

Maintaining Quality and Effectiveness in a Testing System – The RELTA

Mr. Michael Kay
RMIT University
Melbourne, Australia

Language tests developed in response to the strengthening of the ICAO language provisions need to reflect the ICAO requirements in both test design and application. These tests need to be high quality, valid and to be supported by systems which facilitate their effective implementation. This paper provides an overview of how the RMIT English Language Test (RELTA) is designed and implemented to address the standard.

Introduction

The strengthening of the standards established in the ICAO Language Proficiency Requirements (LPRs) requires that states ensure the language proficiency of affected pilots and air traffic controllers (ATCOs) is reported against an international standard (ICAO, 2004). Determining the proficiency of pilots and controllers for these purposes requires the use of an instrument that can effectively assess communicative competence across the six criteria and six proficiency levels. The instrument also needs to address the requirements outlined in the amendments to Annex 1, 6, 10 and 11 to the Convention on International Civil Aviation.

The ICAO Language Proficiency Requirements and their impact on testing

The key amendments to the ICAO conventions underpinning the LPRs state that pilots, air traffic controllers and aeronautical station operations shall demonstrate the ability to speak and understand the language used for radiotelephony communications to the level specified in the language proficiency requirements in the Appendix (Annex 1) (ICAO, 2004). The Holistic Descriptors in the Appendix state that proficient speakers be able to:

- communicate in radiotelephone and face-to-face situations;
- communicate on common, concrete and work-related topics;
- use communicative strategies to exchange messages and resolve misunderstandings in general or work-related contexts;
- manage linguistic challenges presented by an unexpected turn of events that occur within the context of a routine work situation or communicative task which is unfamiliar.

In addition, the Appendix states that the LPRs are applicable to the use of both phraseology and plain English. The construct of proficiency tests designed to successfully address the LPRs must take account of the SARPs and the Appendix to Annex 1.

Thus, tests need to assess speaking and listening proficiency in both radiotelephony (including phraseologies and plain English) and plain English in face-to-face contexts, elicit language associated with non-routine communication contexts and provide situations requiring candidates to apply strategies to repair communication breakdowns. RMIT took the approach that these factors are critical to the design of any test that claims to be ICAO compliant.

The Language of Radiotelephony

ICAO acknowledges communication between pilots and air traffic controllers over the radio requires the use of phraseology for communication in routine and predictable situations and plain English in contexts where phraseology cannot adequately convey meaning (ICAO, 2004).

Successful communication in radiotelephony requires pilots and ATCOs demonstrate communicative competence in the use of highly context-specific phraseology and have broader proficiency in natural English (Enright, 2004). The vast majority of radiotelephony comprises phraseology for communication in routine situations while communication in the very rare but safety-critical situations requires proficiency in plain English. ICAO levels 3, 4 and 5 are differentiated in the descriptors within the scale, by a number of recurring terms such as *concrete*, *familiar/unfamiliar*, *predictable/unpredictable*, *expected/unexpected* and *routine/non-routine*. These terms reflect the language associated with plain English. Therefore, proficiency at ICAO level 4 or above requires proficiency in the use of phraseology and plain English.

RELTA and the ICAO Language Proficiency Requirements

Since the announcement of the ICAO LPRs in 2003, RMIT University has devoted considerable time and resources to the development of tests to evaluate language competence of pilots and ATCOs. Development of the RELTA commenced with the design of a test for licensed pilots. The test development team is made up of a number of test writing experts, aviation personnel, and language experts with extensive experience working in aviation fields. RMIT developed a partnership with QANTAS Airways for ongoing consultation in the development of the RELTA. Experienced international QANTAS pilots are consulted to maximise the authenticity and accuracy of the scenarios.

The language needs of pilots and controllers differ significantly. Pilots predominantly listen to information and instructions transmitted by a single controller at any one time. They also respond to instructions and report information. In contrast, at any given time, ATCOs are required to manage communications with multiple pilots, initiate communications and give instructions. Based on these fundamental differences, the test development team established the need for two forms of the RELTA.

Upon establishing that the RELTA for Pilots was both an effective assessment tool and was positively received by target test-taker populations, the development of the RELTA for Air Traffic Controllers followed in 2005. The RELTA for Air Traffic Controllers has been developed in consultation with subject matter experts (approach, enroute and tower-based air traffic controllers) from Airservices Australia. Both forms of the RELTA are designed to allow proficiency levels to be reported with respect to the ICAO language proficiency rating scale, across the six criteria.

The RELTA is a criterion-referenced proficiency test that assesses proficiency in radiotelephony and face-to-face contexts. The test addresses amendments in Annex 1, 6 and 10 and the Appendix by assessing:

- understanding and production of phraseology for communication in routine voice-only radiotelephone situations (Annex 1, Annex 6, Annex 10 and Holistic descriptors);

- understanding and production of plain English in contexts where phraseology does not suffice in non-routine radiotelephony situations (Annex 1, Annex 6, Holistic descriptors and Appendix to Annex 1)
- understanding and production of natural English in face-to-face situations (Appendix to Annex 1)

Design of the RELTA

The RELTA for Pilots and RELTA for Air Traffic Controllers each consist of a 30-minute listening component (to assess the criterion comprehension in the ICAO rating scale) and a 25-minute speaking component, to assess the remaining criteria in the rating scale related to speaking proficiency, as shown in Table 1. In addition, three forms of the RELTA Speaking have been developed to cater to the specific language needs of approach, enroute and tower-based ATCOs.

Each component of the RELTA consists of three sections. Each of the three sections has a specific purpose in relation to the ICAO LPRs, with the same language domains assessed in the equivalent sections across the speaking and listening components (as shown in Table 1). In Section 1 and 2 the candidate takes part in a roleplay with the examiner who assumes the corresponding role of pilot/ATCO. Section 1 assesses language ability in routine radiotelephony contexts (communicative ability in phraseology). Section 2 assesses language ability in less-predictable and non-routine radiotelephony contexts (communicative ability in plain English, in voice-only situations) and Section 3 assesses language ability in a range of face-to-face contexts, requiring test takers to recognise and use simple and complex natural English (communicative ability in plain English in aviation-related English domains).

RELTA Test component	Criteria from Appendix to Annex 1	Section 1 Use of phraseology in radiotelephony (voice-only)	Section 2 Use of plain English in radiotelephony (voice-only)	Section 3 Use of plain English in conversation (face-to-face)
Speaking	Pronunciation	Pronunciation of in routine radiotelephony contexts.	Pronunciation plain English in non-routine radiotelephony contexts.	Pronunciation plain English in face-to-face aviation-related contexts.
	Vocabulary	Use of range of vocabulary in phraseology.	Use of range of aviation-specific vocabulary.	Ability to produce language fluently and knowledge of discourse markers.
	Structure	Ability to construct transmissions using phraseology.	Grammatical range and accuracy of plain English in non-routine radiotelephony situations.	Grammatical range and accuracy in work-related conversational contexts.
	Fluency	Fluency of phraseology in transmissions.	Fluency of plain English in non-routine radiotelephony contexts.	Fluency of plain English in work-related conversational contexts.

	Interactions	Immediacy and appropriacy of responses and ability to check, confirm and clarify information and deal with misunderstandings using phraseology.	Immediacy and appropriacy of responses and ability to check, confirm and clarify information and deal with misunderstandings during non-routine radiotelephony events.	Immediacy and appropriacy of responses and ability to check, confirm and clarify information and deal with misunderstandings in conversational contexts.
Listening	Comprehension	Comprehension of pilot/controller exchanges associated with phraseology in routine contexts.	Comprehension of pilot/controller exchanges associated with non-routine events involving plain English.	Comprehension of plain English in a work-related conversational context.
Specific-purpose background knowledge required		Knowledge of aircraft operating procedures and associated phraseology in routine situations (not assessed).	Knowledge of aircraft operating procedures and associated phraseology in non-routine situations (not assessed).	Knowledge of flight processes and issues in aviation (not assessed).

Table 1: Language assessed in each section of the listening and speaking components of the *RELTA*

Authenticity is a vital aspect of language tests designed for assessing language for specific purpose (Douglas, 2000). Language in the *RELTA* is presented in radiotelephony-specific contexts in Sections 1 and 2. During these sections, the examiner assumes the role of an air traffic controller, in the *RELTA* for Pilots, or the pilots the test taker is communicating with in the *RELTA* for Air Traffic Controllers. The examiner and test taker communicate via headsets, and candidates' responses are recorded for subsequent rating. To maximise contextualisation, authenticity and test taker engagement, visual and audio information in the listening component and Sections 1 and 2 of the speaking component are delivered via computer. For example, during the speaking test, the test taker is required to respond to or initiate communication with the examiner in response to images/audio presented via the screen/headset. A context image is on the left of the screen, in the form of a view from the flight deck in *RELTA* for Pilots (as shown in Figure 1), or radar screen for the *RELTA* for Air Traffic Controllers. The context image provides a means for the candidate to recognise the context in which the language is occurring, enhancing authenticity.



Figure 1: Example information displayed to the candidate via the screen, with context image on the left (flight deck view) and language prompts on the right.

The RELTA is designed and developed to reflect language testing best practice and adopt the recommendations outlined in the Manual on the Implementation of ICAO Language Proficiency Requirements (Doc 9835). These key elements are as follows:

- The ICAO Language Proficiency Rating Scale informs both test design and scoring processes.
- Only skills directly related to listening and speaking proficiency (not reading, writing and knowledge of grammar) are assessed.
- The test construct is confined to ensure only proficiency in aviation English affects test performance. Test takers are not assessed on their knowledge of systems or procedures.
- Understanding of language where sound quality compromises comprehension is not a part of the test construct, as this requires separate skills that are not related to language proficiency, and can undermine the validity of a language test (Douglas, 2000).
- Computers are utilised to facilitate delivery and maximise authenticity, interactivity, accuracy and uniformity of administration.
- Live examiners are utilised to ensure interactivity and authenticity are high.
- Performance is assessed by at least two independent trained and accredited raters anonymously, maximising reliability, objectivity and fairness.
- The test assesses language in aviation contexts (phraseology, plain English in radiotelephony contexts and natural English in work-related face-to-face contexts).
- All forms and versions of the RELTA are trialled with actual target-user populations (both native speakers and non-native speakers) to allow test tasks and items to be refined and to enhance reliability and validity.
- Data from test trials is analysed using statistical packages to allow the test development team to ensure test tasks/items are meaningful with respect to the overall test construct and score generation process.
- Procedures are established to ensure results are checked for accuracy, fairness and reliability, including the use of computer software to monitor inter- and intra- rater-reliability and trigger the need for third ratings, when required.
- All rating and reporting processes are controlled by RMIT to ensure security, fairness and accountability are maintained.
- Comprehensive examiner and rating programs are provided to ensure test delivery and rating processes are performed consistently, according to RMIT's quality assurance procedures.

Delivery of the RELTA

Managed by RMIT, the RELTA is designed to be administered locally at accredited test centres in locations convenient for test users, and to be administered according to schedules which align to the needs of organisations. A number of options are available for test implementation: it can either be:

- delivered and rated by trained and accredited locally-based personnel;
- delivered by trained and accredited locally-based examiners and rated by RMIT; or
- delivered and rated by RMIT. RMIT oversees test registration, administration and scoring processes.

Delivery of the listening component

The listening component is delivered via a computer and can be administered to a group of candidates simultaneously over a 30-minute period, as shown in Figure 2. Candidates record their answers in a test booklet.



Figure 2: Delivery of the listening component

Delivery of the speaking component

Sections 1 and 2 are computer delivered. The examiner console is linked to a candidate monitor (Figure 3). The speaking component is delivered on a one-to-one basis. The monitors are positioned so that the examiner and candidate do not have direct visual contact, simulating the non face-to-face air/ground communication. The examiner controls how the test progresses (display of visual information on the screen and any audible transmissions from other aircraft or controllers). The candidate and examiner view the same information simultaneously, with the examiner's screen replicated on the candidate's monitor. The candidate and examiner communicate via a microphone and headset. Section 3 involves direct face-to-face communication with the examiner.



Figure 3: Delivery of Sections 1 and 2 of the speaking component, showing the examiner in the foreground.

Test Administration

A secure internet-based test data management system allows the test to be efficiently and effectively controlled, distributed and rated. This system expedites administration processes, reduces workloads and costs and minimises data entry or loss, while maintaining security. The Scoring Record Management System (SRMS) generates and issues results, allowing the test to be delivered and rated in any geographical location while being managed centrally by RMIT. It also acts as a communication portal, allowing raters, examiners and candidates to access applicable information.

The SRMS facilitates test administration by allowing candidates to register for the test, having accredited personnel securely access test files and associated documents and allowing appeals to be lodged and addressed. All candidates remain anonymous, with candidate identification numbers referenced by the system, once generated upon registration.

Scoring processes

Both components of the RELTA are scored independently of delivery. Upon completion of the speaking component, candidates' recorded responses are uploaded securely via the internet to the SRMS. Sound files and rater scripts are then distributed to at least two raters via the SRMS. Raters then enter the scores for each of the three sections into the system. The SRMS collates scores, identifies variations, flags differences and distributes files to third raters, if required. It computes scores for each section, each criterion and the overall ICAO level, from 1 to 6. Following the requirements in the Appendix to Annex 1, the lowest score a candidate receives on any of the six ICAO criteria is the score which is reported as the ICAO level.

Evaluation and Maintenance of the RELTA

Initial trials of a prototype pilot test revealed that the design platform for eliciting language was well received by test takers. Feedback indicated the tests are engaging, relevant and an effective means of assessing language in a context which simulates the way pilots/ATCOs use language in their work.

Although face validity and content validity were high, some elements of the prototype were found to be less effective in the generation of an overall score. The test specifications and design was subsequently altered and updated by: removing the use of any semi-direct delivery (pre-recorded prompts), updating and replacing less effective test tasks in the listening component and shortening the test in response to end-user feedback. New versions of both forms of the RELTA have been subsequently developed, trialled and updated. Trials have occurred in China, Russia, Thailand, Australia, parts of South America and the Middle East. Formal test trials conducted on the first three test versions revealed:

- Correlations between established general proficiency tests and the RELTA indicated it is an effective instrument for measuring language proficiency.
- The test evaluates elements of language proficiency specific to pilots and air traffic controllers that cannot be assessed by general language tests.
- Each section of the RELTA meaningfully and contributes to the overall scores and ICAO level.
- The test identifies strengths and weakens across the six ICAO criteria, in phraseology and in plain English.

Today, the effectiveness of all aspects of the RELTA is continually monitored by RMIT. The test tasks, items, rating and score generation processes are constantly scrutinised, refined and updated. Further, the security and viability of the RELTA are maintained by the ongoing development of new test versions.

Conclusion

The RELTA is currently used throughout China, with over 10 000 pilots to be tested in 2007 and 2008 to conform with domestic and international licensing requirements. The RELTA is also used in parts of the Middle East and Europe, and is scheduled for use parts of South America and South East Asia. Continual reinvestment, research and commitment are required to maintain the effectiveness and viability of the RELTA.

The washback effect the RELTA will have on training is yet to be fully realised, however it is anticipated the effects will be positive. The RMIT Aviation English Program, for example, has been implemented in various regions and. in China, the airlines allocate resources and time to using the full program to develop a wide range of language skills, including those assessed by the RELTA.

Extensive time, ongoing investment and resources, including access to testing and aviation subject matter experts, are required to develop and maintain proficiency tests which satisfy the LPRs. The provision of ongoing support and administration systems is also crucial for successful implementation of these kinds of tests. Implementing tests which are lack the quality, validity or the specificity to assess aviation English or which or are deficient in the infrastructure for adequate administration would not only undermine the intentions behind the ICAO standard but also compromise safety. Test developers and test users have significant obligations to develop and select sound and effective tests.

References

Douglas, D. (2000) *Language testing for specific purposes*. Cambridge: Cambridge University Press.

Enright, A., (2004) Updated test satisfies new requirements to demonstrate English language proficiency.
ICAO Journal, 59 (1), 15-17.

ICAO (2004) *Manual on the implementation of ICAO Language Proficiency Requirements* (Doc 9835), International Civil Aviation Organisation: Montreal.

— — — — —