

Contexts and constructs: Implications for the testing of listening in pilots' communication with air traffic controllers¹

Contextos e construtos: implicações para o teste de compreensão auditiva de pilotos na comunicação com controladores de tráfego aéreo

Angela C. M. GARCIA (Carleton University)²
Janna FOX (Carleton University)³

ABSTRACT

The International Civil Aviation Organization (ICAO) published the Language Proficiency Requirements for pilots and air traffic controllers (ATCOs) in 2003. Research has shown that there is still a lack of clarity regarding what needs to be assessed in terms of the English used by pilots and ATCOs in radiotelephony (DOUGLAS, 2014; EMERY, 2014; KIM; BILLINGTON, 2016; KIM; ELDER, 2015; MONTEIRO, 2019; READ; KNOCH, 2009). The purpose of the present study was to investigate the nature of the listening tasks performed by pilots as an essential step in test development and validation. This explanatory sequential mixed-methods study (CRESWELL, 2015), elicited questionnaire responses from 156 pilots (Phase 1) and subsequently interviewed six aeronautical English experts (Phase 2) to better understand the characteristics of the listening tasks that pilots undertake during radiotelephony communications. Quantitative and qualitative data were analysed, and findings were merged. They provide information that may usefully inform the development of the listening test construct and the test specifications.

Keywords: ESP Assessment, Aeronautical English, Listening Comprehension, Testing Listening, ICAO

RESUMO

A Organização Internacional de Aviação Civil (OACI) publicou os Requisitos de Proficiência Linguística para Pilotos e Controladores de Tráfego Aéreo (ATCOs) em 2003. Pesquisas mostraram que ainda há falta de clareza em relação ao que precisa ser avaliado no que se refere ao inglês usado pelos pilotos e ATCOs na radiotelefonia (DOUGLAS, 2014; EMERY, 2014; KIM; BILLINGTON, 2016; KIM; ELDER, 2015; MONTEIRO, 2019; READ; KNOCH, 2009). O objetivo deste estudo foi investigar a natureza das tarefas de compreensão auditiva realizadas pelos pilotos, como uma etapa essencial no desenvolvimento e

¹ This article is part of a PhD's research conducted by the first author at the School of Linguistics and Language Studies of Carleton University, under the supervision of the second author.

² Carleton University, Ontario, Canada. PhD student in Applied Linguistics & Discourse Studies; School of Linguistics and Language Studies; ORCID: <https://orcid.org/0000-0003-3018-2370>; angelagarcia@cmail.carleton.ca

³ Carleton University, Ontario, Canada. Professor Emeritus at the School of Linguistics and Language Studies; ORCID: <https://orcid.org/0000-0003-4528-4513>; jannafox@cunet.carleton.ca

validação de testes. Este estudo explanatório sequencial de métodos mistos (CRESWELL, 2015), suscitou respostas de 156 pilotos a um questionário (Fase 1) e posteriormente entrevistou seis especialistas em inglês aeronáutico (Fase 2) para um melhor entendimento das características das tarefas de compreensão auditiva que os pilotos realizam durante as comunicações de radiotelefonia. Dados quantitativos e qualitativos foram analisados e os resultados foram integrados. Eles fornecem informações que podem ser úteis para o desenvolvimento do contrato de teste de compreensão auditiva e das especificações de teste para pilotos.

Palavras-chave: *Avaliação De Inglês Para Fins Específicos, Inglês Aeronáutico, Compreensão Auditiva, Avaliação da Compreensão, OACI*

1. Introduction

Since 5 March 2008 (later extended to 2011), pilots and air traffic controllers (ATCOs) have been required to demonstrate their ability to speak and understand the language used for radiotelephony (RT) communications in accordance with the International Civil Aviation Organization's (ICAO) Language Proficiency Requirements (LPRs). Although the adoption of the ICAO LPRs was a tremendous advance, several issues have been identified with the policy (e.g., DOUGLAS, 2014; EMERY, 2014; KIM; BILLINGTON, 2016; KIM; ELDER, 2015; MONTEIRO, 2019; READ; KNOCH, 2009). The main issue is that the test construct is underrepresented. Emery (2014), for example, argued that “the ICAO guidance ... is of little practical use in the definition of the test construct and the development of test specifications” (EMERY, 2014, p. 206), and Douglas (2004) pointed out that “we need much more information about the nature of language use in aviation contexts outside the prescribed phraseologies” (DOUGLAS, 2004, p. 250). The term construct is defined by Bachman (2007) simply as “what it is that we want to assess” (BACHMAN, 2007, p. 42). He argues that understanding the roles of abilities, contexts and their interactions has been a central problem in the field of language assessment, especially because the way these roles are viewed has a significant impact on how a construct is defined. Construct underrepresentation occurs when a test “is too narrow and fails to include important dimensions or facets of focal constructs” (MESSICK, 1996, p. 4), and this represents a threat to validity.

One of the problems regarding the ICAO policy lies in the fact that listening comprehension represents only one of six criteria included in the ICAO rating scale (GARCIA, 2015). The ICAO rating scale consists of descriptors for the assessment of six different categories or skills: pronunciation, structure, vocabulary, fluency, comprehension, and interactions, that range from Level 1 (Pre-elementary) to Level 6 (Expert). Level 4 (operational) is the minimum level required for international operations. However, in the second edition of ICAO's DOC 9835 (2010), *Manual on the implementation of ICAO*

Language Proficiency Requirements, ICAO included the information that “while comprehension is only one out of six skills in the Rating Scale, it represents half of the linguistic workload in spoken communications” (ICAO, 2010, 4-13). Garcia (2015) interviewed expert ICAO language test developers and raters to discuss the rating scale descriptors. One of the participants remarked that he did not think that “the listening construct in the case of RT communications is adequately defined or captured by the criteria as they stand” (GARCIA, 2015, p. 93). He argued that not only was comprehension inadequately measured, but also its importance was undermined by the rating scale:

Comprehension sitting alongside components of spoken language proficiency firstly diminishes the importance of listening comprehension. If we consider listening/speaking to be skills which are equal and they interact and relate very closely together, the way that comprehension is perceived in the rating scale is perceived as one of six things that students should be able to do, when it is not, it's one of two things that students should be able to do, or pilots and controllers should be able to, one being speaking and the second being comprehension. So, I think it misleads us to think that comprehension is a very thin slice of the ability to speak and it is not, it's an extremely important, if not more important, part of the overall proficiency construct in this case. (GARCIA, 2015, p. 38)

As a result, the author concluded that “it is necessary to conduct research to find out if it is interesting to develop an exclusive rating scale for comprehension, and, in case it is, what aspects it should include” (GARCIA, 2015, p. 62). As pointed out by Kim and Elder (2015), “a single piece of unclarified information could have disastrous results in air traffic control” (KIM; ELDER, 2015, p. 133). Thus, it is extremely important to investigate the nature of the listening comprehension skills of pilots in their communications with the ATCOs. Only with a clear, empirically sound understanding will a test construct adequately represent the features of the target language use (TLU) tasks (BACHMAN; PALMER, 1996). As Field (2019) argued, “when designing a test of one of the language skills, we need extensive information about the nature of the phenomenon we are dealing with. This enables us to shape the material and tasks that we devise so as to ensure that they truly measure competence in the skill” (FIELD, 2019, p. 7). Therefore, the goal of the present study is to identify the characteristics of pilots' listening tasks in the context of RT communications in order to better define the test construct. It aims to contribute to the creation of knowledge regarding the construct of the ICAO LPRs for pilots and ATCOs, and also to contribute to a future revision of the rating scale descriptors. The findings from this study may be useful in informing the development of ICAO language proficiency tests, as well as supporting test developers in the design of listening tests and tasks that are more appropriate to the TLU domain.

As background to the study, in the section which follows below, a review of relevant literature in the fields of language assessment and aeronautical English is provided. This is followed by the specific

research questions that guided the study, the methods, and results and discussion. Concluding remarks include the limitations of the study and provide some suggestions for further research.

2. Literature review

Listening is complex, especially in a second or foreign language. Assessing listening comprehension is challenging, particularly in the case of assessment of Languages for Specific Purposes (LSP), which can be understood as:

[...] one in which test content and methods are derived from an analysis of a specific purpose target language use situation, so that test tasks and content are authentically representative of tasks in the target situation, allowing for an interaction between the test taker's language ability and specific purpose content knowledge, on the one hand, and the test tasks on the other. Such a test allows us to make inferences about a test taker's capacity to use language in the specific purpose domain. (DOUGLAS, 2010, p. 19)

This literature review is divided into three subsections: 1) a selective overview of research on the assessment of listening; 2) the nature of listening in the context of pilots/ATCOs RT communications; and 3) the assessment of pilots' listening ability.

2.1 The assessment of listening

As Buck (2001) pointed out, if we want to assess listening, it is important to first understand the process of listening, which is very complex. Listening can be understood as the recognition and comprehension of spoken language. Listening, similarly to reading, but differently from speaking and writing, is an internal process, i.e. a receptive skill, as it is not observable (WAGNER, 2014). As Field (2019) pointed out, "the listening process takes place in the mind of the listener" (FIELD, 2019, p. 1). At the same time, listening is an active process, as listeners are not empty containers into which the information is poured (LYNCH; MENDELSON, 2010). Buck (2001) found that listeners interpret the information they receive according to their linguistic knowledge (e.g., phonology, lexis, syntax, semantics and discourse structure) and their non-linguistic knowledge (e.g., their topical knowledge, their contextual knowledge, their individual characteristics, and their background knowledge). As he pointed out, "when we listen, we use our background knowledge of the world to set up expectations, and then use those expectations to help us comprehend what we hear" (BUCK, 2001, p. 8). Those expectations might be different among listeners, along with their motives for listening. As Buck (2001) explained, different people have different interests and different needs. Furthermore, listening, as discussed by Lynch and

Mendelsohn (2010), “is not merely an auditory version of reading” (LYNCH; MENDELSON, 2010, p.180). It disappears as quickly as it appears, and patterns of rhythm, sound, stress, and intonation are key to understanding it. Moreover, it has the presence of natural fast speech features and the frequent need to interact immediately. In other words, listening is “an interactive process” (BRINDLEY, 1998, p. 172).

The assessment of listening, as Wagner (2014) noted, “has historically been somewhat neglected and even overlooked in the language literature” (WAGNER, 2014, p.1). Test developers and item writers face some challenges when designing listening tests and creating tasks. It is difficult to describe and to assess “an invisible cognitive operation” (BRINDLEY, 1998, p. 171), such as listening. For example, it is challenging for test writers to create tasks that will not only elicit the use of receptive skills, but also result in a situation where the test-taker will have the opportunity to demonstrate that they can use their skills successfully (HUGHES, 2003).

Test developers face some other challenges when writing test specifications, such as, deciding whether or not the candidates will be allowed to listen to the text more than once (TAYLOR; GERANPAYEH, 2011), issues related to task authenticity (ALDERSON et al., 1995; BRINDLEY, 1998; HUGHES, 2003; LYNCH; MENDELSON, 2010; WAGNER, 2014), to memory (BUCK, 2001; WU, 1998), and to the complexities of the cognitive processes involved in listening (TAYLOR; GERANPAYEH, 2011). Therefore, test developers need to think carefully about the purpose of the test and the specific target language situation in order to make informed decisions.

2.2 The nature of listening in pilots/ATCOs RT communications

RT communications between pilots and ATCOs are very different from normal every-day conversations. Pilots and ATCOs use standardized phraseology, a formulaic code made up of specific words and phrases, in order to communicate via the radiotelephone. Whenever phraseology does not suffice, they must resort to plain language, which is defined as “the spontaneous, creative and non-coded use of a given natural language” (ICAO, 2010, p. x). Listening in pilot/ATCO communication consists of a two-way listening mode as it involves both listening and speaking. As pilots and ATCOs talk on the radio, they must take turns. There is no overlapping, which is very common in other conversational contexts. This kind of communication “represents a very specialized and socially significant form of discourse” (READ; KNOCH, 2009, p. 21.3).

In a typical transmission, the ATCO sends a message, the pilot listens to it, repeats it to the ATCO (*readback*), and the ATCO listens to it and confirms it (*hearback*). Research has shown that visual support

helps listeners to understand spoken texts (BUCK, 2001). However, pilot/ATCO RT communications are voice-only interactions, with no facial references. For this reason, it is very important for pilots and ATCOs to be collaborative listeners. Collaborative listening comprises “making appropriate requests for clarification, back-channelling, making responses to interactional language, or taking responsibility for organising turn-taking” (BUCK, 2001, p. 12).

Both native English speakers (NES) and non-native English speakers (NNES) make mistakes when trying to understand an ATCO’s transmission. However, NNES pilots seem to be more challenged by the English used in RT communications (WU et al., 2019)⁴. Wu et al. found that, NNES pilots “make more overall errors in their transmissions than native English sounding pilots” (WU et al., 2019, p. 9).

A number of factors may make it difficult for pilots to understand ATCOs. Emotional stress or increased mental workload may impact language production, as people have the tendency, under these conditions, to talk louder and faster (PRINZO; BRITTON, 1993, p. 16). Moreover, Estival and Molesworth’s (2016) study showed that increased pilot workload has a negative impact on the accuracy of pilots’ readbacks.

Faster speech rate is another factor that might increase communication difficulty for pilots and ATCOs. Prinzo and Britton (1993) claimed that, even under normal conditions, ATCOs tend to speak too fast, and that pilots also talk faster during heavy workload. Research has indicated that “the faster the speech, the more difficult it is to comprehend” (BUCK, 2001, p. 38). Additionally, it is important to point out that there might be a relationship between how much speech rate interferes with understanding speech and the test taker’s proficiency level, although, according to Révész and Brunfaut’s (2013) study, an increased speech rate did not have a significant effect on advanced level test takers’ results in listening tasks. Experience seems to be another important factor. Estival and Molesworth’s (2016) study showed that a faster speech rate, if combined with lower experience, might still be a problem for more advanced speakers. Their results showed that speech rate had a greater impact on communication accuracy of low qualified NES pilots (private pilot licence or lower) and English as a second language pilots than on high

⁴ According to Wu et al.’s (2019) study, “pilots were classified as ‘native English sounding’ if no foreign accent could be detected and if they were flying with an airline registered in a country where the official language is English (e.g., Australia, New Zealand, the United Kingdom, or the United States of America). Pilots were classified as ‘accented English’ if a non-native English accent could be detected and if the aircraft was registered in a country where English is not one of the official languages (e.g., Japan, Korea, China, Chile). Pilots with an English sounding accent, but who were on an aircraft that was registered in a country where English is not the official language or is one of several official languages (e.g., Hong Kong, Singapore, Malaysia, Fiji) were excluded, as it was less certain whether their native language was English” (WU et al., 2019, p. 3).

qualified NES (commercial pilot license or higher), as the former “were adversely affected by faster ATC speech rate” (ESTIVAL; MOLESWORTH, 2016, p. 234).

Furthermore, accent plays an important role in relation to how easily a pilot understands an ATC communication. Pilots and ATCOs are exposed to a wide variety of unfamiliar accents. In many interactions, both participants are NNES and come from different cultural backgrounds. As Kim and Billington (2016) pointed out, “given the likelihood that both pilots and controllers will encounter English in a variety of accents and proficiency levels, their ability to effectively comprehend a range of accents is important” (KIM; BILLINGTON, 2016, p. 139). For Tiewtrakul and Fletcher (2010), “accent is one critical factor, especially where two non-native English speaking parties have to communicate using the English language” (TIEWTRAKUL; FLETCHER, 2010, p. 238). However, accent does not seem to be a critical factor just for NNES. Research has shown that even NES process information more slowly when listening to unfamiliar native accents under adverse listening conditions (ADANK et al., 2009).

Another factor that might interfere with the ease of understanding an ATCO’s transmission is the density of the transmission. Research has shown that information density has an impact on communication performance (ESTIVAL; MOLESWORTH, 2016; TIEWTRAKUL; FLETCHER, 2010). Although it has been recommended that the pilot/air traffic control transmissions should not include more than three pieces of information (BARSHI, 1997, as cited in WU et al, 2019), transmissions sometimes include much more than that. For example, Wu et al. (2019) counted up to eight items in a single transmission. Estival and Molesworth’s (2016) study indicated that “the increased number of items per transmission makes it more difficult for pilots to read back accurately what they have been told” (ESTIVAL; MOLESWORTH, 2016, p. 225), especially for NNES pilots and less experienced NES pilots. Further, issues regarding sound quality and distortion are common features of communication between pilots and ATCOs. Early work by Spolsky et al. (1968) found correlations between proficiency and the ability to understand distorted messages – the so-called ‘white noise’ tests.

Finally, aviation professionals may come across a wide variety of unexpected or unfamiliar topics, and this might also have a negative impact on their ability to understand what they hear. Research has shown that topic familiarity impacts language listening comprehension, regardless of the listener’s proficiency level (SCHMIDT-RINEHART, 1994). This is another important factor that needs to be taken into consideration when identifying the listening comprehension construct for test design and task specification in the aeronautical English workplace.

2.3 The assessment of listening in the context of pilot/ATCOs' RT communications

The ICAO LPRs call for the assessment of plain English. Garcia (2015) suggested that tests should also include the assessment of phraseology. As one of the participants in her study explained:

I believe that the intention of ICAO is to assess language proficiency in the context of radio communications and that target language use domain is made up of two really important components, first being standard radiotelephony phraseology, and the second being plain English where phraseology doesn't suffice. So, I think that ICAO missed an opportunity to combine those two elements ... to separate the two is to artificially divide a single construct which is safe pilot/controller communications into constituent parts which don't necessarily want to be divided. Sometimes it is very difficult, for example, to see where phraseology ends and where plain English begins. The switch between the two happens so fluidly among proficient users. (GARCIA, 2015, p. 34)

Previous research has also shown that the ICAO rating scale descriptors for comprehension are deficient. Knoch (2009) argued that "comprehension could not be accurately measured in a scale designed to assess speaking performance" (KNOCH, 2009, p. 31). Similarly, Kim and Billington (2016) claimed that the comprehension descriptors in the ICAO rating scale seemed to be inadequate, as the ability to comprehend a range of accents is not mentioned in Level 4. Also, in Pfeiffer's (2009) study, the lowest inter-rater reliability result among the six criteria was the one for the assessment of comprehension. She suggested the reason for this might have been that it is difficult to assess some of the descriptors for comprehension. According to Pfeiffer (2009), "the rating scale designers have not properly thought about the pertinency of the features to be included into the scale and hence a scale user could be seduced not to take the scale too seriously" (PFEIFFER, 2009, p. 56). She adds that the comprehension descriptors were "not very enlightening" and that, in her judgment, "they are possibly the least well thought out in the entire rating scale" (PFEIFFER, 2009, p. 57).

As evidenced in the discussion above of literature relevant to the language assessment of pilots' and ATCOs' listening comprehension in the aeronautical English workplace, there is still a lack of clarity with regard to assessment needs. Thus, this study is guided by the following research questions:

- 1) Based on accounts of key stakeholders, what needs to be assessed in relation to the listening performed by pilots in RT communications with the ATCOs?
- 2) Based on accounts of key stakeholders, what are some of the implications of these assessment needs for the development of listening tasks?

Any definitive answers to these questions are exceptionally complex and beyond the scope of the present study. Our intention has been to conduct a preliminary needs analysis of the language that should

be targeted in order to help test developers to design tests and test tasks which may better represent the TLU domain (BACHMAN; PALMER, 2010; WAGNER, 2014). Douglas (2010) refers to needs analysis as “the process of deciding what and how to test” (DOUGLAS, 2010, p. 39). Although the present study focuses on “what to test”, “how to test” is considered in a larger study, of which the current study is a part.

3. Methods

3.1 Overall research design

This mixed-methods study applied a two-phase explanatory sequential research design (CRESWELL, 2015). Having both quantitative and qualitative data, as Creswell (2014) argued, “provides a stronger understanding of the research problem or questions” (CRESWELL, 2014, p. 264). In the first phase of the study, quantitative data (QUAN) were drawn, along with some qualitative data (qual), through an online questionnaire using Qualtrics, a survey software. In the second phase, in-depth qualitative data were collected through semi-structured interviews. The qualitative data were gathered in order to help understand the quantitative results and elaborate on them. Ethics approval was granted by the Carleton University Research Ethics Board.

3.2 Participants

Jacoby and McNamara (1999), Kim and Elder (2009), and Knoch (2009) argued that LSP test developers, when developing criteria for rating, or when conducting validation research, should consider indigenous assessment (defining the assessment criteria based on domain experts’ understanding of communicative effectiveness) because, as Jacoby and McNamara argued, “*language and communication* may mean one thing to linguistically oriented professionals and another to gatekeepers within a professional community” (JACOBY; MCNAMARA, 1999, p. 236). Douglas (2001) also argues that an analysis of the specific language use situation, together with indigenous assessment criteria, should be the base of LSP assessment criteria. Fox and Artemeva (2017) also emphasized the importance of incorporating indigenous criteria in the development of tests, as, according to them, it likely makes tasks more meaningful and useful.

Taking this into consideration, participants in Phase 1 of the present study were airplane and helicopter pilots. 156 pilots participated in the study (151 men and 5 women). All participants were NNES.

Seven were private pilots, 66 commercial pilots, 82 airline transport pilots and one was a military pilot. Among them, 145 participants were airplane pilots, four helicopter pilots and seven were both airplane and helicopter pilots. The majority of participants' first language was Brazilian Portuguese (124), but other first languages included Spanish (6), Azerbaijan (1), French (1), and Thai (1). The participants' experience with international flights (or outside their own country) can be seen in Table 1.

Table 1. Participants' experience with international flights

Experience with international flights	Number of participants
Never flown internationally	55
Up to 50 hours of international flights	17
51-200 hours of international flights	20
201-500 hours of international flights	12
501-1000 hours of international flights	8
1001-5000 hours of international flights	30
Over 5001 hours of international flights	14

139 participants reported having had their aeronautical English formally assessed according to the ICAO's LPRs. Table 2 shows the final levels they were awarded.

Table 2. Reported final levels awarded to participants on aeronautical English language proficiency tests

Final Level	Number of participants
3 – Pre-operational	10
4 – Operational	72
5 – Extended	42
6 – Expert	15
Not assessed	17
Total	156

In Phase 2 of the study there were six participants, four men and two women. All were NNES: four participants came from Brazil, one from France (but had been living in an Anglosphere country for many years), and one from Azerbaijan. They had extensive professional background in language assessment and aeronautical English. At the time of the study,

- four pilots had experience flying internationally and had also worked as ICAO language proficiency test raters. One of them was not only a pilot and a rater, but also held a PhD in Linguistics and was an active researcher in the field;

- one air traffic controller had been working as an ICAO test rater for more than 10 years;

- one expert in aeronautical English had worked for 15 years as a test developer, rater trainer and test administrator, held a PhD in Applied Linguistics and was also an active researcher in the field.

Participants were recruited through email and social media to participate in both phases of the study. Additionally, pilots who in the questionnaire reported having experience as ICAO test raters, were invited to participate in Phase 2.

3.3 Type of data and data collection methods

Alderson et al. (1995) suggested that a needs analysis might involve “questionnaires or interviews with language users, or the consultation of relevant literature or of experts on the type of communication involved” (ALDERSON et al., 1995, p. 22). In Phase 1, the quantitative data were gathered through an online anonymous questionnaire. The questionnaire comprised of 19 questions (nine demographic questions, five closed-ended opinion questions, and one open-ended question at the end, which invited participants to offer comments of their own regarding the topics discussed in the questionnaire). Participants took an average of approximately 18 minutes to complete the questionnaire (Attachment A). It is important to mention that, although participants were required to finish the questionnaire in order to have the data saved at the end, they were not forced to answer all questions. For this reason, for some questions there were fewer than the total of 156 responses.

In Phase 2, one-on-one semi-structured voice-only interviews were carried out through Skype to gather rich qualitative data. All the interviews were conducted in English, and included some demographic questions, a general question on listening assessment (“what do you think should be assessed in terms of pilots’ listening comprehension in the radiotelephony communications?”), followed by the same questions that were asked in the questionnaire. Participants were requested to explain their answers, by answering questions such as “why?”, “why not?”, or “how?”. The duration of the interviews varied from approximately 40 to 75 minutes, with an average of approximately 53 minutes. The interviews were audio-recorded. Participants were assigned alphabetical codes (A to F) for confidentiality. The recordings were transcribed for analysis.

3.4 Methods of data analysis

Descriptive statistical analyses were carried out to investigate the quantitative data. The results of Phase 1 were used to inform Phase 2. To analyze the qualitative data, descriptive coding was performed (SALDAÑA, 2013), which “summarizes in a word or short phrase – most often as a noun – the basic topic of a passage of qualitative data” (SALDAÑA, 2013, p. 70). The codes that reflected the topics of the questions were: speech rate, accents, unfamiliar words, cultural subtleties, information density,

phraseology, complex grammar, main ideas, recognition of words and numbers, and extract specific meaning. Some other generated codes represent topics that were not explicitly discussed, but that were recurrent during the interviews. They were: background noise, radio interference, frequency congestion, workload, expectations, and interactions. In order to validate the coding, both authors performed the coding and consensus was reached. The findings from the quantitative and qualitative phases were integrated and are reported in the next section.

4. Results and Discussion

From the analysis of both qualitative and quantitative data, it is possible to identify what stakeholders perceive as key listening comprehension needs that should be included in the construct definition of a proficiency test that aims to assess pilots' listening in the context of RT communications. Findings are summarized below.

4.1 The need to understand speech at different rates of delivery

79.49% of participants in the survey (n=124) stated that a faster ATCO speech rate interferes with the ease of understanding an ATCO's transmission and that assessment based on varying speeds of delivery is either important (n=90) or extremely important (n=43), with a mean response of 5.05. Table 3 summarizes the pilots' responses to the importance of understanding speech at different rates of speed.

Table 3. How important is it to assess the ability to understand speech at different speeds?

Ability	Extremely irrelevant (1)	Irrelevant (2)	A little irrelevant (3)	A little important (4)	Important (5)	Extremely important (6)	Total	Mean
To Understand speech at different speed	0 (0.00%)	4 (2.56%)	4 (2.56%)	15 (9.62%)	90 (57.69%)	43 (27.56%)	156	

In the survey, participants were asked to listen to four extracts of the same air traffic control message spoken at four different speeds (100, 140, 180, and 220 words per minute [wpm]) and to answer how frequently they hear ATCOs speaking at each of the given speeds. Table 4 shows their answers. By looking at the means, we can see that the speech rate participants reported hearing the most was 180 wpm, and the one they claimed to hear the least was 100 wpm, which is the speech rate recommended by ICAO to be used. One of the survey participants commented that "controllers usually speak too fast". These results confirm Prinzo and Britton's (1993) findings that ATCOs speak very fast.

Table 4. How frequently do pilots hear ATCOs speaking at different speech rates?

Speech rate	Never (1)	Rarely (2)	Sometimes (3)	Frequently (4)	Always (5)	Total	Mean
100 wpm	61 (39.61%)	63 (40.91%)	13 (8.44%)	13 (8.44%)	4 (2.60%)	154	1.94
140 wpm	18 (11.69%)	40 (25.97%)	46 (29.87%)	43 (27.92%)	7 (4.55%)	154	2.88
180 wpm	4 (2.61%)	10 (6.54%)	27 (17.65%)	101 (66.01%)	11 (7.19%)	153	3.69
220 wpm	19 (12.42%)	41 (26.80%)	35 (22.88%)	43 (28.10%)	15 (9.80%)	153	2.96

Five interview participants agreed that ATCOs tend to speak faster than the recommended 100 wpm. For participant A, 100 wpm is not realistic or natural. The other participant said that only American ATCOs speak faster. For him, “if you call the control faster and with very good English the controller answers very quickly, but if you call the control in a slow way, the controller answers in the same way”. Participants B, C and F stated that ATCOs speak very fast especially in busier airspaces, where their workload is heavy. This supports Prinzo and Britton’s (1993) finding that people talk faster when under stress or increased mental workload. ATCOs should be trained to speak at the recommended speech rate. Participant B stated that the speed of delivery might be one of the main challenges that pilots face when trying to understand an ATCO’s transmission. According to her, “this is very important because, even if you have an accent in your pronunciation, but your speed is okay, you sound more intelligible. But when it is the other way around, it is very hard to understand.” The results corroborate Buck’s (2001) claim that when the speech rate is greater, understanding is more difficult.

4.2 The need to understand different accents

64 participants in the survey (43.54%) responded that they have spent more than 60% of their time while conducting international flights listening to NNES, 47 (31.97%) answered that they have spent between 40% and 60% of their time listening to NNES, and 36 (24.49%) reported to have spent less than 40% of their time listening to NNES. Although interviewees claimed that how much a pilot listens to a NES ATCO or to a NNES ATCO will depend on factors such as where they usually fly or the company they fly for, most of them reported that pilots usually speak more to NNES than to NES. As participant A pointed out, because of the numbers of NES and NNES around the world, “it makes sense that most of the interactions would be with non-native English speakers when they are on international flights”. For participant B, “it is common sense that we have more second language speakers of English communicating

with each other than with first language English speakers”. On the other hand, participant C stated “there is a tendency that the pilots have more chance to speak with native speakers ... because the traffic is denser in countries that speak English”. According to participant E, pilots spend 50% of their time talking to NNES and 50% to NES.

The survey participants claimed that they hear unintelligible pronunciation with the frequency described in Table 5. The mean was 3.13, which indicates that the average was between sometimes and frequently. As Buck (2001) pointed out, “when listeners hear an unfamiliar accent ... this can cause problems and may disrupt the whole comprehension process. An unfamiliar accent can make comprehension almost impossible for the listener” (p.35). It is important to note that an unintelligible pronunciation does not necessarily mean that it is a non-native speaker’s pronunciation. As a matter of fact, interview participants A and F reported that understanding a NNES accent might be easier than a NES one.

Table 5. How frequently do pilots hear unintelligible pronunciation in RT communications?

How frequently pilots hear unintelligible pronunciation	Never (1)	Rarely (2)	Sometimes (3)	Frequently (4)	Always (5)	Total	Mean
	2 (1.28%)	30 (19.23%)	75 (48.08%)	44 (28.21%)	5 (3.21%)	156	3.13

72 participants in the survey (46.15%) reported that a difficult accent interferes with the ease of understanding an ATCO’s transmission. For five interview participants, it is common for pilots to struggle to understand ATCOs’ pronunciation, especially, as participant A said, when they are not familiar with the accent. Fortunately, as participant E pointed out, pilots eventually get used to the accents they hear. On the other hand, for participant D, pilots do not struggle so often because “if someone has a strong accent, in the context, it is easy to understand”.

Table 6 shows how much the pilots who participated in the survey considered it important to assess the ability to understand different accents. It is possible to see, by looking at the mean, that the average of the participants considered the assessment of the ability to understand different accents to be of some importance.

Table 6. How important is it to assess the ability to understand different accents?

Ability	Extremely irrelevant (1)	Irrelevant (2)	A little irrelevant (3)	A little important (4)	Important (5)	Extremely important (6)	Total	Mean

To understand different accents	0 (0.00%)	7 (4.49%)	5 (3.21%)	31 (19.87%)	86 (55.13%)	27 (17.31%)	156	4.78
---------------------------------	-----------	-----------	-----------	-------------	-------------	-------------	-----	------

All six interview participants stated that it is important to assess pilots' ability to understand different accents. Five of them argued that it is important for all pilots who are authorized to fly internationally (pilots who are Level 4 and above) to demonstrate this ability because they need to be prepared as they are going to listen to different accents. This corroborates Kim and Billington's (2016) claim that "the criteria should explicitly require pilots and controllers to demonstrate comprehension ability across a broad range of accents, and that this should be included from Level 4" (KIM; BILLINGTON, 2016, p. 154).

4.3 The need to understand the meaning of uncommon words and expressions

51 participants in the survey (32.69%) reported that unknown vocabulary interferes with the ease of understanding an ATCO's transmission. Table 7 shows that the average frequency of hearing words and expressions whose meaning pilots do not understand was between rare and occasional (sometimes).

Table 7. How frequently, in RT communications, do pilots hear words and expressions whose meaning they do not understand?

How frequently pilots hear words and expressions whose meaning they do not understand	Never (1)	Rarely (2)	Sometimes (3)	Frequently (4)	Always (5)	Total	Mean
	7 (4.49%)	60 (38.46%)	73 (46.79%)	15 (9.62%)	1 (0.64%)	156	2.63

Table 8 shows how important the pilots who participated in the survey considered the assessment of the ability to understand the meaning of uncommon words and expressions. The mean of 4.34 indicates that on average participants in the study considered the assessment of the ability to understand the meaning of uncommon words and expressions to be of some importance.

Table 8. How important is it to assess the ability to understand the meaning of uncommon words and expressions?

Ability	Extremely irrelevant (1)	Irrelevant (2)	A little irrelevant (3)	A little important (4)	Important (5)	Extremely important (6)	Total	Mean
To understand the meaning of uncommon words and expressions	1 (0.64%)	13 (8.33%)	17 (10.90%)	47 (30.13%)	57 (36.54%)	21 (13.46%)	156	4.34

Interview participants B, E and F considered it important to assess the understanding of unfamiliar vocabulary, whereas participants C and D did not. Participant A said that it is important only if the words and expressions are related to aviation, remarking, “there are things that we are not able to understand and that are not crucial”. Participant B said that it is important because “unusual situations may happen inside the plane or on the runways”. Participant F agreed that it is important because “even though everything is standardized, in the real life, real world, sometimes things don't happen as the manual says”. He further explained that “some airports have some restrictions that the pilots might not be familiar with and may not understand what they are not supposed to do”. One survey respondent also argued that “although the international standard phraseology suits the needs of air traffic on a normal day, even to the extent that multinational controller-pilot communication is facilitated by it, basic English vocabulary and plain English conversation capability prove to be an obstacle when a given situation goes abnormal”. Participants C and D, on the other hand, noted that, although pilots listen to words that they do not understand, it is not so important to assess this skill. For participant D, “the most difficult is to understand the name of the intersections. I can give you an example, we have an intersection here in Brazil called Utbur. And when American pilots says Utbur, it is completely different. And there are a lot of crazy names.” Although unfamiliar vocabulary is heard in radiotelephony, participants might have not considered it to be extremely important to be assessed because, as three interview participants pointed out, when a pilot listens to an unfamiliar word, that is not a major problem because they can ask for clarification. Participant D argued that “if you are not sure about an unfamiliar word, it is much better to ask for clarification because sometimes you may think one thing, but it is totally wrong”.

In relation to idioms and phrasal verbs, participant A explained that ATCOs use them in radiotelephony, although they should not. As a matter of fact, the greatest difficulty that one of the survey participants reported facing was “when the controller uses some local expressions.” Indeed, research has shown that ATCOs use idiomatic vocabulary when communicating with pilots (PRINZO et al., 2009). Prinzo et al. (2009) found that the use of slang has caused readback errors. In addition, Prinzo’s (2009) results suggested that “colloquialisms, slang; pleasantries, although well meaning, may cause problems for those who are less proficient in the English language” (PRINZO, 2009, p. 11). All interview participants reported that ATCOs use idiomatic vocabulary (three participants emphasized that it happens especially in the USA). As participant B pointed out

We have like a paradox in radiotelephony communications, because if everybody complied with the rules, the procedures and the standard phraseology, we wouldn't have pilots or controllers resorting to slang, to idioms or to more colloquial language over the radio, but sometimes this happens and I think more often than we would desire. ... So, they need to be able to understand but not to use them in radiotelephony (Participant B).

As idiomatic vocabulary is part of the real world, two participants said that the understanding of phrasal verbs and idioms should be assessed. However, they both think that it should only be assessed in tests that assess Level 6 pilots.

4.4 The need to recognize indirect meaning

Table 9 shows how pilots who participated in the survey viewed the importance of assessing the ability to recognize indirect meaning (e.g., politeness, sarcasm, professionalism, anger). 51 participants considered this ability irrelevant, whereas 105 considered it important (the mean of 3.9 suggests most participants thought it was not of great importance).

Table 9. How important is it to assess the ability to recognize indirect meaning?

Ability	Extremely irrelevant (1)	Irrelevant (2)	A little irrelevant (3)	A little important (4)	Important (5)	Extremely important (6)	Total	Mean
To recognize indirect meaning (e.g., politeness, sarcasm, professionalism, anger)	3 (1.92%)	24 (15.38%)	24 (15.38%)	53 (33.97%)	38 (24.36%)	14 (8.97%)	156	3.90

Four interview participants remarked that it is important to assess pilots' ability to understand cultural subtleties. As participant B argued, "we do not communicate without resorting to our cultural frames of reference". Participants E and F indicated it should be tested in Level 6 tests. Participant B argued that even Level 4 speakers should demonstrate this ability. Participant A was not sure if it was important, but she stated that it should be included at least in the pilots' training. Half of the interview participants agreed that the ability to recognize indirect meaning was important to assess, and half agreed it was not. Participant A asserted that it should be assessed only at Level 6. Participant B said interpreting indirect meaning was important because "they need to be able to recognize that in order to deal with the situations in a very professional way". Participants who stated that it was not important argued that it is unusual for ATCOs to be sarcastic or angry, and that they are usually polite and professional. Although

instances of unprofessional tone and attitude, such as impatience, sarcasm, impoliteness, and arrogance might not occur very frequently in RT communications (MONTEIRO, 2019), pilots and ATCOs who participated in Monteiro's survey considered such instances to be potential threats to safety.

4.5 The need to understand transmissions with more than three items

For 81 of the participants in the survey (51.92%), a high number of pieces of information in a transmission interferes with the ease of understanding a transmission. Most participants reported that according to their experience, a single air traffic control (ATC) transmission might include more than the ICAO recommended three pieces of information per transmission. Table 10 shows the answers that the survey participants gave to the question regarding the maximum number of elements they have heard in a single ATC's transmission. The mean was 2.74, which suggests that the average of the participants reported hearing a maximum of four to five items in an ATC transmission. One survey participant mentioned that "usually all controllers speak up to three pieces of information". Another survey participant explained that "depending on the phase of the flight, there might be a higher number of pieces of information in a message. When requesting a clearance before starting the engines, be prepared to write down a lot of information." A different participant commented that "as a pilot, I can assume that it is difficult to listen to a high number of items of information during a critical phase of flight. The most critical phases are during the initial climb-out and approach to land."

Table 10. What was the maximum number of pieces of information participants listened to in a single transmission from an ATCO?

Maximum number of pieces of information in each communication	Number of participants (percentage)
3	29 (19.08%)
4	43 (28.29%)
5	41 (26.97%)
6	24 (15.79%)
7	7 (4.61%)
8 or more	8 (5.26%)

All five interview participants who were pilots reported having heard more than three pieces of information in a single ATCO's transmission. Participant A argued there might be up to eight items. Participant D claimed there might be up to 20 elements in a single transmission, but there are typically around three or four. He explained:

During the clearance, taxi and reclearance, there are a lot of pieces of information. ... In big airports there are a lot of taxiways, a lot of runways. [...] It is very difficult to read back or take note because of the name of the airways, VORs, etc., so the most difficult situation is during those phases because there are a lot of different names, different numbers, different taxiways, and there is a lot of information (Participant D).

Participant C said there are usually around five, participant E said around three or four, sometimes more than that, and participant F reported around three to five.

Although ICAO recommends a maximum of three items per ATC transmission, the reality seems to be that pilots listen to ATCOs' transmissions with more than three items. It is important to note that, although a test construct would need to represent the features of the real-life language use (and this also applies to the use of idiomatic vocabulary and the need to adhere to phraseology), ATCOs should be encouraged to comply with the recommendations, as Wu et al. (2019) also suggested. As participant C explained, "if you give a very long instruction, [...] you could have a misunderstanding". According to Wu et al. (2019), "the number of items in a message, known as information density, has also been found to increase error rate in pilot communications" (WU et al., 2019, p. 3). Additionally, participant F expressed concern in relation to the importance of human memory in this context. Buck (2001) and Wu (1998) also highlighted the role that memory plays in the listening process. According to the interviewee, "because sometimes pilots don't have time to write down the information [...], they rely a lot on their memories and if you give them too much information, it is guaranteed that something is going to be missed". One of the survey participants also commented on this issue: "it is easy to understand. The problem is to remember all the information. Even in Portuguese." As Buck (2001) pointed out, a person whose processing is more controlled than automatic is a less efficient listener (Buck, 2001).

4.6 The need to understand both plain English and phraseology

The findings of the present study correspond with Garcia's (2015), as both studies suggest that not only plain English but also phraseology should be assessed. The vast majority of the survey participants considered that it is very important to assess comprehension of both plain English and phraseology. Table 11 shows that 96.76% of participants somewhat agree, agree or strongly agree that it is important to assess comprehension of plain English, and 96.16% somewhat agree, agree or strongly agree that it is important to assess understanding of phraseology.

Table 11. How important is it to assess comprehension of plain English and phraseology?

	It is important to assess comprehension of plain English		It is important to assess understanding of phraseology	
	Number of participants	Percentage	Number of participants	Percentage
I strongly disagree with it	1	0.65%	2	1.28%
I disagree with it	4	2.60%	3	1.92%
I somewhat disagree with it	0	0.00%	1	0.64%
I somewhat agree with it	17	11.04%	2	1.28%
I agree with it	65	42.21%	55	35.26%
I strongly agree with it	67	43.51%	93	59.62%
Total	156		154	

All six interview participants agreed that it is important to assess comprehension of phraseology. Participant D remarked that assessing comprehension of phraseology is of utmost importance. As interview participant A pointed out, communication in aviation is based on “the comprehension of phraseology and the ability to produce phraseology correctly”, so it should be assessed. Participant B claimed that “what pilots face in their daily routines while they are communicating with air traffic controllers includes both phraseology and plain English. So, if this is part of their reality, [...] it should be assessed because it is part of the construct.”.

4.7 The need to understand complex grammar

Only 26 participants in the survey (16.67%) responded that complex grammar interferes with the ease of understanding an ATCO’s transmission. Table 12 shows the responses given by the survey participants to the question of how frequently pilots hear sentences they cannot understand because of structure. The average respondent said they have this problem between rarely and sometimes.

Table 12. How frequently, in RT communications, do pilots hear sentences they cannot understand because of the way the words are put together (structure)?

	Never (1)	Rarely (2)	Sometimes (3)	Frequently (4)	Always (5)	Total	Mean
How frequently pilots hear sentences they cannot understand because of the way the words are put together (structure)	12 (7.69%)	68 (43.59%)	60 (38.46%)	13 (8.33%)	3 (1.92%)	156	2.53

For four interview participants, pilots do not need to understand complex grammar. One mentioned that it is not important at least for level 4, and another one said that it is not important for Levels 4 and 5,

but that it is for Level 6. Participant D argued that messages are very short and they can be summarized by cutting out words. Participant E claimed that pilots just need to understand the basic verb tenses like present simple, present continuous, past and future, but not future continuous or past perfect. Participant C, on the other hand, argued it is important. As he explained, “when the airport closes and the air traffic controller has to explain the reason why the airport is closed, a special situation, and to give some alternatives for the pilot to decide, I think the pilot should be required to understand complex grammar”, and participant B claimed that, although ATCOs are not supposed to use complex grammar, they sometimes do, so it might be important for pilots to understand it.

4.8 Other important needs

Results suggest other important needs. Table 13 shows how the pilots who participated in the survey viewed the importance of assessing the abilities to understand main ideas, to recognize words and numbers, and to extract specific meaning from both short and long transmissions. By looking at the means, we can see that the average of the participants in the survey considered the assessment of the ability to recognize words and numbers, to understand main ideas, and to extract specific meaning from both short and long transmissions, to be between important and extremely important.

Table 13. How important is it to assess the ability to understand main ideas, the ability to recognize words and numbers and the ability to extract specific meaning from both short and long transmissions?

Abilities	Extremely irrelevant (1)	Irrelevant (2)	A little irrelevant (3)	A little important (4)	Important (5)	Extremely important (6)	Total	Mean
To understand main ideas	1 (0.64%)	0 (0.00%)	1 (0.64%)	1 (0.64%)	59 (37.82%)	94 (60.26%)	156	5.56
To recognize words and numbers	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (1.28%)	38 (24.36%)	116 (74.36%)	156	5.73
To extract specific meaning from both short and long transmissions	0 (0.00%)	2 (1.29%)	1 (0.65%)	14 (9.03%)	84 (54.19%)	54 (34.84%)	155	5.21

Of all the abilities included in the survey items, recognizing numbers was considered the most important by the pilots in this study, as the mean was 5.73. Interview participant B claimed that pilots “have to understand lots of information involving numbers, such as runway numbers, altitudes, speed,

directions, and headings.” Participant F expressed concern in relation to the comprehension of numbers and what they refer to. As he explained:

The most important thing would be numbers and the ability to distinguish them from flight levels or headings or even air speed, as pilots usually get confused with headings and altitude. That is very dangerous. So, I think that the main thing would be precise numbers related to what they are meant to be, for example, the pilot is cleared for flight level 050, and then sometimes we can hear the candidate saying that they are supposed to fly on heading 050 (Participant F).

For participant D, it is very important to understand numbers and aviation words. Recognizing words also seems to be a very important skill. Participant C argued that understanding isolated words is crucial, especially the ones that are more critical to safety, but also it is very important to understand the main idea. Participant B explained that:

They need to recognize some isolated information like numbers, letters and some important words, but they also need to understand the meaning of certain transmissions which is broader. ... For example, if the controller is informing about an emergency situation or unusual situation happening in an airport, or if they are in an emergency situation reporting a problem and the controller asks questions to get more details about this problem (Participant B).

For participant A, it is important for the pilot to be able to extract important information that is presented. As she stated, “some words, some bits of information are not that important, and you don't need to retain that in your head, whereas others are absolutely crucial and you must be able to extract them”. She argued that it is not only important for pilots to extract important information from what is said, but also to understand the cues given by prosodic features of pronunciation, such as intonation and stress.

For interview participant B, pilots’ ability to understand ATCOs “also depends on technical issues, like background noise or radio interference, poor radio quality. Sometimes they face difficulties in understanding. It is not just because they lack the abilities but sometimes this can be exacerbated by other factors.” For 104 participants in the survey (66.67%), bad quality of the radio transmission interferes with the ease of understanding. This supports Prinzo’s (2009) findings that indicated that technical issues such as radio interference, noise and problems in the transmission contribute to communication issues. Participant B pointed out that the quality of the radio transmission may be a big challenge for pilots. However, low quality of the transmission seems to be less of a problem nowadays than it used to be in the past. Participant D stated that “sometimes the radio has a lot of interference”. However, he also said that “the quality of the radios I think nowadays is very clear. I don't have problems with this.” Participant E also argues that the quality of the transmission is “much better than it used to be” and that “you get used

to this distortion”. In relation to the background noise, as Estival and Molesworth (2009) explained, “the cockpit is a noisy environment and the presence of noise makes it more difficult for [NNES] listeners to understand speech” (ESTIVAL; MOLESWORTH, 2009, p. 24.4). Indeed, research has shown that the more noise, the harder it is to understand speech (SHIMITZU et al., 2002, as cited in ESTIVAL; MOLESWORTH, 2009).

Another technical issue that has been discussed in the interviews and that might interfere with the ease of understanding an ATCO’s transmission was radio frequency congestion. Although results from previous research have suggested that “radio frequency congestion does not affect pilots’ ability to communicate effectively, irrespective of their language background, pilot qualification or flying experience” (ESTIVAL; MOLESWORTH, 2016, p. 232), participants have argued that it may have an impact. For participant D, congestion can make communication very confusing as the transmission may become “very blocked”. Additionally, participant A pointed out that pilots need to have the ability to understand when the ATCO is addressing them. As she explained, “on the frequency, when there are a lot of other aircraft, and there is one communication that is directed to you, you want people to be able to recognize it, and respond. ... I think that is really important in testing comprehension. Being able to identify ... what is actually relevant to you and what is addressed to you.” Cushing (1994) gave real examples of this kind of miscommunication which he called an “uncertain addressee” problem of reference (when a pilot complied with a descent clearance intended for another aircraft).

Another recurrent theme was that pilots need to have the ability to deal well with their own expectations. Two survey participants commented on this, as well as five interview participants. One survey participant said that most of what the ATCO tells the pilot before starting the engines, pilots already expect to hear, “but sometimes controllers change the requested route, flight level or even an expected departure procedure.” Another respondent said that most miscommunications happen “due to some out of script requests or information being passed”. Participant A commented that “they expect something and, if the air traffic controller doesn't tell them what they think or they are expecting, that is going to be an issue.” Participant F argued that “the main difficulties are when the pilot is expecting something, and, out of nowhere, something changes, and they get instructions that are totally different from what they were expecting to hear”. According to participant E, “you should be very vigilant” when expecting instructions from the ATCO in order not to have a miscommunication. Thus, pilots need to be aware of their expectations in order to be able to identify when the situation does not happen according to what they were expecting. It seems to be important to assess this ability.

A theme that has emerged from three interviews was the need for pilots to be able to understand communications while having an increased workload or when under stress. Participant C argued that “the pilot has lots of things to do inside the cockpit so if the air traffic control gives him much information, he won't be able to process this information”. For participants B and E, coping with heavy workload is a challenge for pilots.

A recurrent theme in the interviews was the need for pilots to interact well. All interview participants emphasized the importance of pilots being able to use strategies in order to ensure accurate communication. As Wagner (2014) pointed out, assessing listening together with speaking is challenging, “but it is a necessary and advisable goal” (WAGNER, 2014, p. 6). Participant A discussed the importance of testing “the ability to respond appropriately”, and participant E talked about the importance of assessing “how the pilots receive and respond”. As participant E argued, “if you do not understand something, it is much better to ask to repeat the information”. Participant C pointed out that “they won't be able to know everything, so they need to develop skills that can lead them to a different way to understand what is going on if they don't understand a word, so they have to be skilled at asking for repetition and paraphrasing”. For participant F, not understanding a transmission is not a major problem as long as the pilot is able to interact well. As he explained, “if a pilot is not capable of understanding, I am pretty sure that he is capable, at least, to say that, ‘hey, I'm not a great speaker of this language. So please calm down. I need just the basic information’”. Field (2019) argued that a competent listener must be “capable of using strategies effectively to compensate for gaps in understanding” (FIELD, 2019, p. 14). Both NES and NNES pilots and ATCOs need to be trained in order to learn strategies that will help them to interact appropriately and effectively. The importance of pilot training was emphasized by four interview participants. A survey participant complained that “native speakers usually don't care if listeners are native or not”. As interview participant F pointed out, NNES “try to use simple words for conversations, and they do make an effort to understand you and to be understood. Sometimes native speakers just don't have that kind of feeling. They speak the way they are used to, and they don't care about how to express themselves, so they are not worried about it.” Participant A also claimed that “some air traffic controllers don't necessarily try to make it as clear as they should.” It is important to note that three interview participants mentioned the difficulty in understanding American ATCOs. They suggested that this might be the case because Americans ATCOs do not seem to follow the standardized phraseology. It seems to be important for pilots and ATCOs to have the ability to accommodate to less proficient users of the

language. As argued by O'Malley et al. (1989), learners can be taught to apply some strategic resources which can help them to become more effective listeners.

Conclusion

This study responded to two research questions that explored key stakeholders' perceptions of what needs to be assessed with regard to listening comprehension:

- 1) Based on accounts of key stakeholders, what needs to be assessed in relation to the listening performed by pilots in RT communications with the ATCOs?
- 2) Based on accounts of key stakeholders, what are some of the implications of these assessment needs for the development of listening tasks?

Investigating stakeholders' needs may help to inform test development and support the definition of clear, coherent and well-articulated test constructs and concomitant specifications, thereby improving the overall technical quality of the required proficiency testing in this specialized domain. The second research question examined the implications of these assessment needs in developing assessment tasks. Table 14 provides a summary of the needs that were identified as an outcome of the present study and the implications for the development of listening assessment tasks.

Table 14. Summary of pilots' listening comprehension needs and implications for testing

Needs	Implications for the development of listening assessment tasks
To understand speech at different speeds	The speech rate of the recordings should vary from 100 to 220 wpm. Most of them should be 180 wpm. At the same time, ATCOs should be trained to speak slower.
To understand different accents	Level 4 pilots should be tested on their ability to understand different accents, not only from Level 5 up. About half of the recordings should be made by NES and half by NNES. Only some accents (from both NES and NNES) should be more challenging.
To understand meaning of uncommon words and expressions	It should be tested, but it does not seem to be very important, as pilots have the benefit of asking for clarification when they do not understand a word. Idiomatic vocabulary should be assessed in tests designed to test Level 6.
To recognize indirect meaning	It seems to be important to be tested at least in tests designed to assess Level 6 pilots. It seems to be very important to be trained.
To understand transmissions with more than 3 items	Some recordings may have more pieces of information, but the average number should be around three or four.
To understand both plain English and phraseology	Understanding of not only plain English but also phraseology should be assessed.
To understand complex grammar	It does not seem to be very important to be tested. However, it should be tested in tests designed to assess Level 6 pilots.
To understand main ideas	It seems to be very important to test this ability.
To recognize words and numbers	It seems to be very important to test this ability.

To extract specific meaning from both short and long transmissions	It seems to be very important to test this ability.
To understand low quality transmissions	Low quality transmission does not seem to be a major problem anymore. Some interference may be included, but it should not be very serious.
To understand transmissions in spite of background noise	Aircraft noise should be included in the recordings as background noise.
To understand transmissions during increased workload	How to assess this ability should be further investigated.
To understand transmissions in congested radio frequency	How to assess this ability should be further investigated. It seems to be important for pilots to be able to recognize when they are being addressed.
To understand messages that differ from what was expected	It seems to be important to test this ability. Research should be conducted to find out how to do this.
To use strategies in order to understand	It seems to be very important to test this ability. It is also very important to be trained.

Some limitations of this study should be acknowledged. Firstly, the study only shows the perceptions of NNEs who are mainly native Portuguese speakers, most of whom were residing in Brazil at the time of the study. Additionally, 35% of the survey participants reported that they had not flown internationally. Their opinions were based on their general knowledge and on their experience with domestic flights. More importantly, the generated data need to be treated as an indirect reflection of the participants' understanding of the issues (what the participants claim may be different from the reality or even from what participants actually think).

With these caveats in mind, it is useful to report some other themes that emerged from the qualitative data that warrant further exploration, and are currently under investigation in a large-scale study: the types of recordings that should be included in the test (authentic audios versus scripted studio recordings); the topics that the recordings should cover; how much interference and background noise to include in the recordings; how many times test takers should be allowed to listen to the recordings; how cultural subtleties can be assessed; the importance of assessing comprehension separately and also during real interactions; how increased workload can affect pilots' ability to understand ATCOs' transmissions; how to include in the listening test the challenges posed by traffic congestion; the need to revise the ICAO rating scale and what categories to possibly include in a scale to assess comprehension specifically. Further, additional research should also be conducted in order to identify the most appropriate listening tasks, test methods and procedures.

As Read and Knoch (2009) pointed out, "the whole topic of oral communication in the aviation context is likely to engage the attention of language testers and other applied linguists for some time to come" (READ; KNOCH, 2009, p. 21.10). Although there is still a lot to be investigated in relation to the

assessment of the English used by pilots and ATCOs in the context of RT communications, it is hoped that this study will contribute to improving flight safety by helping language test developers to design tests and tasks that better represent the listening comprehension construct.

REFERENCES

- ADANK, P.; B. G. EVANS; J. STUART-SMITH; S. K. SCOTT. 2009. Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 35.2: 520-529. doi: 10.1037/a0013552
- ALDERSON, J. C.; C. CLAPHAM.; D. WALL. 1995. *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- BACHMAN, L.F. 2007. What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In FOX, J. et al. (Eds.), *Language Testing Reconsidered*, pp. 41-71. Ottawa, ON: University of Ottawa Press.
- BACHMAN, L.F.; A. S. PALMER. 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- _____. 2010. *Language assessment in practice*. Oxford: Oxford University Press.
- BARSHI, I. 1997. Effects of linguistic properties and message length on misunderstandings in aviation communication. Doctoral thesis. University of Colorado, Boulder, CO, USA.
- BRINDLEY, G. 1998. Assessing listening abilities. *Annual Review of Applied Linguistics*, 18.1: 171-191. doi: 10.1017/S0267190500003536
- BUCK, G. 2001. *Assessing Listening*. Cambridge: Cambridge University Press.
<http://dx.doi.org/10.1017/CBO9780511732959>
- CRESWELL, J. W. 2015. *A Concise Introduction to Mixed Methods Research*. Los Angeles, CA: Sage.
- _____. 2014. *Research design: Qualitative, quantitative and mixed method approaches*. Thousand Oaks, CA: Sage Publications.
- CUSHING, S. 1994. *Fatal words: Communication clashes and aircraft crashes*. University of Chicago Press.
- DOUGLAS, D. 2001. Language for specific purposes assessment criteria: Where do they come from? *Language Testing*, 18.2: 171-185. doi:10.1177/026553220101800204
- _____. 2004. Assessing the language of international civil aviation: Issues of validity and impact. *Proceedings from the International Professional Communication Conference*, IEEE Professional Communication Society, p. 248-252. Minneapolis: IEEE.
- _____. 2010. *Understanding language testing*. London: Hodder Education.
- _____. 2014. Nobody seems to speak English here today: Enhancing assessment and training in aviation English. *Iranian Journal of Language Teaching Research*, 2.2: 1- 12.
- EMERY, H. J. 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
- ESTIVAL, D.; B. R. C. MOLESWORTH. 2009. A Study of EL2 Pilots Radio Communication in the General Aviation Environment. *Australian Review of Applied Linguistics*, 32: 24.1-24.16.
- _____. 2016. Native English speakers and EL2 pilots: an experimental study. In ESTIVAL, D.; C. FARRIS.; B. MOLESWORTH. *Aviation English: A lingua franca for pilots and air traffic controllers*. New York: Routledge.

- GARCIA, A. C. M. 2015. What do ICAO language proficiency test developers and raters have to say about the ICAO language proficiency requirements 12 years after their publication? A qualitative study exploring experienced professionals' opinions. Master's dissertation. Lancaster University, Lancaster, UK.
- FIELD, J. 2019. Rethinking the Second Language Listening Test – From Theory to Practice. Equinox.
- FOX, J.; ARTEMEVA, N. 2017. From diagnosis toward academic support: Developing a disciplinary, esp-based writing task and rubric to identify the needs of entering undergraduate engineering students. *ESP Today*, 5.2: 148–171. <https://doi.org/10.18485/esptoday.2017.5.2.2>
- HUGHES, A. 2003. *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- INTERNATIONAL CIVIL AVIATION ORGANIZATION. 2010. *Manual on the Implementation of ICAO Language Proficiency Requirements*, 2nd Edition (Doc. 9835). Montreal, Canada: Author.
- JACOBY, S.; T. MCNAMARA. 1999. Locating competence. *English for Specific Purposes*, 18.3: 213–241.
- KIM, H.; R. BILLINGTON. 2016. Pronunciation and comprehension in English as a Lingua Franca Communication: Effect of L1 influence in international aviation communication. *Applied Linguistics*, first published online January 14, 2016. doi:10.1093/applin/amv075
- KIM, H.; C. ELDER. 2009. Understanding Aviation English as a lingua franca: Perceptions of Korean aviation personnel. *Australian Review of Applied Linguistics*, 32.3: 23.1-23.17.
- _____. 2015. Interrogating the construct of aviation English: Feedback from test takers in Korea. *Language Testing*, 32.2: 129–149.
- KNOCH, U. 2009. Collaborating with ESP stakeholders in rating scale validation: the case of the ICAO rating scale. *Spain Fellow Working Papers in Second or Foreign Language Assessment* 7: 21-46.
- LYNCH, T.; D. MENDELSON. 2010. Listening. In N. Schmitt (Ed.), *An introduction to applied linguistics* (2nd ed.), p. 180-196. London: Hodder Arnold.
- MESSICK, S. 1996. Validity and washback in language testing. *ETS Research Report Series*, i–18. doi:10.1002/j.2333-8504.1996.tb01695.x
- MONTEIRO, A. L. T. 2019. *Reconsidering the measurement of proficiency in pilot and air traffic controller radiotelephony communication: from construct definition to task design*. Doctoral thesis. Carleton University, Ottawa, Canada.
- O'MALLEY, J. M.; A. CHAMOT; L. KUPPER. 1989. Listening comprehension strategies in second language acquisition. *Applied Linguistics*, 10: 418–37.
- PFEIFFER, A. 2009. *Inter-rater reliability in an aviation speaking test*. Master's dissertation. Lancaster University, Lancaster, UK.
- PRINZO, O.V. 2009. *The ICAO English language proficiency rating scale applied to enroute voice communications of U.S. and foreign pilots*. (DOT/FAA/AM-09/10). Washington D.C., Federal Aviation Administration.
- PRINZO, O.V.; T. W. BRITTON. 1993. *ATC/Pilot Voice Communications: A Survey of the Literature*. (DOT/FAA/AM-93/20). Washington, DC: Federal Aviation Administration.
- PRINZO, O.V.; A. M. HENDRIX; R. HENDRIX. 2009. The outcome of ATC message length and complexity on En Route pilot readback performance. DOT/FAA/AM-09/2. Washington, DC: Federal Administration.
- READ, J.; U. KNOCH. 2009. Clearing the air: Applied linguistic perspectives on aviation communication. *Australian Review of Applied Linguistics*, 32.3: 21.1–21.11. doi: 10.2104/ara10921.
- RÉVÉSZ, A.; T. BRUNFAUT. 2013. Text Characteristics of Task Input and Difficulty in Second Language Listening Comprehension. *Studies in Second Language Acquisition*, 35: 31-65. doi:10.1017/S0272263112000678.
- SALDAÑA, J. 2013. *The coding manual for qualitative researchers qualitative data analysis: a methods sourcebook* (2nd ed.). London: Sage Publications.

- SCHMIDT-RINEHART, B. 1994. The effects of topic familiarity on second language listening comprehension. *Modern Language Journal*, 18: 179–89.
- SHIMITZU, T.; K. MAKISHIMA; M. YOSHIDA; H. YAMAGISHI. 2002. Effect of background noise on perception of English speech for Japanese listeners. *Auris Nasus Larynx*, 29: 121–125.
- SPOLSKY, B.; SIGURD, B.; SATO, M.; WALKER, E.; ARTERBURN, C. 1968. Preliminary studies in the development of techniques for testing overall second language proficiency. *Language Learning*, 18: 79–101. <https://doi.org/10.1111/j.1467-1770.1968.tb00224.x>
- TAYLOR, L.; A. GERANPAYEH. 2011. Assessing listening for academic purposes: Defining and operationalising the test construct. *Journal of English for Academic Purposes*, 10.2: 89–101.
- TIEWTRAKUL, T.; S. R. FLETCHER. 2010. The challenge of regional accents for aviation English language proficiency standards: A study of difficulties in understanding in air traffic control–pilot communications, *Ergonomics*, 53.2: 229-239. DOI: 10.1080/00140130903470033?
- WAGNER, E. 2014. Assessing listening. In A. KUNNAN (Ed.), *Companion to language assessment*, 1: 47-63. Oxford, UK: Wiley-Blackwell.
- WU, Q.; B. MOLESWORTH; D. ESTIVAL. 2019. An investigation into the factors that affect miscommunication between pilots and air traffic controllers in commercial aviation. *The International Journal of Aerospace Psychology*, p. 53-63.
- WU, Y. 1998. What do tests of listening comprehension test? – A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing*, 15: 21-44.





Angela Carolina de Moraes Garcia is in her first year of the PhD program in Applied Linguistics and Discourse Studies at Carleton University. She has been working as a language test developer and raters' trainer at the Brazilian Civil Aviation National Agency (ANAC) since January 2008. Angela holds a Master's degree in Language Testing from Lancaster University. Her research interests include English for Specific Purposes and Language Testing, especially aeronautical English testing, validity, reliability, construct definition and rating scale development. E-mail: angelagarcia@cmail.carleton.ca

Janna Fox, PhD, is Emeritus Professor of Applied Linguistics, in the School of Linguistics and Language Studies at Carleton University, Ottawa, Canada. Her current research interests include validation theory, language testing, classroom based assessment, and pre-service/in-service teacher development. Email: Janna.Fox@carleton.ca

Attachment A – Questionnaire

1	I identify my gender as: () man () woman () non-binary Other: _____ () do not wish to respond
2	Are you a native or a non-native speaker of English? I am a native speaker of English I am a non-native speaker of English
3	What is your first language? My first language is _____.
4	Have you ever had your aviation English formally assessed according to the ICAO (International Civil Aviation Organization) Language Proficiency Requirements? () yes () no
5	(If 'yes' in question 4) What final Level were you awarded? 1 – Pre-elementary 2 - Elementary 3 – Pre-operational 4 - Operational 5 - Extended 6 - Expert
6	What aircraft do you fly? a) Airplane. b) Helicopter. c) Both airplane and helicopter.
7	What license do you hold? a) I am a private pilot.

	b) I am a commercial pilot. c) I am an airline transport pilot. d) I am a military pilot. e) I am not a pilot (<i>If letter "f" was chosen, the survey would be ended</i>)					
8	How many hours have you flown internationally ? a) I have never flown internationally b) Up to 50 hours of international flights c) 51-200 hours of international flights d) 201-500 hours of international flights e) 501-1000 hours of international flights f) 1001-5000 hours of international flights g) Over 5001 hours of international flights					
9	To what extent do you agree or disagree with the following statement? <i>"It is important to assess understanding of phraseology."</i>					
	() I strongly disagree with it	() I disagree with it	() I somewhat disagree with it	() I somewhat agree with it	() I agree with it	() I strongly agree with it.
10	Knowing that <i>Plain English</i> is the English used in radiotelephony communications when phraseology does not suffice, to what extent do you agree or disagree with the following statement? <i>"It is important to assess comprehension of plain English."</i>					
	() I strongly disagree with it	() I disagree with it	() I somewhat disagree with it	() I somewhat agree with it	() I agree with it	() I strongly agree with it.
11	To what extent do you consider it to be important to assess the pilots' abilities in radiotelephony communications listed below?					
a	Understand main ideas					
	() Extremely irrelevant	() Irrelevant	() A little irrelevant	() A little important	() Important	() Extremely important
b	Recognize words and numbers					
	() Extremely irrelevant	() Irrelevant	() A little irrelevant	() A little important	() Important	() Extremely important
c	Understand different accents					
	() Extremely irrelevant	() Irrelevant	() A little irrelevant	() A little important	() Important	() Extremely important
d	Recognize indirect meaning (e.g., politeness, sarcasm, professionalism, anger, etc.)					
	() Extremely irrelevant	() Irrelevant	() A little irrelevant	() A little important	() Important	() Extremely important
e	Extract specific meaning from both short and long transmissions					
	() Extremely irrelevant	() Irrelevant	() A little irrelevant	() A little important	() Important	() Extremely important
f	Understand speech at different speed					
	() Extremely irrelevant	() Irrelevant	() A little irrelevant	() A little important	() Important	() Extremely important

	g	Understand the meaning of uncommon words and expressions				
		() Extremely irrelevant	() Irrelevant	() A little irrelevant	() A little important	() Important
12	How frequently, in radiotelephony, do you hear _____ ?					
	a	Unintelligible pronunciation				
		() Never	() Rarely	() Sometimes	() Frequently	() Always
	b	Words and expressions whose meaning you do not understand				
		() Never	() Rarely	() Sometimes	() Frequently	() Always
	c	Sentences you cannot understand because of the way the words are put together (structure)				
		() Never	() Rarely	() Sometimes	() Frequently	() Always
13	How frequently do you hear air traffic controllers speaking at the speed below? (Click on the “play” icon on the left)					
a		() Never	() Rarely	() Sometimes	() Frequently	() Always
	(1)					
b		() Never	() Rarely	() Sometimes	() Frequently	() Always
c		() Never	() Rarely	() Sometimes	() Frequently	() Always
d		() Never	() Rarely	() Sometimes	() Frequently	() Always
* (a) speech rate of approximately 100 words per minute (b) speech rate of approximately 140 words per minute (c) speech rate of approximately 180 words per minute, and (d) speech rate of approximately 220 words per minute						
14	Speaking of your experience with international flights, approximately how much time did you spend listening to non-native speakers of English?					
a) More than 60% of the time listening to NON-NATIVE SPEAKERS OF ENGLISH.						
b) Between 40% and 60% of the time listening to NON-NATIVE SPEAKERS OF ENGLISH.						
c) Less than 40% of the time listening to NON-NATIVE SPEAKERS OF ENGLISH.						
15	Who do you consider to be EASIER to understand in radiotelephony communications, a native speaker of English or a non-native speaker?					
a) It is usually easier to understand a NATIVE SPEAKER OF ENGLISH in radiotelephony communications.						
b) It is usually easier to understand a NON-NATIVE SPEAKER OF ENGLISH in radiotelephony communications.						
16	Speaking of your experience with international flights, what was the highest number of pieces of information you listened to in a single transmission from an air traffic controller? For example, the following communication has 4 pieces of information: “ANAC 123, [maintain runway heading], [climb to 7,000 feet], [squawk 4437]. [Contact Miami Departure on frequency 122.4]”					
a) Up to 3 pieces of information in each communication.						
b) 4 pieces of information in each communication.						
c) 5 pieces of information in each communication.						

	<p>d) 6 pieces of information in each communication.</p> <p>e) 7 pieces of information in each communication.</p> <p>f) 8 or more pieces of information in each communication.</p>
17	<p>What makes it difficult for you to understand an air traffic controller communication? Choose all that apply.</p> <p>A difficult accent A high number of pieces of information Unknown vocabulary Complex grammar When the controller speaks too fast Situational complication (for example, going through an abnormal situation) Lack of background knowledge Bad quality of the radio transmission Other: _____</p>
18	<p>Do you have experience working as an ICAO test rater? () yes () no</p>
19	<p>Please leave any comments regarding the topics discussed in this questionnaire.</p>