lower score than 4 for any one skill area indicates inadequate proficiency. For example, a pilot with Operational Level 4 ratings in all areas except pronunciation may not be understood by the air traffic controllers with whom that pilot should communicate. In summary, an individual should demonstrate proficiency to at least Level 4 in all skill areas of the ICAO Rating Scale in order to receive an overall Level 4 rating.

3.3 Test Validity and Reliability

3.3.1 A statement of evidence for test validity and reliability should be accessible to all decision-makers, in plain, layperson language.

— **What it means.** In language testing, fairness is interpreted in terms of validity and reliability. Validity refers to the degree a test measures what it is supposed to measure. Reliability refers to the degree that the test produces consistent and fair results. TSPs should supply documented evidence of the validity and reliability of their testing methods.

— **Why it is important.** Aviation language tests have high stakes. It is important for safety and for the integrity of the industry, particularly the operators and for test-takers themselves, that language tests be fair and accurate. Testing systems that are not supported by documented validity and reliability may not provide, or may not seem to provide, fair and accurate results.

— **Additional information.** It is important that evidence for test validity and reliability be written in plain, layperson language. The primary target audience of documents outlining test validity and reliability should be civil aviation authority or licensing personnel rather than language testing experts. Because aviation communication safety is very much in the public interest, it is also appropriate for aviation language testing organizations to make information about the validity and reliability of their tests publicly available.

3.3.2 A description of the development process that includes the following information should be accessible to all decision-makers:

- a) a summary of the development calendar; and
- b) a report on each development phase.

— **What it means.** The TSP should document the entire development process.

— **Why it is important.** Before a decision is made to use a test, its quality should be examined carefully, and documentation of the development process is essential to that examination. A development calendar and report will provide information about the nature and depth of analysis that went into the test development. If it is obvious that a test was developed hastily and without the required expertise, that test may not provide, or seem to provide, valid and reliable results. The same is true of tests with incomplete documentation.

3.3.3 An appraisal of expected test washback effect on training should be accessible to all decision-makers.

— **What it means.** Test washback refers to the effect a test has on a training programme or on students' behaviour. TSPs should demonstrate that their test will have a positive effect on training and that their test will not encourage training that focuses on memorization and test preparation rather than on building proficiency.

— **Why it is important.** The goal of aviation operational language testing is to ensure that flight crews and air traffic controllers have adequate language proficiency for the conduct of safe operations. Robust language training programmes are an essential component of a programme to enable pilots and controllers to achieve ICAO Operational Level 4 language proficiency. High-quality testing will encourage high-quality training.

— **Additional information.** Test-takers naturally will want to prepare for a test. While aviation language test-takers can memorize phraseology, they cannot acquire language proficiency as described in the ICAO LPRs simply by memorizing words and phrases. If pilots or controllers think that certain types of narrow learning or practice activities will best and most readily prepare them for a test, they will be inclined to direct their energies to such activities, potentially at the expense of activities that can genuinely improve their language proficiency.

In the aviation environment, an example may be found in an aviation language test that focuses on the use of phraseology, to the exclusion of plain aviation language. In such a case, learners may focus their learning energies on memorizing ICAO standardized phraseology rather than on genuine language learning activities that will actually improve their English language proficiency.

3.4 Rating

3.4.1 Whether rating is conducted “live” during the assessment, or after the test using recordings of the test performance, the rating process should be documented.

— **What it means.** Some speaking and listening tests rate performance during the test. Others record the test performance and rate performance later. Both rating methods are acceptable, but whichever method is used, the rating process should be explained in test documentation.

— **Why it is important.** Rating is one of the most important steps in language proficiency testing. It is critical to explain how rating is conducted in the testing process to ensure that it is transparent to all stakeholders.

— **Additional information.** One advantage of rating test-takers after the test is that the test-taker's statements can be repeated as necessary for closer analysis by the raters. Another advantage of this method is that the raters do not have to be physically present for the test; in fact, raters can reside in an entirely different location, provided they can receive an audio or video recording of the test and submit their rating reports effectively, for example, electronically. A potential advantage of rating live during the assessment may be greater efficiency.

3.4.2 To fulfil licensing requirements, rating should be carried out by a minimum of two raters. A third expert rater should be consulted in the case of divergent scores.

— **What it means.** Best practice in language proficiency assessment calls for at least two trained and calibrated raters, at least one of whom is a language expert.

— **Why it is important.** Using at least two raters reduces the possibility of rater error and helps to ensure a comprehensive evaluation of each test-taker.

— **Additional information.** Ideally, an aviation language test will have two primary raters — one language expert and one operational expert — and a third rater who can resolve differences between the two primary raters' opinions. For example, there could be a situation where the primary raters agree that in five of the six skill areas a test-taker demonstrates Level 4 proficiency; however, the first rater assigns the test-taker a score of 3 on pronunciation (thereby making the test-taker's overall language proficiency level “3”) and the second rater assigns the test-taker a “4” for pronunciation. A third rater would make a final determination for that skill area and, in doing so, would determine the overall score for that test-taker. A third rater would likely be involved in the process only in cases in which a test-taker may obtain an overall rating of 3 or 4, since the difference between these two levels is the most critical distinction for ICAO language proficiency licensing testing.

3.4.3 Initial and recurrent rater training should be documented; the rater training records should be maintained, and audits of raters should be conducted and documented periodically.

— **What it means.** Language proficiency test raters need to be trained, and the raters need to be trained together to ensure they apply the rating scale consistently. Audits should be conducted periodically to check rater performance to ensure it is consistent over time.

— **Why it is important.** When evaluating language proficiency tests, consistency in the rating process is Critical. Unlike other forms of testing, in which one response to a question is correct and another response is incorrect, evaluating language proficiency relies upon subjective judgements by raters. In this context, consistency is achievable through training and experience but easy to lose without regular audits of raters and rating teams. The reliability of test results, and of the test process as a whole, depends on the consistency achieved in the rating process. Audits provide a mechanism for checking consistency and, where consistency has been lost, making adjustments as necessary.

— **Additional information.** Consistency is measured in terms of reliability. Reliability has two components:

- a) **Intra-rater reliability.** The extent to which a particular rater is consistent in using a proficiency scale. In other words, does the rater apply the proficiency scale in a consistent way to all test takers whom that rater is evaluating?
- b) **Inter-rater reliability.** The level of agreement between two or more independent raters in their judgement of test-takers' performance. In other words, are different raters in agreement in the scores that they assign to individual test-takers? Raters' assessments should be monitored, both individually and comparatively, on an ongoing basis. Senior raters should formally evaluate the test-rater staff periodically. Periodic cross-rating by members of different rating teams is also highly recommended as a means to prevent gradual divergence in the interpretation of the rating scale by different teams.

3.4.4 If rating is conducted using new technology, including speech recognition technology, then the correspondence of such rating to human rating, on all aspects of the Rating Scale, should be clearly demonstrated, in layperson language, and be accessible to all decision-makers.

— **What it means.** If a testing organization uses a new technology, such as speech recognition technology, to evaluate the speaking and listening proficiency of a test-taker, then that organization has a responsibility to clearly and plainly demonstrate that the ratings are valid and correspond to the ICAO Rating Scale.

— **Why it is important.** Until now, best practice in testing speaking and listening proficiency has involved the use of experienced and trained raters, who evaluate a person's proficiency based on criteria established in a rating scale. In the context of language testing, the use of speech recognition technology to evaluate human speech is a very new method. The validity and reliability of such testing should be clearly and plainly demonstrated.

— **Additional information.** The ICAO language proficiency requirements will require large-scale testing programmes. If technology can assist by making the test process easier and more cost-effective than person-by-person human rating, then it will be useful. Such testing may be particularly appropriate as a pre-test screen to determine generally those who may be ready for a licensing test and those who require more training.

3.5 Test Administration and Security

Administration

3.5.1 A complete sample of the test should be published, including the following:

- a) test-taker documents (paper instructions, screen display, etc.);
- b) interlocutor instructions or prompts;
- c) rater documentation (answer key, rating scale, instructions);
- d) one complete sample of audio recordings (for listening sections or semi-direct prompts); and
- e) a demonstration of test-taker/interlocutor interaction.

— **What it means.** Decision-makers have a right to examine a complete sample of a test before they adopt, use, take or buy the test. Because of the high-stakes nature of aviation language testing, it is appropriate for testing organizations to make a complete sample of their test publicly available.

— **Why it is important.** Seeing a complete sample of a test is essential for evaluating it. Information about a test, such as a description of the test or a marketing brochure, is not sufficient for determining the test's validity, reliability, practicality and washback effect.

— **Additional information.** It is important to note that for instructors in a training programme, being familiar with the structure and format of a test is not the same thing as “teaching to the test.” Paragraph 3.3.3 cautions against test designs that might provoke test-takers to try to prepare specifically for the test by memorizing phraseology or by memorizing test answers.

Becoming familiar with the format of a test is good practice for both instructors and test-takers; it helps to ensure that test-takers are not unduly surprised or intimidated by the format of the test or the types of interaction it involves. For example, if the test interaction includes a voice-only segment that is conducted by telephone, it is beneficial for test-takers to be aware of this.

Such knowledge does not provide them with anything they can memorize in preparation for the test; it will simply make them comfortable with the test format and the types of interaction they can expect to have during the test.

3.5.2 The test rating process should be documented, and the documentation should include instructions on the extent and nature of evidence that raters should collect.

— **What it means.** Raters should be given clear instructions on the kind of evidence they need to collect to justify and support their evaluations.

— **Why it is important.** Language is complex, and one simple statement by a person can be analysed in many different ways. Raters need to understand the depth of analysis that is expected of them in order to make and justify a rating. Documenting and supporting evaluations of test-takers are also essential in order to later review a test, either to address an appeal or complaint by a test-taker or to audit a rater or rating team (as described in 3.4.3). For such reasons, a documented set of scores alone is not sufficient; evidence and support for that score are required.

Evidence in this context would typically include examples of language use by the test-taker that indicate strengths or weaknesses: several instances of incorrect use of verb tenses, for example, might support a particular structure rating; or a problem pronouncing certain sounds might be documented as evidence for a pronunciation score.

3.5.3 The instructions to the test-taker, the test administration team and test raters should be clearly documented.

— **What it means.** Clear instructions for each part of the test process and for each stakeholder should be available and unambiguous.

— **Why it is important.** Clear instructions demonstrate that the testing organization has thoroughly considered all aspects of the testing process. Test users, test administrators and test raters all need clear, easy-to-understand instructions for their involvement to be effective. In addition, clear instructions are an important feature to ensure tests are administered in a consistent and therefore reliable manner.

3.5.4 The equipment, human resources and facilities necessary for the test should be included in the instructions.

— **What it means.** The administration of tests may require a variety of equipment (computer, videotape, tape recorder), the support of different personnel (information technology personnel or sound technicians) and facilities that can accommodate the required equipment and personnel. Clear instructions for each part of the test process should be available.

— **Why it is important.** Clear descriptions and instructions for the equipment, human resources and facilities required demonstrate that the testing organization has thoroughly considered all aspects of the testing process.

Test users, test administrators and test raters all need clear, easy-to-understand instructions for their involvement to be effective and to ensure that the test is administered in a consistent and therefore reliable manner.

— **Additional information.** These requirements include the room where the test will be conducted, furniture, equipment for playing audio prompts used during the test, headsets (if used) and/or any other resources required by the test.

3.5.5 The testing location should offer moderate comfort, privacy and quiet.

— **What it means.** The testing location should not be uncomfortable or noisy.

— **Why it is important.** Aviation language testing is important. TSPs have an obligation to ensure a fair outcome to the test. This obligation includes eliminating undue distractions during the test.

— **Additional information.** Examples of inappropriate locations would be a staff kitchen, cafeteria, coffee lounge or hallway where people are gathering and talking. Such settings could violate the test taker's privacy and potentially introduce distractions during the test. Similarly, a testing room that is extremely cold or hot could introduce an artificial and distracting condition to the test that could impact the test-taker's performance.

3.5.6 A full description of test administration policies and procedures should be available to all decision-makers, including information about the following:

- a) policies and procedures for retaking the test;
- b) score reporting procedures (who receives the results of tests?);
- c) record-keeping procedures; and
- d) plans for quality control, test maintenance, ongoing test development and purchasing conditions.

— **What it means.** Policies and procedures concerning scores, records, quality control, future development, and purchasing conditions need to be clearly and readily available to decision-makers and test users.

— **Why it is important.** One of the considerations in test development and/or test selection is whether or not there is adequate infrastructure to support and maintain the test goals.

3.5.7 A documented appeals process should be established, and information about it should be available to test takers and decision-makers at the beginning of the testing process.

— **What it means.** All testing programmes should have an appeals process. In some cases, a re-examination may be needed. Test-takers who feel their scores are not accurate may request that their tests be re-rated or that they have the opportunity to take the test again.

— **Why it is important.** Even if the testing process follows best practices, errors may occur. While every appeal should not be expected to result in a complete re-scoring or re-examination, the procedures for an appeal should be clearly documented so that they can be fairly applied when appropriate.

— **Additional information.** An appeals process should address issues such as, but not limited to:

- a) extenuating circumstances that affect the test-taker's performance. Test-takers who claim that they were having a bad day or were nervous should not be allowed an appeal since they will need to communicate in operational situations when they are having a bad day or feeling nervous. But a test-taker who suffers a family tragedy in the days prior to the test, or who is ill on the day of the test, should be at least considered for an appeal;
- b) steps test-takers should take to initiate an appeals process and the communication that they can expect to receive during that process;
- c) the period of time (for example 30 days or 60 days) within which the employer or licensing authority commits to resolving an appeal — either in the form of a re-review of the test, a re-examination or a rejection of the appeal.

Test security

3.5.8 A full description of security measures required to ensure the integrity of the testing process should be documented and available to all decision-makers.

— **What it means.** Test security refers to the ability of the testing organization to protect the integrity of the testing process. Testing organizations should ensure that people do not have access to specific test content or questions before the test event. In addition, TSPs should ensure that test scores are kept confidential.

— **Why it is important.** The ongoing reliability, validity and confidentiality of a language proficiency testing system will depend heavily on the test security measures that are in place.

— **Additional information.** Testing organizations should protect test-item databases and provide secure storage of scores and test materials. They should require, establish and maintain formal commitments to confidentiality and integrity from test developers, administrators, raters, information technology personnel and any other staff who are involved in any aspect of the testing process.

Other necessary security measures during test administration should prevent:

- a) communication between test-takers;
- b) communication between test-takers and people elsewhere during the test (for example, by use of a mobile telephone);
- c) impersonation of others; and
- d) the use of false identities.

Finally, security measures should ensure the authenticity of test result data, including databases and certificates.

3.5.9 In the case of semi-direct test prompts (which are pre-scripted and pre-recorded), there should be adequate versions to meet the needs of the population to be tested with respect to its size and diversity.

— **What it means.** Tests with specific pre-recorded or pre-scripted questions or prompts require multiple versions. Decision-makers need to know that there are adequate versions of the test to ensure security for their particular testing needs.

— **Why it is important.** Once test items have been used, there is the possibility that people may repeat or share the prompts with other test-takers; this would violate the security and validity of the test.

— **Additional information.** It is not practical to prescribe the number of versions or test prompts required for any specific test situation. The determination of what is adequate in any situation is dependent on specific circumstances. Examples of variables that impact adequacy are:

- a) the number of test-takers;
- b) the geographic and organizational proximity of the test-takers. The closer the individuals within the test-taking population, the more likely it is that they will share their testing experience with each other. If people share test information and that same information is used in another test, test-takers have the opportunity to prepare a response for a known test prompt. This is an example of negative test washback; and
- c) the variability inherent in the test design. A test that contains very little variability in prompts (in other words, all test-takers are asked the same questions or very similar questions) will require more frequent version changes than a test in which the interlocutor can, for a particular item, ask the test-taker a variety of questions. It is common in large testing initiatives for a testing service to use a version of a test only once before retiring it. In other cases, a testing service develops a number of versions,

then recycles them randomly. Test-takers may then generally know the sorts of questions and prompts they will encounter during a test, but will be unable to predict the specific questions and prompts they will encounter during a particular testing interaction.

One security measure that testing organizations may take is to always include at least one completely new prompt or question in every version. A pattern of test-takers achieving high scores on most or all test prompts or questions, but failing the new prompt, may indicate a breach in test security.

3.5.10 Test questions and prompts should be held in confidence, and not be published or provided to test-takers prior to the test event.

— **What it means.** Test-takers should not have access to test questions or prompts before they take the test.

— **Why it is important.** Authorities and organizations that make test items publicly available negatively impact the integrity of the testing process. Test takers' prior knowledge of specific test content does not allow them to "recognize and resolve misunderstandings" and to "handle successfully and with relative ease the linguistic challenges presented by a complication or unexpected turn of events" in accordance with the ICAO language proficiency requirements. This approach will lead test-takers to memorize items and responses.

— **Additional information.** As mentioned in A.3.5.1, one sample version of the test should be provided to decision-makers so that they are familiar with the format of the test and the general test procedures. Specific test questions or prompts from actual tests should not be available in any way.

3.5.11 A documented policy for all aspects of test security should be accessible to all decision-makers.

— **What it means.** TSPs should clearly describe in publicly available documents how they establish and maintain all required aspects of test security.

— **Why it is important.** A testing process with inadequate or unknown safeguards for test security will not be recognized as generating valid results or ensuring a test-taker's confidentiality.

— **Additional information.** All test materials, including paper documents and electronic versions, should be stored securely at all times by all stakeholders involved in test administration processes. Periodic reviews, in the form of physical inspections, should be conducted by testing management personnel to verify that security procedures, including storage of all test materials, are being followed.

3.6 Record-Keeping

3.6.1 All proficiency tests of speaking ability involving interaction between the test-taker and interlocutor during the test should be recorded on audio or video media.

— **What it means.** Because of the high-stakes nature of aviation language testing, it is critical that test organizations maintain either video or audio recordings of all speaking tests.

— **Why it is important.** Test recordings provide a safeguard against charges of subjective judgements and unfairness. Recordings allow:

- a) review or re-rating by different raters in case of uncertainty or an appeal; and
- b) confirmation of assessments in case of appeals by test-takers or their employers.

3.6.2 Evaluation sheets and supporting documentation should be filed for a predetermined and documented period of time of sufficient duration to ensure that rating decisions can no longer be appealed.

— **What it means.** In addition to preserving the actual recording of each speaking test, for each test taker, all score sheets and supporting documentation, including electronic data, should be filed and retained for an appropriate duration of time.

— **Why it is important.** Records are important in the case of appeals, for internal analysis related to auditing, for establishing an individual training plan and for establishing recurrent testing schedules.

— **Additional information.** At a minimum, the records should be maintained through the validity period of the licence's language proficiency endorsement requirement. Part 61 requires that a validity period of three years for those evaluated at Level 4 or 5. The record-keeping process should be adequate for the scope of the testing and should be documented.

3.6.3 The record-keeping process should be adequate for the scope of the testing and should be documented.

— **What it means.** A testing service should document how a test-taker's performance can be captured and securely stored.

— **Why it is important.** Decision-makers need to know if the record-keeping processes are adequate.

— **Additional information.** The outcome of the operational language assessment should comprise written comments on language performance in each skill area of the ICAO Rating Scale as well as the test result in terms of the demonstrated level of proficiency. In case of uncertainty, documentation should include a recommendation for assessment by a specialized language test or by another rating team.

3.6.4 The score-reporting process should be documented and scores maintained for the duration of the licence.

— **What it means.** The method of scoring and the persons to whom scores are reported should be clearly documented. When a test has been rated and the results documented, the process for reporting should be clear to all decision-makers.

— **Why it is important.** This practice is important to ensure that those individuals in the organization who need to know receive test result information and to ensure that the privacy of the test-taker and the security of the information are maintained.

3.6.5 Results of testing should be held in strict confidence and released only to test-takers, their sponsors or employers, and the civil aviation authority, unless test-takers provide written permission to release their results to another person or organization.

— **What it means.** The licensing authority should ensure that a policy concerning the release of test results is established. The TSP should have documented procedures on how it manages recordkeeping and the confidentiality of test results.

— **Why it is important.** A confidentiality policy on test results is a key measure the licensing authority should use to manage the impact of aviation language testing on the career of a flight crew or controller and the safety of passengers. A TSP should provide documented evidence on how it manages confidentiality of test results through every step of the testing process, including how it intends to transmit test results to the licensing authority.

3.7 Organizational Information and Infrastructure

3.7.1 An aviation language TSP should provide clear information about its organization and its relationships with other organizations.

3.7.2 All associations or links with other organizations should be transparent and documented.

— **What it means.** In developing and administering their aviation language tests, TSPs may partner with other organizations in order to enhance their credibility with the aviation community. TSPs should provide documentation of any and all organizational links to other organizations.

— **Why it is important.** In any high-stakes testing environment, relationships between a TSP and other organizations can compromise the integrity of the testing process. For example, the CASA might reject a TSP because it does not follow good testing practices; subsequently, that provider could change its name or form another organization, re-package its test and sell the same testing system (which still does not conform to good testing practices) to the CASA via deceptive marketing practices.

In order to prevent this type of deception, the provider should be required to document any other names under which it is conducting business or has conducted business in the past. The CASA should, in any case, conduct inquiries into all TSPs whose services are being considered in order to establish their legitimacy. A related issue concerns claims made by TSPs about their relationships with leading industry entities. TSPs might, for example, make claims such as “Our test is endorsed by FAA” or “Advised by NASA.”

In such cases, the provider should be required to supply documentation that explains and supports the claim, and the decision-makers should contact the related organization to validate the claim.

— **Additional information.** The assessment of language proficiency for the endorsement of licences is the responsibility of Contracting States. ICAO does not accredit, certify or endorse language TSPs.

3.7.3 If a TSP is also a training provider, there should be a clear and documented separation between the two activities.

— **What it means.** A clear separation between testing and training activities should be documented by an organization that provides both services.

— **Why it is important.** Typically, in high-stakes testing situations, testing and training should be clearly separated in order to avoid conflicts of interest. Two examples of conflicts of interest follow. An organization that provides both training and testing services could award higher scores to students in its training programme since low scores for those students could reflect badly on the training they have received. Conversely, the organization could assign lower scores to test-takers, if additional training for those test-takers would result in increased revenues for the organization’s training programme.

Another concern regarding organizations that provide both training and testing services is the potential for training staff to also serve as interlocutors and raters in the testing process. It is never acceptable for instructors to also be testers of their own students. There is a natural inclination for instructors to develop sympathies toward some students while perhaps regarding others less favourably. Such perceptions could interfere with the objectivity that is required of interlocutors and raters in the testing process.

3.7.4 The TSP should employ sufficient numbers of qualified interlocutors and raters to administer the required tests.

— **What it means.** In addition to developing tests and new test versions, it is important that testing services have enough staff members to administer and rate the tests.

— **Why it is important.** Raters and interlocutors administering or evaluating speaking proficiency tests are usually effective only five to six hours per day. After that, tester fatigue is likely to have an impact on their effectiveness, and their interactions and ratings may become less reliable. Testing organizations should provide evidence that they have enough trained and qualified staff to manage the volume of required tests.

3.7.5 Documentation on how the test is maintained, including a description of how ongoing test development is conducted, should be provided.

— **What it means.** A testing organization should plan not only for the development of an initial test, but it should plan and budget for ongoing test development.

— **Why it is important.** An effective test that is not supported by adequate ongoing test development will not remain effective for very long. In a short period of time, test-takers will be able to predict the test items they will be presented with and memorize responses to those items.

— **Additional information.** New test versions will constantly need to be developed. Ongoing test development should also include the creation and maintenance of a database containing all questions that have appeared on each version of a test. This practice will help to ensure that test items, or whole test versions, are not accidentally recycled as subsequent versions are developed. This practice will also enable the testing team to analyse which test items were most successful in eliciting appropriate language responses from the test-taker and those that were less successful and thus develop improved tests subsequently.

3.8 Testing-Team Qualifications

3.8.1 Guidance on the required qualifications for test development, design and administration teams as well as for organizations that aim to contract TSPS is provided below. Within a testing team, the same person may possess several areas of expertise or play several roles. The testing team should include test designers, developers, administrators, interlocutors and raters.

Familiarity with ICAO documentation

3.8.2 All members of the testing team should be familiar with the following ICAO publications:

- a) the relevant Standards and Recommended Practices of Annex 1;
- b) holistic descriptors (Appendix 1 to Annex 1) and the ICAO Rating Scale (Attachment A to Annex 1);
- c) Doc 9835; and
- d) ICAO Rated Speech Samples CD.

Test design and development team

3.8.3 The test design and development team should include individuals with the operational, language test development, and linguistic expertise described below:

- a) Operational expertise:
 - 1) radiotelephony experience as a flight crew member, air traffic controller or aeronautical station operator;
 - 2) experience in aeronautical operations and procedures and working knowledge of current practices.
- b) Language test development expertise:
 - 1) specialization in language test development through training, education or work experience; 2) working knowledge of the principles of best practice in language test development.
- c) Linguistic expertise:
 - 1) working knowledge of the principles of theoretical and applied linguistics;
 - 2) working knowledge of the principles of language learning;
 - 3) experience in language teaching.

— **Why it is important.** A test design and development team that includes all the above types of expertise offers the best foundation for a successful test development project.

Test administration team (Administrators and interlocutors)

3.8.4 Test administrators (the people who supervise and manage the administration of tests) and interlocutors should have a working knowledge of the test administration guidelines published by the test organization.

3.8.5 Interlocutors should demonstrate language proficiency of at least ICAO Extended Level 5 in the language to be tested and proficiency at Expert Level 6 if the test is designed to assess ICAO Level 6 proficiency.

3.8.6 Interlocutors should have successfully completed initial interlocutor training.

3.8.7 Interlocutors should successfully complete recurrent interlocutor training at least once each year.

3.8.8 Interlocutors should have appropriate aviation operational or language testing expertise, or both.

Rater team

3.8.9 In 3.4.2 it is recommended that at least two raters should evaluate language tests: one with operational expertise and the other with language specialist expertise.

- a) **Operational expertise.** The involvement of operational experts such as pilots, controllers and flight instructors or examiners in the rating process will add operational integrity to the process. Operationally experienced raters can also assist by making informed judgements from an operational perspective on such aspects of language use as conciseness (exactness and brevity) in speech and intelligibility of accents and dialects that are acceptable to the aeronautical community.
- b) **Language specialist expertise.** Because language testing for licensing requirements will have an impact on the professional careers of the test-takers as well as the reputations of operators and service providers and, ultimately, the safety of passengers and flight crews, test raters should be able not only to correctly interpret the descriptors of the Rating Scale but also to accurately identify strengths and weaknesses in a test-taker's performance. Only qualified language specialists serving as raters can identify and describe these strengths and weaknesses.
It may be true that laypersons or inexperienced raters (people with no academic training or qualifications in language teaching or testing) can make informal judgements about language proficiency, particularly in a pass/fail sense. However, test takers who do not pass a high-stakes test will demand, and will deserve, accurate information about how their performance did not meet the target performance (in this case, Level 4 language proficiency) and the areas in which they should focus their efforts to improve performance.
Likewise, detailed justifications for giving a test-taker a passing score (in this case, an overall language proficiency score of 4, 5 or 6) will need to be documented and archived.

3.8.10 Raters should demonstrate language proficiency of at least ICAO Extended Level 5 in the language to be tested. If the test is designed to assess ICAO Level 6 proficiency, raters should demonstrate language proficiency at ICAO Expert Level 6.

— **What it means.** In order to credibly and effectively evaluate test-takers' language proficiency, raters should at least demonstrate the highest level of proficiency that test takers can achieve during assessment.

— **Why it is important.** To ensure safety, pilots and air traffic controllers expect the examiners and inspectors that assess them during operational training, and periodically thereafter, to meet stringent requirements. The assessment of language proficiency should follow the same practice given the high stakes involved. In addition, test-takers may question the validity and reliability of the test and testing process if they have doubts concerning the credibility and qualifications of the rater.

3.8.11 Raters should be familiar with aviation English and with any vocabulary and structures that are likely to be elicited by test prompts and interactions.

— **What it means.** In order to credibly and effectively evaluate test-takers' language proficiency, raters should be familiar with the vocabulary and structures that test-takers are likely to use during the test.

— **Why it is important.** Communication between pilots and controllers is highly specialized; it includes terms that are specific to aviation (approach fix, hold position, etc.) as well as everyday words and structures that have singular and distinctive meanings for pilots and controllers (e.g. approach, cleared). A rater who is unfamiliar with these terms may be confused or distracted by them during a test interaction; similarly, a rater who does not understand how pilots and controllers interact with each other may have difficulty comprehending statements made by test-takers. In cases such as these, the rater may be unable to effectively evaluate the language proficiency of test-takers in this environment.

— **Additional information.** The rater training process should include an aviation familiarity component, so that raters can comprehend, as much as their role requires, technical aspects of the language they will hear during tests.

3.8.12 Raters should have successfully completed initial rater training.

3.8.13 Raters should successfully complete recurrent rater training at least once each year.

— **Why it is important.** Initial and recurrent training aiming to standardize rater behaviour is vital to objectivity. As a language testing standard, raters should undergo approximately 40 hours of initial rater training and 24 to 40 hours of recurrent training per year.