

# ○ PLANES, POLITICS AND ORAL PROFICIENCY

## TESTING INTERNATIONAL AIR TRAFFIC CONTROLLERS

*Carol Lynn Moder, Oklahoma State University*

*Carol Lynn Moder is an Associate Professor of Linguistics and TESL at Oklahoma State University. From 2004 to 2006, she held contract grants with the International Training Division of the U.S. Federal Aviation Administration to develop an Aviation English curriculum for International Air Traffic Controllers. She served as a consultant to ICAO in 2005, participating in the development of the CD of ICAO rating samples.*

*Correspondence to Carol Lynn Moder: carol.moder@okstate.edu*

*Gene B. Halleck, Oklahoma State University*

*Gene B. Halleck is Professor of Linguistics and TESL at Oklahoma State University. She was a co-Principal Investigator of a grant funded by the FAA (2004–2005) to establish Aviation English curricula for pilots and air traffic controllers. During that time she designed placement tests, as well as different versions of pre- and post-tests known as Oral Proficiency Tests for Aviation (OPTA), for the two courses that she and her colleagues at Oklahoma State University were creating.*

*Correspondence to Gene B. Halleck: gene.halleck@okstate.edu*

This study investigates the variation in oral proficiency demonstrated by 14 Air Traffic Controllers across two types of testing tasks: work-related radio telephony-based tasks and non-specific English tasks on aviation topics. Their performance was compared statistically in terms of level ratings on the International Civil Aviation Organization (ICAO) scale. The results demonstrate significant differences in the performance of the test-takers across task types, differences that were not fully predictable across subjects. The differences between general English proficiency and specific purpose proficiency were even greater than those we would expect for other LSP situations. We discuss the implications of these findings for fairly and safely assessing Aviation English using ICAO standards in a politicized context.

## INTRODUCTION

In specific purposes performance testing, best practices demand that test tasks be representative of those in the target language use situation, based on a needs analysis (McNamara, 1996; Douglas, 2000). In this paper, we discuss test design in a distinctive Language for Specific Purposes domain, Aviation English.

In the aviation domain, English is used internationally primarily for radio communication between pilots and air traffic controllers in contexts where one of the parties does not speak the local language of air traffic control (ATC). The language of air traffic

control is distinct from many English for Specific Purposes domains, because it is used for a very restricted set of functions and has a prescribed phraseology with reduced syntax and vocabulary for routine actions. The communication involves extensive shared information concerning the aircraft in the area, the parameters of the airport or airspace, and the expected actions at particular points in the flight. To illustrate the nature of the language used in this domain, we present two examples of authentic ATC communication. Example (1)<sup>1</sup> is a routine radiotelephony exchange between an area controller and a pilot of a small plane with the call sign AJT 142 (Alpha Juliet Tango one-four-two).

### **Example (1)**

CC: Alpha Juliet Tango one four two, radar service terminated, maintain flight level 390, contact Peachtree Controller 125.5, So long  
AJT: 125.5, Alpha Juliet Tango one four two

The exchange in this example models prescribed ATC phraseology, with its distinctive turn-taking identifiers for speakers and addressees, prescribed number and letter readout format, limited syntax, and restricted vocabulary. The exception to standard phraseology is the signoff greeting ‘So long.’ Although not prescribed in standard phraseology, greetings are common in ATC communication to open and close interactions. Overall, the exchange is short, abbreviated, and efficient. This brevity is facilitated by the large amount of given information that the pilot and the controller share. Each is aware in advance of the call sign of the plane, its intended flight path, the navigation points along the path, and their radio frequencies. In this exchange both parties hear the expected information, and there is no need for any negotiation of meaning. As exemplified here, most ATC exchanges proceed as expected, but it is critical to safety that both the pilot and the controller maintain focus on assuring the information is accurately delivered and acknowledged and on correcting any discrepancies.

### **Example (2)**

A2220: Airline 2220 at 350 and about forty miles ahead we’ll need to deviate for weather  
CC: Airline 2220, roger contact, flight level 350, confirm what side, left or right?  
A2220: uh deviation uh left for weather in about uh probably start in about 3-0 miles  
CC: 2220, Center Control, roger, left deviation approved, after deviation fly direct to Casa  
A2220: left deviation approved afterwards cleared direct to Casa Airline 2220  
CC: 2220 affirm

Example (2) illustrates a common situation in which phraseology is mixed with ‘plain language’.<sup>2</sup>

Here the pilot of Airline 2220 informs the controller that he intends to deviate from the prescribed flight path because of bad weather. Although standard phraseology exists for such a request (‘request weather deviation’), the native English-speaking pilot frames the request in plain language, ‘we’ll need to deviate for weather’. Since this request does not specify the direction of the deviation, the non-native English speaking ATC uses plain language to request a clarification, ‘what side, left or right?’

The tension between the use of standard phraseology and plain language is an ongoing issue in the aviation community. Operational training facilities and international policies emphasize that standard phraseology should be used in all cases, and that the use of plain language should be restricted only to those situations in which there is no standard phraseology (ICAO, 2004, pp. 2–3). However, discourse-based studies of pilot-air traffic controller communication have consistently found plain language like that exemplified above to be common, and furthermore, to contribute to the local coherence and effectiveness of the interactions (Sänne, 1999; Neville, 2004; Wyss-Bühlmann, 2005). The ongoing debate about the proper balance between phraseology and plain language is an important political aspect of the culture of aviation that has major ramifications for test design. The way this issue has affected the development of international aviation testing policy will be discussed further below.

Early aviation English tests incorporated the distinctive interactional language features of radiotelephony in their design. For example, Teasdale (1996) conducted a detailed discourse-based needs analysis which included the transcription and analysis of twelve hours of authentic air traffic communication, supplemented with a questionnaire on language use. Based on the needs analysis, Teasdale outlined test objectives that clearly linked test characteristics to important recurring air traffic functions, such as understanding pilot readbacks and checking, confirming, or clarifying information when necessary.

More recently, because of a major international policy change regarding English language proficiency (see Introduction to this volume), the aviation language testing situation has been driven more by politics and expediency than by best practices in language test design and validation procedures. The ICAO English proficiency standard is an international political policy, based on collaboration between governments, aviation professionals, and applied linguists. As Brindley (2001) and McNamara & Roever (2006) have discussed, the interaction of politics and language testing can create negative effects

for some stakeholders, potentially leading to testing decisions based on criteria other than those stipulated by best practices. In the case of the ICAO proficiency requirement, the political aspects are writ large.

An example of the political negotiations inherent in ICAO language proficiency decision-making processes can be seen in the discussion of the use of two languages in an airspace in the ICAO manual on language proficiency. The discussion acknowledges that the use of more than one language in radiotelephony has been a contributing factor in a number of accidents because it reduces the situational awareness of aviation personnel. However, it goes on to say that ‘significant national, cultural, economic and organizational impediments’ preclude establishing a single language requirement worldwide. The section then notes that a number of non-native English-speaking member nations have voluntarily established English-only requirements and that other member states are free to establish similar policies that are more stringent than ICAO standards (ICAO, 2004, pp. 7–4).

A further example of the political nature of the process is the position that ICAO has taken with respect to those member states that were unable to comply with the standard by March 2008 (cf. Alderson, 2009). In September 2007, the 36th Assembly of ICAO asked contracting states to accept into their airspaces until March 2011 flights from contracting states that have posted implementation plans and interim safety measures to the ICAO Flight Safety Exchange website (ICAO FLS, n.d.). Thus, the decision about whether to strictly enforce the deadline or to extend it an additional three years is left ultimately to each contracting state.

Against the background of the politics of international standard-setting, ICAO’s decision to implement a standard defined by a set of levels with holistic descriptors, but without a corresponding test, is perhaps understandable, especially since each contracting state is administratively and financially responsible for implementing policy. However, the information available for member states about the domain to be tested and the characteristics of an appropriate test is equivocal with respect to some key issues.

The ‘Manual on the Implementation of ICAO Language Proficiency Requirements, Document 9835’ (ICAO, 2004, pp. 6–5) specifies that an appropriate language test must:

- a. be a proficiency test of speaking and listening;
- b. be based on the ICAO Rating Scale and holistic descriptors;
- c. test speaking and listening proficiency in a context appropriate to aviation, and;
- d. test language use in a broader context than in the use of ICAO phraseologies alone.

An examination of the ICAO rating scale and holistic descriptors shows that they primarily reference general concepts of oral proficiency, making little or no specific mention of the Aviation English domain. The scale measures six separate linguistic areas: pronunciation, structure, vocabulary, fluency, comprehension, and interaction. Among these areas, only the descriptors for vocabulary and comprehension refer to a specific purpose context, mentioning the ability to communicate about or comprehend 'common, concrete, or work-related topics' (ICAO, 2004, pp. A-8-9). The main information relating these general proficiency descriptors to aviation is provided in the ICAO Manual in the brief statements preceding the explanation of each of the scales. For structure, the statement specifies that 'relevant grammatical structures and sentence patterns are determined by the task,' referring the reader to a separate chapter of the manual about 'relevant communicative aeronautical language functions' (ICAO, 2004, p. A-10). For vocabulary the statement indicates: 'An appropriate methodology is to train and test in a broad aviation-related context, in order to ensure that proficiency will be fully adequate in the usually narrow constraints of aeronautical radiotelephony communications' (ICAO, 2004, p. A-11). For comprehension, the statement specifically references the ability of pilots to understand air traffic controller instructions, especially when they differ from expectations, and the need for air traffic controllers to handle all communications, even those not expected (ICAO, 2004, p. A-13). If the member states carefully read the manual and attend to the prefatory statements, the scales are more likely to be contextualized for the aviation domain. However, if they rely on the rating scale alone, they are more likely to focus on general English proficiency. In either case, they must still determine what kind of test to use to elicit a sample. What official information do they have to aid them in making this determination?

In introducing the new proficiency requirement in the *ICAO Journal*, Mitsutomi and O'Brien (2004) divide the necessary English competency for safe pilot-ATC communication into three components: Air Traffic Control phraseology, English for Specific Purposes, and English for General Purposes. They state that the new English proficiency guidelines target English for General Purposes, on the assumption that 'everyone in the cockpit and tower has mastered at least the basics of ATC communication' (7-9). They go on to specify that English for General Purposes in the aviation context 'will consist mostly of aviation-specific topics and vocabulary' (9) and that accurate and safe pilot-controller communication requires an ability to 'ask and answer questions, follow instructions, narrate events, describe situations, and paraphrase information' (27).

These specifications are remarkable in several respects. First, even if the assumption that all aviation personnel have mastered ATC communication is justified, we cannot

assume that they have mastered such communication in English. As suggested above in the ICAO information on airspace languages, many aviation personnel operate regularly in contexts in which they use the local language for all or most ATC communication, switching to English only when their interlocutor does not share the local language. For example, in recordings of several hours of ATC communication from some international airports we analyzed, the vast majority of the communication was in the local language with only a few sporadic exchanges in English. It is not obvious that pilots or controllers operating in such a context can be assumed to have mastered ATC communication in English. An additional concern is what precisely one is to understand to be the distinction between English for Specific Purposes and English for General Purposes in aviation contexts. This ambiguity is at the heart of difficulties in determining what ICAO has designated to be appropriate test characteristics.

In the same issue of the *ICAO Journal*, Mell (2004) takes a clearly domain-based approach. His article strongly emphasizes the importance of contextualizing various functions and grammatical structures within the context of effective pilot-ATC communication. He illustrates specific plain language structures that might be needed in emergency situations and the domains of general English vocabulary that might be relevant. He also argues the need for further ‘linguistically informed selection’ of features in order to adequately focus the testing and language training of aviation professionals (14).

Juxtaposing these two presentations of the appropriate test characteristics, we are left with considerable ambiguity. Is the test to focus on English for General Purposes about aviation topics or more narrowly on English relevant to effective pilot-controller communication?

Contributing to the problem of determining what the tests are to sample is the domain-specific issue of the relation between radiotelephony and plain language mentioned earlier. Many aviation professionals strongly advocate limiting the use of plain language in work-related contexts. In addition, most believe that non-aviation personnel are unqualified to evaluate the use of ATC phraseology, even for the purposes of establishing English proficiency. This concern is reflected in statements about testing on the ICAO website. As a response to the question, “Could a language proficiency test contain radiotelephony and technical questions?”, the website states:

Because of the high stakes involved, pilots and air traffic controllers deserve to be tested in a context similar to that in which they work and test content should therefore be relevant to their roles in the workplace. The descriptors for Vocabulary and Comprehension for ICAO

Operational Level 4 refer to "work-related topics". Tests should provide test-takers with sufficient and varied opportunities to use plain language in aviation work-related contexts in order to demonstrate their ability with respect to each descriptor in the Language Proficiency Rating Scale and the Holistic Descriptors. To achieve this, the design of tests should be undertaken by a team of linguistic and operational subject matter experts to ensure validity, reliability and operational relevance.

The Note found in the Appendix to Annex 1 indicates that the Holistic Descriptors and Rating Scale apply to the use of phraseology as well as plain language. Just as testing of ICAO phraseology cannot be used to assess plain language proficiency, neither can English language proficiency tests be used to test ICAO standardized phraseology.

It is acceptable that a test contain a scripted test task in which phraseology is included in a prompt. The test task may be used as a warm up or an ice-breaker and elicit a plain language response from the test taker. Test prompts should not be intended to evaluate specific technical knowledge concerning operations.

*(ICAO FLS)*

This statement appears to encompass both of the views put forth in the *ICAO Journal*. It clearly endorses testing plain language in aviation-related contexts. With respect to phraseology the statement is less clear. It acknowledges the need to test in a context similar and relevant to the workplace. However, it then reflects concerns of some stakeholders about having non-aviation specialists evaluate radiotelephony exchanges by suggesting that phraseology tasks be restricted to warm-up or ice-breaker questions.

Because of the ambiguity in the ICAO policy over the types of language and test tasks that should be included in aviation tests, it is essential to analyze the performance of aviation professionals on both work-related and general English tasks in the aviation context. Such information would increase the awareness of the aviation community and language testing professionals of the consequences of adopting one set of test specifications or another.

This study investigates whether there is significant variation in the oral proficiency demonstrated by fourteen air traffic controllers across two types of testing tasks: work-

related radio telephony-based tasks and general English tasks on aviation topics. We compare their performance statistically in terms of level ratings on the ICAO scale.

## **METHOD**

### **SUBJECTS**

Subjects were fourteen air traffic controllers from Asia: 6 from Korea, 5 from China, 2 from Mongolia, and 1 from Japan. All subjects were licensed and practicing air traffic controllers in their home countries who had enrolled in an Aviation English training course in the United States in 2005.

### **INSTRUMENTS**

The general English proficiency of the subjects was measured using a variety of oral proficiency tests, including the SPEAK test and Ordinate's SET 10. The SPEAK test is the institutional version of the Test of Spoken English (TSE), produced by the Educational Testing Service. It is a semi-direct test administered via audio-recording through headphones, designed to measure the ability of nonnative speakers of English to communicate orally in English in an academic or professional environment. The Set-10 is a ten-minute test of oral proficiency for adult nonnative speakers of English, which is administered over the telephone and automatically scored by computer. It is now incorporated in the Versant test program (see Van Moere et al., this volume).

In addition, subjects took the Oral Proficiency Test for Aviation (OPTA), a test designed by the authors based on the ICAO standards to place students in aviation courses. In order to preserve the characteristics of the target language use situation, the OPTA was administered via digital recording over headsets. Test-takers heard the prompts through their headsets and responded into the headset microphones. Two raters who had been trained in the context and form of radiotelephony and in the use of the ICAO scale rated each of the eight tasks on the OPTA. They rated each of the language areas of the scale separately and averaged these ratings to provide an overall level rating for each task.<sup>3</sup> The level ratings of the two raters were entirely consistent for the tasks reported here.

The scores of the subjects on the three oral proficiency tests are provided in Table 1. Medians and ranges of the scores of the 14 subjects are provided in order to give a general measure of central tendency and variation for each test, not for comparative purposes. It should be noted that both the SPEAK and the OPTA are scored on a rating scale of six levels. For the OPTA the scores are reported with decimals and for the SPEAK the



scores are reported with the customary ten point per level increments. The Set-10 is evaluated on a scale with 80 as the maximum. Because the tests are rated on ordinal scales and because the sample is not normally distributed, the relative rankings of the test-takers on each test was compared using a Friedman test. The Friedman test indicated that the scores of each of the three tests generated significantly different rankings (Friedman (14,2)= 28.00,  $p<.05$ ). As expected, the overall scores on the OPTA, which includes aviation-specific tasks, generated different test-taker rankings than did the general oral proficiency tests.

	Mean Rank	Median	Range
OPTA	1.00	3.25	2.54 - 4.0
SPEAK	2.00	35	30 - 45
Set-10	3.00	43	34 - 55

**Table 1** Performance on three oral proficiency measures  
Friedman (14,2) = 28.00,  $p<.05^*$ , All two-way differences significant

To provide the broadest possible profile of oral English proficiency for instructional purposes in the context of the Aviation English course, the OPTA was designed to include a variety of general English, Aviation English, and radiotelephony tasks.

In this paper we will discuss the subjects' performance on three tasks from the OPTA. The prompts for all three tasks are given in the Appendix. The first task (Common Occurrence) asked for a description of an unfortunate, but common, occurrence in the aviation context, an aircraft off the runway in a ditch adjacent to a highway. The second task (Less Expected) required a description of a less expected context, a plane at the gate whose nose had hit and penetrated the wall of the terminal.

These two tasks were designed to allow assessment of the test-takers' abilities to communicate on 'common topics' and in 'unusual or unexpected circumstances,' as required by the ICAO scale. The third task (Radiotelephony) was designed to reflect communication in the work-related context within a radiotelephony communication. Test takers listened to an ATC exchange and were asked to respond as if they were the air traffic controller communicating with the pilot. To model the shared information normally found in the workplace, the prompt provided specific information on the airspace, the plane, and its flight details. The task required the controllers to hear and correct the pilot's incorrect read-back of the radio frequency, a critical aspect of safe and effective communication in the workplace context.

RESULTS

We now examine the test-takers' performance on the three OPTA items, with a focus on the contrast between the radiotelephony task and the two tasks that require general English in an aviation context. Table 2 shows the Spearman correlations for the three tasks, the radiotelephony task and the two tasks involving general English in the aviation context.

	Common Occurrence	Less expected
Radiotelephony	.408 ns	.396 ns
Common Occurrence		.887*

**Table 2** Comparative performance on three OPTA tasks: Spearman Rank Order Correlations  
\*p. < .05, ns=not significant

As the table illustrates, the two general English tasks correlate moderately highly,  $\rho = .887$ . On the other hand, the radiotelephony task does not correlate significantly with either of the two general English tasks.

Table 3 shows the pairwise comparison of the three tasks using a Wilcoxon Signed Ranks Test, which indicates that the two general English tasks do not differ significantly, but the scores on the radiotelephony task are significantly different from the scores on each of the two general English tasks.

	z
Common Occurrence- Less expected	.582 ns
Radiotelephony Task (median = 4.0) Common Occurrence (median = 3.0)	-3.066*
Radiotelephony Task (median = 4.0) Less expected (median = 3.12)	-2.953*

**Table 3** Wilcoxon Signed Ranks Test pairwise comparison of three OPTA tasks.  
\*p<.05 (with Bonferroni correction), ns = not significant

These comparisons of the overall performance of the group on the three tasks demonstrate that the radiotelephony task ranks the test-takers significantly differently

from each of the general English tasks on aviation topics, with scores on the radiotelephony task higher overall than scores on the other two tasks.

Because the ICAO guidelines will be used to make high stakes decisions about the careers of individual aviation professionals, it is important to examine not only the overall group performance, but also how the items divide individuals with respect to the cutoff score, Operational Level 4. Table 4 shows the percentage of test-takers measured at each of the ICAO levels on each of the three tasks.

	Radiotelephony	Common Occurrence	Less expected
Level 2	1 (7%)	3 (21%)	4 (29%)
Level 3	4 (29%)	9 (64%)	9 (64%)
<b>Level 4</b>	<b>7 (50%)</b>	<b>2 (14%)</b>	<b>1 (7%)</b>
<b>Level 5</b>	<b>2 (14%)</b>	<b>0</b>	<b>0</b>

**Table 4** Percentage of individuals rated at ICAO levels by task

The table shows that 64% of the air traffic controllers were rated at Operational Level 4 or higher on the radiotelephony question, whereas only 14% were rated at Level 4 for the common occurrence question and just 7% for the less expected context. These numbers indicate that even on the radiotelephony question, which closely mirrors their work-related language use context, not all practicing controllers scored at the minimum level according to the new ICAO standards. In addition, the majority of the controllers scored below the minimum level on the questions that involve description of aviation-related events, a task that does not typically require English in their job context.

We now turn to an examination of the performance of three representative individuals to further consider the ways these tasks place individuals on the ICAO scale.

ATC	Common Occurrence ICAO Level	Less Expected ICAO Level	Radiotelephony ICAO Level
RW(Mongolia)	4.0	3.25	5.0
MH(Korea)	4.0	3.75	3.75
TF(Japan)	3.0	3.25	5.0

**Table 5** ICAO level ratings of selected individual by task

In Table 5 we illustrate the three response patterns found in the data with individual examples. As the group results suggested, for most of the controllers, scores for the common occurrence and less expected context questions were lower than their scores on the radiotelephony task. Many controllers scored at or above the Operational Level on the radiotelephony task and below the minimum for both of the other tasks, as illustrated in Table 5 by TF. These individuals were rated at Operational Level 4 for the radiotelephony task and at Level 3 for both general English tasks. For some controllers, there was also a difference in level for the two general English tasks, as illustrated by RW, whose performance on the radiotelephony question is at Level 5, for the common occurrence is at Level 4 and for the less expected context is at Level 3. Thus, he meets the minimum operational level for all tasks, except the description of the unexpected occurrence. Of particular interest is that the pattern of obtaining higher scores on the radiotelephony task than on the description tasks was not true for all of the air traffic controllers. For example, MH was at Operational Level 4 in his response to the common occurrence question, but his scores for responses to the less expected context and the radiotelephony task were both rated at Level 3. Contrary to expectations, he failed to score at Operational Level for the radiotelephony task, but did score at Level 4 for the description of a common occurrence.

## CONCLUSIONS

The political nature of the decision-making involved in the establishment of the ICAO policy has left considerable ambiguity about precisely how targeted to the specific workplace communicative context an appropriate proficiency test should be. Since the primary concern of ICAO in establishing the new proficiency requirement was to address important issues in world-wide aviation safety, it is critical that ICAO member states and the testing professionals they engage take steps to make sure that the tests they use to implement the policy accurately predict the performance of pilots and air traffic controllers in critical workplace circumstances.

First, it is worth noting that a third of the practicing aviation professionals in our sample scored below Operational Level 4 on the radiotelephony task. Given that the test task was adapted from actually occurring radiotelephony communication and that this type of exchange has been repeatedly cited as critical to aviation safety, this result suggests that one cannot safely assume that all aviation personnel have fully mastered ATC communication in English. Second, strong performance on general English tasks in the

aviation context did not necessarily entail strong performance on an ATC communication task.

If common English tasks related to aviation topics could accurately predict language performance on the job, we would have expected that air traffic controllers' responses to the common occurrence task or the less expected task would be similar to their response to the radiotelephony task. In fact, we found that the radiotelephony task and the two general English tasks resulted in significant differences in the ranking of the controllers and, most importantly, in their performance with respect to the established minimum required proficiency, Operational Level 4. Furthermore, the results were variable across individuals in not entirely predictable ways. Because the two general English description tasks are not a normal part of the controller's work domain, we might expect all individuals to perform better on the familiar radiotelephony task. The overall group results suggested that this was the case. The majority of the practicing air traffic controllers that we tested did perform better on the radiotelephony task, often by one or two ICAO levels. In these instances, the description in the aviation context tasks might underestimate the controllers' ability to adequately perform in the workplace, resulting in the potential loss of employment. Of greater concern from a public safety perspective is the finding that some controllers performed at Operational Level on one of the general description in the aviation context tasks and failed to demonstrate minimum proficiency on the radiotelephony task. In such a case, the general aviation English task would have inaccurately predicted the controller's performance level on a critical workplace task. Although the sample size in this study is small, the findings suggest potentially critical implications.

The communication requirements of pilot-air traffic controller communication overlap little with those found in general English or other English for Specific Purpose contexts. ATC communication is highly distinctive in structure and discourse organization. For that reason, the ability to narrate or describe events outside the ATC domain does not show a direct relationship to workplace communication. The ICAO guidelines require aviation professionals to be able to deal with 'unexpected circumstances.' In this context that means aircraft emergencies. What is critically required in emergencies is the ability to assess the accuracy of the information conveyed and understood, to ask questions or make statements that clarify the details of the emergency, to interact and negotiate meaning effectively. In ATC communication this is usually done with a combination of phraseology and plain language (cf. Cushing, 1994; ICAO, 2004, pp. 7–3). Tasks set outside of the ATC communication setting are unlikely to accurately assess the ability of aviation professionals to effectively combine these two modes of communication. Furthermore, one-way tasks that require the test-taker to listen to a general English

prompt and provide a one-way response will not provide adequate opportunities to assess critical interactional competencies.

The findings of this study strongly indicate that both phraseology and unexpected work-related contexts must be fully included in ICAO proficiency tests. General English proficiency items on aviation topics will not provide an adequate sample of on-the-job proficiency. In order to meet the concerns of aviation professionals regarding the evaluation of ATC communication, language testers and raters must be trained to effectively evaluate the use of English in these contexts, independent of the professional correctness of the phraseology. Where practical, they should work together with aviation professionals. In this very high stakes situation, best practices in Language for Specific Purposes must engage with the political process to ensure fairness for the test-taking pilots and air traffic controllers and safety for those who will travel on their planes and through their airspaces.

## **APPENDIX: OPTA TASK PROMPTS**

### **Common Occurrence Question**

The picture below shows a plane in an unusual position. Describe where it seems to be, how you think it may have gotten there and the condition of any passengers who may have been aboard at the time of the accident.

### **Less expected Occurrence Question**

Accident photos are often quite horrible. But when I first saw the photo below, my reaction was to laugh. Imagine that you are on the phone with someone who is unaware that there has been an accident. Try to explain what has happened to the plane in the photo and suggest a possible explanation of how the plane ended up in this position.

### **Radiotelephony Question**

Below is an exchange between an Air Traffic Controller at Center Area Control and the pilot of a small plane with call sign JAW 4124. The Center Controller hands the plane off and advises the pilot to contact Peachtree Control at the frequency 125.5 The pilot then calls in to the Center Controller again. Listen to the pilot and respond appropriately.

CC: JWA 4124 radar service terminated, maintain flight level 390,  
contact Peachtree Controller 125.5. So long.

JWA: 125.5 JWA 4124

JWA: Center Control JWA 4124

CC:

JWA: Confirm the frequency is 125.5?

CC:

JWA: 135.5 will monitor and try them again over WPP JWA 4124

CC:

JWA: Thanks for your help.

## ENDNOTES

- <sup>1</sup> Examples come from transcripts of air traffic communication. The callsigns of the planes and the names of the controller stations and navigation points have been altered to preserve anonymity.
- <sup>2</sup> The term ‘plain language’ is used in the aviation community to refer to the use of general English syntax and vocabulary, as opposed to standard phraseology.
- <sup>3</sup> The ICAO policy does not mandate individual task rating. It requires a rating of the overall performance. It also mandates that the overall level rating be the *lowest* rating obtained on any of the language areas. We rated individual tasks as described for diagnostic and research purposes.

## REFERENCES

- Alderson, J. Charles (2009). Air safety, language assessment policy, and policy implementation: The case of Aviation English. *Annual Review of Applied Linguistics*, 29, 168–187.
- Brindley, Geoff (2001). Outcomes based assessment in practice: Some examples and emerging insights. *Language Testing*, 18(4), 393–407.
- Cushing, Steven (1994). *Fatal Words*. Chicago: University of Chicago Press.
- Douglas, Dan (2000). *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press.
- International Civil Aviation Organization (ICAO) (2004). *Manual on the implementation of ICAO Language Proficiency Requirements*. Montreal: Author.
- International Civil Aviation Organization (ICAO) FLS. n.d. Flight Safety Section. *Personnel Licensing. Frequently Asked Questions*. Accessed June 9, 2009. Available from: [www.icao.int/icao/en/trivia/peltrgFAQ.htm](http://www.icao.int/icao/en/trivia/peltrgFAQ.htm).
- McNamara, Tim (1996). *Measuring Second Language Performance*. London: Longman.

- McNamara, Tim; Roever, Carsten (2006). *Language Testing: The Social Dimension*. Malden, MA: Wiley-Blackwell.
- Mell, Jeremy (2004). Language training and testing in aviation need to focus on job-specific competencies. *ICAO Journal*, 59(1), 12–14, 27.
- Mitsutomi, Marjo; O'Brien, Kathleen. (2004). Fundamental aviation language issues addressed by new proficiency requirements. *ICAO Journal*, 59(1), 7–9, 26–27.
- Neville, Maurice (2004). *Beyond the Black Box: Talk in Interaction in the Airline Cockpit*. Aldershot, England: Ashgate.
- Sänne, Johan M. (1999). *Creating Safety in Air Traffic Control*. Lund: Arkiv Förlag.
- Teasdale, Alex (1996). Content validity in tests for well-defined LSP domains: An approach to defining what is to be tested. In Milanovic, M. and Saville, N. (Eds.), *Performance Testing, Cognition and Assessment: Selected Papers from the 15<sup>th</sup> Language Testing Research Colloquium*, pp. 211–230. Cambridge: Cambridge University Press.
- Wyss-Bühlmann, Eveline (2005). *Variation and Co-operative Communication Strategies in Air Traffic Control English*. Bern: Peter Lang.

Cite this article as: Moder, Carol Lynn; Halleck, Gene B.. (2009). 'Planes, politics and oral proficiency: Testing international air traffic controllers'. *Australian Review of Applied Linguistics* 32 (3), 25.1–25.16. DOI: 10.2104/aral0925.