Developments in LSP Testing 30 Years On? The Case of Aviation English

Henry John Emery

To cite this article: Henry John Emery (2014) Developments in LSP Testing 30 Years On? The Case of Aviation English, Language Assessment Quarterly, 11:2, 198-215, DOI: 10.1080/15434303.2014.894516

To link to this article: http://dx.doi.org/10.1080/15434303.2014.894516

Published online: 21 May 2014.

Article views: 556

Citing articles: 1 View citing articles
Developments in LSP Testing 30 Years On? The Case of Aviation English

Henry John Emery

Latitude Aviation English Services, Plymouth, United Kingdom

The proceedings of the first Language Testing Forum in 1980 were published in *ELT Documents 111: Issues in Language Testing* (Alderson & Hughes, 1981). Discussants at the 1980 Forum raised a number of questions on Language for Specific Purposes (LSP) testing relating, notably, to test specificity, test content, the relationship between subject matter knowledge and language knowledge and predicting real-life language performance. The 2010 Language Testing Forum looked back at the last three decades in language testing to reflect on what developments, if any, have occurred. Following the 2010 Forum, this article addresses the questions raised in 1980 with reference to testing for a very specific purpose—the International Civil Aviation Organisation’s Language Proficiency Requirements for pilots and air traffic controllers. In analysing the testing context—aeronautical radiotelephony communications—the author argues that, in spite of theoretical and methodological advances in LSP testing, these questions are still as relevant to testing LSP today as they were in the early 1980s.

### INTRODUCTION

To mark the 30th anniversary of the Language Testing Forum, the theme of the 2010 Language Testing Forum, held at Lancaster University, was Issues in Language Testing Revisited. The purpose of the forum was to present reactions from today’s perspective to the relevant account of the 1980 discussions with particular reference to the *ELT Documents 111: Issues in Language Testing* (Alderson & Hughes, 1981) and to reflect on developments since then. This article covers reactions to one of the three main issues raised in 1980, namely, *Testing for Specific Purposes*. This article draws on the experience of working in language testing in the domain of aviation English, or more specifically, the development of a test of English language proficiency for pilots. In the process of developing such a test—analysis of the target language use domain and language policy context, the development of test and task specifications and the development and trialling of the test blueprint—test developers are faced with a multitude of theoretical and practical challenges which formed the basis of reactions to the discussions of Language for Specific Purposes (LSP) in the early 1980s.

On reviewing Section 3 of *Issues in Language Testing* (Alderson & Hughes, 1981), it could easily be said that much has happened in the field of LSP testing in the three decades since
Brendan Carroll and his colleagues were faced with the Herculean challenge of developing a test of English for entrance into higher education in Britain. Our understanding of specific purpose language use is unarguably much deeper than it was in 1981, and LSP test developers of today are equipped with more robust theoretical frameworks, more detailed and clearly defined procedures for LSP test development, and a broader and more sophisticated range of techniques for quantitative and qualitative investigations of test validity and reliability. With the benefit of hindsight, one could be forgiven for reading *Issues in Language Testing* and quickly concluding, on first sight, that the issues raised at the first Language Testing Forum are no longer relevant to today’s debate and that the discipline has long since moved on.

However, things are never that simple. Chapter 3 of *Issues in Language Testing* contains a summative discussion of Carroll’s paper on English for Academic Purposes with regard to the British Council’s English Language Testing System (ELTS), the predecessor of IELTS, and reactions to it. An examination of this chapter reveals a number of issues fundamental to the project of testing LSP, which are perhaps as valid today as they were 30 years ago. They are, at least, valid issues in the development of language tests for pilots and air traffic controllers. These issues are as follows:

1. In testing language for specific purposes, how specific is specific?
2. How does one decide what is to be tested?
3. Can one be relatively sure that one is not testing subject-matter knowledge rather than linguistic or communicative abilities?
4. How can one predict from one performance on a specific test to performances in real life?

Before turning to a discussion of these questions with reference to LSP testing for pilots and air traffic controllers, I describe the specific-purpose language domain of aeronautical radiotelephony communication and the international language policy framework that governs it.

**RADIOTELEPHONY COMMUNICATIONS**

All aircraft in controlled airspace should adhere to specific procedures and manoeuvres, and each of these procedures and manoeuvres has a corresponding phrase. These phrases, documented and sanctioned by aviation authorities, form the language of routine air–ground communications and are together known as standard radiotelephony phraseology. On busy radio frequencies, where time is limited, radiotelephony phraseology is a code that allows for the brief, concise, and accurate transmission of specific information pertaining to the flight. This prescribed code of predetermined phrases separates itself from the wider English language by its standardized and nonidiomatic forms and usage. The following exchange between a controller and the flight crew of an aircraft on approach to an airfield demonstrates this:

**Air Traffic Control Officer:**

*Fastair 345, cleared straight in ILS approach runway 28, descend to altitude 3000 feet QNH 1011, report established on the localiser.*

**Pilot:**

*Cleared straight in ILS approach runway 28, descend to altitude 3000 feet, QNH 1011, Wilco, Fastair 345.*
As a core component of professional training, all pilots and air traffic control officers undergo training in the use of the radio and radiotelephony phraseology. Owing to the restricted, repetitive, and situationally dependent nature of radiotelephony phraseology, it is relatively straightforward to learn. As Mitsutomi and O’Brien (2004) pointed out, “Those who undertake aviation studies with the goal of participating in international operations will memorize this standard phraseology in English, whether they speak the language as a native or not” (p. 8).

One of the reasons that aviation communications are as safe and efficient as they are is due to the repetitive and predictable nature of flight operations and associated radiotelephony phraseology. In the commercial aviation system of today, thankfully, things rarely go wrong; millions of flights are conducted routinely and safely each year, resulting in the transmission of millions of radio messages between thousands of pilots and air traffic control officers, the majority of whom do not have English as their first language. As things rarely break away from the routine in-flight operations, radiotelephony phraseology serves its purpose very safely and effectively. That is not to say that the system is foolproof; on the contrary, a number of studies show that miscommunication does occur (Barshi & Farris, 2013; Cushing, 1994; Morrow, Lee, & Rodvold, 1993; Prinzo, 1996; Van Es, 2004; Van Es, Wever, & Verbeek, 2006), although “many instances of miscommunication and communication problems are apparently caught and solved by the controllers and pilots, leaving them only with momentary confusion or annoyance” (Van Es et al., 2006, p. 7). Generally speaking, given the sheer volume of communication and the relatively low rate of reported occurrences, the system is regarded as “robust” (Van Es et al., 2006, p. 7).

Owing to its formulaic nature, radiotelephony phraseology serves the purposes of routine communications very well. However, it is because of this very nature that radiotelephony phraseology cannot serve all pilot–controller communications, especially in nonroutine situations. (The term “nonroutine” for the purposes of this article describes a range of situations, from those that do not directly affect the safety of a flight, such as a passenger with a medical problem, to a full-scale emergency, such as an aircraft with a hydraulic failure in flight.) In nonroutine situations, pilots and controllers are required by force of circumstance to move beyond standard radiotelephony phraseology into “plain” language in order to pass and receive messages. The following exchanges are what may be heard on the radio frequency in nonroutine situations:

Pilot:

*Control. Redline 253. We have a passenger on board with suspected heart attack. We'd like to divert to the nearest available airfield. Request full medical assistance on arrival.*

or

*Genesis 1415. Understand you are having problems with your ailerons. Are you able to make left turns?*

In a discussion of English as a Specific Language, Mackay and Mountford (1978) acknowledge that “the language of international air-traffic control . . . is strictly limited and can be accurately determined situationally. Such restricted repertoires are not languages . . . knowing a restricted ‘language’ would not allow the speaker to communicate effectively in novel situations” (pp. 4–5).

Similarly, Davies (2001) suggested that “where the target LSP is formulaic (for example, the English of air traffic control), it must depend on a broader proficiency in order to deal with emergencies which no ritualised code can encompass” (p. 138). This broader proficiency
is the objective of the International Civil Aviation Organization (ICAO) Language Proficiency Requirements to which the discussion now turns.

**THE ICAO LANGUAGE PROFICIENCY REQUIREMENTS**

In 2003, ICAO introduced the Language Proficiency Requirements following a series of high-profile accidents in commercial aviation in which insufficient language proficiency on the part of the flight crew and/or the air traffic controller was found to be a contributory factor. The Language Proficiency Requirements strengthened the existing provisions for English language communications in the 1944 Convention on Civil Aviation in order “to ensure that the language proficiency of pilots and air traffic controllers is sufficient to reduce miscommunication as much as possible and to allow pilots and controllers to recognize and solve potential miscommunication when it does occur” (ICAO, 2010, Section 4.2.1) and “that all speakers have sufficient language proficiency to handle non-routine situations” (ICAO, 2010, Section 4.2.2). Under the Language Proficiency Requirements, all pilots flying across international borders, and all air traffic control officers managing international traffic are required to demonstrate the ability to speak and understand the English language as used for international radio communications.

Some key features of the Language Proficiency Requirements are as follows.

**Language**

First, in recognizing that the use of nonstandard radiotelephony phraseology poses a threat to flight safety, the Language Proficiency Requirements stipulate the use of ICAO standard radiotelephony phraseology specifically. Second, ICAO recognizes that it is not possible to develop radiotelephony phraseology to cover every conceivable situation, and so the Language Proficiency Requirements make explicit the need for “plain” language proficiency as a fundamental component of aeronautical communications. ICAO defines “plain” language as “the spontaneous, creative and non-coded use of a given natural language” (ICAO, 2010, p. x) and

---

1 Until March 2003, language provisions were addressed through recommendations that English be made available whenever an aircraft station was unable to communicate in the language used by the station on the ground. It was also stipulated that air traffic controllers demonstrate knowledge of “the language or languages nationally designated for use in air-ground communications and ability to speak such language or languages without accent or impediment which would adversely affect radio communication” (ICAO, 2010, Section 1.3.1). ICAO acknowledges that there were no similar requirements for flight crew and that, in the absence of a clearly defined required proficiency level, harmonisation was difficult and assessment uneven.

2 “In Annex 10, Volume II, it is stipulated that radiotelephony communications shall be conducted either in the language of the station on the ground or in English, and that English shall be made available when pilots are unable to use the language of the station on the ground” (ICAO, 2010, Section 4.3.5). A problem often cited by pilots operating in international airspace is the use of English and local languages on the same air traffic control frequency. For example, a French-registered aircraft flown by a French-speaking crew is number one to land at Paris Charles de Gaulle and the crew are communicating with local control in French. Number two to land is a Chinese-registered aircraft flown by a Chinese-speaking crew communicating with the controller in English. Should a nonroutine event occur for the French aircraft and communications ensue in French, the Chinese crew would experience a loss of what is known as “situational awareness” (unless, of course, they can speak French). The same would apply to any other non-French-speaking crew on the same frequency.
states that plain language shall be used “only when standardized phraseology cannot serve an intended transmission” (ICAO, 2010, Appendix A).

When plain language is required, it should be delivered in the same clear, concise and unambiguous manner as standardized phraseology in emergencies or unusual situations, to clarify or elaborate on instructions or when the need to negotiate information or instructions arises. (ICAO, 2010, Section 4.3.4)

Applicability

The Language Proficiency Requirements apply equally and exclusively to pilots, air traffic controllers, and aeronautical station operators. They do not provide for intracoacht communications or communications between air traffic control officers on the ground, for example, between different air traffic control units via telephone.

Availability of English

Although it is recognised that the use of more than one language on the radio is a threat to safety, the Language Proficiency Requirements do not specify an English-only communicative environment. “ICAO provisions do not in any way limit the use of a national, regional or local language but recognize the practical requirement for English to be available for the many pilots who do not speak the national language of a particular State” (ICAO, 2010, Section 4.3.5). Thus, beyond the airspace of the English-speaking world, it is common to hear local languages used in radio communications. For example, a Spanish-speaking crew operating a flight from Santiago, Chile, to La Paz, Bolivia, is likely to conduct communications exclusively in Spanish.

Rating and Personnel Licensing

To make explicit the level of language proficiency required for personnel licensing, ICAO developed a Rating Scale with six language criteria: Pronunciation, Structure, Vocabulary, Fluency, Comprehension, and Interactions. The scale runs from Level 1 (Elementary) to Level 6 (Expert), the threshold level for personnel licensing being Level 4 (Operational). A test taker’s overall proficiency rating is determined by the lowest rating assigned in any of the criteria of the rating scale.

Recurrent Testing

Under the ICAO Language Proficiency Requirements, all personnel must be retested periodically. ICAO recommends that those demonstrating Level 4 proficiency are retested every 3 years, those demonstrating Level 5 proficiency are retested every 6 years, and those demonstrating Level 6 are exempt from further retesting.

Testing Radiotelephony Phraseology and Plain English

Although recognising that test tasks may prompt and elicit test taker responses that contain radiotelephony phraseology, ICAO recommends that language assessment for either pilots or air
traffic control officers for the ICAO Language Proficiency Requirements should focus on plain English. Furthermore, as each national aviation authority has established procedures for training and testing radiotelephony phraseology, and this training and testing is the preserve of operational subject matter experts, ICAO strongly recommends that language proficiency tests should not be designed to assess radiotelephony phraseology.

**ISSUES IN LANGUAGE TESTING**

Having introduced the language context for this article, that is, testing English language proficiency for pilots, we may now turn to a discussion of the first of the questions raised in *Issues in Language Testing* (Alderson & Hughes, 1981).

1. **In Testing Language for Specific Purposes, How Specific Is Specific?**

   In terms of language test specificity, the logical conclusion for LSP testing is one specific test for one specific person at one specific time. For obvious reasons, this is not possible. So, as ICAO (2008) recommends that “general-purpose language tests are not appropriate for the specialized domain of aviation language testing” (Section 3.1.3), test developers need to decide where on the general-specific continuum to place an LSP test for pilots.

   The ELTS presented by Brendan Carroll and colleagues as an example of LSP testing in practice was a test of English for Academic Purposes (Carroll, 1981). Naturally, the resulting reactions and discussion tended to centre on the English for Academic Purposes endeavour, a tendency seen in the research focus and debate in language testing due to the expansion in numbers of those seeking English medium higher education. Alderson (1981) suggested that “the level of specificity chosen for the test is inevitably arbitrary” (p. 127), and although this point does hold water, it is possible to apply some logic when deciding on the specificity for tests of aviation English. To begin, let us consider first the target population for the ICAO Language Proficiency Requirements: pilots, and air traffic control officers involved in international civil aviation.

   In participating in radiotelephony communications, pilots and air traffic control officers occupy two ends of the same communicative channel. Clearly each must be equally proficient in the language of aviation; each must be able to speak to and understand the other, and there are many commonalities in terms of language use for the two participants—setting, participants, purposes, form and content, tone. However, beyond the radio interface where the two professions meet, the similarities end; flying aircraft and controlling traffic are very different roles, and within the parameters of aircraft performance, operating procedures, airspace, and aerodrome organisation and infrastructure, the pilot and air traffic control officer pursue different objectives. The pilot’s concern is transporting the aircraft and its contents from origin to destination as quickly, comfortably, and above all as safely as possible. “Most air carriers operate under a system of prioritized goals including safety, customer service (on-time departures and arrivals) and operating economics” (Midkiff, Hansman, & Reynolds, 2004, p. 4). Air traffic control officers, on the other hand, are primarily concerned with issuing “such ATC [Air Traffic Control] clearances as are necessary to prevent collisions and to expedite and maintain an orderly flow of air traffic” (ICAO, 2001, Section 4.5.1.7).
How do these roles translate to communications? Broadly speaking, air traffic control officers issue commands and instructions to pilots, and pilots acknowledge and comply. However, as the two parties have different and sometimes exclusive objectives, pilots and air traffic control officers routinely engage in negotiation regarding the assignment of altitude, speed and direction. ICAO states that “while pilots and controllers are communication partners, they approach the task from different perspectives, and therefore their communication differs in purpose and standpoint” (ICAO, 2010, Section 3.3.2). Therefore, to engage both language knowledge and strategic competence, which Douglas (2000) argued is “central to the LSP enterprise” (p. 38), developing separate tests for pilots and air traffic control officers would seem a fairly obvious distinction in terms of test specificity. ICAO advises that “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. Test content should, therefore, be relevant to their work roles” (ICAO, 2010, Section 6.2.8.3). Indeed, if the purpose of a test of LSP is to make inferences about a test taker’s ability to perform on-the-job language tasks in English, that is, the ability to communicate from the air (in the case of pilots) and communicate from the ground (in the case of air traffic control officers), it is simply not possible to elicit a domain specific language performance using the same test tasks.

Licensed operations personnel are those who have successfully completed theoretical and practical training and have earned the right to exercise the privileges extended under their licence. As such, specific purpose language testing for the ICAO Language Proficiency Requirements will necessarily exclude ab-initio pilots and air traffic control officers, that is, those still in training and not yet licensed. Thus, the student–professional distinction would appear to be a fairly clear second line of specificity to draw, a point also made by ICAO: “The degree of aviation specificity, will be determined in part by at what point in the training or career the test is given” (ICAO, 2004, Section 6.8.7).

Let us now look separately at the two professional domains. In terms of air traffic control, there are three main positions: aerodrome (e.g., controlling traffic within a cylinder-shaped sector of airspace in the immediate vicinity of an airfield), approach (e.g., controlling departing and arriving traffic from the aerodrome limit up to, e.g., 30 miles horizontally and up to 12,000 ft vertically), and en route (e.g., a sector of airspace 200 miles wide from 12,000 ft to 50,000 ft). Then there is the type of control service offered, for example, procedural control or radar control, and the class of airspace, from Class A to Class E. Different positions and services have accompanying standard phraseology and differences in the nonroutine scenarios that may occur. To illustrate, it is possible that an en-route air traffic control officer managing Class A airspace from 18,000 ft under radar surveillance may not know how to manage a situation, either operationally or in terms of communications, in which a light aircraft in Class D airspace has a landing gear failure on final approach.

In testing for pilots, there are a number of ways that a developer could draw the line of specificity. A test could be specified according to aircraft class or category, for example, Multi-Engine Piston, Single-Engine Piston Sea, or Helicopter. Likewise, it could be specified according to aircraft type: Embraer ERJ 145 or Boeing 747-400. Another alternative would be to specify according to pilot licence and develop tests for Private Pilot Licence holders, Commercial Pilot Licence holders, or Airline Transport Pilot Licence holders.

Holding a pilot licence can mean different things. On a spectrum we may have at one end a Private Pilot Licence holder with several hundred hours’ experience who rents a single engine light aircraft to visit friends in the neighbouring country for fun on weekends. At the other, we
may have an Airline Transport Pilot Licence holder, a captain flying a multiengine heavy jet in transoceanic airspace from continent to continent. Clearly, the rules and procedures covering these two extremes are very different, as is the breadth and depth of training required to exercise these flying privileges. But are the communicative contexts the same?

If flying under Instrument Flight Rules in controlled airspace, which is very often the case with international operations, then essentially they are the same. However, there are differences in the risks associated with flying at these two ends of the spectrum. For example, the light aircraft is unpressurised and rarely flies at altitudes where oxygen may become an issue, whereas at 40,000 ft, cabin pressurisation is essential to the lives of those on board. Consequently, the pilot of the jet aircraft ought to be well trained in how to manage a scenario in which cabin pressurisation fails, including how to communicate with air traffic control in such a situation. Conversely, pilot incapacitation or engine failure is a much more serious situation where there is only one pilot or one engine—as could be the case with the private pilot. However, it is quite possible that heavy jets and light aircraft operate in the same airspace—indeed, they often do.

To this, geographic, meteorological, and procedural context need to be added. Heathrow is not subject to dust storms, and it does not pour with rain in Abu Dhabi. The terrain and meteorological issues associated with Alpine flying will not be relevant to an Australian bush pilot, and although turtles on the runway are a common hazard in Costa Rica, they are not in Zurich. Procedures for North Atlantic crossings are specific and complex given the sheer density of traffic flying at high speed, the absence of radar coverage and the availability of High Frequency radio (rather than Very High Frequency). If a pilot does not, and will never, fly across the North Atlantic at altitude, then it would seem inappropriate to bring such content into tests.

Douglas (2000) warned that “language use, even in highly restricted domains...is so complex and unpredictable that coverage, or sampling of tasks, will be inadequate” (p. 12). However, in terms of developing nonroutine scenarios specific to pilots, there are only a certain number of nonroutine situations that can be treated in a language test beyond which content moves into the realm of fantasy or tragedy. Take, for example, an incident in 1990 in which a British pilot was sucked out of the cockpit of a BAC-111 during climb due to the build up of cabin pressure on a badly fitted windscreen, his life saved only by the crew hanging onto his legs during an emergency descent and landing (Air Accident Investigation Branch, 1992). Or the incident in New York in 2009 in which an A320 suffered double engine failure due to bird ingestion on initial climb and the commander elected to perform a ditching in the Hudson River, resulting in only relatively minor injuries sustained by those on board (National Transportation Safety Board, 2010). Both events, superb stories of heroism and airmanship as they are, would probably not make workable scenarios in a language test. At the same time, nonroutine events in flight operations are, by their very definition, unpredictable and can be extreme. Furthermore, we are reminded that test takers “may find relatively infrequent activities very difficult and of crucial importance” (Alderson, 1981, p. 130). To conclude, the crux of the issue as regards test specificity seems to lie in the development of tasks and sampling that are appropriate to pilots’ licensing requirements (i.e., internationally, and very often, under Instrument Flight Rules), while maintaining sufficiently diverse yet plausible content. Language performance generated in such tasks may then be interpretable as evidence of the specific purpose language ability. As we have now moved into a discussion of sampling and content, this would be a good time to raise the second question.
2. How Does One Decided What Is to Be Tested?

At first glance, this question appears to be relatively straightforward to tackle in the case of aviation English. After all, radiotelephony communication is a very restricted target language use domain, and a vital component of the language itself, radiotelephony phraseology, is documented and officially sanctioned. Whether in radiotelephony phraseology or in plain English, radio communications are wholly contingent on the status of the aircraft and the procedures and environment in which it is operating. As such, aviation English ought to be specifiable with great exactness. However, as Davies (2001) pointed out, a historical perspective teaches us that “language varieties have defied discrete description” (p. 42), which throws the theoretical basis for LSP testing into question. Indeed, LSP boundary definition is a slippery subject that leads test developers to ask fundamental questions in the development of specifications. In 1984, Skehan wrote that

although, at first sight, “waiter behaviour” might seem to be a straightforward affair, we soon need to ask questions like: what range of customers needs to be dealt with? What range of food is to be served? Once one probes a little, the well-defined and restricted language associated with any role is revealed to be variable, and requiring a range of language skills. (p. 216)

If we substitute some of Skehan’s words with those relevant to aviation, the assertion is equally true for the testers of aviation English of today.

Although at first sight “pilot behaviour” might seem to be a straightforward affair, we soon need to ask questions like, What range of nonroutine situations needs to be dealt with? What range of flight conditions is the pilot likely to encounter? Once one probes a little, the well-defined and restricted language associated with flight is revealed to be variable and requiring a range of language skills. Development of a test for pilots, then, requires precise specification of the test takers and analysis of the domain in which they operate.

In turning to the literature for help, one finds that, fortunately, commercial aviation is a much-studied industry, and there is a volume of literature on the subject of air–ground communications produced in a variety of fields and disciplines, from aviation psychology to applied linguistics (Barshi & Farris, 2013; Bürki-Cohen, 1995; Cardozi, 1993, 1994; Cardozi, Brett, & Han, 1996; Global Aviation Information Network, 2004; Lopez, Condamines, Jesselin-Leray, O’Donoghue, & Salmon, 2013; McGrath, 2011; Mell, 1992; Morrow et al., 1993; Nübold & Turner, 1983; Prinzo, 1996; Prinzo & Thomson, 2009; Rubenbauer, 2009; Sassen, 2005; Van Es, 2004; Van Es et al., 2006).

Training materials for RT communications are numerous and readily accessible, for example, see Gardner (2010), Leech (2006), Robertson (2009), Thom (2003), and Ward (2006). Very important, ICAO has published a series of detailed guidance documents for stakeholders in aviation English testing, notably Manual on the Implementation of the Language Proficiency Requirements (ICAO, 2010) and the Rated Speech Samples Training Aid (ICAO, 2012), as well as the rating scales by which language proficiency is to be assessed.

However, in spite of what would appear to be a pretty sound starting point for test development, things are not so straightforward. Unfortunately, although the ICAO guidance material is very useful in laying out the context of and purpose for aviation language testing, it is of little practical use in the definition of the construct and the development of test specifications. The assessment criteria themselves present a collection of language abilities organised by criteria, and
although the second edition of document 9835 (ICAO, 2010) provides some accompanying notes on the rating scale, not all of these abilities are clearly defined, and they require interpretation. In addition, the operationalisation of some of the abilities would appear to be problematic; for example:

Is sensitive to verbal and non-verbal cues and responds to them appropriately (ICAO Rating Scale, Interactions, Level 6).

The scales have been in use in aviation language testing worldwide for more than 7 years, though there is no empirical evidence to support their validity for the high-stakes context in which they are used. Prinzo and Thomson (2009) highlighted challenges associated with using the scale in assessing radiotelephony communications, in particular “the language used within the ICAO descriptors themselves and the difficulty in applying them when awarding a grade” (p. 11) and suggested that further research into the use of the scale is needed. Post hoc research is emerging, which suggests that the scale descriptors for Level 4 do not describe a language level adequate for operational flying (Knoch, 2009).

Speaking on criterion referencing in 1984, Skehan (1984) commented that “one immediately is led to ask here what the basis is for the grouping of such language features and whether there is any evidence for the claim that the different features conjoin to define this level” (p. 217). The same questions might be asked of the ICAO rating scale.

In terms of the very objective of the ICAO Language Proficiency Requirements—plain English in radio communications—there is a dearth of research. Some have turned their attention to language use in nonroutine situations. For example, Mell (1992) analysed and categorised pilot–controller communications in routine and nonroutine situations, and Sassen (2005) analysed a corpus of aircraft cockpit voice recorder (also known as the black box) recordings for Head-driven Phrase Structure Grammar for radiotelephony communications. Research on plain language in radiotelephony communications is emerging (e.g., Lopez et al., 2013), though at the time of writing, there was no published research into language use in nonroutine situations in which the careful and successful negotiation of meaning in plain language was part and parcel of the safe resolution of a nonroutine situation.

Therefore, in terms of deciding what must be tested (and taught), we are in a position where aviation English course writers and test developers are working largely on expert intuition. Of course, the expertise of experienced pilots and air traffic control officers gives integrity and credence to test development, and we cannot overlook the value of a commonsense, intuitive approach to the creation of test specifications. Moreover, at just 8 years old, the implementation of this global, industry-wide language proficiency requirement is arguably still in its infancy. That said, empiricism in construct definition is lacking and, in the absence of published needs analyses, one might presume that test developers are operating on not much more than informed guesswork. Some years ago, Skehan (1984) posited that “the basis on which we select particular skills as important for particular ESP roles, is, to say the least, ad hoc” (p. 210). Just a few years later, Alderson (1988) cautioned that “problems of inadequate or biased sampling abound, and criteria for the inclusion and exclusion of potential items, skills or abilities are often of dubious validity” (p. 221). Alarmingly, in spite of the critically high-stakes environment of aviation English testing, these warnings still ring true.

Researching plain English in radiotelephony communications presents a considerable challenge for several reasons. Unlike other specific language use domains, for example, finance or law,
one cannot simply observe a nontest language use situation armed with a recorder and a notepad. In the post-9/11 environment, flight decks and air traffic control facilities are subject to tight security measures, often leaving the researcher dependent on postfactum transcripts and recordings. Moreover, such material is often released only when a catastrophic accident has occurred, resulting in an accident investigation, and therefore, the researcher cannot be sure of the extent to which language use contributed, or at least did not hinder, the efforts of personnel managing the situation. Gaining access to the authentic day-to-day plain English exchanges in situations where there is no direct threat to flight safety, and those more serious incidents, which occur but which conclude safely, is very difficult as recordings are rarely made public and are commonly subject to legal and proprietary restrictions.

There are, indeed, some very useful resources available to the researcher, for example, http://www.liveatc.net. This website provides free access to live air traffic control streams. LiveATC also hosts a repository of prerecorded radio communications of a nonroutine nature. Although these resources are very useful, they tend to contain, in the case of live streaming, communications that are almost exclusively routine radiotelephony phraseology and, in the case of prerecorded communications, have a bias toward North American native-speaking personnel. Finally, and of importance, radiotelephony communications are governed by the KISS principle (EUROCONTROL, n.d.)—Keep It Short and Simple—which is absolutely crucial in busy airspace. Pilots and controllers are routinely trained to keep their transmissions precise, concise, and brief and are actively trained not to move away from standard phraseology unless absolutely necessary. Then, when something does go wrong on the flight deck, controllers are trained to maintain silence on the frequency, giving the pilot the time space to communicate while managing a very high workload (EUROCONTROL, 2003). The result is that the actual quantity of plain English that occurs in authentic communications is relatively low, belying its crucial importance for flight safety.

3. Can One Be Relatively Sure That One Is Not Testing Subject-Matter Knowledge Rather Than Linguistic or Communicative Abilities?

The extent and nature of the relationship between subject matter knowledge and performance on language tests and the threat this represents to the validity of test scores is a key issue in LSP testing. In the 30 years since the publication of *Issues in Language Testing*, researchers have set about the task of investigating this relationship with results (Alderson & Urquhart, 1985a, 1985b; Clapham, 1996), which have greatly informed language testing and have led to a more comprehensive understanding of the LSP construct. Often the question is treated with respect to tests used for decisions about entrance into education and training programmes, such as the ELTS presented at the first Language Testing Forum in 1980, a discussion of which led Alderson (1981) to note that “[a] . . . problem is that of lack of knowledge: a specific test might well assume or presuppose subject knowledge that the testees do not have” (p. 127).

When the ICAO Language Proficiency Requirements were introduced in 2003 with an implementation date of 2008, the vast majority of those identified for testing in that 5-year period were already holders of pilot and air traffic control officer licences. Furthermore, as English language proficiency is a recurrent testing requirement, a large proportion of the future test-taker population will be current licence holders. At the same time, language testing is also a requirement for those personnel earning a licence and entering the system upon completion of training. For
test developers such as Oxford Aviation Academy, whose core business is to supply trained flight crew to the airline industry, a test for pilots must be appropriate not only to experienced pilots but also to those newly trained and licensed pilots too. As with all test development projects, decisions have to be made about test tasks and content based on assumptions about levels of subject matter knowledge among the test-taker populations. To shed light on these assumptions, it is perhaps useful here to describe training for pilot licensing. Next is a description of a typical integrated pilot training programme according to the European Aviation Safety Agency (2011) syllabus for Airline Transport Pilot License in Europe. Such programmes, offered by many flight training organisations in Europe, lead cadets from “zero flight hours” to readiness for the position of airline First Officer.

Pilot cadets begin with 6 months of theoretical ground training covering 14 subjects ranging from airframes and navigation to aircraft performance and meteorology. Once cadets have passed all 14 theoretical examinations, approximately 6 months of foundation flying training commences, during which cadets gain flying experience in single-engine piston aircraft. Upon completing this phase, cadets then continue for a further 6 months with advanced multi-engine flying training, which concludes with the demanding Instrument Rating examinations. Now holding a “frozen” Airline Transport Pilot Licence, the pilot undergoes Multi-Crew Cooperation training in a jet simulator before moving on to Type Rating, in which the pilot undergoes classroom and simulator training on a particular aircraft type, a great deal of which involves rehearsing and practicing nonroutine situations in flight. Once Type Rating is complete, the pilot joins the airline and typically undergoes base and line training involving familiarisation with the airline’s policy and standard operating procedures. Finally, the pilot will take the right-hand seat on the flight deck as first officer for a “check ride,” flying under the observation of a training captain. Every 6 months for the rest of their flying career, the pilots will return for recurrent training in the aircraft simulator to practise handling the aircraft in nonroutine and emergency situations. Language proficiency testing may be conducted at various points in the training pathway depending on where pilot training takes place, with, as previously mentioned, a periodic retesting requirement once a licence has been endorsed at ICAO Level 4 or 5. This rigorous, systematic, closely regulated, harmonised, and continuous process of pilot training is undoubtedly one of the reasons why aviation is the safest form of transport. In short, one can be reasonably confident that licensed pilots will be experts in their aircraft, in operating procedures, in radiotelephony phraseology communications and in flight safety.

Five of the six criteria in the ICAO rating scale refer to “common, concrete and work-related topics” and “familiar topics.” Therefore, the job of the test developer is to select topics that are familiar to licensed pilots while being mindful of the fact that “the more specialised test content becomes, the greater the influence of specific purpose background knowledge” (Douglas, 2000, p. 92). ICAO (2008) cautioned,

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. (Section 1.2.2)

The need to strive toward a balance in the selection of topics that are at once universally familiar to pilots and are yet not too subject specific highlights again the importance of involving subject specialist informants, in this case, licensed pilots, flight instructors and examiners, and air traffic control officers in LSP test development.
ICAO document 9835 states that

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 deserve to be tested in a context similar to that in which they work. Test content should, therefore, be relevant to their work roles. (ICAO, 2010, Section 6.2.8.3)

It follows, then, that test tasks that are designed to simulate the “real-life situation,” that is, radiotelephony communications, will logically engage the test taker in radiotelephony phraseology and plain English in nonroutine situations. Although ICAO recommends that the test should not be designed to assess phraseology, knowledge of radiotelephony phraseology, and the whole range of aircraft manoeuvres and procedures that radiotelephony phraseology represents, is implicit. Thus, performance in a test of LSP for pilots is thoroughly dependent on subject matter knowledge. To turn the question on its head, if a test taker did not possess any such knowledge, they would be as unable to perform in such test tasks as they would be unable to fly an aircraft.

Thirty years ago, the following question was asked about subject matter knowledge in LSP testing with regard to the ELTS: “How can one possibly avoid involving prior knowledge, since comprehension and presumably production must depend upon the prior existence of some set of knowledge?” (Alderson, 1981, p. 127). It would seem, in the case of aviation English, that the question of whether it is possible or even desirable to separate subject matter knowledge from language knowledge is perhaps less relevant for three reasons. First, the test population is made up of licensed professionals known to be expert in their field. Second, the language policy itself requires that test content be field specific. Third, it is now generally accepted that the construct for LSP testing is one that allows for interaction between the test taker’s language ability and specific purpose content knowledge. That is not to say that we need not be concerned with the relationship between subject matter knowledge and language knowledge. On the contrary, as long as we know that background knowledge impacts LSP test performance, further research with regard to the aviation domain is essential; otherwise, as Douglas (2000) pointed out, “true LSP test development, authenticity in test performance and valid interpretation of language test results will be elusive goals” (p. 27).

In the next section, we turn to the fourth and final question.

4. How Can One Predict From One Performance on a Specific Test to Performances in Real Life?

The predictive validity of specific purpose tests of aviation English is very difficult to determine. However, this difficulty is not confined to tests of aviation English. It is a commonly cited problem in the validation of proficiency tests of English for Academic Purposes, where the criterion for predictive validity is often said to be the test taker’s performance on academic courses in the 1st year of study. Such a criterion, needless to say, is hugely complex and relates to matters of academic knowledge, skills, hard work, and much more than it does to language proficiency. Moreover, predictive validation is only one sort of validity evidence and other sources of evidence (construct, content, concurrent, consequential . . .) might be more easily gathered and more convincingly related to language ability. It was suggested in Issues in Language Testing that “many
variables enter into a student’s ultimate performance, in addition to whatever the particular test is measuring” (Alderson, 1981, p. 126), a point that could not be more true than in the case of LSP testing for pilots.

When something goes wrong on the flight deck, workload increases dramatically. The crew is typically alerted to a problem by physical sensations resulting from changes in the way the aircraft flies, or a visual or aural prompt from one of the many indicators on the flight deck. The priority is to verify the source and nature of the problem and find a solution as quickly as possible. In the case of commercial jet operations, this will typically involve both members of the crew running through one or a series of checklists presented on screen or in a quick-reference handbook while controlling the flight; maintaining awareness of position; and, if time permits, communicating the situation to air traffic control. Broadly speaking, the sequence of activities follows the pattern commonly introduced to pilots in flight training—aviate, navigate, communicate. The safe resolution of a nonroutine situation will involve a multitude of technical, environmental, and human factors, of which communication is just one. Indeed, where incidents and accidents occur, they are usually the result of a series of events rather than one single event. Even in the well-publicised accidents commonly cited as giving rise to the ICAO Language Proficiency Requirements, such as the head-on collision of two Boeing 747 aircraft in Tenerife in 1977, language proficiency is usually identified as only one in a chain of contributory factors. Furthermore, communication is itself a complex issue in which second language proficiency is just one factor: “Most communication problems have causal factors associated with human performance limitations. Factors often mentioned are controller workload, frequency congestion, non-standard phraseology, readback/hearback errors, similar call signs, message complexity, speech rate, language proficiency and accent” (Van Es et al., 2006).

LSP testing theory has in more recent years been underpinned by McNamara’s (1996) weak performance hypothesis, that is, we make inferences on the basis of test scores about the underlying ability to use language in the target situation rather than the ability to do particular tasks in the target language use situation. This is in contrast with the “strong” sense of the term second language performance test where “strong” performance test tasks “represent real-world tasks and performance will primarily be judged on real-world criteria, that is, the fulfilment of the task set. Aspects of language ability may or may not be assessed at all” (McNamara, 1996, p. 43). This is a very important position to maintain in the case of LSP testing for pilots and air traffic control officers. Speaking about LSP testing for air traffic control officers, Douglas (2000) illustrated this point:

The raters would not be interested in whether or not the test-taker was able to prevent a mid-air collision, which might be outside his or her control, owing to pilot error, for instance; rather, they would be interested in what the performance could tell them about the nature of the language ability underlying it. (p. 27)

In a sense then, one question we might ask of LSP tests for pilots today is “based on language test performance, does the test-taker demonstrate the ability to communicate to the described ICAO standard?” rather than “Does the test-taker communicate safely in real-life language tasks?”

In a practical sense, researching predictive validity in language testing for the ICAO Language Proficiency Requirements is problematic. As previously discussed, when situations do occur in which English language communications may play a crucial role in the safe resolution of a nonroutine situation, we rarely, if ever, hear about it. Gaining access to the material that would assist
in the research of predictive validity—for example, recordings of test takers engaged in real-life language use—is fraught with difficulty. Tracing a test taker’s interlocutors—the air traffic control officers at the various air traffic control units along the route of an international flight—is impractical. More frequently one hears of aviation authority ramp inspectors being unable to communicate at all in English with flight crew during routine aircraft inspections, in spite of licence endorsement at Level 4. Similarly, stories of licensed pilots being taken “off the line” by flight examiners for reasons of language proficiency are becoming more common. However, such evidence is anecdotal and could easily be more to do with regulatory, political, and economic factors than with issues in language testing. Research is needed.

In any case, no matter how adequately a test covers the content of the domain and no matter how real life the test tasks are, we have yet to understand the extent to which the threshold level of proficiency described in the ICAO rating scale is commensurate with the level of proficiency required for safe radiotelephony communications. For this reason alone, the promise of establishing the predictive validity of LSP tests for the ICAO requirements remains distant, and, for the time being, the requirements for validity will need to be satisfied with other measures of test quality, as previously indicated.

CONCLUSION

To return to the theme of the article, that is, a reflection on developments (if any) in LSP testing in the last 30 years, we might conclude that much and yet little has changed. The four issues that I selected—those relating to test specificity, what to test, background knowledge versus language knowledge, and predictive validity—are still as valid for the enterprise of LSP as they were 30 years ago, though perhaps it is the questions that we need to ask that have changed rather than any resolution of the issues themselves.

I began by discussing test specificity, putting forward the idea that because flying aircraft and controlling air traffic are different jobs, pilots and air traffic control officers require different test instruments in order to allow for appropriate “interaction between the test taker’s language ability and specific purpose content knowledge, on the one hand, and the test tasks on the other” (Douglas, 2000, p. 92). Beyond the pilot–air traffic control officer distinction itself, test specificity for both professional domains can be approached from a number of angles. In whatever way the lines are drawn, careful consideration needs to be given to test specificity, or precisely which type of test taker to target in language test specifications. Obviously, there is a need to develop tests that are fit for purpose for the specified range of test takers. However, more specific tests will inevitably result in smaller test populations, which may make test development impractical.

I then moved on to the issue of deciding what must be tested, which, even in the very restricted context of the ICAO Language Proficiency Requirements, represents a significant challenge. This is due in large measure to the fact that the plain English component of radio telephony communications is underresearched. Plenty is known about the target language use domain. Indeed, flying aircraft and controlling air traffic are possibly two of the most widely studied professional activities. However, a better understanding of the language itself is urgently required, the pursuit of which will require tenacity on the part of the researcher in gaining access to the restricted, sensitive, and geographically disperse environments and language data that would form the basis
of research. Until then, LSP testers may simply follow the advice that “it seems better do what ESP teachers have usually done—react to each particular set of circumstances intuitively in the manner that seems most appropriate” (Skehan, 1984, p. 218), advice that is still as valuable today as it was nearly thirty years ago.

With regard to the subject-matter knowledge debate, what has changed significantly in more recent years is the acceptance that it is neither possible nor desirable to separate language knowledge from subject matter knowledge; in fact, subject matter knowledge has come to occupy a central position in today’s construct of LSP testing. With LSP testing for the ICAO requirements, we know that the test-taker population is trained and licensed professionals, and it has been suggested that “when test takers are known to possess a high level of expertise in the specific field . . . then it will not be necessary to disambiguate language and background knowledge” (Douglas, 2000, p. 39). Nevertheless, LSP testers will need to continue to explore the relationship between subject matter knowledge and language knowledge so long as we know the relationship has an effect on test performance and scores.

Understanding whether tests designed to assess language proficiency for safe radiotelephony communications have any degree of predictive validity is problematic to say the least. The staggering complexity of the flight and air traffic control environment, particularly in nonroutine situations—the equipment and the human factors (cognitive workload, decision making, behaviour, and communications)—is such that determining second language proficiency among the multitude of variables may prove to be an impossible task. Again, in a professional domain that is highly sensitive to issues of security, gaining access to conduct investigations of predictive validity presents serious challenges. In any event, until we have an established degree of criterion validity and therefore a tangible, robust link between the test and the non-test situation, one might suggest that to explore predictive validity is perhaps to build castles in the sand.

Notwithstanding the theoretical and practical challenges associated with proficiency testing for the ICAO Language Proficiency Requirements, the pursuit of LSP testing remains, at least in this context, wholly worthwhile. Davies (2001) suggested that “if LSP testing has a beneficial impact on learners and teachers, on test candidates and on subject specialists” (p. 134), then in the final analysis, it is the practical need driven “by what learners need to do in the language” (ICAO, 2010, Section 7.4.2) and the pragmatic effect—improvements in the safety of aeronautical communications—that are the justification for LSP testing.

REFERENCES


