

Review research paper

## INVESTIGATING THE CONSTRUCT OF AERONAUTICAL ENGLISH LISTENING TESTING: A QUALITATIVE ANALYSIS OF THE ICAO RATING SCALE

Angela C. de M. Garcia

Carleton University, Canada

**Abstract.** *Since the publication of the International Civil Aviation Organization (ICAO)'s language proficiency requirements, a number of different tests have been developed and implemented around the world to assess pilots and air traffic controllers' proficiency in English. Meanwhile, researchers have questioned the clarity and appropriateness of the policy, and the reliability of tests (e.g., Alderson 2011; Douglas 2004; Emery 2014). ICAO has recently acknowledged that, over the years, multiple interpretations of the policy have led to practices that might undermine the meaningfulness of aviation English tests (ICAO 2022b). However, the ICAO Rating Scale remains as the instrument to be used in assessing pilots and air traffic controllers' aeronautical language proficiency. Thus, this article explores the construct of aeronautical English listening tests stated in the comprehension descriptors of the ICAO rating scale, as well as the elements of the other descriptors that may inform the definition of this construct. An in-depth content analysis of the rating scale was conducted by using the "interview technique", as described by O'Leary (2021). Results provide useful information for the development of listening tests in the aeronautical context. A better interpretation of the construct informed by the policy can help to reduce the differences among test implementations around the world and further contribute to more standardized and meaningful testing practices.*

**Key words:** *language testing, ICAO rating scale, listening comprehension, construct definition*

### 1. INTRODUCTION

In 1998, after a tragic accident that led to 349 fatalities in 1996, India submitted a working paper asking the International Civil Aviation Organization (ICAO) Assembly to consider the lack of language proficiency of pilots and air traffic controllers (ATCOs) with a high degree of priority (Popa 2019). In 2003, the ICAO Council adopted the Amendment 164 to the Annex 1 (entitled *Personnel Licensing*) to the Convention on International Civil Aviation, requiring pilots and ATCOs to "demonstrate the ability to

---

Submitted February 20<sup>th</sup>, 2023, accepted for publication March 29<sup>th</sup>, 2023

Corresponding author: Angela C. de M. Garcia, Carleton University, Canada

E-mail: [angelagarcia@cmail.carleton.ca](mailto:angelagarcia@cmail.carleton.ca)

*speak and understand* [emphasis added] the language<sup>1</sup> used for radiotelephony communications” (ICAO 2022a: paragraph 1.2.9.1). For pilots and ATCOs to demonstrate their speaking and listening abilities, they need to be tested, unless they are native speakers of English. Appendix 1 to Annex 1 includes a set of *holistic descriptors*, which describe five abilities which pilots and ATCOs should demonstrate, and, Attachment A includes the *ICAO Language Proficiency Rating Scale*, which details six analytic criteria (pronunciation, structure, vocabulary, fluency, comprehension, and interactions) for the six proficiency levels mentioned in paragraph 1.2.9.1 (level 1, *Pre-elementary*, to level 6, *Expert*).<sup>2</sup> For pilots and ATCOs to meet the ICAO LPRs, they must demonstrate compliance with the holistic descriptors and the ICAO Operational Level (Level 4), detailed in the rating scale. In other words, to be allowed to fly internationally or to control international flights, pilots and ATCOs should be awarded at least level 4 in all six criteria of the rating scale. To help civil aviation authorities and testing service providers to develop tests to assess pilots and ATCO’s aeronautical language proficiency, ICAO published, in 2004, the first edition of the DOC 9835 – *Manual on the Implementation of ICAO Language Proficiency Requirements* (ICAO 2004). In that time, Douglas argued that in the aviation English context it would be very important to have a clear picture of the language which is being assessed, as well as a “clear, complete and unambiguous definition of the construct<sup>3</sup> to be measured in relation to the purposes for which the measurement is being made” (Douglas 2004: 250). After the publication of the ICAO language proficiency requirements (LPRs), the so-called *aviation English tests* (which I call *aeronautical English tests* [see Tosqui-Lucks and Silva 2020]) started to be developed around the globe by different organizations. In 2010, Alderson (2010) conducted a survey on aviation English tests for pilots and ATCOs and concluded that “little or no confidence can be had in the meaningfulness, reliability and validity of several of the aviation language tests currently available” (p. 63). He then interrogated if the ICAO scales are explicit, relevant, and adequate.

In 2016, during the 39<sup>th</sup> session of the ICAO Assembly, Brazil (2016) presented a working paper inviting the Assembly to review the ICAO LPRs, arguing that the construct underpinning the policy is unclear and under-represented. The response to this working paper was that “it [the working paper] did not present sufficient evidence that existing language proficiency requirements posed a safety threat”, and that “the need to revise the language proficiency requirements could be considered once additional implementation data was [*sic*] collected through the different initiatives of ICAO” (ICAO 2016, para. 35.155). Unfortunately, over the years, as ICAO itself remarked, multiple

---

<sup>1</sup> A note to the ICAO Annex 1 paragraph 1.2.9 explains that the language to which paragraph 1.2.9 refers may be English or the language “normally used by the station on the ground” (ICAO 2022a, p. 1-17). In practice, the language that is commonly evaluated is English, because when pilots and ATCOs use the language used by the station on the ground, they usually use their native language. As another note to paragraph 1.2.9 informs, pilots and ATCOs who demonstrate expert language proficiency (e.g., native speakers) do not need to be formally evaluated.

<sup>2</sup> Due to a limitation on the number of words of this article, the holistic descriptors and the rating scale were not included in the appendix, but Annex 1 is available at <https://elibrary.icao.int/home>.

<sup>3</sup> A construct can be understood as “the theoretical entity that the test developers and test users intend the test to measure, the quality or qualities of the test takers we wish to make inferences about” (Douglas 2010: 33).

interpretations of the policy have led to practices that might have undermined the meaningfulness of aviation English tests (ICAO 2022b). The lack of clarity, appropriateness, and fairness of the ICAO policy, as mentioned by many (e.g., Douglas 2004, 2014; Emery 2014; Kim and Billington 2016; Kim and Elder 2015; Knoch 2014; Read and Knoch 2009), most likely contributed to such consequences. In spite of the criticism towards the policy, the LPRs have not changed and are still what Contracting States need to comply with, unless ICAO is notified of any differences between national regulations and the ICAO requirements. Thus, as Emery (2007) points out, “every testing programme and test instrument developed to measure the language proficiency of aviation operations personnel will employ the Rating Scale and Holistic Descriptors in each of the 190 ICAO member states” (p. 1). Therefore, test developers need to have a clear and deep understanding of the ICAO policy. Knoch and Macqueen (2020) argue that policy analysis “is a crucial aspect of work that needs to be completed before an assessment is developed or adopted” (Knoch and Macqueen 2020: 87)

Many aspects related to the policy could be explored. As Wodak (2006) points out, “there are obviously many relevant research issues and a variety of genres and public spaces where a precise linguistic analysis of oral, visual, or written texts will provide differentiated knowledge on aspects of language politics/policies” (Wodak 2006: 170). The focus of the present study is to investigate one specific aspect of the ICAO policy: the construct of the listening in isolation test represented in the ICAO rating scale. Criterion 3 of the ICAO test design guidelines<sup>4</sup>, which were developed by the International Civil Aviation English Association (ICAEA) in partnership with ICAO, recommends that “test instruments need to contain tasks dedicated to assessing listening comprehension separate from tasks designed to assess speaking performance” (ICAEA n.d.). It is necessary to assess listening in isolation in order to minimize a major threat to the validity of the interpretations and uses of tests scores: *construct irrelevant variance*, “that is, the test is too broad and contains excess reliable variance associated with other distinct constructs as well as method variance making items or tasks easier or harder for some respondents in a manner irrelevant to the interpreted construct” (Messick 1989: 14). ICAEA points out that “assessing comprehension at the same time as speaking compromises the validity of the result for comprehension”, and that “test developers need to be mindful of ensuring interference of ability in other skills do not unfairly influence the assessment results” (ICAEA n.d.). The ICAO test design guidelines adds that “this means test-takers are required to listen to prescribed recordings and then complete follow up comprehension tasks. Such tasks could be on paper, require test-takers to summarise information or answer prescribed written questions asked orally or provided on a test paper/computer screen” (ICAEA n.d.). It is important to point out that, although it is necessary to assess listening in isolation, the assessment of interactive listening is also essential in this context, as most of the listening performed by pilots and ATCOs happen as part of an interactive conversation between them. Field (2020) argues that the listening processes employed in conversations might be different from those employed when listening to a recording, and also more cognitively demanding. Lam (2021) also

---

<sup>4</sup> At the time this article was written, this author was participating in the meetings of the ICAO Exploratory Group - Language Proficiency Requirements (EG-LPRs/03), which was created to revise the ICAO test design guidelines in order to have them published by ICAO as a handbook.

emphasizes that interactive listening “needs to be assessed outside the boundaries of (receptive) listening tests” (Lam 2021: 20). However, the focus of the present study is on the construct of a test that aims to assess listening in isolation. Hence, the research questions that this study addresses are the following:

**Research Question 1 (RQ1):** *What is the listening in isolation construct represented in the descriptors for comprehension detailed in the ICAO Language Proficiency Rating Scale?*

**Research Question 2 (RQ2):** *How may the descriptors for the other criteria help to define the construct of a test to assess listening in isolation?*

## 2. LITERATURE REVIEW<sup>5</sup>

The ICAO rating scale was developed by a committee appointed by ICAO known as the Proficiency Requirements in Common English Study Group (PRICESG). ICAO (2010) describes that “this study group brought together, from Contracting States and international organizations, operational and linguistic experts with backgrounds in aviation (pilots, air traffic controllers and civil aviation authority representatives), aviation English training and applied linguistics” (para. 1.4.2). Estival, Farris, and Molesworth (2016) observe that this study group did not include language testing expertise. McNamara, Knoch, and Jason (2019) adds that this group “was dominated by representatives from English-speaking nations” (McNamara, Knoch, and Jason 2019: 19).

The ICAO policy has been criticized from a number of different perspectives. One of the main criticisms is related to the fact that native speakers do not need to be formally evaluated. Many authors (e.g., Borowska 2017; Douglas 2014; Estival, Farris, and Molesworth 2016; Kim 2013; McNamara, Knoch, and Jason 2019; Monteiro 2019; Read and Knoch 2009; Trippe 2018) argue that native speakers should have their ability to communicate evaluated. This test should assess their ability to accommodate to non-native speakers when necessary by, for instance, using simpler vocabulary or by speaking at a slower rate. I believe the main reason why Brazil’s paper to ICAO was not accepted was because the working paper highlighted that it was necessary to assess native speakers of English. Having the paper accepted would mean that some countries would have to spend money and time to assess and train their pilots and ATCOs. Asking pilots to pay a two-dollar fee to receive a new license with an “English Proficient” endorsement, as the United States were doing (Alderson 2011) would not be enough, and this is probably why the United States voted against the Brazilian request. As it has been argued, the politics involved in this context have a huge impact on the decisions that are made, and their agenda is often hidden (see Aragão 2018; Alderson 2011).

Another recurrent criticism to the ICAO policy lies in the fact that the guidelines advice that the focus of the language assessment should be on plain English proficiency, without taking into consideration the incorrect use of phraseology or the lack of technical knowledge of operations (ICAO 2010). DOC 9835 does mention that it is important to adhere

---

<sup>5</sup> This literature review focuses on studies that discuss the construct of aeronautical English tests for ATCOs and/or licensed civilian pilots. For a through discussion on the appropriacy of the ICAO scale for assessing *ab initio* pilots, see Treadaway (2022), and for a detailed comparison between the assessment of civilian and military pilots in the Brazilian context, see this issue’s article by Silva (2023).

to ICAO standardized phraseology. It also acknowledges that 70% of the radiotelephony speech acts do not comply with it. However, although a somewhat contradicting note to Annex 1's Appendix 1 says that "the language proficiency requirements are applicable to the use of both phraseology and plain language" (ICAO 2022a), the policy does not require phraseology to be assessed. DOC 9835 explains that "it is acceptable that a test contains a scripted task in which phraseology is included in a prompt, but the test should not be designed to assess phraseology" (ICAO 2010: 6.3.2.9). Korean pilots and ATCOs who participated in Kim's (2013) study believe that the non-observance of radiotelephony conventions plays a more important role on safety than proficiency in plain English. Indeed, underutilization of phraseology has shown to increase problems in communication (e.g., Howard 2008). Additionally, DOC 9835 says that "the test should not be designed to evaluate technical knowledge of operations" (ICAO 2010: 6.3.2.10). This is understandable, as the tests should not ask questions 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" (ICAO 2010: 6.3.2.10). However, from analysing indigenous assessment criteria (the criteria adopted by domain language users to assess the effectiveness of communication [Jacoby and McNamara 1999]), Aragão (2018) argues that ATCOs consider non-linguistic elements, such as psychological aspects and operational knowledge, to be significant contributors to effective communication. McNamara, Knoch, and Jason (2021) also point out that "experienced pilots and air traffic controllers know that technical knowledge is an inextricable part of language use" (McNamara, Knoch, and Jason 2021: 17). Research findings (e.g., Kim 2013; Knoch 2009, 2014; Aragão 2018; McNamara, Knoch, and Jason 2021) support Douglas's (2001) notion of specific purpose language ability in Language for Specific Purposes (LSP) testing, which he defined as "a construct that results from the interaction between specific purpose background knowledge and language knowledge" (Douglas 2001: 50). Unfortunately, it seems that, as Knoch, Deygers, and Khamboonruang (2021) point out, when the ICAO rating scale was developed, indigenous assessment criteria were not taken much into consideration.

Not much is known about the development and validation of the ICAO rating scale (Kim and Elder 2009; Knoch 2009) and few studies have been conducted to investigate the construct represented in the ICAO rating scale. Knoch (2009) conducted a validation study to investigate how test developers, administrators, and raters viewed the rating scale. Although participants responded that they were generally satisfied with the rating scale descriptors, their responses to open-ended questions indicated a range of problems that stakeholders identified in the ICAO scale. For comprehension, the most cited problem was that "comprehension could not be accurately measured in a scale designed to assess speaking performance" (Knoch 2009: 31). This confirms Pfeiffer's (2009) findings. Pfeiffer (2009) investigated inter-rater reliability in a German speaking test, and found it to be low, the lowest being for comprehension. She points out that "the ICAO descriptors are often incomplete and therefore need amendment, however with comprehension the user could easily have the impression that 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: 56). She adds that "the wording of the level descriptors for comprehension is not very enlightening. According to my judgement, they are possibly the least well thought out in the entire rating scale" (Pfeiffer 2009: 57).

Similarly to Knoch (2009), Garcia (2015) interviewed very experienced ICAO LPRs test developers and raters to investigate their perceptions on the ICAO policy in general and, more specifically, on the ICAO rating scale. She reported a number of recurring themes criticized

by participants, such as the presence of contradictions within the policy, the lack of fit between the policy and the target language use (TLU) domain, the need to assess pilots and ATCOs ability to communicate effectively (not only proficiency in plain English), and the importance of following standard phraseology. Participants were asked to discuss the strengths and weaknesses of the ICAO rating scale descriptors for each criterion. One participant complained about the fact that comprehension is only one category out of six in the rating scale. He argued that “it makes us think that comprehension is less than 20% of the overall ability to communicate on the radio. It is not, it is 50%, at least”. Indeed, according to Feyten (1991), as cited by Buck (2001), it seems that people spend about 45% of the total time of communication using listening skills. The ICAO requirement itself emphasizes this importance by saying, as mentioned, that pilots and ATCOs must demonstrate their ability to do *two* things: *speak* and *understand* the language used for radiotelephony communications. The same participant in Garcia (2015)’s study advocated that comprehension “is an extremely important, if not more important, part of the overall proficiency construct in this case (Garcia 2015: 38). For him, comprehension should have its own rating scale. This belief is upheld by Knoch (2009) when she points out, as mentioned, that using a scale that was designed to assess speaking in order to assess comprehension is problematic. However, differently from Knoch’s (2009) results, that suggest that comprehension of cultural subtleties seems irrelevant, participants in Garcia’s (2015) study pointed out that the main strength of the comprehension descriptors was this reference to comprehension of cultural subtleties. One of the participants even argued that this reference should have been included in other levels of proficiency, not only in level 6. Monteiro’s (2019) findings uphold Garcia’s (2015), in opposition to Knoch (2009). Monteiro (2019) investigated the proficiency construct of intercultural radiotelephony communications in aviation and specified the communicative demands of pilots and ATCOs within a construct framework. She points out that pilots and ATCOs perceive that intercultural factors can impact the safety of flights. The results of her study indicate that the ICAO policy does not include important components of the construct. Monteiro (2022) highlights that

Effective RT communications require competencies not addressed in prevailing models of communicative competence. They do require specific purpose language ability and background knowledge (AE), the need to speak English as a lingua franca and to adjust to the communicative needs at hand (ELF), to accommodate and negotiate sociocultural differences (ICA), and to solve misunderstandings between members of different cultures, while at the same time sharing responsibility for successful communication (IC). And most importantly, this applies to both first language (L1) speakers of English, and those who speak English as a second (L2) or additional language. (Monteiro 2022: 239)

As seen in this brief literature review, there has been a quite prolonged debate on the appropriacy of the ICAO policy. As Knoch, Deygers, and Khamboonruang (2021) point out, the rating scale “includes features of construct irrelevance and construct under-representation of the TLU domain” (Knoch, Deygers, and Khamboonruang 2021: 618). We can then conclude that “the ICAO policy has not met its intended goals, and these seem unlikely to be met in the future unless the policy and its underlying construct are modified” (Kim 2013: 108).

### 3. METHODS

#### 3.1. What Document was Selected?

Merriam (1988) argues that “documents of all types can help the researcher uncover meaning, develop understanding, and discover insights relevant to the research problem” (Merriam 1988: 118). Policy documents carry valuable information that can offer insights as to how such policies should be implemented.

The source of data for this study was the ICAO Rating Scale, included in the Annex 1. When planning this project, the intention was to analyse not only the rating scale, but the whole ICAO policy, including the Holistic Descriptors and the rest of ICAO Annex 1 (ICAO 2022a), the second edition of the ICAO DOC 9835 (ICAO 2010), the ICAO Circular 318 (ICAO 2009), *Language Testing Criteria for Global Harmonization* (ICAO 2009), and the ICAO test design guidelines (ICAEA n.d.). However, the ICAO Annex 1 is the only document that has the Standard and Recommended Practices (SARPs), or, in other words, the requirements that Contracting States need to comply with, a decision was made to only analyse this fundamental document. The manual, the circular, and the test design guidelines contain guidelines, which are very relevant but unfortunately not mandatory for Contracting States to follow. However, when discussing the results, I occasionally mention DOC 9835. However, the most relevant piece of the policy to define the construct is the rating scale, which every test instrument developed to assess pilots’ and ATCO’s ability to speak and understand the language used in radiotelephony communications must employ.

#### 3.2. What Method was Applied?

The method applied in this qualitative study was document content analysis. Bowen (2009) defines document analysis as “a systematic procedure for reviewing or evaluating documents” (p. 27), which “entails finding, selecting, appraising (making sense of), and synthesising data contained in documents” (p. 28). This procedure is recommended by Cardno (2018) to analyse the organization and content of educational policy documents. As Cardno (2018) explains, “as a research tool, policy document analysis is a method for investigating the nature of a policy document in order to look at both what lies behind it and within it” (Cardno 2018: 625). In spite of the traditional quantitative nature of content analysis (Merriam, 1998), the focus of the present analysis was not on the quantitative aspects of the policy content, such as frequencies, but on its qualitative nature, such as “the presence or absence of certain content characteristic” (George 2009: 145).

Bowen (2009) explains that document analysis is an iterative process which includes superficial skimming, careful reading, and interpretation. This document analysis was done through a careful reading of the document, and involved focused re-reading and review of the documentary data. Thus, an “interview technique”, as described by O’Leary (2021), was conducted to interrogate the rating scale as if the text was being interviewed. As O’Leary (2021) points out:

In ‘interviewing’ your documents, you are, in a sense, treating each document as a respondent who can provide you with information relevant to your enquiry. The questions you ask will be dependent on the nature of your enquiry and on the document type. As with an interview, you will need to determine what it is you want to know, and whether your document can provide you with the answers. You then need to ‘ask’ each question and highlight the passages in the document that provide the answer. (O’Leary 2021: 200)

Having RQ1 in mind, the following questions were asked towards the document:

- Q1) *What are the common elements in the comprehension descriptor levels?*
- Q2) *What elements are not recurrent among the comprehension descriptor levels?*
- Q3) *How do the comprehension descriptors differentiate the different levels?*
- Q4) *Do the comprehension descriptors make a difference between a test to assess interactive listening and a test to assess listening in isolation?*

While highlighting the common elements in the comprehension descriptors and comparing the differences between the levels, some other questions arose:

- Q5) What may be considered a *common* topic?
- Q6) What may be considered a *concrete* topic?
- Q7) What may be considered a *work-related* topic?
- Q8) What *accents or varieties* may be considered *sufficiently intelligible for an international community of users*?
- Q9) What can be considered a *linguistic* complication?
- Q10) What can be considered a *situational* complication?
- Q11) What is an *unexpected turn of events*?
- Q12) What is meant by *nearly all contexts*?

#### 4. RESULTS AND DISCUSSION

##### 4.1. What is the Listening in Isolation Construct Explicitly Stated in the ICAO Language Proficiency Rating Scale?

The common element in the comprehension descriptors from level 3 to level 5 is that they talk about comprehension in common, concrete, and work-related topics. Another common element is that the speaker might be confronted with a linguistic or situational complication or an unexpected turn of events. Table 1 shows the differences between comprehension in these three levels. We can see that comprehension in common, concrete, and work-related topics seems to be easier, as comprehension of test takers who will not even pass the test (level 3) is often accurate and comprehension of level 4 test takers is mostly accurate (although it allows a few misunderstandings). When the speaker is confronted with a linguistic or situational complication or an unexpected turn of events, the descriptors seem to be harsher on test takers. For test takers to pass the test (get a level 4), although it might take them some time or the use of clarification strategies, they seem to need to understand the communication, as only level 3 test takers may fail to understand.

Table 1 Comprehension differences among levels 3 to 5

Level	Comprehension in common, concrete, and work-related topics	Comprehension when the speaker is confronted with a linguistic or situational complication or an unexpected turn of events
Extended 5	Accurate	Mostly accurate
Operational 4	Mostly accurate	May be slower or require clarification strategies.
Pre-operational 3	Often accurate	May fail to understand



We can see that the difference between comprehension in levels 3, 4, and 5 is how accurate the comprehension on common, concrete, and work-related topics of the test taker is (level 5, accurate, level 4, mostly accurate, and level 3, often accurate). For levels 3 and 4, the scale includes an observation that for comprehension in common, concrete and work-related topics to be either mostly accurate (level 4) or often accurate (level 3), the accent or variety used should be sufficiently intelligible for an international community of users. The absence of this observation in the level 5 descriptors may imply that, at this level, comprehension must be accurate even when the accent or variety used is *not* sufficiently intelligible for an international community of users.

As mentioned, during the analysis, questions 5 to 12 above were risen. Some of these questions had been raised in previous work. Douglas (2004), for example, asked two important questions: “Does the phrase ‘common, concrete, and work-related topics’ refer to three topic categories or one?” and “how is the intelligibility of the various English dialects and accents in use internationally to be determined?” (Douglas 2004: 250). Knoch (2009) suggested that this reference to intelligibility for an international community of users could be deleted from the rating scale because it would be problematic to select speakers that would satisfy this requirement. Garcia (2015) recommends that this issue should be further researched.

The descriptors for level 6, on the other hand, instead of describing comprehension in the two mentioned scenarios, talk about comprehension *in nearly all contexts*. What is meant by “nearly all contexts”? Is it more than comprehension in both mentioned scenarios? Apparently, yes, as the word *all* implies. But why *nearly* all contexts? What contexts are not included?

DOC 9835 explains that

Work-related context can accommodate different interpretations. A narrow interpretation would aim to closely replicate radiotelephony communications, including the extent of plain language needed in unusual, unexpected or emergency situations. A broad interpretation of the holistic descriptors and Rating Scale would aim to elicit plain language on various topics that are related to radiotelephony communications or aviation operations, without replicating radiotelephony communications specifically. ... Both interpretations are valid. (ICAO 2010: 6.2.8.9)

Not having the “right” interpretation to be followed determined by ICAO gives too much flexibility for test developers to define the construct they want to measure. The fact that the DOC 9835 allows for different interpretations contributes to the significant differences in test design which have led to uncertainties in relation to what tests measures, what results mean, and overall quality of tests worldwide (ICAO 2022b). Unfortunately, although this issue of having a possibility for either a broad or a narrow interpretation was frequently brought to discussion in the ICAO EG-LPRs/03 meetings mentioned in Note 3, this will remain an unresolved problem for some time to come.

Furthermore, some elements in the comprehension descriptors are only mentioned in one of the levels. These are: the reference to the ability “to comprehend a range of speech varieties ... or registers”, which are only included in level 5, and the ability to comprehend linguistic and cultural subtleties, which are only mentioned in level 6. The issues in the ICAO rating scale reported here, such as the terminology problems, inconsistencies, and lack of definition of concepts do not only happen in this scale.

Alderson et al. (2004) has also listed similar problems regarding the Common European Framework. These problems make it difficult to fully understand the construct to be measured, but, within our limitations, we can conclude that a test that aims to assess pilots and ATCOs listening comprehension in isolation must include the abilities listed in Table 2. The comprehension descriptors seem to apply to the assessment of both listening in isolation and interactive listening.

Table 2 Listening construct represented in the comprehension descriptors

Item	Construct	Applicability
1	Comprehension on common, concrete, and work-related topics	Levels 3, 4, 5
2	Comprehension when the speaker is confronted with a linguistic or situational complication or an unexpected turn of events.	Levels 3, 4, 5
2.1	In assessing 1, include accents or varieties sufficiently intelligible for an international community of users.	Levels 3, 4
3	Ability to ask for clarification when comprehension fails	Level 4
4	Comprehension of a range of speech varieties (dialect and/or accent) or registers.	Level 5
5	Comprehension in nearly all contexts	Level 6
6	Comprehension of linguistic and cultural subtleties	Level 6

Test developers need to have clear definitions for the concepts that are not defined. An analysis of the ICAO documents that include guidelines for the implementation of the ICAO LPRs, such as the ones mentioned in 3.1, may help to further understand ICAO's intentions. The second edition of DOC 9835 includes an explanation of the rating scale descriptors from level 3 to 6 which may be useful in this process.

#### 4.2. How May the Descriptors for the Criteria other than Comprehension Help to Define the Construct of a Test to Assess Listening in Isolation?

Questions 5 to 12 were addressed towards the rating scale in order to investigate how the descriptors for the other criteria could help to define the construct of a test to assess listening in isolation. The rating scale does not provide a clear answer to questions 5, 6, and 7, as it does not explain what can be considered a common, concrete, and work-related topic. However, these terms also appear in the descriptors for vocabulary from level 3 to 5. In the vocabulary descriptors, the frequency in which vocabulary range and accuracy are sufficient to communicate on common, concrete, and work-related topics should be evaluated. According to Knoch (2009), "reference to 'common, concrete and work-related topics' is not clear to stakeholders as the whole test should be in the aviation domain" (Knoch 2009: 43). DOC 9835 states that

Context is an important consideration in communications, and an individual's language proficiency may vary in different contexts. This holistic descriptor limits the domain of the communicative requirements to work-related topics; that is, air traffic controllers and pilots are expected to be able to communicate about issues in their field of professional practice. Language proficiency should not be limited to standardized phraseology and should range across a relatively broad area of work-related communicative domains. (ICAO 2010: 4.5.3)

Very high lexical familiarity seems to be essential for good comprehension (Bonk 2000). According to Rost and Brown (2022), word recognition is the basis of spoken language comprehension, and lexical knowledge seems to be, in second language listening, “the most significant variable contributing to listening proficiency” (Rost and Brown 2022: 241). Thus, test takers’ lexical knowledge should also be assessed in a listening test. Elements of the vocabulary descriptors, such as knowledge of vocabulary used on common, concrete, and work-related topics, of idiomatic vocabulary, of vocabulary used on a wide variety of familiar and unfamiliar topics, and of nuanced vocabulary can inform the definition of the construct of a test to assess listening in isolation. It is important to point out that the purpose of the present study is to try to understand the construct represented in the ICAO rating scale, not to evaluate or criticize it. However, it is good to mention that criticism (e.g. Garcia 2015; Knoch 2009; Pfeiffer 2009) has been made in relation to the fact that the rating scale descriptors include idiomatic and nuanced vocabulary since communications between pilots and ATCOs should always be “clear, concise, and unambiguous”, even when using plain language (ICAO 2010: 4.3.4). Knoch (2009), for example, argues that “any references to idiomatic language should be deleted as this is not appropriate in the TLU domain” (Knoch 2009: 43). I have also strongly argued that “idiomatic vocabulary should never have been included in the rating scale” (Garcia 2015: 22). Today I think differently because unfortunately real-life communications are not always according to standards. As a matter of fact, ICAO acknowledges, as mentioned earlier, that 70% of the radiotelephony speech acts do not comply with the recognized standards (ICAO 2010). Thus, I argue that if idiomatic expressions are used in real-life radiotelephony communications, as the findings of Prinzo, A. Hendrix, and R. Hendrix (2009) indicate, they should be included in a test that intends to reflect the language used in real-life. Although I believe pilots and ATCOs should have the ability to comprehend idiomatic vocabulary assessed, they should be encouraged not to use them when speaking. Also, the vocabulary descriptors for level 6 may help to understand what the comprehension descriptors for level 6 mean by “a wide variety of familiar and unfamiliar topics.”

The descriptors for pronunciation, which include both segmental and suprasegmental features of pronunciation, may help to address question 8. From level 2 to 6, “pronunciation, stress, rhythm, and intonation” might be “influenced by the first language or regional variation”. However, the pronunciation of a level 2 test taker is heavily influenced, while the pronunciation of a level 6 might be influenced or not. The main difference between the levels is the frequency in which pronunciation interferes with ease of understanding. This frequency may “usually” (level 2), “frequently” (level 3), “only sometimes (level 4), “rarely (level 5), or “almost never” (level 6) interfere with ease of understanding. The accents and varieties that may be considered *sufficiently* intelligible for an international community of users, as mentioned in the comprehension descriptors for levels 3 and 4, might be the ones whose pronunciation either rarely (level 5 in pronunciation) or almost never (level 6 in pronunciation) interferes with ease of understanding. Thus, a recording used in a listening test recorded by a pilot or controller who was awarded level 5 or 6 in pronunciation might be considered an accent or variety that is *sufficiently* intelligible for an international community of users, whereas recordings which were recorded by pilots or controllers who were awarded level 4 in pronunciation might be considered within a range of speech varieties, which is mentioned in the comprehension descriptors for level 5. Test developers may even consider having pilots or controllers who were awarded 3 or less in pronunciation to make the recordings

for the listening test, since ICAO's report on the implementation of the LPRs showed that 32.25% of member states had not provided information about their implementation status (ICAO 2013). However, it is important to point out that "although strength of foreign accent is indeed correlated with comprehensibility and intelligibility, a strong foreign accent does not necessarily cause L2 speech to be low in comprehensibility or intelligibility" (Munro and Derwing 1999: 305).

Moving on, what could be considered a linguistic complication (question 5 from Table 2)? As linguistic is a broad term, many linguistic factors may be considered a linguistic complication. The interference of pronunciation, stress, rhythm, or intonation with ease of understanding may be considered linguistic complications. As Rost and Brown (2022) argue, "unexpected speaker accents, an unfamiliar phonotactic pattern, rhythm and intonation systems, length of input, number of speakers, rapid speech rate and lack of pauses, and connected speech phenomena (reductions and assimilations)" may pose challenge to linguistic processing (Rost and Brown 2022: 249). Listening to unfamiliar or uncommon vocabulary may also cause a linguistic complication. Even factors related to the fluency descriptors may add a complication, as "lack of 'orality' features (such as pauses and redundancy)" may also pose challenges. Furthermore, lack of knowledge of syntax may also be considered a linguistic complication. Rost and Brown (2022) argue that for a listener to have a detailed comprehension of a message, "a thorough syntactic processing needs to take place" (Rost and Brown 2022: 242). The descriptors for structure talk about basic and complex grammatical structures. Although there has been a discussion whether the scale's reference to complex structures reflects the real-world of radiotelephony communications (e.g., Prado 2015), there is no doubt that knowledge of grammatical structure plays an important role in comprehension. Rost and Brown (2022) list complexity of grammatical structures as one factor that poses challenges to second language listeners.

Now, what is a situational complication (question 10)? According to DOC 9835, "it is during complications in aviation that communications become most crucial, with a greater reliance upon plain language" (ICAO 2010: 4.6.6). The descriptors for vocabulary and interactions may shed some light on this issue. First, when the descriptors for vocabulary mention common topics, they imply topics might also be uncommon topics. They also talk about unfamiliar topics. Differently from uncommon topics, which may be understood as topics that do not happen frequently in radiotelephony communications, unfamiliar topics are the ones that pilots and ATCOs were not familiar with. When topics are uncommon or unfamiliar, the situation will likely be more difficult to solve. The level 4 descriptors for interactions talk about the ability to deal adequately with apparent misunderstanding. The occurrence of a misunderstanding could also be considered a situational complication. The topic of question 7 ("an unexpected turn of events?") may also be considered a possible situational complication, as it describes a situation in which events do not occur according to what is expected. DOC 9835 points out that

One of the more challenging events in all communications, including those involving the use of a second language, is when the unexpected happens. Human Factors experts have emphasized the threat of letting our expectations hinder our interpretation of reality. Sometimes, a complication or an unexpected event can lead to a communication breakdown. (ICAO 2010: 4.5.3)

The opposite of "an unexpected turn of events", according to the level 3 descriptors for structure and interactions, seems to be "predictable situations." Moreover, if a

response is not *immediate, appropriate, or informative* (as the descriptors for interactions from level 3 to 5 describe), a situational complication may arise.

Table 3 shows a summary of elements test developers may consider when developing a listening test to assess pilots and ATCOs' listening in isolation, which was based on the present analysis of how other descriptors other than comprehension may help to inform the definition of the construct.

Table 3 Elements test developers may consider when developing a listening test to assess pilots and ATCOs' listening in isolation

Include pronunciation, stress, rhythm, and intonation both influenced and not influenced by the first language or regional variation
The accents and varieties that may be considered <i>sufficiently</i> intelligible for an international community of users, as mentioned in the comprehension descriptors for levels 3 and 4, might be the ones whose pronunciation either rarely (level 5 in pronunciation) or almost never (level 6 in pronunciation) interferes with ease of understanding
Include a range of speech varieties (one parameter might be pilots and ATCOs who were awarded level 4 or below in pronunciation)
Include the assessment of the ability to understand vocabulary used on common, concrete, and work-related topics, idiomatic vocabulary, vocabulary used on a wide variety of familiar and unfamiliar topics, nuanced vocabulary
Comprehension of a wide variety of familiar and unfamiliar topics may help to understand what the comprehension descriptors for level 6 mention as comprehension "in nearly all contexts"
A linguistic complication might be caused by: <ul style="list-style-type: none"> <li>▪ Interference of pronunciation, stress, rhythm, or intonation, a linguistic complication on ease of understanding</li> <li>▪ The use of complex structures</li> <li>▪ The presence of uncommon or unfamiliar vocabulary</li> <li>▪ Inappropriate phrasing and pausing, slowness in producing language, use of too many fillers, or use of inappropriate discourse markers or connectors</li> </ul>
An unexpected turn of events is one situational complication, and might be caused by a misunderstanding, or by a response which is not immediate, appropriate, or informative, among other possibilities.
Include predictable situations (to contrast with unexpected situations)

## 5. CONCLUSION

Investigating the construct of the ICAO rating scale is only one piece of the puzzle to define the construct of a test to assess listening in isolation. Investigating the policy is one of the five different areas test developers of Language Assessments for Professional Purposes (LAPPs) should consider when developing a test, as Knoch and Macqueen (2020) recommend. They should also analyse the needs and motivations of test takers, test requirements needs, the availability of resources for test development, administration, and validation, and, most importantly, the characteristics of the TLU domain. Furthermore, the present study only analysed part of the policy, as the focus was on the rating scale.

Knoch and Macqueen (2020) suggest including in the policy analysis, an analysis of the policy environment and of the professional registration environment.

As Upshur and Turner (1995) point out, “in general, ... rating scales present major problems of reliability and validity” (Upshur and Turner 1995: 5). Not differently, the ICAO rating scale is problematic and revision should be considered by ICAO (Aragão 2018; Garcia 2015; Knoch 2009; Pfeiffer 2009). The fact that the rating scale was developed to assess speaking makes it difficult to be used to assess listening (Knoch 2009). As Garcia (2015) suggests, the development of a specific rating scale for the assessment of comprehension seems to be necessary. Future studies of the ICAO rating scale can inform its revision and contribute to the development of a more valid and reliable scale to assess speaking and of a new rating scale to assess listening. Researchers and test developers could look into other dimensions of construct as described by Knoch and Macqueen (2020): the perceived construct (e.g., how stakeholders understand the construct in the rating scale) and the operationalized construct (what is actually being assessed by tests). Further analysis of other aspects of the ICAO policy may offer insights into understanding problems related to the implementation of the ICAO LPRs. For example, researchers and test developers could also look at aspects related to the construction of the policy, its context, and impact, including its strengths and concerns. Moreover, researchers and test developers could also investigate the values embodied in the policy (see Shohamy 2001). Furthermore, they could conduct an evaluation of the policy using the policy evaluation framework provided by Knoch and Macqueen (2020).

The purpose of this study was to improve the understanding of the ICAO policy in order to inform the development of tests that aim to assess pilots and ATCOs’ listening in isolation fairly, as required by the ICAO test design guidelines (ICAEA n.d.). Having a common understanding of the construct may contribute to reduce the variation of quality of tests worldwide. To conclude, test developers need to keep in mind that

In such a ‘high-stakes’ environment, language testing needs to be accountable to the stakeholders in the aviation industry. State regulators, managers of airline and air traffic management service providers, trainers, pilots and controllers, and ultimately, the flying public, need to be able to trust global language assessments and to have confidence that licensed operations personnel are competent communicators in the English language. As those who are obliged to learn and use English on the frequency, pilots and controllers deserve to have their language proficiency assessed fairly, and to know that their counterparts around the world have been assessed according to the same standard. (Emery 2007: 1)

#### REFERENCES

- Alderson, J. Charles, Neus Figueras, Henk Kuijper, Guenter Nold, Sauli Takala, and Claire Tardieu. “Analysing Tests of Reading and Listening in Relation to the Common European Framework of Reference: The Experience of The Dutch CEFR Construct Project.” *Language Assessment Quarterly* 3, no. 1 (2006): 3–30.
- Alderson, J. Charles. “A Survey of Aviation English Tests.” *Language Testing*, 27(1), (2010): 51-72. doi: 10.1177/0265532209347196

- Alderson, J. Charles. "The Politics of Aviation English Testing. Language." *Assessment Quarterly*, 8:4 (2011): 386-403. doi: 10.1080/15434303.2011.622017
- Aragão, Beatriz F. "O Uso de Critérios Autóctones no Contexto Aeronáutico: Contribuições para uma Nova Escala de Proficiência para Controladores de Tráfego Aéreo" [The Use of Indigenous Assessment Criteria in the Aeronautical Context: Contributions to a New Proficiency Rating Scale for Air Traffic Controllers]. In Matilde V. R. Scaramucci, Patricia Tosqui-Lucks, and Silvia M. Damião (Eds.), *Pesquisas sobre Inglês Aeronáutico no Brasil*. Campinas: Pontes Editores, 2018: 243–269.
- Bonk, William J. "Second Language Lexical Knowledge and Listening Comprehension." *International Journal of Listening* 14, no. 1 (2000): 14–31.
- Borowska, Anna P. *Avialinguistics: The Study of Language for Aviation Purposes*. Frankfurt am Main: Peter Lang Edition, 2017.
- Bowen, Glenn A. "Document Analysis as a Qualitative Research Method." *Qualitative Research Journal* 9, no. 2 (2009): 27–40.
- Brazil. *The Need for Revision of Annex 1 Language Proficiency SARPS for Pilots and Air Traffic Controllers*. International Civil Aviation Organization Working Paper A39-WP/249, 2016. Retrieved 14 December, 2022, from [https://www.icao.int/Meetings/a39/Documents/WP/wp\\_249\\_en.pdf](https://www.icao.int/Meetings/a39/Documents/WP/wp_249_en.pdf)
- Buck, Gary. *Assessing Listening*. Cambridge: Cambridge University Press, 2001.
- Cardno, Carol. "Policy Document Analysis: A Practical Educational Leadership Tool and a Qualitative Research Method." *Educational Administration: Theory and Practice* 24, no. 4 (2018): 623-640.
- Douglas, Dan. "Three Problems in Testing Language for Specific Purposes: Authenticity, Specificity and Inseparability." In *Experimenting with Uncertainty: Essays in Honour of Alan Davies*. Cambridge: Cambridge University Press, 2001: 45–52.
- Douglas, Dan. "Assessing the Language of International Civil Aviation: Issues of Validity and Impact." Proceedings from the *International Professional Communication Conference*, IEEE Professional Communication Society. IEEE (2004): 248-252
- Douglas, Dan. *Understanding Language Testing*. London: Hodder Education, 2010.
- Douglas, Dan. "Nobody Seems to Speak English Here Today: Enhancing Assessment and Training in Aviation English." *Iranian Journal of Language Teaching Research*, 2(2), (2014): 1-12.
- Emery, Henry J. "Standardization in Language Rating: An Association of Language Raters." *Second ICAO Aviation Language Symposium*. (May, 2007), available at [https://www.icao.int/Meetings/AMC/MA/Second%20ICAO%20Aviation%20Language%20Symposium%20\(IALS-2\)/23.Emery.pdf](https://www.icao.int/Meetings/AMC/MA/Second%20ICAO%20Aviation%20Language%20Symposium%20(IALS-2)/23.Emery.pdf)
- Emery, Henry J. "Developments in LSP Testing 30 Years On? The Case of Aviation English." *Language Assessment Quarterly* 11, no. 2 (2014): 198–215.
- Estival, Dominique, Candace Farris, and Brett Molesworth. *Aviation English: A Lingua Franca for Pilots and Air Traffic Controllers*. New York: Routledge, 2016.
- Field, John. "Idle Chatter: What Really Goes On in Tests of Interactive Communication?," *CRELLA Symposium*. (2020), available at <https://www.youtube.com/watch?v=ONM7xcJTPD0&feature=youtu.be>
- Garcia, Angela C. M. *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*. Unpublished Master's Thesis. Lancaster University, 2015.

- George, Alexander. "Quantitative and Qualitative Approaches to Content Analysis." In Klaus Krippendorff and Mary A. Bock (Eds.), *The Content Analysis Reader*. Thousand Oaks, CA: Sage Publications, 2009: 144-155.
- Howard, John W. "'Tower, Am I Cleared to Land?': Problematic Communication in Aviation Discourse." *Human Communication Research* 34, no. 3 (2008): 370-391.
- ICAEA. *Guidelines: ICAO LPR Test Design Guidelines*. International Civil Aviation English Association. Retrieved December 14, 2022, from <https://www.icaea.aero/icao-lpr-tdg/guidelines/>
- ICAO. *ICAO Doc 9835. Manual on the Implementation of ICAO Language Proficiency Requirements*. International Civil Aviation Organization. 2004.
- ICAO. *ICAO Cir 318. Language Testing Criteria for Global Harmonization*. International Civil Aviation Organization. 2009.
- ICAO. *ICAO Doc 9835. Manual on the Implementation of ICAO Language Proficiency Requirements* (2nd ed.). International Civil Aviation Organization, 2010.
- ICAO. *Implementation Status of English Language Proficiency Requirements*. Working Paper A38.WP/37. International Civil Aviation Organization, 2013. Retrieved 08 March, 2023, from [https://www.icao.int/Meetings/a39/Documents/WP/wp\\_514\\_en.pdf](https://www.icao.int/Meetings/a39/Documents/WP/wp_514_en.pdf)
- ICAO. *Report of the Technical Commission on Agenda Item 35*. Working Paper A39.WP/514. International Civil Aviation Organization, 2016. Retrieved 14 December, 2022, from [https://www.icao.int/Meetings/a38/Documents/WP/wp037\\_en.pdf](https://www.icao.int/Meetings/a38/Documents/WP/wp037_en.pdf)
- ICAO. *Annex 1 to the Convention on International Civil Aviation – Personnel Licensing* (14th ed.). International Civil Aviation Organization, 2022a.
- ICAO. ICAO Language Proficiency Requirements: Achieving Standardization | Catalin Popa [Video]. *Skytalks*. ICAO TV. 2022b October. Retrieved from <https://www.icao.tv/skytalks/season:1/videos/a41-skytalk-icao-language-proficiency-requirements-achieving-standardization>
- Jacoby, Sally, and Tim McNamara. "Locating Competence." *English For Specific Purposes* 18, no. 3 (1999): 213-241.
- Kim, Hyejeong, and Catherine Elder. "Understanding Aviation English as a lingua franca: Perceptions of Korean aviation personnel." *Australian Review of Applied Linguistics* 32(3), (2009): 23.1-23.17.
- Kim, Hyejeong, and Catherine Elder. "Interrogating the Construct of Aviation English: Feedback from Test Takers in Korea." *Language Testing* 32, no. 2 (2015): 129-149.
- Kim, Hyejeong. "Exploring the Construct of Radiotelephony Communication: A critique of the ICAO English testing policy from the perspective of Korean aviation experts." *Papers in Language Testing and Assessment* 2(2), (2013): 103-110
- Kim, Hyejeong, and Rosey Billington. "Pronunciation and Comprehension in English as a Lingua Franca Communication: Effect of L1 Influence in International Aviation Communication." *Applied Linguistics* 39, no. 2 (2018): 135-158.
- Knoch, Ute. "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 (2009): 21-46.
- Knoch, Ute. "Using Subject Specialists to Validate an ESP Rating Scale: The Case of the International Civil Aviation Organization (ICAO) Rating Scale." *English for Specific Purposes* 33 (2014): 77-86.
- Knoch, Ute, Bart Deygers, and Apichat Khamboonruang. "Revisiting Rating Scale Development for Rater-Mediated Language Performance Assessments: Modelling



- Construct and Contextual Choices Made by Scale Developers.” *Language Testing* 38, no. 4 (2021): 602–626.
- Knoch, Ute., and Susan Macqueen. *Assessing English for Professional Purposes*. Milton: Routledge, 2020.
- Lam, Daniel M. K. “Don’t Turn a Deaf Ear: A Case for Assessing Interactive Listening.” *Applied Linguistics* 42, no. 4 (2021): 740–764.
- McNamara, T. F., Ute Knoch, and Fan Jason. *Fairness, Justice, and Language Assessment: The Role of Measurement*. Oxford, United Kingdom: Oxford University Press, 2019.
- Merriam, Sharan B. *Case Study Research in Education: A Qualitative Approach*. 1st ed. San Francisco: Jossey-Bass, 1988.
- Merriam, Sharan B. *Qualitative Research and Case Study Applications in Education*. 2nd ed. San Francisco: Jossey-Bass Publishers, 1998.
- Messick, Samuel. “Validity.” In *Educational Measurement* (3rd ed.), ed. Robert L. Linn, 13–103. Phoenix: Macmillan, 1989.
- Monteiro, Ana L. T. *Reconsidering the Measurement of Proficiency in Pilot and Air Traffic Controller Radiotelephony Communication: From Construct Definition to Task Design*. PhD’s Dissertation. Carleton University, 2019.
- Monteiro, Ana L. T. “The Assessment of English in Aeronautical Radiotelephony Communications: A Mixed Methods Study.” In *Aviation English: A Global Perspective*, ed. Patricia Tosqui-Lucks. Sao Paulo: Bookerfield, 2022, 224-245.
- Munro, Murray, and Tracey Derwing. “Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners.” In *Language Learning*, 49 (sup. 1), (1999): 285-310.
- O’Leary, Zina. *The Essential Guide to Doing Your Research Project*. 4th edition. London: SAGE Publications, 2021.
- Pfeiffer, Anja. *Inter-Rater Reliability in an Aviation Speaking Test*. Master’s Thesis. Lancaster University, 2009.
- Popa, Catalin. “ICAO Language Proficiency Requirements: Occurrence, Developments and Associated Documents.” *Uniting Aviation*. 05 April 2019. Retrieved December 14, 2022, from <https://unitingaviation.com/news/safety/english-language-proficiency-requirements-developments-and-associated-documents/>
- Prado, Malila C. A. Levantamento de Padrões Léxico-Gramaticais do Inglês para Aviação: Um Estudo Vetorado pela Linguística de Corpus [Survey on the Lexicogrammatical Patterns in Aviation English: A Study Conducted by corpus linguistics]. Master’s Thesis. Universidade de São Paulo, 2015.
- Prinzo, O. Veronika, A. M. Hendrix, and R. Hendrix. *The Outcome of ATC Message Length and Complexity on En Route Pilot Readback Performance*. DOT/FAA/AM-09/2. Washington, DC: Federal Administration, 2009.
- Read, John, and Ute Knoch. “Clearing the Air: Applied Linguistic Perspectives on Aviation Communication.” *Australian Review of Applied Linguistics*, 32(3), (2009): 21.1–21.11.
- Rost, Michael, and Steven Brown. “Second Language Listening.” In *Handbook of Practical Second Language Teaching and Learning*, ed. Eli Hinkel. Routledge, 2022.
- Shohamy, Elana G. *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. Harlow, England: Longman, 2001.

- Silva, Ana L. B. C. "Language Proficiency Test Design for a Group of Military Pilots: Criteria, Construct and Purpose." *The Journal of Teaching English for Specific and Academic Purposes*, 2023.
- Tosqui-Lucks, Patricia, and Ana L. B. C. Silva. Da Elaboração de um Glossário Colaborativo à Discussão sobre os Termos Inglês para Aviação e Inglês Aeronáutico. [From Writing a Collaborative Glossary to Discussing the Terms "Aviation English" and "Aeronautical English"]. *Estudos Linguísticos*, 48(4), (2020): 97–116. doi:10.21165/el.v49i1.25610
- Treadaway, Maria. "The ICAO Scale and Language Testing for *Ab Initio* Cadets: Is There a Fit?" In *Aviation English: A Global Perspective*, ed. Patricia Tosqui-Lucks. Sao Paulo: Bookfield, 2022, 246-261.
- Trippe, Julia E. (2018). *Aviation English is Distinct from Conversational English: Evidence from Prosodic Analyses and Listening Performance*. ProQuest Dissertations Publishing.
- Upshur, John A, and Carolyn E Turner. "Constructing Rating Scales for Second Language Tests." *ELT Journal* 49, no. 1 (1995): 3–12.
- Wodak, Ruth. "Linguistic Analyses in Language Policies." In T. Ricento (Ed.), *An Introduction to Language Policy: Theory and Method*. Oxford: Blackwell, 2006: 170-193.