

Cir 318
AN/180



Language Testing Criteria for Global Harmonization

**Approved by the Secretary General
and published under his authority**

**Cir 318
AN/180**



Language Testing Criteria for Global Harmonization

**Approved by the Secretary General
and published under his authority**

International Civil Aviation Organization



Documents
Document Sharing Hub

Published in separate English, Arabic, Chinese, French, Russian and Spanish editions
by the INTERNATIONAL CIVIL AVIATION ORGANIZATION
999 University Street, Montréal, Quebec, Canada H3C 5H7

For ordering information and for a complete listing of sales agents
and booksellers, please go to the ICAO website at www.icao.int

ICAO Cir 318, *Language Testing Criteria for Global Harmonization*

Order Number: CIR318

ISBN 978-92-9231-271-8

© ICAO 2009

All rights reserved. No part of this publication may be reproduced, stored in a
retrieval system or transmitted in any form or by any means, without prior
permission in writing from the International Civil Aviation Organization.

FOREWORD

1. Assembly Resolution A36-11, *Proficiency in the English language used for radiotelephony communications*, directed the Council to support Contracting States in their implementation of the language proficiency requirements by establishing globally harmonized language testing criteria.
 2. Organizing aviation language testing is one of several steps necessary to effectively implement the ICAO language proficiency requirements. ICAO has previously published detailed "Implementation Guidelines" for aviation language requirements; these are available at <http://www.icao.int/fsix/lp.cfm>.
 3. While the *Manual on the Implementation of ICAO Language Proficiency Requirements* (Doc 9835), published in September 2004, provided some guidance on testing, users of the manual — including licensing authorities, air operators, air navigation service providers, and language training and testing services — have indicated that more detailed guidance on language testing is needed to effectively implement the language proficiency requirements. The purpose of this circular is to address that need.
-

TABLE OF CONTENTS

	<i>Page</i>
Chapter 1. Introduction	1
1. Overview	1
2. Background	2
3. Introduction to Language Testing	4
4. Aviation-Specific Language Testing Issues	8
 Chapter 2. Recommended Criteria for Aviation Language Testing	 11
1. Test Design and Construct	11
2. Test Validity and Reliability	16
3. Rating	17
4. Test Administration and Security	19
5. Organizational Information and Infrastructure	25
6. Testing Team Qualifications	26
 Chapter 3. Checklist	 30
1. Test Design and Construct	30
2. Test Validity and Reliability	31
3. Rating	31
4. Test Administration and Security	32
5. Organizational Information and Infrastructure	34
6. Testing Team Qualifications	34
 Glossary of Language Proficiency and Language Testing Terms and Acronyms	 36

Chapter 1

INTRODUCTION

1. OVERVIEW

1.1 Purpose

The purpose of this circular is to provide guidance to civil aviation authorities and test service providers in processes for testing candidates in accordance with the ICAO language proficiency requirements. In particular, it provides recommended criteria to guide the development or selection of aviation language testing programmes, as well as additional guidance material in that regard. The recommended aviation language testing criteria outlined in this circular were drawn from principles of best practices by the Proficiency Requirements in Common English Study Group (PRICESG) in 2005. They are intended to support the harmonization of global aviation language testing.

1.2 Target Audience

1.2.1 This publication will be useful to civil aviation and licensing authorities that oversee language testing and will provide State authorities, airlines and air navigation service providers with a set of practical tools. Civil aviation and licensing authorities may use these criteria:

- a) as a guide to development, should they decide to assign national resources to develop aviation language testing; and
- b) as a checklist against which to compare and assess externally developed aviation language tests;

1.2.2 Language testing organizations may use the criteria as a guide to provide the information and evidence to document that they comply with the criteria and to establish the integrity of their test.

1.3 Applicability to all languages

The ICAO language proficiency requirements (LPRs), which apply to all languages used in international radiotelephony communications, create a significant testing requirement. This is particularly true with respect to English which is the language for which most training and testing programmes need to be developed. While this circular focuses on criteria guiding the development or selection of language tests in English, the principles apply equally to tests developed for any language used for international radiotelephony communications.

1.4 Context

The recommended criteria in this guidance material are considered appropriate to the diverse contexts in which aviation language testing occurs. The principles underlying these criteria suit various operational and regulatory needs at various points of application within each particular administration.

1.5 Structure

Chapter 1 – Introduction. Presents the background and context, with references to other appropriate ICAO publications and guidance material.

Chapter 2 – Recommended Criteria for Aviation Language Testing. Presents the criteria. Each criterion includes: what it means, why it is important and, where applicable, additional information.

Chapter 3 – Checklist. Provides the recommended criteria in checklist format

Glossary of Language Proficiency and Language Testing Terms and Acronyms

2. BACKGROUND

2.1 ICAO language proficiency requirements

2.1.1 The decision to address language proficiency for pilots and air traffic controllers was first made by the 32nd Session of the Assembly in September 1998 as a direct response to several fatal accidents, including one that cost the lives of 349 persons, as well as to previous fatal accidents in which the lack of proficiency in English was identified as a contributing factor.

2.1.2 Subsequently, the Air Navigation Commission initiated the development of language provisions in the following Annexes to the Convention on International Civil Aviation and PANS:

- a) Annex 1 — Personnel Licensing;
- b) Annex 6 — Operation of Aircraft;
- c) Annex 10 — Aeronautical Telecommunications, Volume II; and
- d) Annex 11 — Air Traffic Services
- e) PANS-ATM

2.1.3 In March 2003, the ICAO Council adopted a comprehensive set of Standards and Recommended Practices (SARPs) that strengthened language proficiency requirements for pilots and air traffic controllers involved in international operations. These language proficiency requirements affirmed that ICAO standardized phraseology should be used whenever possible and required that when phraseology is not applicable, pilots and air traffic controllers should demonstrate a minimum level of proficiency in plain language. The effective use of plain language is vital in routine operational situations in which phraseology provides no “ready-made” form of communication and is especially critical in unusual or emergency situations. The minimum skill level requirements are embodied in the ICAO language proficiency rating scale and the holistic descriptors that appear in Attachment A and Appendix 1 of Annex 1, respectively.

2.1.4 As of 5 March 2008, the ability to speak and understand the language used for radiotelephony that is currently required for pilots, air traffic controllers and aeronautical station operators should be demonstrated based on the holistic descriptors and language proficiency rating scale to at least Level 4. Level 4 is considered the minimum level of proficiency to ensure an acceptable level of safety. Additionally, since November 2003, Annex 10, Volume II, has required the availability of English at all stations on the ground serving designated airports and routes used by international air services.

2.2 ICAO support and guidance

ICAO has provided Contracting States with implementation support in a number of ways, primarily through symposia, regional seminars and workshops, and the publication of guidance material:

- a) First ICAO Aviation Language Symposium — September 2004, Montréal;
- b) Doc 9835, *Manual on the Implementation of ICAO Language Proficiency Requirements*, published in September 2004;
- c) Regional seminars — Japan (2004); Argentina, Azerbaijan, Belgium, Ukraine (2005); China (Hong Kong), France, Mexico, Senegal (2006); Egypt, France, Germany (2007);
- d) Rated Speech Samples CD — A set of Rated Speech Samples developed and published by ICAO in 2006. Samples of speech at ICAO Pre-Operational Level 3, Operational Level 4, and Extended Level 5 are provided, with full explanations and rationale for the assignment of each level. This CD is available for purchase from ICAO (contact: sales@icao.int);
- e) Second ICAO Aviation Language Symposium — May 2007, Montréal;
- f) Regional workshops on the development of States' implementation plans for language proficiency requirements — Belarus, Peru (2007); El Salvador, Mexico, Senegal, Thailand, Uganda, United Arab Emirates (2008);
- g) Implementation of language proficiency requirements website (<http://www.icao.int/fsix/lp.cfm>) — Further to the adoption of Assembly Resolution A36-11, States can find information concerning States' level of compliance with the language proficiency requirements and their implementation plans, as well as other implementation guidance, on this website; and
- h) ICAO Frequently Asked Questions (FAQs) website (<http://www.icao.int/icao/en/trivia/peltrgFAQ.htm>) — updated with information regarding the ICAO language proficiency requirements.

2.3 ICAO Assembly Resolution A36-11

Assembly Resolution A36-11, *Proficiency in the English language used for radiotelephony communications*, directed the Council to support Contracting States in their implementation of the language proficiency requirements by establishing globally harmonized language testing criteria. The circular is a direct outcome of A36-11.

2.4 Frame of reference

2.4.1 While some regional and national language testing certification programmes exist and some testing programmes are self-regulated, no universal system of aviation language testing certification has yet been developed.

2.4.2 The recommended language testing criteria presented herein are aligned with other ICAO guidance material. This circular should be considered a supplement to and not a replacement for other ICAO publications. Users should be familiar with the following:

- a) Doc 9835, *Manual on the Implementation of ICAO Language Proficiency Requirements*;
- b) ICAO Language Proficiency Requirements Rated Speech Samples; and
- c) the Frequently Asked Questions page at <http://www.icao.int/icao/en/trivia/peltrgFAQ.htm>.

2.5 Scope

2.5.1 The recommended criteria and introductory text in this chapter are intended as a guide only. The scope of this circular does not allow an exhaustive treatment of language testing. Neither this circular nor Doc 9835 is intended to replace the more extensive language testing standards, guidelines, and principles of ethics and good practice that can be found in the literature on language testing.

2.5.2 Language testing is a specialized discipline. Expert professional input is recommended at every level of aviation language test implementation and selection but is essential for test development.

3. INTRODUCTION TO LANGUAGE TESTING

3.1 Language testing standards

3.1.1 Information about generic international language testing standards can be found on the websites of a number of testing associations such as:

- a) Association of Language Testers in Europe (ALTE) — <http://www.alte.org>; and
- b) International Language Testing Association (ILTA) — <http://www.iltaonline.com>

3.1.2 However, it is important to recognize that existing academic or general-purpose language tests are not appropriate for the specialized domain of aviation language testing. The specific needs for aviation language testing are described below.

3.2 High stakes

3.2.1 A number of factors make language proficiency testing for compliance with Annex 1 licensing requirements a case of exceptionally high-stakes testing. Inadequate aviation language testing could result in either serious safety gaps or highly negative social and economic consequences.

3.2.2 The results of language testing seriously impact both individuals and organizations. A pilot or controller operating internationally who does not demonstrate compliance with the ICAO language proficiency requirements may be denied a licence to operate internationally, a consequence that may severely impact the career of that individual, as well as the staffing requirements of the airline or air traffic service provider for whom the individual works.

3.2.3 In addition, the safety of airline passengers depends, among other issues, on the effectiveness of pilot and air traffic controller communications. Efficient transfer of operational information is vital. When the language used in radiotelephony communications is English, then reliable, effective, and valid testing systems are required to ensure that pilots and controllers have adequate levels of English language proficiency.

3.2.4 Finally, there are economic factors to consider. State authorities, airlines, and service providers have no funds to waste on inadequate or unproven tests, neither can they afford to lose otherwise competent staff as an outcome of inadequate testing. Ultimately, they cannot afford accidents attributable to ineffective pilot/controller communication.

3.2.5 For all of these reasons, it is vital that language testing for licensing purposes comply with best practices and address the specific requirements of aviation operations.

3.3 Fundamental principles of language testing

3.3.1 There are, however, well-established principles and practices on which there is widespread professional agreement. These principles and practices, which have been incorporated into this circular, provide the recommended framework for the development and administration of aviation language tests.

3.3.2 The overriding concern of high-stakes test developers should be fairness which, in language testing, is interpreted in terms of validity and reliability. Practicality is also a fundamental test consideration. All tests should be evaluated in terms of their validity, reliability, and practicality based on documented evidence.

- **Validity.** Validity indicates the degree to which a test measures what it is supposed to measure. To this end, testers should gather and provide evidence to support the conclusions that are made about an individual's English language proficiency based on test performance.
- **Reliability.** Reliability refers to the stability of a test. Evidence should be provided that the test can be relied upon to produce consistent results. Reliability is usually reported in the form of a coefficient that can range from 0.0 to 1.0. Although no test will achieve a perfect reliability (1.0), tests with reliability coefficients as close to 1.0 as possible would be the most advantageous.

There are a number of standard measures used in language test development to evaluate the reliability of a test. One method is to compare two versions of a test: the version used by one test-taker with the version used by a different test-taker. If the test is reliable, the two sets of items should be equal in difficulty and complexity. Another method of evaluating the reliability is to compare the results of a group of test-takers on a test with the results of the same group of test-takers on another established test.

For more information about evaluating validity and reliability, refer to the document "Principles of Good Practice for ALTE Examinations", which is available at http://www.alte.org/quality_assurance/code/good_practice.pdf.

- **Practicality.** Practicality refers to the balance between the resources required to develop and support a test (including the funds and the expertise) and the resources available to do so.

3.4 Test washback

3.4.1 Another important consideration related to test design is the negative or positive washback effect on training. The washback effect of testing can be described as the influence of testing on teaching and learning. It is reflected in the way trainers tend to model their curriculum around the focus areas, form and content of an examination or a test; or in the way that learners modify their learning strategies in order to succeed on a particular form of test rather than concentrating on mastering the content and skills addressed in the test.

3.4.2 A valid test designed to match the construct and content being taught (i.e. communicative language skills as defined in the rating scale) will foster a positive washback effect.

3.4.3 In contrast, an example of negative washback can be found in the older forms of the Test of English as a Foreign Language (TOEFL). The TOEFL includes a large number of discrete-point grammar questions (multiple choice or error recognition). As a result, students often neglect to learn the full range of communicative skills contained in the syllabus and prefer to spend time completing TOEFL practice tests, believing this will be an easier way to achieve a good test score. However, research indicates that such activities do not, on average, improve proficiency levels.

3.4.4 Test designers have a particular responsibility to foster positive washback as the testing process may have a considerable impact upon:

- a) the validity of the test itself (are test results purely a consequence of practicing for the test, or are they a true reflection of an ability to use the language?); and
- b) the way in which training is provided for the level and breadth of proficiency required to meet the standards defined by ICAO in the rating scale.

3.4.5 In summary, well-designed aviation language proficiency tests will encourage learners to focus on proficiency-building language learning activities.

3.5 Test purpose and test types

3.5.1 Tests may serve a number of different purposes which would have an influence on the test development process. Some types of common language tests and their purposes include the following:

- a) *Diagnostic*. To identify strengths and weaknesses and to assess gaps.
- b) *Placement*. To place students into the appropriate level of a training programme.
- c) *Progress*. To measure learning progress.
- d) *Achievement*. To measure what students have learned.
- e) *Aptitude*. To assess an individual's ability to acquire knowledge or learn new skills.
- f) *Proficiency*. To evaluate overall ability against a set of criteria.

3.5.2 A need for language testing may occur at a number of points in time in the career of a pilot or air traffic controller:

- a) as a screen for pre-training selection;
- b) as a diagnostic tool in a training programme;
- c) as a progress check during training;
- d) as a licensing requirement in fulfilment of Annex 1 requirements; or
- e) as a periodic re-evaluation of proficiency.

3.5.3 The ICAO Language Proficiency Standards and Recommended Practices (SARPs) in Annex 1 require proficiency testing to fulfil licensing requirements.

3.5.4 Proficiency testing is different from progress or achievement testing in that proficiency tests do not correspond directly to a training curriculum; that is, it should not be possible for test-takers to directly prepare or study (by memorizing information, for example) for a proficiency test. Proficiency tests require test-takers to demonstrate their ability to do something representative of the full spectrum of required knowledge and skills, rather than to simply demonstrate how many of a quantifiable set of curriculum learning objectives they have learned. In an aviation context, proficiency testing should establish the ability of test-takers to effectively use appropriate language in operational conditions.

3.5.5 Proficiency tests are the only tests that are suitable for licensing purposes in the aviation community. Because licensing plays a critical role in the safety of aviation operations, this circular focuses on proficiency testing.

3.6 Delivery method

3.6.1 Speaking and listening proficiency tests can be delivered through direct or semi-direct testing. The primary difference between these two testing techniques lies in how speech samples are elicited: that is, in how the "prompts to speak" are delivered to the test-taker. Direct speaking tests involve face-to-face or telephonic interactions between the test-taker and the interlocutor, who may also serve as a rater. In semi-direct testing, test prompts and questions are pre-recorded, and test-takers' responses are recorded for evaluation at a different time and, in some cases, a different place.

3.6.2 Despite their different attributes, both live and recorded testing procedures share a common purpose: the assessment of an individual's speaking and listening abilities.

Direct testing

3.6.3 In direct testing procedures, the test-taker interacts with a "live" interlocutor, who may also be an assessor or rater. The person-to-person interaction in a direct testing procedure may be directly observed and assessed in real time by a rater or can be recorded for subsequent rating. Test-takers are asked to perform language tasks based on a set of elicitation prompts. A prompt may be a question asked by, or a topic given by, an interlocutor. The test-taker may be asked, for example, to engage in a conversation-like interview with the interlocutor or may be asked to perform in a role play.

3.6.4 One benefit of direct testing is that the test tasks can be made more natural or more communicative, as the test-takers interact with an interlocutor. Another benefit is that there is an infinite supply of test prompts available because each test is a unique interaction between the interlocutor and the test-taker. For example, if a test-taker mentions during a test that his father was an air traffic controller, the interlocutor could ask the test-taker questions related to that information — questions which the interlocutor may not ask any other test-taker. In a direct test, there is also less likelihood of a test-taker responding with rehearsed speech samples in an effort to convince an examiner of a higher level of proficiency than actually attained.

3.6.5 Direct tests require particular attention to the standardization of design and administration procedures, notably with regard to the management of time, the nature and content of language input, and overall interlocutor behaviour. This is to avoid any bias that may inadvertently arise due to the human element of the test interaction. For example, an interlocutor may, without realizing it, ask more demanding questions of one test-taker than another; or one interlocutor may speak more clearly or more slowly than another interlocutor.

3.6.6 Because direct testing requires person-to-person interactions, the administration or delivery of the test tends to be more time-consuming and human resource-intensive than semi-direct testing.

Semi-direct testing

3.6.7 In semi-direct testing, speech samples are elicited through pre-recorded and therefore standardized prompts. This is a significant benefit in that every test-taker receives the same or similar prompts, facilitating fairness. Another advantage is that the test can be administered in an audio or computer laboratory so that a larger number of test-takers can be tested at the same time.

3.6.8 However, the inflexibility arising from the use of standardized, pre-recorded prompts may limit the scope of evaluation available from semi-direct tests. This limitation may be particularly critical in the ability of the test to assess the full range of abilities covered by the "interactions" descriptors of the ICAO rating scale. Role plays and simulations conducted in this mode may be short, unnatural and restricted to the most routine aspects of language use.

Overall

3 6 9 Whether direct or semi-direct testing methods are used, it is important that test-takers are evaluated in their use of language related to routine, as well as unexpected or complicated, situations as evidence of their level of proficiency. Both direct and semi-direct tests, if well constructed, can elicit speech samples that may be assessed for proficiency in speaking and listening. Each test method has advantages and disadvantages.

4. AVIATION-SPECIFIC LANGUAGE TESTING ISSUES

4 1 Beyond the best practices of generic language testing, there are fundamental constraints specific to the ICAO language proficiency testing requirements. These concern the following:

- a) test focus,
- b) test content, particularly concerning the role of standardized phraseology in aviation language testing,
- c) test tasks, and
- d) testing for Expert Level 6 proficiency.

Test focus

4 2 The ICAO language proficiency requirements focus on speaking and listening. Therefore, testing for compliance with Annex 1 licensing requirements should focus on speaking and listening proficiency.

Test content

4 3 The purpose of a language proficiency test is to assess test-takers' use of language based on their performance in an artificial situation in order to make generalizations about their ability to use language in future real-life situations. Because of the high stakes involved, pilots and air traffic controllers need to be tested in a context similar to that in which they work. Test content should, therefore, be relevant to their work roles.

4 4 Radiotelephony communication requires not only the use of ICAO standardized phraseology, but also the use of "plain" language. Phraseology is the formulaic code made up of specific words that, in the context of aviation operations, have a precise and singular operational significance. Plain language is defined in ICAO documents as "the spontaneous, creative and non-coded use of a given natural language." In simple terms, plain language can be thought of as the non-phraseology language that is used in radiotelephony communications when standardized phraseology is not appropriate.

4 5 The provisions of the ICAO language proficiency requirements that directly address test content are:

- a) *Annex 1, Appendix 1*, where holistic descriptors refer to "work-related topics", "work-related context", and "routine work situation", and
- b) *Annex 1, Attachment A*, under Vocabulary and Comprehension, which refers to "work-related topics".

4 6 The use of ICAO standardized phraseology is an operational skill that is taught by qualified aviation operational specialists and is acquired to the required level of proficiency by trainee pilots and controllers during operational training. Teaching and testing standardized phraseology is an operational issue, not a language proficiency

issue. It follows that a test designed to evaluate knowledge or use of standardized phraseology cannot be used to assess plain language proficiency.

4.7 Before the ICAO language provisions were adopted in 2003, assessments of standardized phraseology were based on technical accuracy and appropriateness within the operational context and, with respect to delivery technique, only on generic "good practice". Since the adoption of the language provisions in 2003 and the publication of the language proficiency rating scale, it is recommended that assessments of ICAO standardized phraseology should, in addition to the existing guidelines in the PANS-ATM, take into account the descriptors for pronunciation and fluency of Operational Level 4.

4.8 It is acceptable that a test of plain language in a work-related context could contain a scripted test task or a prompt in which standardized phraseology is included. The test task may be used as a warm-up or as a means of setting a radiotelephony context in which to elicit plain language responses from the test-taker.

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

Test tasks

4.10 There are many kinds of test tasks or prompts that can be used to elicit speech samples. In general, tasks that resemble real-life activities are the most suitable.

4.11 It is important to keep in mind that the idea of a "work-related context" can accommodate different interpretations. A "narrow" interpretation would aim to closely replicate radiotelephony communications, including the extent of plain language needed in unusual, unexpected or emergency situations. A "broad" interpretation of the holistic descriptors and rating scale would aim to elicit plain language on various topics that are related to radiotelephony communications or aviation operations, without replicating radiotelephony communications specifically. Examples may include question and answer routines, problem-solving exchanges, briefings, simulations and role plays. Both interpretations are valid.

Testing for Expert Level 6 proficiency

4.12 The Level 6 descriptors in the ICAO rating scale refer to features of language use that go beyond the work-related context indicated in descriptors at lower levels. Formal evaluation of Level 6 using a specialized language test would follow an exhaustive procedure involving tasks and contexts that go beyond the subject matter of radiotelephony communications. Furthermore, since language proficiency at "both ends" of a proficiency scale is relatively easy to evaluate, it is not difficult to recognize "Expert" (including "native" or "native-like") proficiency. For these reasons, assessment at Level 6 should be carried out by a trained and qualified rater but not necessarily by a language testing specialist or require the use of a fully developed, specialized language test.

4.13 Monolingual native speakers of the language should be considered as "probable expert speakers". However, probable expert speakers may also include multilingual speakers who include the language as one of their "native" languages, and foreign-language speakers who have acquired a high level of proficiency. A test-taker who is tentatively considered to be a Level 6 speaker of the language may be evaluated through informal assessments (such as interviews or oral interactions with licensing authorities, recruitment officers or flight examiners), supported by documented evidence about that individual's linguistic history. This history, to be determined by state authorities, could include:

- a) place of birth and early residence;
- b) the language(s) used during childhood in the family, in the community and in education;