Reconsidering the Measurement of Proficiency in Pilot and Air Traffic Controller Radiotelephony Communication: From Construct Definition to Task Design

by

Ana Lúcia Tavares Monteiro

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Applied Linguistics and Discourse Studies

Carleton University
Ottawa, Ontario

Abstract

The high-stakes context of international radiotelephony (RT) communication, in which pilots and air traffic controllers (ATCOs) use aviation English (AE) as a *lingua franca*, requires a robust testing policy that is clear and fair to all stakeholders. The International Civil Aviation Organization Language Proficiency Requirements have been criticized for their lack of fit with pilots' and ATCOs' real-life communicative needs, for both native and non-native speakers of English (Douglas, 2014; Kim, 2012). This dissertation investigates and specifies the proficiency construct (awareness, knowledge, skills, and attitudes) in pilot-ATCO intercultural RT, following Fulcher and Davidson's (2007) test development framework, and drawing on theoretical and empirical studies in the domains of Aviation English, English as a Lingua Franca, Intercultural awareness, and Interactional competence. It is guided by the following overarching research question: What are the communicative demands of pilots and ATCOs involved in intercultural RT communications; how can they be specified within a construct framework and operationalized as test tasks? In order to address this question, a multiphase mixed methods (MM) research design (Creswell, 2014) with three integrated phases is applied. In phase one, the intercultural RT context is explored (MM exploratory study). In phase two, models of language use relevant to this aviation workplace are proposed; and, key construct components are specified and validated by 128 aviation stakeholders (qualitative study). In the third/final phase, draft tasks that operationalize the identified RT construct are designed and pilot tested with AE testing experts (convergent parallel MM study). Integration of all findings allows for *meta-inferences* regarding the research question and underscores 1) the value of MM research designs in addressing

such complex questions, and 2) the role of domain experts in: exploring a target testing context, specifying and validating a construct, and designing and pilot testing tasks to generate the evidence needed to make valid inferences about test-takers' specific purpose language ability. A broader view of professional communicative competence for intercultural RT suggests an encouraging perspective in a highly specialized occupational domain, has implications for aviation safety, and serves as an example of how non-linguistic competencies could be operationalized in contexts other than aviation.

Acknowledgements

On the occasion of finalizing my master's thesis, I began to write the acknowledgements, but I must confess that, among all, this was the most difficult task. At that time, I lacked the words that should have given the measure of how grateful I was to those people who, with their knowledge and intelligence, contributed so much to the enrichment of that work.

Now, ten years later, the moment has come that is very dear to my heart, because to thank those who have benefited me is not only a duty of justice but, above all, a wealth that is incorporated into my life.

The present doctoral work could not have existed if I had not counted on the capacity, expertise, and even apparent small contributions, of others. A doctoral work depends on the help of a set of skills and professionals, just as the performance of a symphony depends on a harmonious and tuned orchestral ensemble.

With the joy of one who has had an experience of true fraternity, I would like to express my deepest gratitude:

- to God, for having given me a wonderful family and all the strength and capacity I needed to conclude this work; for answering my prayers, opening doors and making my dream come through;
- to my dearest husband Guilherme, I have no words to thank him for his immense support and unconditional love; for agreeing to come to Canada with me and changing completely his personal and professional life to fulfill my dream of pursuing a PhD program at Carleton University; for understanding the many times I was not able to give him the attention he deserved in order to read articles, write reflections, collect and

analyze data, and especially during these last days when I was totally immersed in writing this dissertation;

- to my daughter Ana Carolina and my son-in-law Pedro, to whom I am particularly indebted for their valuable support throughout these four years and especially for the precious hours they spent doing the hard work of formatting and checking the final text;

- to my son José Guilherme and my daughter-in-law Mônica, for their continuous encouragement and for understanding my absence and distance for such a long time;

- to my father Odilon, to whom I am particularly grateful for supporting me all the time and for agreeing to come to Ottawa to assist me in the final task of revising my text, tables and figures and providing insightful comments and suggestions; for the precious time he spent with me, giving me courage to move on and making me believe that I would be victorious in this research endeavor;

- to my mother Marlene and my sister Ângela, for their prayers and support at a distance, and for always believing in my aptitude to conclude this Doctoral program;

- to the Civil Aviation National Agency (ANAC, Brazil), for having approved my leave of absence from Brazil to pursue the Doctoral Program in Applied Linguistics and Discourse Studies at Carleton University (Ottawa, ON, Canada);

- to my supervisor, Professor Janna D. Fox, for the opportunities she has given me to deepen with confidence my knowledge of Applied Linguistics, and for the safe guidance I have had during the four years of my Doctoral Project, with permanent encouragement demonstrated by her interest and attention to my work; as a notable

language testing expert, Janna has been a perfect match to my research interests. Her vast experience in high-stakes test development and a plethora of publications in this field confirm how privileged I am of being supervised by her. She is definitely an inspiration; Janna proved to be more than a supervisor to me, but a real friend who helps to overcome academic and personal challenges as well. She always valued my work and research in a way no one has done before, and in difficult times she made me believe all efforts were worth it;

- to the Advisory Committee Members of my Doctoral Project, Professor Natasha Artemeva, Professor Trudy O'Brien and Professor Dan Douglas, to whom I am particularly thankful for a number of valuable and stimulating discussions, for forwarding articles and studies related to my research topic, and for suggesting revisions to improve my final text;
- to the organizers of the two international conferences (ICAEA, Croatia and GEIA, Brazil), who authorized me to announce my research study and recruit participants;
- to all pilots, air traffic controllers, Aviation English teachers, examiners, test developers, regulators and researchers who volunteered to participate in this study. Without their accounts and experience, it would not have been possible to conduct this research;
- to those who contributed to this research as second or third coders of qualitative data and also to those who provided comments to my initial findings, as part of the member-checking strategy;
- to the Language Proficiency Group Coordinator (ANAC, Brazil), for having authorized me to conduct the pilot testing of the draft tasks during ANAC accredited examiners' recurrent course in São José dos Campos-SP;

- to the pilots who worked with me in the development of the two draft tasks, I am particularly thankful for their time and expertise;
- to the pilots and ATCOs who volunteered to play the roles of test-takers and interlocutors in the pilot testing of the task, my special thanks;
- to the ICAEA Board members and members of the ICAEA Research Group, for the continuing support and encouragement;
- to Joan Grant, to whom I am indebted for her kind attention to all my needs as the Degree Programs Administrator of the School of Linguistics and Language Studies at Carleton University;
- to all the people not mentioned here who, directly or indirectly, gave their contribution to the accomplishment of this work, my sincere acknowledgment.

Table of Contents

List of Tables

List of Illustrations

List of Appendices

List of Abbreviations and Acronyms

AE – Aviation English

AETEs – Aviation English Testing Experts

ANAC – Agência Nacional de Aviação Civil (Civil Aviation National Agency – Brazil)

AT − Attitude

ATC – Air traffic control

ATCOs – Air traffic control officers (Air traffic controllers)

ATIS – Automatic Terminal Information Service

AUM – Anxiety/Uncertainty Management

AW – Awareness

CAA – Civil Aviation Authority

CLT – Communicative Language Testing

CRM – Crew Resource Management

CUREB – Carleton University Research Ethics Board

EAP – English for Academic Purposes

EFL – English as a Foreign Language

EGP – English for General Purposes

EIL – English as and International Language

ELEs – English Language Experts

ELF – English as a Lingua Franca

ERB – Ethics Review Board

ESL – English as a Second Language

ESP – English for Specific Purposes

FE – Flight engineer

FG – Focus group

FO – First Officer

IA – Interactional authenticity

IC – Interactional competence

ICA – Intercultural awareness

ICAEA – International Civil Aviation English Association

ICAO – International Civil Aviation Organization

ILS – Instrument Landing System

IMP − Impact

JFK – John F. Kennedy International Airport (Airport Code), New York, US

K − Knowledge

L1 – First language

L2 – Second language

LPRs – Language Proficiency Requirements

LSP – Language for Specific Purposes

M – Mean

MM – Mixed methods (sometimes MMR = Mixed Methods Research)

NAV – Navigation

NNSs – Non-native speakers

NSs – Native speakers

OET – Occupational English Test

P – Practicality

PELA − Proficiency Test in English Language for Air Traffic Controllers

QUAL – Qualitative (research approach)

QUAN (quan) – Quantitative (research approach)

R – Reliability

REB – Research Ethics Board

RQ – Research question

RT – Radiotelephony

S – Skills

SA – Situational authenticity

SARPs – Standards and Recommended Practices

SD – Standard Deviation

SDEA – Santos Dumont English Assessment (Brazilian Language Proficiency Test for civil aviation pilots)

SE – Standard English

SMEs – Subject matter experts

TLU − Target language use

UCLES – University of Cambridge Local Examination Syndicate

V − Validity

VOR − VHF omnidirectional range (short-range radio navigation system)

Glossary of Terms

Term	Definition
Accommodation strategies	Communicative strategies that include being aware of the challenges faced by speakers of English as a foreign language (EFL) and of an interlocutor's possible linguistic difficulties; replacing high-risk (possibly unclear or ambiguous) features of the language to increase communicative efficiency; focusing on keeping their intonation neutral and calm, admittedly difficult at busy control areas, but a good strategy to calm the language anxiety of an EFL speaker; among others (ICAO, 2010).
Accommodation in speaking	"A natural process of adapting speech habits to the constraints of the context and the perceived ability of the hearer to understand" (ICAO, 2010, p. 2-6).
Anxiety/uncertainty management theory	The theory that is based on the assumption that "strangers cannot communicate effectively with hosts if their uncertainty and anxiety are too high" (Gudykunst, 1998, p. 229). Interlocutors experience uncertainty about attitudes, feelings and behaviors of the other, and anxiety relates to "the tension or apprehension that strangers have about what will happen when they communicate with hosts" (p.229), which is generally based on negative expectations.
Attribute coding	Essential information related to sites of data collection, time frame, data format and number, participants` characteristics and the context of each scenario (Saldaña, 2009).
Aviation English	The language used by pilots and ATCOs in radiotelephony, which encompasses both the English-based standardized phraseology prescribed by ICAO and plain English used for the specific purpose of aeronautical communications.
Codes	"A word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data" (Saldaña, 2013, p. 3).
Communication accommodation theory	The theory that assumes that "accommodation is the process through which interactants regulate their communication (adopting a particular linguistic code or accent, increasing or decreasing their speech rate, avoiding or increasing eye contact, etc.) in order to appear more like (accommodation) or distinct from each other (non-accommodation)" (Gallois, Ogay & Giles, 2005, p. 137).
Communication styles	Directness, i.e., being explicit and straight to the point, or indirectness, i.e., being implicit, vague and speak in a confusing way. Can be associated with conflict communication styles, such as dominating, avoiding,

	accommodating, compromising, and integrating (Ting-Toomey, 2005).
Construct	"A construct is some postulated attribute of people, assumed to be reflected in test performance. In test validation, the attribute about which we make statements in interpreting a test is a construct" (Cronbach & Meehl, 1955, p. 283).
Construct-irrelevant variance	A threat to validity in which "the assessment is too broad, containing excess reliable variance that is irrelevant to the interpreted construct" (Messick, 1996, p.4).
Construct underrepresentation	A threat to validity in which "the test is too narrow and fails to include important dimensions or facets of focal constructs" (Messick, 1996, p.4).
Construct validity	A quality of test usefulness that "refers to the extent to which we can interpret a given test score as an indicator of the ability (ies), or constructs, we wish to measure" (Bachman & Palmer, 1996, p. 21).
Convergent parallel MM study	Used when "a researcher collects both quantitative and qualitative data, analyzes them separately, and then compares the results to see if the findings confirm or disconfirm each other" (Creswell, 2014, p. 269).
Conversational constraints theory	The theory that encompasses five conversational constraints (Kim, 2005): concern for clarity, concern for minimizing imposition, concern for avoiding hurting the hearer's feelings, concern for avoiding negative evaluation by the hearer, and concern for effectiveness.
Culture	"Neither relatively static nor ever-changing, but both" (Kesckes, 2014, p. 4), having *a priori* elements (ethnic or cultural marking in communicative behavior) and emergent features (co-constructed in the moment of interaction), which should be combined to approach culture in a dialectical and dynamic way.
Dimensions	In this study, they refer to major aspects of the construct that structured the matrix of construct specification, i.e., awareness, knowledge, skills, and attitudes.
Distributed cognition	A social theory proposed by Hutchins (1995a) in which the outcomes of participants' tasks depend on high levels of cooperation and coordination with artifacts and technological tools.
Domains	In this study, they refer to the four key areas of interest from which theoretical and empirical research were selected for review: Aviation English (AE), English as a lingua franca (ELF), Intercultural Awareness/competence (ICA) and Interactional Competence (IC).
English as a Lingua Franca	"An additionally acquired language system which serves as a common means of communication for speakers of

	different first languages" (Jenkins, Cogo & Dewey, 2011, p. 283).
Evidence-centered design	"A methodology for designing assessments that underscores the central role of evidentiary reasoning in assessment design" (Mislevy, Almond & Lukas, 2003, p. 20).
Evidence models	A sub-layer in the test development process that refers to the evidence "we need to collect in order to make inferences from performance to underlying knowledge or ability. Therefore, the evidence model answers the question: what evidence do we need to test the construct(s)?" (Fulcher & Davidson, 2007, p.66).
Expectancy violation theory	The theory that "assumes that expectancies (and beliefs) entail both a predictive and a prescriptive component" (Burgoon & Hubbard, 2005, p. 151), that's to say, one reflects typical communicative acts in a given culture while the other refers to idealized standards of conduct, respectively. However, the authors emphasize that each culture's expectancies, and the way one reacts to them, will vary significantly along Hofstede's (1991) cultural dimensions, communication styles and degrees of face concern.
Face	The concept of face is tied up "with notions of being embarrassed or humiliated, or 'losing face'. Thus face is something that is emotionally invested, and that can be lost, maintained, or enhanced, and must be constantly attended to in interaction" (Brown & Levinson, 1987, p. 63).
Face work	"The specific verbal and nonverbal behaviors that we engage in to maintain or restore face loss and to uphold and honor face again", where face is related to "identity respect and other-identity consideration issues within and beyond the actual encounter episode" (Ting-Toomey, 2005, p. 73).
Face-negotiation theory	[Integrated] "cultural-level dimensions and individual-level attributes to explain face concerns, conflict styles, and face work behaviors" (Ting-Toomey, 2005, p. 9).
Factor	In this study, the word 'factor' is used generically, as a synonym for feature, associated with QUAL studies. It does not refer to the outcome of factor analysis, associated with QUAN studies.
First cycle coding methods	Processes that occur during the initial coding of data (e.g., Grammatical, Elemental, Affective, Literary and Language, Exploratory, and Procedural coding, as well as Theming the data (Saldaña, 2013).

High vs. low power distance	Hofstede's (1991) cultural dimension related to "*the extent to which the less powerful member of institutions* [family, school, community] *and organizations* [places where people work] *within a country expect and accept that power is distributed unequally"* (Hofstede & Hofstede, 2005, p. 46).
High vs. low uncertainty avoidance	Hofstede's (1991) cultural dimension, associated with feeling threatened by ambiguous or unknown situations, i.e., (non) acceptance of ambiguity.
Impact	A quality of test usefulness that refers to "the various ways in which test use affects society, an education system, and the individuals within these" (Bachman & Palmer, 1996, p. 39).
Impoliteness strategies	Strategies that seek to cause social disruption, through hostile communication or confrontational discourse (e.g., being unsympathetic, seeking disagreement, making the other feel uncomfortable, and associating the other with a negative aspect) (Culpeper, 1996).
Individualism vs. collectivism	Hofstede's (1991) cultural dimension, which refers to individual interests versus group interests.
Initial Coding	A coding method that involves "breaking down qualitative data into discrete parts" (Saldaña, 2013, p. 100) as a starting point for further exploration. It is appropriate "for virtually all qualitative studies" (p. 101) and one of its goals is to remain open to possible theoretical directions.
Interactional authenticity	A quality of good testing practice that "involves the interaction of the test taker's specific purpose language ability with the test task. The extent to which the test taker is engaged in the task, by responding to the features of the target language use situation embodied in the test task characteristics, is a measure of interactional authenticity" (Douglas, 2000, p.17).
Interactional competence	Kramsch (1986) states that "successful interactions presuppose not only a shared knowledge of the world, the reference to a common external context of communication, but also the construction of a shared internal context or 'sphere of inter-subjectivity' that is built through the collaborative efforts of the interactional partners" (p. 367).
Intercultural awareness	"A conscious understanding of the role culturally based forms, practices and frames of reference can have in intercultural communication, and an ability to put these

	conceptions into practice in a flexible and context specific manner in real time communications" (Baker, 2011, p. 202).
Intercultural communicative competence	The ability to "interact with people from another country or culture in a foreign language…to negotiate a mode of communication and interaction which is satisfactory to themselves and the other and …to act as mediator between people of different cultural origins" (Byram, 1997, p. 71).
Interculturality	"A phenomenon that is not only interactionally and socially constructed in the course of communication but also relies on relatively definable cultural models and norms that represent the speech communities to which the interlocutors belong" (Kesckes, 2014, p. 14).
Likert scale questions	"A scale often used in questionnaires that asks participants to rate some idea using a range of numbers, usually no more than 10. A typical Likert scale may have five points, where 1 = strongly agree, 2 = somewhat agree, 3 = neutral, 4= somewhat disagree, 5 = strongly disagree" (Larson-Hall, 2016, p.478).
Linguaculture	The linguistic and cultural backgrounds of participants, used by Baker (2009) to highlight "the language-culture connection and the importance of different languages and cultures in communication" (p. 569).
Magnitude Coding	A coding method that is used to add an extra code (alphanumeric or symbolic) to a category or piece of data already coded in order to signal its intensity, frequency, direction, presence or evaluative content (Saldaña, 2013).
Main Trial	A formal pretesting of an examination or task before it becomes operational, "presented under the same circumstances as in the live exam" (Alderson, Clapham & Wall, 1995, p.76).
Masculinity vs. femininity	Hofstede's (1991) cultural dimension, which refers to a continuum of attitudes related not only to gender roles within a culture, but also to values regarding who has a greater investment in social relations and nurturing, regardless of gender.
Matrix of construct specification	A framework that details the dimensions of interest (awareness, knowledge, skills and attitudes) across the domains of Aviation English, English as a lingua franca, Intercultural awareness/competence, and Interactional competence (see Table 8.7, this study).

Meta-inferences	"An overall conclusion, explanation, or understanding developed through an integration of the inferences obtained from the qualitative and quantitative strands of a mixed method study" (Tashakkori & Teddlie, 2008, p. 101).
Multiphase MM design	When a researcher "examines a problem or topic through an iteration of connected quantitative and qualitative studies …, with each new approach building on what was learned previously to address a central program objective" (Creswell & Plano Clark, 2011, p. 100).
Mutual-face concern	Face orientation, related to "the concern for both party's images and/or the 'image' of the relationship" (Ting-Toomey, 2005, p. 74).
Negative Washback	"Negative washback is said to occur when a test's content or format is based on a narrow definition of language ability, and so constrains the teaching/learning context" (Taylor, 2005, p.1).
Pattern coding	A coding method that is appropriate for the second cycle of coding. Pattern codes "not only organize the corpus but attempt to attribute meaning to that organization" (Saldaña, 2013, p. 209), by identifying emerging themes or constructs.
Phase	The major parts of this MM research study (e.g., Phase 1, Phase 2 and Phase 3). Some authors refer to them as stages.
Pilot testing	Less formal pretesting of an examination or task before launching, aiming to "iron out the main problems before the major trials" (Alderson, Clapham & Wall, 1995, p.74).
Plain language	It is "the spontaneous, creative and non-coded use of a given natural language". However, in this specific context, plain language is also "constrained by the functions and topics (aviation and non-aviation) that are required by aeronautical radiotelephony communications, as well as by specific safety-critical requirements for intelligibility, directness, appropriacy, non-ambiguity and concision." (ICAO, 2010, p. 3-5).
Positive Washback	"[t]he extent to which the test influences language teachers and learners to do things that they would not necessarily otherwise do" (Messick, 1996, p. 243).
Practicality	The quality of test usefulness that refers to "the relationship of the resources that will be required in the design, development, and use of the test, and the resources that will

	be available for these activities" (Bachman & Palmer, 1996, p. 39).
Process coding	A coding method that uses “ing” words (gerunds) to indicate action in the data, from simple observable activities to more general conceptual actions (Saldaña, 2013).
Provisional coding	A coding method that involves the application of a set of pre-defined codes, developed from the literature review or theoretical framework, previous research findings or prior knowledge (Saldaña, 2013).
Quantitizing	‘Quantitizing’ is an expression used by Saldaña (2009) and also by Ziegler and Kang (2016) when referring to the quantification of qualitative data, as a method of data transformation for analysis.
Radiotelephony (RT) triad	To be clear, concise and unambiguous.
Reliability	A quality of test usefulness which refers to "a function of the consistency of scores from one set of tests and test tasks to another" (Bachman & Palmer, 1996, p.19).
Second cycle coding methods	Methods that are employed during a second step of coding, requiring “such analytical skills as classifying, prioritizing, integrating, synthetizing, abstracting, conceptualizing, and theory building” (Saldaña, 2013, p. 58).
Self-face concern	Face orientation, related to “the protective concern for one’s own image when one’s own face is threatened in the conflict situation” (Ting-Toomey, 2005, p. 74).
Sentiment analysis	A type of coding that allows the researcher to “grasp the range of sentiment across … content by categorizing over four streams: very positive, moderately positive, moderately negative or very negative” (NVivo, 2019).
Simultaneous coding	A coding method that is used when there is a need to apply more than one code to the same segment of data (e.g., data suggest multiple meanings, rich and complex events, several areas of interest) (Saldaña, 2013).
Situational authenticity	A quality of good testing practice that “can be demonstrated by making the relationship between the test task characteristics and the features of tasks in the target language use situation explicit” (Douglas, 2000, p. 17).

Standard phraseology	"A restricted or coded use of language comprising fixed standard phrases or lexical and syntactical routines, developed either by consensus for highly repetitive communications (e.g. everyday exchanges of greetings) or formally prescribed for special or professional purposes." (ICAO, 2010, p. ix).
Strand	Separate studies undertaken within a phase of a MM research study (e.g., QUAL and QUAN strands).
Task models	A sub-layer in the test development process that describes "the situations in which test takers respond to items or tasks that generate the evidence we need" (Fulcher and Davidson, 2007, p. 67).
Taxonomy of intercultural factors	A categorization of culturally-related factors that may impact international RT communication (see Table 8.10, this study).
Test specifications	"Usually called 'specs', they are generative explanatory documents for the creation of test tasks. Specs tell us the nuts and bolts of how to phrase the test items, how to structure the test layout, how to locate the passages, and how to make a host of difficult choices as we prepare test materials. More importantly, they tell us the rationale behind the various choices that we make." (Fulcher & Davidson, 2007, p. 52). Test Specifications "are generative blueprints or plans for a specific test." (Fulcher & Davidson, 2007, p. 36).
Test task	"A test task is essentially a device that allows the language tester to collect evidence" (Fulcher & Davidson, 2007, p. 62)
Test Validation	"Test validation is empirical evaluation of the meaning and consequences of measurement, taking into account extraneous factors in the applied setting that might erode or promote the validity of local score interpretation and use." (Messick, 1996, p. 6)
Validity threats	Potential problems that "might compromise the merging or connecting of the quantitative and qualitative strands of the study and the conclusions drawn from their combination" (Creswell & Plano Clark, 2011, p. 239).
Values coding	A coding method that is used when data reflect participants' values, attitudes and/or beliefs, based on their perceptions and views of the world. This method is mostly employed in studies that "explore cultural values, identity,

	intrapersonal and interpersonal participants' experiences and actions in case studies" (Saldaña, 2013, p. 111).

Chapter 1 – Introduction

Over the past decades, the international aviation industry has dramatically increased in complexity. More often than not, flight crews and ground staff are working in multilingual and multicultural environments, where a large proportion of radiotelephony (RT) communications take place in the English language. Consequently, the communicative demands of pilots and air traffic controllers (ATCOs) involved in intercultural RT communications, both native (NSs) and non-native speakers (NNSs) of English[1], call for a complex array of knowledge, skills and attitudes that go beyond language proficiency.

1.1 The research problem – rationale for the study

Effective communication and collaboration are essential in the multicultural, complex and dynamic context of international aeronautical communications, in which pilots and ATCOs use aviation English (AE) to interact over the radio. The International Civil Aviation Organization (ICAO) established Language Proficiency Requirements (LPRs) for pilots, ATCOs and aeronautical station operators involved in international communications as of March 2008 (ICAO, 2004a). This occupation-specific context requires that aviation professionals involved in RT communications adhere to a set of standardized expressions, known as *Air Traffic Control Phraseology*, but also to demonstrate a minimum level of *plain English*, Operational Level 4, in order to negotiate meaning while communicating through the radio in expected and unexpected situations.

[1]Although I acknowledge that in the field of English as a lingua franca (ELF) distinctions between native speakers (NSs) and non-native speakers (NNSs) are considered problematic, these terms are used in this study because the International Civil Aviation Organization (ICAO) testing policy for pilots and ATCOs clearly make a distinction between NSs and NNSs concerning formal language testing requirements.

Some of the issues that are addressed in this dissertation are evident in the two excerpts of RT communications provided below. Excerpt 1 exemplifies the standard protocol envisioned by the policy-makers who set the standards for communication in aviation, i.e., the International Civil Aviation Organization (ICAO). It shows an example of a standard communication between a pilot and a ground controller who giving him taxi instructions, and illustrates some of the linguistic features of the language used for RT communication, such as reduced vocabulary, short sentences, the deletion of determiners, auxiliary and linking verbs, subject pronouns and some prepositions (ICAO, 2010, p. 3-4):

Excerpt 1

PILOT: Georgetown Ground, Fastair 345 Heavy. Request taxi. Information Charlie.

ATCO: Fastair 345, taxi to holding point runway 27. Give way to B747 passing left to right. QNH 1019.

(Source: ICAO, 2007, p. 4-3)

This example can be juxtaposed with a type of interaction that occurs with greater frequency than many in the field would like to admit. Excerpt 2 also shows an interaction between a pilot and a ground controller. However, in this scenario, when the pilot tries to get instructions to go to gate 8, the ground controller violates his expectancies in the sense that, by resorting to non-standard phraseology, he emphasizes in an conflictual style who is in control of the situation while, at the same time, explicitly associating his interlocutor with a negative aspect, using the pronouns “I” and “you”.

Excerpt 2:

PILOT: Ground, [ ]S6?

ATCO: [ ]S6, Kennedy.

PILOT: [ ]S006, gate 8 is becoming open.

ATCO: OK, when it becomes available and looks like….to get the chance to get in, we'll start bringing you that way. Otherwise, hold where you are for now.

(some time later)

PILOT: Ground, [ ]S006 SUPER, gate 8 is available for us.

ATCO: No, it is not. They lied to ya. So, just hold there. I'll call you when it is available. I don't need you to tell me what I can see and you can't.

PILOT: Okaaay.

(Source: https://www.youtube.com/watch?v=UCtcpXvsuiM.h)

It is important to stress that this kind of tension and friction occurs in the aviation workplace, which although not envisioned by the policy-maker, is part of the lived experience of professionals communicating via radiotelephony, even between speakers of English as a first language (L1). As a result, non-compliance with existing standards coupled with language and cultural issues can lead to misunderstandings, compromising safety.

Research in the field of pilot-ATCO communication suggests that crucial features of the aviation RT-specific construct, that is, what needs to be measured in a language proficiency test for this occupational context, may be absent in the assessment of these

professionals (e.g. Douglas, 2014; Kim, 2012, 2018; Monteiro, 2017). The fact that the construct of international RT communication might be underrepresented in the ICAO testing policy[2], may also lead to questions regarding the validity of inferences drawn from current testing practices (Messick, 1996). As a result, Kim and Elder (2015) remind us that "questions of justice may arise when the construct espoused by a particular policy, and reflected in tests used to implement this policy, fails to reflect the real-life situation or to accord with the views of relevant stakeholders" (p. 2).

Not surprisingly, an issue faced by test developers in this professional domain is the lack of a clear definition of the construct to be measured (Emery, 2014; Farris, 2016; Garcia, 2015; Moder & Halleck, 2009). This challenge is coupled with conflicting information on what should be privileged in the assessment of pilots and ATCOs, as found in the ICAO testing mandate, i.e., Annex 1 Language Proficiency Requirements (LPRs) and in the ICAO guidance material, i.e., ICAO Doc 9835 (Douglas, 2014; Farris, 2016). Fulcher and Davidson (2009) comment on the relevance of this step in test design, explaining that "the process of selection [of constructs] limits the purpose of the test, places boundaries upon the claims that can be associated with test scores, and removes design and validity chaos" (p. 127). In addition, the concept of washback, considered within the scope of the impact of tests on stakeholders (Cheng, 2005; Wall, 2005), reflects the connection of teaching, learning and assessment. Therefore, it means that if the construct of a test fails to address any important feature of the target language use (TLU) domain, most probably it will not foster a positive impact on teaching and

[2] The ICAO testing policy was introduced by Amendment 164 to the Standards and Recommended Practices (SARPs) in Annex 1 to the Convention on International Civil Aviation. It includes the ICAO Rating Scale and the Holistic Descriptors (ICAO, 2004a).

learning. Thus, as mentioned by Cheng and Fox (2013), the principle of 'what is assessed becomes what is taught' appears as a way to provoke changes in this chain of events.

The assessment of language proficiency for professional certification is an area of language assessment that has been relatively less researched (Cheng & Fox, 2013). An exception to this is the body of research on the *Occupational English Test* (OET), a specific purpose English language test for health professionals (see Elder, 2016 for an overview of current articles). As for the language assessment studies in the context of aviation English tests for pilots and ATCOs, few attempts have been made so far in order to address this gap in research "conducted using aviation professionals as informants despite the high stakes of the ICAO language proficiency requirements" (Knoch, 2014, p. 78). It's worth mentioning the work by Kim (2012, 2018) and Kim and Elder (2015), who explored test-takers' perceptions of the test administered in Korea as a way of interrogating the underlying construct of both the test and the ICAO policy. In the same way, the study carried out by Knoch (2014) explored the utility of using pilot informants in the context of post hoc validation of an aviation-related LSP rating scale. In Brazil, aiming to contribute to a revised rating scale to assess ATCOs' language proficiency, Aragão (2018) conducted a study to investigate the characteristics of effective RT communication based on the perceptions of Brazilian air traffic controllers. As a result, the Language for Specific Purpose (LSP) testing community in the context of aeronautical communications could benefit from knowing more about what aviation stakeholders and domain experts consider as central in their professional communicative practices.

In a timely manner, the proposed study will also add to current discussions among the aviation industry on the ICAO testing policy. First, it is worth mentioning the difficulty of construct re-definition when constructs are defined by policy mandates. As McNamara (2011) has stated in relation to the broader field of language testing, "once the standards as initially formulated have been adopted—they are cast in stone" (p. 44) and cannot easily be changed. This is also the case in the context in which the ICAO testing policy is embedded. A formal attempt to change this state of affairs was undertaken in the 39th ICAO Assembly in September, 2016. A working paper entitled 'The need for revision of Annex 1 Language Proficiency SARPs for pilots and air traffic controllers' (ANAC, 2016) was proposed by Brazil, one of the ICAO Contracting States and a member of the ICAO Council, calling for a revision of the ICAO Language Proficiency SARPs and highlighting "the importance of having a clear definition of the abilities that should be assessed as well as the necessity of native speakers also being formally evaluated on their abilities to use the language effectively and to manage communication successfully" (p. 1). However, the ICAO Technical Commission "was of the view that it [the working paper] did not present sufficient evidence that existing language proficiency requirements posed a safety threat" (ICAO, 2016, Section 35.15). Yet, the fact that the construct to be measured is not clear (e.g., Emery, 2014; Moder & Halleck, 2009) to all stakeholders involved in designing and approving tests for licensing purposes and that it might as well be underrepresented in the ICAO rating scales (e.g. Douglas, 2014; Kim, 2012, 2018), might lead to questionable decisions made on the basis of test scores with serious consequences. Added to that, the fact that not all participants involved in aeronautical RT communications, i.e., native or expert speakers of English, are required

to be tested (ICAO, 2004a) and also that fatal accidents happened in the past due to factors that go beyond language proficiency alone (e.g. Helmreich & Merritt, 1998), may be an indication that the ICAO rating scales and testing policy need to be reconsidered to address all the competencies required for effective and safe communications. Therefore, this study is an attempt to draw on current research and generate empirical data in order to address the issues highlighted by ANAC.

Second, an international conference organized by the International Civil Aviation English Association (ICAEA) took place in Croatia, in April 2017, under the theme 'The ICAO LPRs – 10 Years on: Progress or Pain?', in which a number of workshops elicited participants' reflections and perceptions on the following main topics: (a) Have LPR tests been doing their job?; (b) What have the washback effects of tests been on training?; (c) How have the LPRs affected perceptions and attitudes?; and (d) Have air-ground communications improved? Although those questions will not be directly addressed in this project, they reflect similar concerns to the ones that triggered the development of this research study. The outcome of this ICAEA conference was a call for the harmonization of LPRs internationally, in terms of clearer guidance to allow effective comparisons to be made between tests around the world. In response to that, ICAEA started a project called *ICAO LPR Test Design Guidelines, to assist "*Civil Aviation Authorities and organizations involved in the design of LPR tests recognize and understand key issues related to the design of tests and their impact on overall testing practices" (ICAEA, n.d.).

Third, aiming to investigate the perceptions of Brazilian pilots on the ICAO testing policy and on the Brazilian test, i.e., *Santos Dumont English Assessment* (SDEA),

in March 2017 ANAC sent a survey to more than 5000 pilots, receiving a total of 1172 complete responses. The results indicate a call for the development of new test tasks that better represent the construct of international RT communications (Garcia, 2017). Therefore, a research study to underpin this endeavor and improve testing practices in the Brazilian context is well timed.

1.2 The researcher standpoint

My interest in this particular subject comes from my professional experience as a civil aviation regulation specialist at the Civil Aviation National Agency (ANAC – Brazil) and as an Applied Linguist. Having worked with the implementation of the LPRs for 15 years, and more specifically with the development of the language proficiency test for Brazilian pilots – *Santos Dumont English Assessment* (SDEA), questions arising from my own practice and from my sense as an AE test developer prompted me to conduct this study in the context of the specific language assessment of pilots and ATCOs.

In order to qualify to work in such a specific field, I took several courses which helped me to understand the complexity of international RT communications and some of the communicative demands of pilots and ATCOs in their occupational domain. These comprise: (a) International Air Traffic (including RT phraseology) – first, at VARIG Training Center, and later a second course at Galeão Air Force Base, both in Rio de Janeiro, Brazil; (b) Instructor development: Aviation English – Federal Aviation Administration (FAA) Academy, Oklahoma, US; (c) Course for English Language Examiners using ICAO language descriptors – Mayflower College, Plymouth, UK ; (d) Applying the ICAO Assessment Criteria – Oxford Aviation Academy, Oxford, UK; (e) Crew Resource Management – Department of Civil Aviation (DAC), Rio de Janeiro,

Brazil; among others. In addition, I participated in a joint simulator session in a B737 full-flight simulator at VARIG Training Center, in Rio de Janeiro, Brazil, in which the objective of the training was to familiarize an ATCO with the realities of the pilots inside the cockpit. I also had the opportunity to talk to several ATCOs in my visits to control towers and approach control facilities, as well as to a great number of pilots that I had the opportunity to work with or to examine.

Furthermore, I attended and presented in a number of international events related to the implementation of the LPRs, most of them hosted by ICAO, and others related to challenges in the testing and training of pilots and ATCOs, organized by ICAEA. Since 2013, I have been an active member of the ICAEA Board and contributed to the association projects in a variety of ways, including the ICAEA Research Group. I am also part of an interinstitutional research group named GEIA – *Grupo de Estudos em Inglês Aeronáutico* (Aviation English Research Group), linked to the Institute of Airspace Control, in Brazil. Research being conducted in these two groups include, but are not limited to: (a) LSP training and testing; (b) Aviation English as lingua franca; (c) Intercultural communication; (d) Corpus Linguistics; and (e) Description and analysis of the language within the Aviation English environment.

In addition, in my Master's thesis I focused my research on the investigation of "possible threats to the oral comprehension related to the use of the English language by Brazilian pilots and controllers in a multicultural context, in which subjects with different language abilities interact" (Monteiro, 2009). On that occasion, I proposed a taxonomy of linguistic, discursive-interactional and intercultural factors that might lead to misunderstandings in radiotelephony communications. Therefore, given the present

scholarly debates on the use and assessment of ELF in multicultural contexts, I was also motivated to further explore the intercultural factors I had previously identified in my research.

1.3 The research focus and scope

Responding to these needs, the present study addresses a combination of research-based problems or needs and real-life problems or difficulties in the high-stakes context of the specific purpose English assessment of pilots and ATCOs. First, it focuses on the investigation of LSP testing for professional certification and engages experienced aviation stakeholders in important steps throughout the test development process, cited as less researched areas by Cheng and Fox (2013) and Douglas (2014), respectively. Second, it explores a multicultural occupational context that requires a clear definition of the test construct, i.e., what needs to be measured, and also the design of test tasks that operationalize these constructs, in order to draw valid inferences based on test results. Thus, the main objective of this study is to increase the safety of pilot-ATCO intercultural radiotelephony through the exploration of the awareness, knowledge, skills, and attitudes required for effective communication, relying on the perceptions of a range of international stakeholders. Based on a social-construction perspective (e.g., Berger & Luckmann, 1966; Hall, 1999; Kramsch, 1986) this project aims to increase the validity drawn from the results of specific purpose aviation English tests through the design and pilot testing of tasks that operationalize the identified RT construct. As a result, to address the research problems described above, the overarching research question that guides this study is as follows:

What are the communicative demands of pilots and ATCOs involved in intercultural RT communications that go beyond language proficiency; how can they be specified within a construct framework and operationalized as test tasks?

At this point, it is useful to present the overall conceptualization of the thesis. Figure 1.1 illustrates the focus and delimits the scope of the study (included within the red boundaries), while at the same time provides an organizational framework for the direction I chose in investigating the research question.

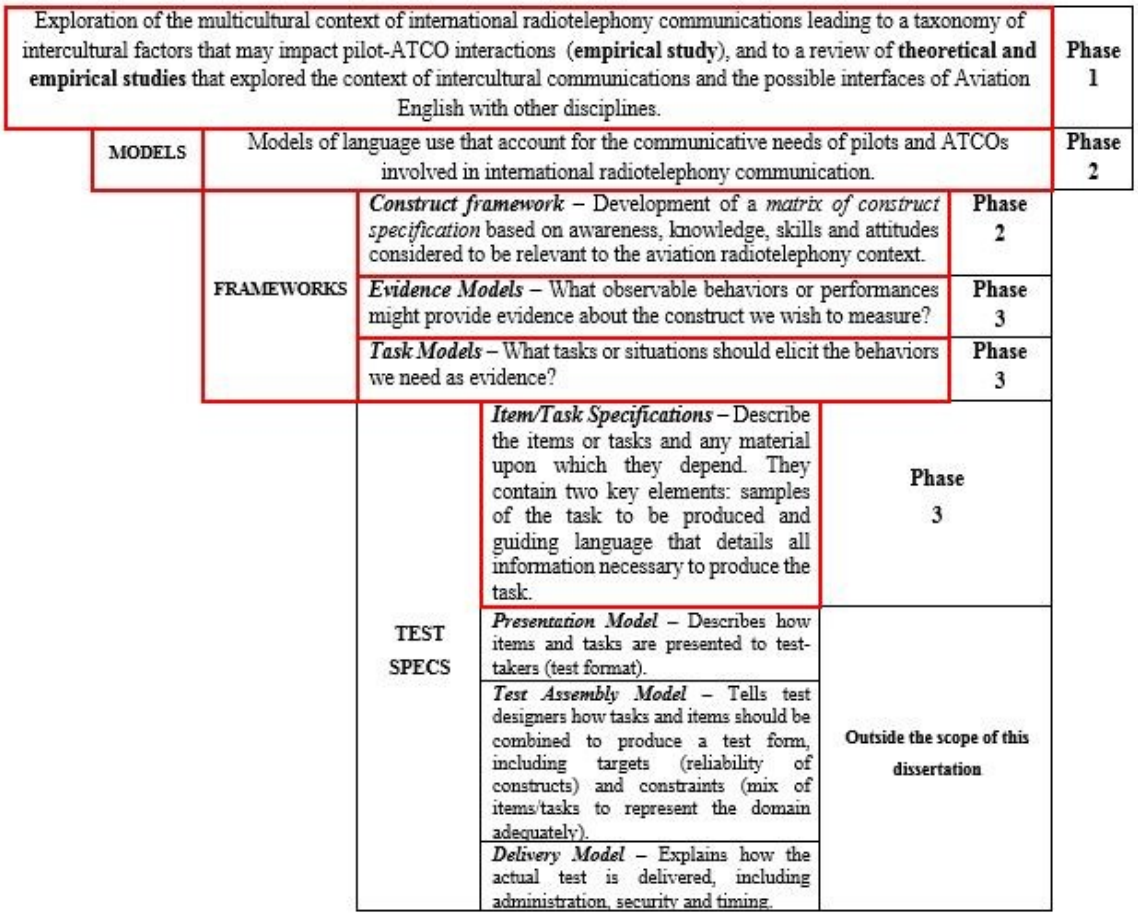


Figure 1.1 Overall conceptualization of the thesis based on the layers and sub-layers of architectural documentation (adapted by Monteiro from Fulcher and Davidson, 2007, 2009)

It is based on Fulcher and Davidson's (2007, 2009) test development framework, which includes layers and sub-layers of architectural documentation that articulate design decisions (see Chapter 3 for a detailed discussion of this framework and Chapter 6 for study design and research questions specific to each Phase of the research).

Additionally, by giving voice to domain experts from different 'linguaculture'[3] backgrounds, this study aims to address cultural diversity, include multiple viewpoints and provide a more comprehensive analysis of the complex ways to operationalize the aviation RT-specific construct in test design. The research outcomes could be used as a way to raise awareness among aviation stakeholders of the need to explore other features of communication that are relevant for successful interactions between pilots and ATCOs. Added to that, a clearer definition of the RT construct appears as fundamental to inform subsequent steps in test development and/or revision processes. The investigation of those issues through different perspectives is of paramount importance due to their impact on individuals, on teaching and learning activities and on testing policies and practices, not to mention the safety implications of this particular setting.

Having explained the relevance and motivation to undertake this research, the lens through which I will approach and interpret my data, as well as the overall conceptualization of the thesis, in the next chapter I will provide the necessary background to further contextualize the present study.

[3] The expression *linguaculture* was first used by Jenkins (2006a), in her definition of English as a lingua franca (ELF), but Baker (2009) reinforces the relevance of the term "to highlight the language-culture connection and the importance of different languages and cultures in communication" (p. 569).

Chapter 2 – Contextual Background

Having introduced the rationale for this doctoral research in Chapter 1, in this chapter I will provide background essential to understanding the context of the study. First, I will discuss some characteristics of pilot-ATCO radiotelephony communications, highlighting their complex and dynamic nature. Then, I will present the ICAO testing policy and finally, introduce the context of LSP testing of Brazilian pilots.

2.1 International RT communications in aviation

The multicultural context of international radiotelephony communications between pilots and air traffic controllers encompasses high-risk situations in which safety is paramount and communication is key. A multitude of variables related to equipment and human factors (e.g., cognitive workload, decision making, behavior, and communications) contribute to "the staggering complexity of the flight and air traffic control environment, particularly in non-routine situations" (Emery, 2014, p. 213). ICAO Doc 9835 – *Manual on the Implementation of ICAO Language Proficiency Requirements* (2010), specifies some characteristics of aeronautical RT communications that make this workplace multifaceted and even more challenging:

- they require speaking and listening skills, but not reading and writing (although data link applications will undoubtedly require these additional skills in the near future). Receptive, productive, interactive, and mediating activity (in the case of relayed messages) are all required;
- they are highly context-dependent since they rely on a great deal of specific technical knowledge related to aviation themes or topics such as aircraft, navigation, air traffic control procedures, and equipment;

- the absence of a visual/kinetic channel puts increased reliance on clear and accurate speech, since the usual conversational supports of gesture, posture, gaze, etc., are unavailable;
- the separation of speakers in space, and the resulting absence of common points of reference, mean that much more information needs to be exchanged in order to establish common ground;
- only one speaker can transmit a message at any one time. Speakers are therefore unable to interject remarks or comments that may serve to monitor effective mutual understanding;
- the acoustic conditions under which communication takes place is generally poorer than in face-to-face communications due to the narrow bandwidth which can obscure some sounds (for example "s" and "f"), background noises such as static interference or the cockpit working environment. Imperfect microphone technique on the part of speakers, who may, for example, switch their microphone on some moments after commencing a message, will "clip" part of that message. (p. 3-2)

Added to that, participants have distinct levels of language proficiency and potentially conflicting perspectives, values, beliefs, and attitudes. Moreover, they operate in busy airports and airspaces that demand expeditious communications, while at the same time engage in tasks that require a great deal of coordination and cooperation. Therefore, the complexity of this occupational context requires from pilots and ATCOs a joint cooperative effort, including distribution of knowledge and high levels of coordination with artifacts and technological tools (Hutchins, 1995a). This concept comes from the

social theory of distributed cognition, proposed by Hutchins (1995a), who investigated the activity of ship navigation and also the "distributed, socio-technical system" that exists in the cockpit of a commercial airliner (1995b). Distributed cognition also applies to the air traffic control system, where the coordination between sectors makes it possible for aircraft to depart from one airport and land safely in a different one. What these contexts have in common is that the outcomes of their tasks depend on high levels of cooperation and coordination. While interacting with ATCOs, who are operating within their complex system, pilots are also engaged in their activities within their own system, which has interfaces with the environment, hardware and software. This makes their interactions even more challenging, especially because each participant has different expectations, values, priorities, training, among other aspects. However, as Hutchins (1995a) points out, "sometimes the larger system may fail for reasons having to do with the interactions of the units rather than with any particular unit" (p. 10). Finally, I would like to mention Hutchins' (1995a) statement that "human error is inevitable in human systems" (p. 272), which is very true for aviation and specifically for pilot-ATCO communication, still highly dependent on the human factor.

In the ICAO *Human Factors Digest Nº 16* (ICAO 2004b), the concept of cultural interfaces in aviation is explored and characterized in the following way:

> […] as long as we stay within the bounds of our own culture, all of the advantages of cultural membership hold: Fellow members and the environment are predictable, thereby making daily routines easier and quicker. But as soon as we encounter members or artefacts (aircraft, procedures, regulations) from other cultures, these cultural efficiencies are challenged and the opposite occurs: The

> environment becomes less predictable, more uncertain, and requires more cognitive effort (ICAO, 2004b, p. 2).

This is the case with pilots and ATCOs, who belong to different professional cultures, possessing different roles, objectives and who are constantly performing their activities under distinct, and sometimes conflicting, types of pressure. Emery (2014) explains that "beyond the radio interface where the two professions meet, the similarities end; flying aircraft and controlling traffic are very different roles, and … the pilot and air traffic control officer pursue different objectives" (p. 203).

In regards to cultural interfaces and cross-cultural factors in aviation safety, Monteiro (2012) discussed two conceptual frameworks or models explored by ICAO (2004b), with a specific focus on pilot-ATCO communications: the SHEL Model and the Reason Model. The SHEL Model was introduced by ICAO (1998) as "a conceptual framework to help understand Human factors" (p. 2-1-5). It refers to an individual perspective that presents the interfaces of the Human element, i.e., *Liveware*, with the *Environment*, *Hardware*, *Software,* and other *Liveware*. The Reason Model, on the other hand, introduces a systemic perspective, to account for "several causal and contributing factors [that] converge in time and space to create a situation that is particularly vulnerable to one or more unexpected unsafe acts" (ICAO, 2004b, p. 16). In the case of intercultural communication between pilots and ATCOs using AE as a lingua franca, both models add to our understanding of the multiple cultural interfaces in this context and of the multiple factors that may cause misunderstanding over the radio.

2.2 ICAO language testing policy

Aiming to ensure efficient and safe communications, the International Civil Aviation Organization (ICAO) developed policies regarding the use of a standardized phraseology, a type of formally prescribed formulaic speech[4] for pilots and ATCOs involved in aeronautical radiotelephony communications. In order to highlight the phraseologies' level of specialization, ICAO (2010) considers "aviation language, radiotelephony language, and phraseologies as increasingly smaller subsets within the larger category of 'language' " (p. 3-2).

In more detail, Estival (2016) provides a linguistic description of Aviation English, both in terms of standard phraseology and plain English, including dialogic, syntactic, lexical and phonological levels (for more information see also ICAO Annex 10 – Vol. 2 (ICAO, 2014), ICAO Doc 4444 – Chapter 12 (ICAO, 2007) and ICAO Doc 9432 (ICAO, 2007).

In addition, ICAO (2004a) established another set of policies, namely the Language Proficiency Requirements (LPRs) for pilots, ATCOs and aeronautical station operators involved in international communications. The Standards and Recommended Practices (SARPs) in ICAO Annex 1 – *Personnel Licensing* to the Convention on International Civil Aviation were introduced by Amendment 164[5], and stipulate the following:

---

[4] A restricted or coded use of language comprising fixed standard phrases or lexical and syntactical routines, developed either by consensus for highly repetitive communications (e.g. everyday exchanges of greetings) or formally prescribed for special or professional purposes (ICAO, 2010, p. ix).

[5] ICAO SARPs became applicable on 27 November 2003. For more information, please refer to Annex 1, Chapter 1, paragraph 1.2.9, which is now in its 11th edition (ICAO, 2011).

> As of 5 March 2008, aeroplane, airship, helicopter and powered-lift pilots, air traffic controllers and aeronautical station operators shall demonstrate the ability to speak and understand **the language used for radiotelephony communications** [emphasis added] to the level specified in the language proficiency requirements in Appendix 1. (ICAO, 2010, p. 4-4)

These requirements established a minimum level of language proficiency skills in what was called 'plain language', which shall be used when standard phraseology does not suffice. It is "the spontaneous, creative and non-coded use of a given natural language" (ICAO, 2010, p. 3-5) required not only during unusual or emergency situations, but also in many routine situations, as pilots and ATCOs frequently need to negotiate a variety of matters. However, in this specific context, plain language is also "constrained by the functions and topics (aviation and non-aviation) that are required by aeronautical radiotelephony communications, as well as by specific safety-critical requirements for intelligibility, directness, appropriacy, non-ambiguity and concision" (p. 3-5).

Although the LPRs apply to all languages used in radiotelephony, either the language of the station on the ground or English, they stipulate that "English shall be made available when pilots are unable to use the language of the station on the ground" (ICAO, 2010, 4-3). Nevertheless, considering the rapid growth of international air travel and the increasing number of multicultural crews, a large proportion of radiotelephony communications take place in the English language. Despite the fact that English is also widely used among other aviation professionals, such as flight dispatchers, flight attendants, maintenance and ground service personnel, "voice-based communication between pilots and air traffic controllers, so-called *radiotelephony*, is regularly referred to

as *Aviation English*" (Bieswanger, 2016, p. 68). Moreover, the widespread focus on training and research related to pilot-ATCO communications (Moder, 2013) led a number of scholars to adopt the expression Aviation English to characterize its specific use in international radiotelephony communications (e.g., Bieswanger, 2016; Douglas, 2014; Estival, 2016; Estival & Farris, 2016; Intemann, 2008; Kim & Elder, 2009; Moder, 2013; Read & Knoch, 2009).

Bieswanger (2016) investigated the status of AE from the perspective of register research. According to Biber and Conrad (as cited in Bieswanger, 2016, p. 71), "a register is a variety associated with a particular situation of use (including particular communicative purposes)". Drawing on a framework that considers situational, linguistic and functional characteristics of registers, his analysis confirmed that 'plain Aviation English', as he defined it in order to make a distinction from plain English used in everyday conversations, and standardized phraseology are indeed two distinct specialized registers. He concluded that "they are both fundamentally different from the very general register of conversation, and they are distinct because they differ in their degree of specificity" (p. 83). This is in line with the previous ICAO (2010) assertion that, for the purposes of radiotelephony communications, pilots and ATCOs share two distinct registers of language: standardized phraseology and plain language (p. 3-6).

The assessment criteria to judge whether aviation professionals are 'safe' to fly or control air traffic in international settings were developed by ICAO and are included in the SARPs. They include the ICAO Rating scale (see Appendix A) and the holistic descriptors, and can be referred to as the testing mandate, developed as a result of international policies on language use for aviation radiotelephony. As such, they cannot

be easily changed or improved according to "updated and current theories and definitions of language knowledge as these change over time" (Shohamy, 2017), and as the ongoing and iterative nature of the test development processes requires (Cheng & Fox, 2017; Davidson & Lynch, 2002).

ICAO (2010) further clarifies that aviation personnel shall comply with the LPRs at least at the Operational Level 4 in the ICAO Rating Scale. Licensing authorities shall determine the way in which language proficiency is to be demonstrated; however, ICAO specifies that recurrent evaluations for personnel who demonstrate Operational Level 4 and Extended Level 5 shall be formal (p. 4-4). On the other hand, the assessment at Level 6 does not need to be carried out by a language testing specialist nor does it require the use of a fully developed specialized language test (p. 6-7), but rather expert speakers may be evaluated through informal assessments. Actually, Alderson (2011) comments that a simple conversation with NSs or expert speakers of English is enough in order to certify them as being at Level 6, without a formal test nor a language assessor, which "in effect, means that native speakers of English are automatically regarded as English proficient for the purposes of aviation" (p. 396).

The ICAO Rating Scale defines six levels of language proficiency ranging from Pre-elementary (Level 1) to Expert (Level 6) across six skill areas of linguistic performance: pronunciation, structure, vocabulary, fluency, comprehension and interactions. The five holistic descriptors mentioned above provide more general characteristics of proficient speakers and establish the context for communications. Specifically, they state that proficient speakers shall be able to:

1. communicate effectively in voice-only (telephone/radiotelephone) and in face-to-face situations;
2. communicate on common, concrete and work-related topics with accuracy and clarity;
3. use appropriate communicative strategies to exchange messages and to recognize and resolve misunderstandings (e.g., to check, confirm, or clarify information) in a general or work-related context;
4. handle successfully and with relative ease the linguistic challenges presented by a complication or unexpected turn of events that occurs within the context of a routine work situation or communicative task with which they are otherwise familiar; and
5. use a dialect or accent which is intelligible to the aeronautical community (see ICAO, 2010, Section 4.5.3 for additional details).

Since the adoption of the LPRs, different tests for aviation personnel have been developed in order to implement those requirements and comply with the assessment criteria designed by ICAO (ICAO, 2010). However, lack of standardization is still prevalent in this language for specific purpose (LSP) testing field, mainly due to different interpretations of the ICAO guidance material and the absence of a clearer definition of the construct to be measured. Besides that, the assessment criteria still place a great emphasis on NSs norms and on linguistic-oriented components, which do not take into consideration what domain experts value for effective communication in this occupational context (Elder, McNamara, Kim, Pill & Sato, 2017; Harding & McNamara, 2017; Kim, 2018; Kim & Elder, 2015).

2.3 LSP testing of Brazilian civil aviation pilots

As mentioned in Chapter 1, one of the objectives of this study is to contribute to the ongoing improvement of the language proficiency test for pilots in Brazil. Therefore, as Phase 3 of this research was carried out in the Brazilian context of LSP testing with Brazilian stakeholders, it is important to provide some background in relation to the test currently in use, so as to contextualize participants' accounts that will appear in Chapter 9.

In order to comply with the ICAO Language Proficiency Requirements, the Brazilian National Civil Aviation Agency - ANAC (former Department of Civil Aviation) implemented the SARPs in the national regulations and developed its own test to assess the language proficiency of civil aviation pilots. The test was named *Santos Dumont English Assessment* (SDEA)[6], in honor to Alberto Santos Dumont, the Brazilian inventor, designer, and pilot of the aircraft called *14-Bis*, the first aircraft heavier than air to fly autonomously, in 1906.

A general description of the test is provided at the organization website (ANAC, 2018), and reads as follows:

> Santos-Dumont English Assessment (SDEA) is the exam developed by the National Civil Aviation Agency (ANAC) to verify pilots' level of proficiency in English, according to the provisions of Annex 1 of the International Civil Aviation Organization (ICAO) and RBAC 61[7]. Following ICAO's recommendations, SDEA assesses candidates' ability to speak and understand the English language within work-related contexts. These include routine, unexpected and emergency situations, all of which appropriate to the operational context, which allow candidates to demonstrate their ability to interact efficiently and safely in radiotelephony communications using the English language. Thus, topics presented in all parts of the examination are related to civil aviation, and many

[6] More detailed information about the SDEA and related documents are available at http://www.anac.gov.br/en/licenses-and-language-proficiency/language-proficiency.
[7] RBAC 61 refers to the Brazilian Civil Aviation Regulation that prescribes the requirements and procedures for the issuance of pilot licenses, aircraft ratings and certificates.

> occur in the context of radiotelephony and standard air traffic phraseology. It is important to emphasize that the candidate's oral production is not assessed in terms of its technical or operational precision, since the exam aims at verifying only the candidate's proficiency in English. ICAO rating and the holistic descriptors are the basis for assessing the candidate's performance. The scale details six areas of linguistic production: pronunciation, structure, vocabulary, fluency, comprehension, and interactions. For each of these areas, the rater awards the candidate a level, from level one (pre-elementary) to level five (advanced). The final level is determined by the lowest rating assigned in any of the six language proficiency skills of the rating scale. In order to be authorized to perform international flights, the pilot must obtain at least final level four - Operational. The complete result, detailing the level reached by the pilot in each evaluated area, is sent to the candidate via e-mail by ANAC, in up to 30 working days after the exam. (https://www.anac.gov.br/en/licenses-and-language-proficiency/santos-dumont-english-assessment-instructions-for-candidates)

The current test format consists of an oral interview divided into four parts, which takes approximately 40 minutes, including versions for airplane and helicopter pilots:

1. Part I – Aviation Topics;
2. Part II – Interacting as a Pilot;
3. Part III – Unexpected Situations
4. Part IV – Picture Description and Discussion

Brazilian pilots can also find at ANAC website the Candidate's Manual[8], in which detailed information about each part of the test is provided, as well as what is expected of the test-taker in each type of task, features of the test setting, the roles of the English Language Expert (ELE) and of the Subject Matter Expert (SME), among other information, including a complete sample version[9].

[8] The Candidate's Manual (a condensed version in English) is available at http://www.anac.gov.br/en/licenses-and-language-proficiency/santos-dumont-english-assessment-instructions-for-candidates

[9] A sample version of the SDEA is available at https://www.anac.gov.br/assuntos/setor-regulado/profissionais-da-aviacao-civil/arquivos/sdeamodelo.pdf

ANAC, more specifically the Language Proficiency Group, is responsible for the development of testing materials, interlocutor and rater training, accreditation of institutions to administer the test, and continuous surveillance of the entire testing process, all leading to a language proficiency endorsement in the pilot's license. Apart from that, the ongoing nature of the test development process requires a continuous effort to monitor, collect evidence, evaluate, and improve the SDEA and all inter-related activities and processes required for the language proficiency certification of Brazilian pilots.

Having provided background to the study, in the next three chapters (Chapters 3, 4, and 5) a review of the literature that is relevant to the research questions will be presented.

Chapter 3 – Literature Review: First Layer -- Models

In order to provide a firm foundation for the research design and a rationale for the research questions, an extensive literature review has been conducted. It is presented in a way that hopefully facilitates the readers' understanding of how the overall topic chosen, namely, the test development process in the occupation-specific context of aviation RT communications, relates to a number of differing fields of inquiry. Specifically, I have organized the literature review according to the three layers of the test development process, defined as "layers of architectural documentation" by Fulcher and Davidson (2009, p. 126). They can be represented in the form of an inverted pyramid, starting with *Models*, in the higher level, moving to *Frameworks,* and then to *Test Specifications*. The authors explain that "the higher levels of the architectural documentation are generalized and can be applied across different tests, while other layers are unique to specific test purposes and contexts of test use" (p. 126).

The existing relation of test purpose, test use and validity is highlighted by Fulcher and Davidson (2009), who argue that "a critical component in any validity argument is the relationship between test purpose, test architecture, the claims that we wish to make about the meaning of test scores, and hence the use of the test for decision making" (p. 140). The authors' use of architecture as a metaphor for test development proves to be helpful in identifying the layers and sub-layers of architectural documentation that articulate design decisions. *Models*, as the authors define the first layer, provide "a theoretical overview of what we understand by what it means to know and use a language" (p. 126). The second layer, *Framework*s:

> lays out the constructs to be tested, selected from models, because they are shown to be *relevant* to the specific context in question, and *useful* in the decisions that need to be made. These all contribute to the intended effect of the test. The process of selection limits the purpose of the test, places boundaries upon the claims that can be associated with test scores, and removes design and validity chaos. (Fulcher & Davidson, 2009, p. 127)

Finally, the third layer includes *Test Specifications*, "where we find the detail that is specific to a particular test for use in the context specified in the [construct] framework" (p. 128).

Fulcher and Davidson (2007, 2009) further explore the Conceptual Assessment Framework within the methodology called Evidence Centred Design[10] (Mislevy, Almond & Lukas, 2003), and argue that its six elements or 'models' can be considered as sub-layers in their test architecture, some operating within the layer of frameworks and others within the layer of test specifications. A graphic representation, as shown in Figure 3.1, integrates and illustrates my reading of different authors' perspectives on the process of test development.

[10] Evidence-Centered Design (ECD) "is a methodology for designing assessments that underscores the central role of evidentiary reasoning in assessment design" (Mislevy et al., 2003, p. 20). Fulcher and Davidson (2007) refer to evidentiary reasoning as "a validity argument. The argument shows the reasoning that supports the inferences we make from test scores to what we claim those scores mean" (p. 64).

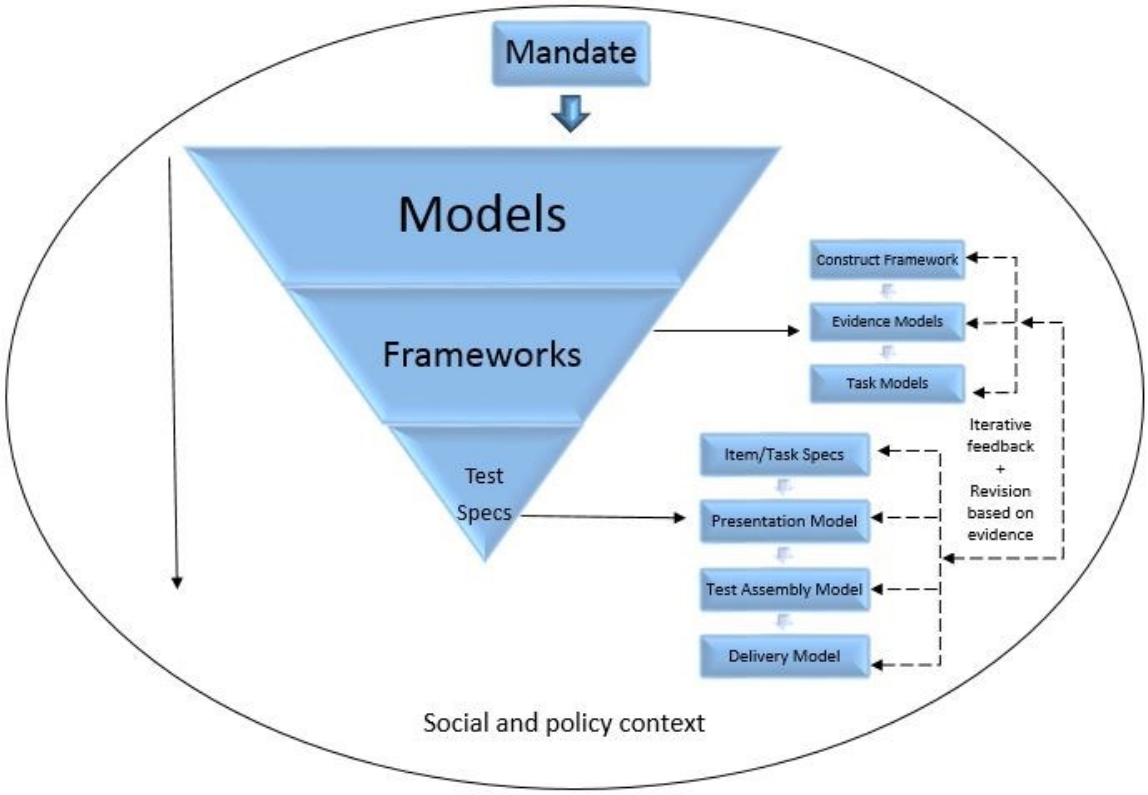


Figure 3.1. The test development process including layers and sub-layers of architecture documentation (adapted by Monteiro from Fulcher and Davidson, 2007, 2009)

First, it includes Fulcher and Davidson's (2007, 2009) understanding of how Mislevy et al.'s (2003) models operate across the layers of architectural documentation.

Second, it shows the test mandate "as the starting point for the development process" (Cheng & Fox, 2017, p.109), which is also defined by Davidson and Lynch (2002) as "the combination of forces which help to decide what will be tested and to shape the actual content of the test" (p. 77). According to Cheng and Fox (2017), a test mandate may arise internally, for example, within a program of instruction, or externally, as a result of international organizations or government policies and legislation. As such, "the mandate motivates the purpose of the test and provides parameters for the definition of useful constructs in the test" (p. 110). However, as discussed in Chapters 1 and 2, the

ICAO test mandate and related documents seem not to address the important interplay between test mandate, test purpose and construct definition. Therefore, Alderson (2007) poses the following question: "…when a mandate for an assessment procedure has been established, and it had been determined that a test or testing procedures would best address the mandate, the critical first question is 'do we know what we are measuring?'" (p. 21).

Third, Figure 3.1 also depicts the ongoing nature of the test development process based on iterative feedback for test revision (Davidson & Lynch, 2002), and on continued collection of "evidence of test function and use [which] is the essential requirement for *validity*" (Cheng & Fox, 2017, p. 109).

Finally, the consideration of the social and policy context within which test design processes are located is central to be included in this figure, circumscribing the entire process. As McNamara and Roever (2006) argued about the ECD, "Mislevy does not consider the context in which tests are commissioned and, thus, cannot problematize the determination of test constructs as a function of their role in the social and policy environment" (p.24). Therefore, McNamara (2007) further explains that an awareness of tests as "site[s] of social recognition and control" (p. 135) appears as a way to understand the values implicit in test constructs.

The representation of the test development process proposed in Figure 3.1 proves suitable to the purpose of this research study, which is to explore the construct of pilots and ATCOs' international RT communications in order to operationalize it as test tasks, moving from the highest layer of the test development process (see Fig 3.1, *Models*) to the lowest layer (*Test specifications)*.

In the following sections, I will first present a review of the literature that informs the development of models of language use relevant to the context of intercultural RT communications in aviation, as part of the first layer of architectural documentation. Then, I will provide a discussion of existing models of communicative competence to highlight how the present study adds to what is already known in the field.

3.1 Literature that informs First layer Models

In this first layer, the literature review includes previous theoretical, empirical and practical studies that explore the context of pilot-ATCO intercultural communications and the possible interfaces of Aviation English with other disciplines, in order to underpin the development of models of language use that account for the communicative demands of international RT communication. Models can be represented in different ways, depending on the purpose the researcher wants to achieve by focusing on different aspects of the communicative context.

First and foremost, it was important to review policy documents that established the language-related standards for flight crews, ATCOs, and aeronautical station operators (ICAO, 2004a, 2010, 2011), which stipulate the use of ICAO standardized phraseology specifically, clarify that both phraseology and plain language proficiency are required, and strengthen the provisions that English be made available in international operations. Added to these, studies that describe the language used for radiotelephony communications have been appraised (e.g. Bieswanger, 2016; Estival, Farris, & Molesworth, 2016; ICAO, 2007; Intemann, 2008; Moder, 2013; Read & Knoch, 2009). These studies are briefly mentioned here, as they have already been discussed in more detail in Chapter 2.

Second, radiotelephony communication problems have been considerably investigated by individual researchers and organizations (e.g. Cushing, 1995; Eurocontrol, 2006; McMillan, 1998; Morrow & Rodvold, 1998; Tajima, 2004). However, ICAO reported that a number of accidents and incidents have occurred in aviation history due to misunderstandings caused by communication breakdowns (ICAO, 2004a), despite the standardized phraseology that pilots and ATCOs must adhere to. The underlying rationale is that each professional belongs to a number of social groups or cultures, "carrying several layers of mental programming within themselves, corresponding to different levels of culture" (Hofstede, 1991, p. 10). These differences will influence what they say, how they say it, the responses they expect, and how they react to them. As the international aeronautical community is multicultural, Kim (2012) also claims that "there are different and potentially conflicting conceptions and degrees of awareness amongst its members. These differences arise from the particular experience or expertise, first languages, attitudes, and organizational culture or cultures of the smaller communities to which pilots and air traffic controllers belong" (p. 48).

Third, the challenges of communicating in such an intercultural environment are numerous. Possible threats to oral comprehension related to the use of AE by pilots and air traffic controllers range from linguistic, discursive-interactional and intercultural factors (Monteiro, 2012). As some of the categories identified by the author may be culturally influenced, such as non-cooperation, reluctance to declare emergency, inference and false suppositions, incorrect interpretation of speech acts, power distance and face-work, among others, there was a call for further research regarding the intercultural dimension.

Furthermore, in recent years a range of presentations in international conferences have addressed the topic of cross-cultural communications in aviation (e.g., Bieswanger, 2018; Kay, 2015, 2018; Lloyd Evans, 2013; Monteiro, 2013, 2014, 2017b, 2018a; Pacheco, 2018); however, the impact of cultural background on radiotelephony communications is still underestimated, requiring training in intercultural communicative strategies for both NSs and NNSs of English, pilots and ATCOs, as suggested by Douglas (2014). Douglas argues that "we have a professional/ethical responsibility to continue to study the phenomenon of aviation radiotelephony and the role of both native and non-native speakers of English in maintaining communication" (p. 10).

Therefore, considering the need to further investigate the potential threats posed by intercultural factors to the safety of international RT communications in aviation, below I provide an overview of studies in this field.

3.2 Aviation English and Intercultural Awareness/Competence

In the aviation context, a number of studies (e.g., Helmreich & Merritt, 1998) have applied Hofstede's (1991) cultural dimensions in order to understand the influences of culture on professional's behavior and communications. Hofstede's (1991) work with IBM employees in 40 countries followed a positivist paradigm and classified national cultures across four dimensions, which have been expanded and defined later by Hofstede and Hofstede (2005) in the following way:

- Individualism vs. collectivism – individual interests versus group interests;
- High vs. low power distance – relates to the acceptance of power differences in society;

- Masculinity vs. femininity – a continuum of attitudes related not only to gender roles within a culture, but also to values regarding who has a greater investment in social relations and nurturing, regardless of gender;
- High vs. low uncertainty avoidance – the extent to which the members of a culture feel threatened by ambiguous or unknown situations, i.e., (non) acceptance of ambiguity; and
- Long- vs. short-term orientation – virtues oriented towards future rewards versus virtues related to the past and present.

As observed by T. O'Brien (personal communication, August 8, 2019), Hofstede's cultural dimensions focus on acceptance of the features more than a rigid presentation of social norms. For example, in relation to power distance, Hofstede and Hofstede (2005) clarify that it refers to "*the extent to which the less powerful member of institutions* [family, school, community] *and organizations* [places where people work] *within a country* ***expect*** *and* ***accept*** [emphasis added] *that power is distributed unequally"* (p. 46). Therefore, what characterizes a high power distance culture is not the existence of inequality, but the group's compliance and acceptance of power rankings. As the authors explain, in high power distance countries, subordinates display a considerable dependence on superiors; a large emotional distance is perceived between the two, which means that subordinates most likely will not approach and contradict their superiors directly; and the hierarchical system is based on an existential inequality between superiors and subordinates. On the other hand, in low power distance countries, there is limited dependence of subordinates on superiors and a preference for consultation; a relatively small emotional distance between them, which makes it easy for subordinates

to approach and contradict their superiors; and the hierarchical system is just a matter of inequality of roles, which may be changed, as subordinates and superiors consider each other as "existentially equal" (p. 56).

Concerning the cultural dimension of masculinity/femininity, Hofstede and Hofstede (2005) further explain that in masculine cultures, male and female responsibilities are clearly distinct, in terms of being assertive vs. modest, tough vs. tender, focused on material success vs. concerned with life quality. There is also a sense of competitive authority, aiming for superiority of a single gender, and an orientation more towards tasks and accomplishments than social equality. Conversely, in feminine cultures gender roles overlap, where both are supposed to be modest, tender, and concerned with the quality of life (p. 120). In other words, "feminine cultures tend to hold the overall belief that all persons require and should provide support and nurturing to others equally, without competition or hierarchical ranking or division of roles based on gender" (T. O'Brien, personal communication, August 8, 2019).

Hofstede's research influenced a number of other fields exploring cultural differences. Noteworthy are the studies that investigated pilots' behavior inside the cockpit and others that included the impact of culture on aircraft incidents and accidents (Helmreich, 1994; Merritt & Helmreich, 1996; Helmreich & Merritt, 1998; Merritt, 2000; Hazrati, 2015; Monteiro, 2016a; Ragan, 2004). What they all have in common is that, in some way or another, they reflected upon Hofstede's cultural dimensions. As an example, Helmreich and Merritt (1998) stress that uncertainty avoidance and power distance were critical factors that contributed to the fatal disaster of Avianca flight 052 in 1989. When analyzing this accident, Ragan (2004) reports that the first officer's communication style

with the captain and the air traffic controller showed a "certain facility in speaking indirectly, non-aggressively, and politely despite the desperate urgency of the situation" (p. 59). In a discussion of the cultural aspect of language and ICAO standards related to the LPRs, Hazrati (2015) argues that "attempts should be made to include intercultural communicative competence requirements instead of the sole language proficiency criterion to ensure that the standards are properly met" (p. 250). He adds that "what is lacking from this framework is the intercultural knowledge which is complementary to appropriate meaning construction, and which proves especially important in unexpected circumstances, as shown in the past events" (p. 250).

Aiming to understand the impact of different cultural backgrounds on interpersonal behavior in the context of RT communications, an empirical study (Monteiro, 2016a) investigated intercultural encounters involving pilots and ATCOs communicating through AE. Applying a Cultural Discourse Analysis approach based on Carbaugh's[11] (2007) five modes of inquiry, Monteiro began with the theoretical mode, followed by the modes of description, interpretation, comparison, and critical evaluation of two scenarios of RT communication. She uncovered instances of culturally influenced behavior that affected pilot-ATCO communications and the different ways participants reacted in the moment of interaction. Despite the criticism of national, essentialist views of culture and of the use of dichotomous dimensions in cultural research (Baker, 2012, 2015; Holmes & Dervin, 2016), drawing on the national cultural dimensions (Hofstede,

[11] Carbaugh (1988) highlights some features or qualities captured by the concept of 'culture' explaining that it is best used in empirical studies when "it describes communication patterns of action and meaning that are deeply felt, commonly intelligible, and widely accessible, and when it explores situated contexts of use through conceptual frames, treats cultural terms as focal concerns, and exploits the benefits of comparative studies" (p. 40).

1991) and theories of intercultural communication (Gudykunst, Lee, Nishida & Ogawa, 2005) proved helpful in the exploration of culturally influenced factors that may bring additional challenges to pilot-ATCO interactions. In this study, Monteiro (2016a) argues that communicative competence is not enough for the safety of international RT exchanges over the radio and that the cultural expectations of both native and non-native speakers require consideration in such an intercultural environment.

Additionally, Mathews and Albritton (2014) point out that other aspects of human factors in aviation have been more widely addressed by the industry than language related issues, requiring further studies on factors that impact communication with expert input from applied linguists. Likewise, an independent research report (Clark, 2017) commissioned by the UK Civil Aviation Authority (CAA), mentioned issues related to language and cultural awareness and to politeness in aviation discourse. Two recommendations were clearly stated:

- Incorporate cultural factors in future research on language-related miscommunication between pilots and controllers. This could involve ethnography, questions in surveys or interviews, or some other means (p. 72); and
- Incorporate awareness of politeness markers into future research on miscommunication between pilots and controllers (p. 73).

The present research addresses the less explored issues mentioned above. To that effect, foundational studies in the field of discourse and pragmatics, theories of cross-cultural communication, frameworks of intercultural communicative competence (ICC), and

concepts of intercultural awareness (ICA) stand out as crucial to inform this investigation.

3.2.1 Theoretical concepts.

Before drawing on intercultural theories and conceptualizations, it is important to explain how culture is approached in this study. Following Kesckes (2014), culture is "neither relatively static nor ever-changing, but both" (p. 4), having *a priori* elements (ethnic or cultural marking in communicative behavior) and emergent features (co-constructed in the moment of interaction), which should be combined to approach culture in a dialectical and dynamic way. In addition, interculturality is considered here as "a phenomenon that is not only interactionally and socially constructed in the course of communication but also relies on relatively definable cultural models and norms that represent the speech communities to which the interlocutors belong" (p. 14).

Another important concept that adds to the understanding of the multicultural context under investigation is the notion of intercultural communication as a discourse approach (Scollon & Scollon, 2001):

> Each of us is simultaneously a member of many different discourse systems. We are members of a particular corporate group, a particular professional or occupational group, a generation, a gender, a region, and an ethnicity. As a result, virtually all professional communication is communication across some lines which divide us into different discourse groups or systems of discourse. (p. 3)

Moreover, in order to investigate the impact of culture on international RT communications, this study privileges the view of language as action (Austin, 1962) and communication as socially constructed (e.g., Berger & Luckmann, 1966; Hall, 1999;

Kramsch, 1986). The present investigation also relies on concepts of face, as defined by Brown and Levinson (1987), whose notion of face "is derived from that of Goffman and from the English folk term, which ties up face with notions of being embarrassed or humiliated, or 'losing face'. Thus face is something that is emotionally invested, and that can be lost, maintained, or enhanced, and must be constantly attended to in interaction" (p. 63). For example, one ATCO talks to several pilots in the same radio frequency, which may sometimes raise the anxiety levels of participants, even more of NNSs of English, who, according to Estival (2016), "are aware that ATC and other pilots are judging their performance over the radio. A number of cases of miscommunication have been attributed to this pragmatic level factor and to the fear of losing face" (p. 43). In addition, politeness strategies (Brown & Levinson, 1987) and Culpeper's (1996) theory of impoliteness are equally important. Contrary to the first, the latter includes strategies that seek to cause social disruption, through hostile communication or confrontational discourse.

Some theories of cross-cultural communications classified by Gudykunst et al. (2005) as theories of cultural variability in communication are also central to the present investigation. Ting-Toomey (2005) presents the *face-negotiation theory* and explains that face-work refers to "the specific verbal and nonverbal behaviors that we engage in to maintain or restore face loss and to uphold and honor face again", where face is related to "identity respect and other-identity consideration issues within and beyond the actual encounter episode" (p. 73). The author further discusses five thematic clusters of face-work: (a) face orientations/concern, (b) face movement patterns, (c) face-work interaction strategies, (d) conflict communication styles, and (e) face content domains.

Conversational constraints theory (Kim, 2005) presents five culture-based conversational constraints. Two of them correlate to the culture-level dimension of individualism and independent self-construal (separate from others, self-sufficient): concern for clarity and concern for effectiveness. The other three correlate to collectivism and interdependent self-construal (connected with others, dependent on others): concern for minimizing imposition, concern for avoiding hurting the hearer's feelings, and concern for avoiding negative evaluation by the hearer. Next, the *expectancy violation theory* (Burgoon & Hubbard, 2005), informs that "every culture has guidelines for human conduct that carry associated anticipations for how others will behave" (p. 149), which vary significantly along Hofstede's cultural dimensions, communication styles and degrees of face concern. In many types of communication, the problems arise when those expectations are violated, because the perception of a deviant behavior as a violation is also culture-dependent, as it is the way people react to them. Concepts and propositions from the *anxiety/uncertainty management theory* (Gudykunst, 2005) also aid understanding to pilot-ATCOs intercultural communications, because although pilots and controllers communicate over the radio, other pilots in the same frequency listen to all interactions in order to maintain situational awareness. The author states that "intergroup interactions have the potential to create intense social anxiety because we do not want to appear prejudiced or perceived as incompetent communicators" (p. 287). Moreover, the *communication accommodation theory* (Gallois, Ogay & Giles, 2005) provides a useful concept in the context of this study, especially what concerns the cognitive function of convergence, which involves "speakers organizing their output to take account of the requirements of listeners, and hence facilitating comprehension" (p. 125).

Further, a review of frameworks of intercultural communicative competence (ICC) (e.g., Byram, 1997; Deardorff, 2006; Fantini, 2000) has been carried out. Byram (1997) explains what ICC entails in the following way:

> … someone with Intercultural *Communicative* Competence is able to interact with people from another country or culture in a foreign language. They are able to negotiate a mode of communication and interaction which is satisfactory to themselves and the other and they are able to act as mediator between people of different cultural origins. (p. 71)

The author privileges the inclusion of the intercultural dimension to the traditional models of communicative competence. His proposed model of ICC encompasses five components or *savoirs*: *savoir être* (intercultural attitudes), *savoirs* (knowledge), *savoir comprendre* (skills of interpreting and relating), *savoir aprendre/faire* (skills of discovery and interaction) and *savoir s'engager* (critical cultural awareness). Figure 3.2 portrays these five components, underscoring the central role of critical cultural awareness.

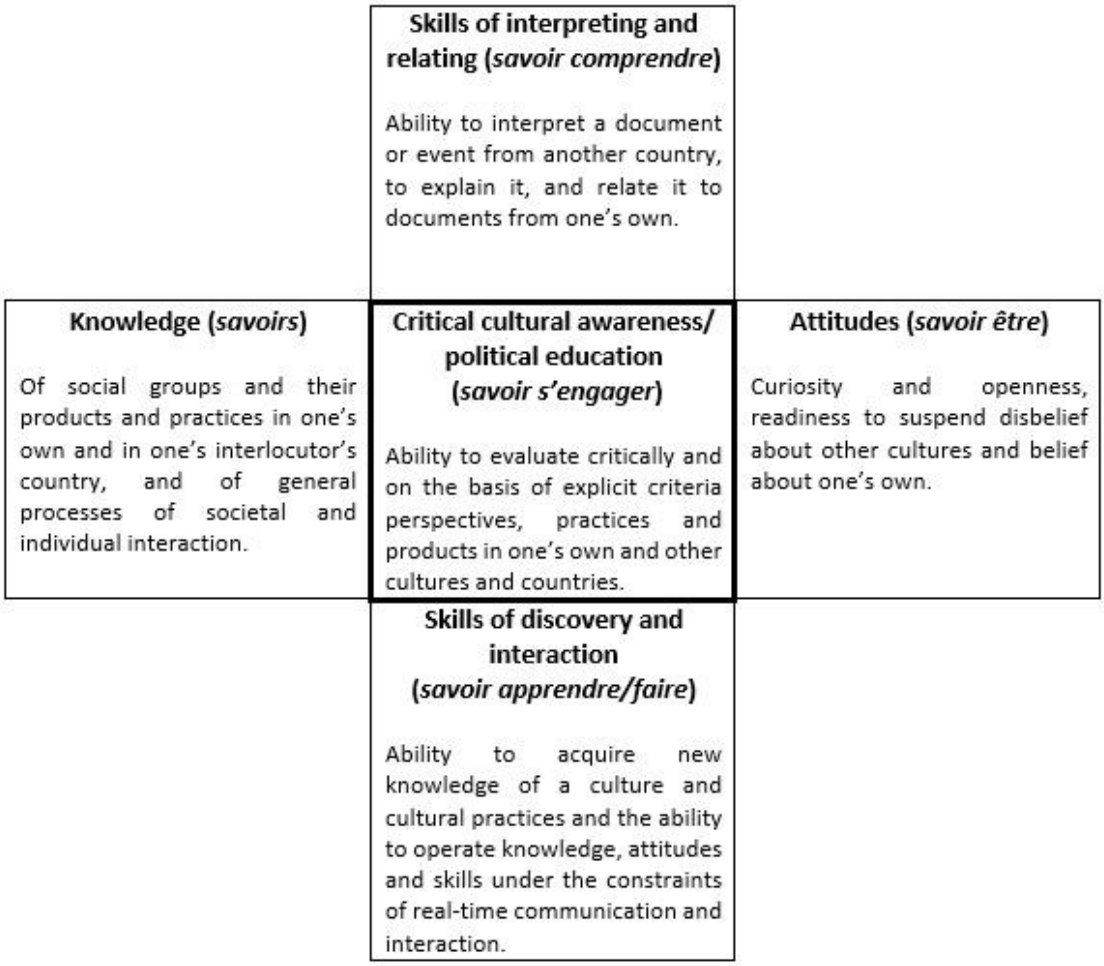


Figure 3.2. Byram's model of Intercultural Communicative Competence

(Adapted from Byram, 1997)

The Process Model of Intercultural Competence proposed by Deardorff (2006, 2009) begins with attitudes, from the individual level – attitudes, knowledge, comprehension (awareness), and skills – to achieve the desired internal outcome (i.e., adaptability, flexibility, ethno relative view, empathy), moving to the interaction level, to achieve the desired external outcome (i.e., effective and appropriate communication and behavior in an intercultural situation).

Additionally, Fantini (2000) considers ICC as a complex phenomenon with multiple components:

- three domains of ability: to develop and maintain relationships; to communicate effectively and appropriately with minimal loss or distortion; and to attain compliance and obtain cooperation with others;
- a variety of traits: respect, empathy, flexibility, patience, curiosity, openness, a willingness to suspend judgement, among others;
- four dimensions: awareness, attitude, skills, and knowledge;
- proficiency in the host language; and
- may be viewed as a developmental process. (p. 27-29)

The relevance of these models to the present study lies in the fact that they will contribute to the design of models of language use and to the specification of the construct appropriate to the context of intercultural RT communications in aviation.

Baker (2011, 2012, 2016) adds to this discussion and recognizes that cultural awareness (CA), a fundamental concept in Byram's (1997) framework of ICC, is important for preparing language learners to participate in intercultural communications by developing an awareness of cultural differences. However, as it is still rooted in predefined target communities or cultures, he acknowledges that ICC also has some limitations, especially in global lingua franca settings, where the notions of heterogeneity, fluidity and dynamism most likely shape cultural influences. Therefore, Baker (2011) argues that what is needed in lingua franca contexts is not just CA. He proposes the notion of intercultural awareness (ICA) as an expanded and dynamic framework for intercultural competence. He defines ICA as "a conscious understanding of the role culturally based forms, practices and frames of reference can have in intercultural communication, and an ability to put these conceptions into practice in a flexible and

context specific manner in real time communications" (p. 202). In addition, the author identifies several components of ICA, including knowledge, skills and attitudes that are organized in three consecutive levels: (a) level 1 refers to basic cultural awareness, (b) level 2 refers to advanced cultural awareness, and (c) level 3 is related to intercultural awareness, in which "the ability to mediate and negotiate is combined with an awareness of the emergent nature of cultural forms, references and practices in intercultural communication" (p. 205).

Thus, the goal for aviation personnel involved in intercultural communications in a lingua franca context would be to begin with the development of basic cultural awareness and then keep moving towards the more comprehensive notion of intercultural awareness. Although Baker (2016) explains that this more sophisticated characterization of ICA is likely to be acquired only through an extensive experience of intercultural communication and/or appropriate education practices, he stresses the importance of developing "an awareness of how the cultural background of participants and cultural contexts can influence communication in complex and multifarious ways" (p. 83). Hazrati (2015) corroborates this view and emphasizes that "intercultural awareness and intercultural communicative competence also need to be enhanced in aviation frontline personnel including pilots and air traffic controllers" (p. 250), in order to prevent tragedies caused by a single misunderstanding. Hence, this study proposes to raise pilots and ATCOs' awareness of the intercultural factors that may impact their communications, thus contributing to the development of their intercultural awareness/competence.

One might argue that as international radiotelephony communications rely on a set of standardized phraseologies and fixed procedures, they would be minimally affected by cultural interference. However, Baker (2017) emphasizes that "whether judged important or not, culture and identity are always present" (p. 27). Therefore, it is paramount that aviation professionals become aware of the fact that there is no such thing as neutral communication, for "all communication, intercultural or otherwise, involves participants whose identities will be present in the interaction in one way or another" (p. 27).

In contexts where there are likely to be disparities in linguistic proficiency and shared frames of reference among participants, as is the case of AE as a lingua franca, Negotiation is not an option, but a top priority (Zhu, 2015). In fact, the ability to negotiate was included in the notion of intercultural awareness, as described above. Zhu (2015) defines Negotiation, with a capital 'N', as "the key to the process whereby participants adjust their (cultural) ways of speaking, apply and refine their cultural schemata, and orient to, assign, or reject social, cultural, or situational categorizations" (p. 64). In addition, he asserts that Negotiation is "the most important means of engagement in intercultural and lingua franca communication in which participants work collaboratively towards making sense of ongoing interactions and making contributions" (p. 64).

At the interface of intercultural communication and ELF studies, Baker (2016) explains that ELF research has adopted postmodernist approaches to the relationship between language and culture, involving fluid, dynamic and multiple viewpoints. Yet the author acknowledges that tensions exist between this approach and national, essentialist positions and argues that "indeed, the tension between normativity and creativity, fixity

and fluidity has been a major part of the discussion of ELF communication" (p. 76). Nevertheless, in relation to research conducted in the fields of ELF and intercultural communication, Baker (2017) observes that although "ELF is deeply intercultural both as a means of communication and as a research field" (p. 25), little crossover has been noticed between the two.

The interfaces of Aviation English and intercultural communication, as indicated above, suggest points of contact with other disciplines, such as English as a Lingua Franca (ELF) and interactional competence (IC). Although ICAO alludes to these concepts and related communicative strategies in the LPRs guidance material (2010), the testing policy reveals its contradictory nature, as it exempts NSs or expert speakers from being formally assessed on these strategies (Douglas, 2014; Farris, 2016). Therefore, an additional group of studies have been selected for review in order to better understand the communicative demands of the international RT context.

3.3 Aviation English and English as a Lingua Franca

Considering the use of AE in cross-cultural and multi-lingual environments, Kim and Elder (2009) state that "it may be more helpful to think of aviation English as a lingua franca than as a restricted specific purpose code" (p. 14). The authors explain that "characterizing English as a lingua franca in radiotelephony implies that what is critical is not so much mastery of the English language . . . but also interactional competence whereby users' responsibilities for communication are shared across participants" (p. 14). However, I argue in this study that ELF is only part of the broader ESP construct of international RT communications in aviation, which includes a range of competencies

relevant for effective interactions over the radio in this intercultural and professional-specific context (see Section 3.5 below for further discussion of this view).

ICAO (2010) refers to the concept of lingua franca in the LPRs guidance material, and seems to recognize its applicability to aeronautical radiotelephony communications: "We are now seeing the emergence of English as an international language (EIL) or lingua franca, which sets its own standards of proficiency to ensure mutual understanding between multi-cultural users with different levels of proficiency" (p. 2-6). However, the ICAO Standards and Recommended Practices (SARPs) related to the LPRs treat NSs and NNSs of English differently and state that "formal evaluation is not required for applicants who demonstrate expert language proficiency, e.g., native and very proficient non-native speakers with a dialect or accent intelligible to the international aeronautical community" (ICAO, 2010, p. A-2). This apparent mismatch between ICAO guidance material and the actual requirements (SARPs in ICAO Annex 1 to the Convention on International Civil Aviation) calls for a revision of the latter, in order to include "features of ELF and the testing of all international pilots and air traffic controllers" (Douglas, 2014, p. 5). The identification of these ELF features and other communicative strategies that are relevant to international RT communications is one of the goals of the present research.

Studies in the field of ELF have grown considerably in the last decade since the publication of the first seminal works (Jenkins, 2000; Seildhofer, 2001). Definitions of ELF in its 'purest form' consider that members of different languages or cultures use English as a foreign language, presupposing no NSs (Seildhofer, 2004). Nonetheless, the author highlights that ELF interactions often also include interlocutors who speak English

as a first or second language. This is also true in the context of AE as a lingua franca, where communications between pilots and ATCOs involve NS-NS, NS-NNS and NNS-NNS of English. Thus, an expanded definition of ELF that does not exclude NSs of English seems more appropriate, as the one proposed by Jenkins, Cogo and Dewey (2011): "an additionally acquired language system which serves as a common means of communication for speakers of different first languages" (p. 283). In this sense, NSs of English should also acquire AE as an additional language system (Bieswanger, 2016; Estival, 2016; Intemann, 2008). As Bieswanger (2016) explained, when referring to plain Aviation English and standardized phraseology, "neither of these specialized registers is among the many registers native speakers acquire 'automatically' without any extra effort" (p. 83).

Estival and Farris (2016) make a distinction between "English as a lingua franca (ELF) and Aviation English, which is a lingua franca and a variety of English, but is not ELF" (p. 16). They explain that ELF represents a much broader construct and includes a wider variety of contexts, situations and speakers. In addition to aviation, other specific domains where English is used for specific purposes (ESP) have also engaged in ELF research, including business, education (both school and university settings), tourism, politics, technology and the media (Jenkins, Cogo & Dewey, 2011).

According to Seidlhofer (2009), the concept of 'communities of practice', referred to by Lave and Wanger (1991) as a theory of situated learning and defined as "a set of relations among persons, activity, and the world, over time and in relation with other tangential and overlapping communities of practice" (p. 98), is well suited to describe ELF interactions. Evoking characteristics of communities of practice, Seidlhofer (2009)

states that “interactants can best carry out their joint enterprise by mutual engagement through a shared repertoire of negotiable resources, namely by communicating via ELF” (p. 210). Considering that members of these communities share a domain, function as a community, engage in collaborative activities and are practitioners in their field, we can say that pilots and ATCOs belong to communities of practice. As Kim (2012) confirms, “participants in international aviation have multiple memberships in several different but interrelated communities of practice” (p. 45).

Baker (2017) reports that in his analysis of different communicative events, he noticed “ELF users drawing on multiple cultural frames of reference in the same conversation, and moving between and across local, national and global contexts in dynamic ways” (p. 29), emphasizing not only the critical role of adaptation, negotiation and co-construction, but also the close links between ELF and intercultural communication.

Likewise, Cogo and House (2017) explain that recent studies in the pragmatics of ELF explore the intercultural nature of lingua franca communication, “where more attention is dedicated to working towards pragmatic understanding, focusing on how speakers construct and negotiate understanding and how they solve miscommunication problems” (p. 210). The authors highlight the important role of a number of strategies, such as comprehension checks, paraphrasing, self-initiated repair, repetition, rephrasing, in engaging both speakers “in a constant joint effort” (p. 213).

As discussed in Section 3.2 and in this section, ELF has points of contact with intercultural communication research (Baker, 2012; 2015; 2016; 2017), as does Aviation English. Both fields of inquiry and research related to international RT communications (e.g., Douglas, 2014; Hazrati, 2015; Kim, 2012; Monteiro, 2016a) have demonstrated that

successful communication is dependent upon a number of competencies other than linguistic features alone. On top of that, the concepts of negotiation, adaptation, co-construction, and joint effort underline the existing connections between ELF and the use of interactional strategies. These accommodative behaviors will be further explored in the next section, in a discussion of the role of interactional competence in pilot-ATCO international radiotelephony communications.

3.4 Aviation English and Interactional Competence

According to Young (2011), interactional competence (IC) builds on theories of communicative competence that preceded it, which were considered characteristics of a single individual. However, he argues that IC is a different notion, mainly for two reasons. First, because it adds additional components to the ones of communicative competence, such as: (a) identity resources, related to participation frameworks; (b) linguistic resources, such as register and modes of meaning; and (c) interactional resources, encompassing speech acts, turn-taking, repair, and boundaries (p. 429). Second, because the knowledge and use of these resources depend on what other participants do, varying in different practices. He argues that "the most fundamental difference between interactional and communicative competence is that IC is not what a person *knows*, it is what a person *does* together with others" (p. 430).

Young (2011) further acknowledges that Kramsch's (1986) view of interactional competence is another clear sign of its departure from previous theories of competence. Kramsch (1986) states that "successful interactions presuppose not only a shared knowledge of the world, the reference to a common external context of communication, but also the construction of a shared internal context or 'sphere of inter-subjectivity' that

is built through the collaborative efforts of the interactional partners" (p. 367). Thus, in addition to following rules and conventions, a lot of negotiation is necessary in intercultural and lingua franca communication, as discussed above. In this regard, Kramsch (1986) mentions that "communication is not one-way, not the sound of one hand clapping, but a two-way negotiative effort" (p. 368), or as Hall (1999) described our involvement in interactive practices, "a movement between the two, a dialogue" (p. 143). The author explains that "all of our [interactive] practices are sociocultural constructions, developed, maintained, and modified by the members of the groups to which we belong as we together engage in these practices" (p. 139).

In their discussion of Interactional Competence, Plough, Banerjee and Iwashita (2018) highlight the "inherent effects of context (which includes content knowledge and an understanding of appropriate norms and conventions of a particular context)" (p. 429) as central to a speaking construct. Drawing a comparison with the concept of pragmatics, the authors emphasize that although "both integrate various competences … for meaningful and purposeful communications … IC is *necessarily* about building and maintaining relationships, an aspect of the co-constructed nature of speech" (p. 442)

Taking into consideration the key concepts of inter-subjectivity, collaborative efforts, negotiation, dialogue, sociocultural construction, and shared responsibility, the interfaces of interactional competence and the use of AE as a lingua franca are becoming apparent. Kim and Elder (2009) also recognize that, in this particular context, the communicative needs of pilots and ATCO go beyond language proficiency, requiring interactional strategies and a shared responsibility for communication:

> These participants, whatever their language background, need to be able to adapt to the situation at hand and enlist a range of communicative resources to participate in and make sense of messages delivered by speakers with differing levels of English competence in situations which may range from routine to highly unpredictable. The ICAO language testing policy, on the other hand, focuses only on language proficiency, with the implication that the onus rests only on the non-native English speaking pilots and controllers to 'lift their game'. (p.14)

Nevertheless, in the implementation document (ICAO, 2010) the role of native speakers in increasing communication safety is considered fundamental and, in addition, it is argued that "native speakers of English, in particular, have an ethical obligation to increase their linguistic awareness and to take special care in the delivery of messages" (p. 5-4). ICAO (2010) mentions the need, especially for highly proficient speakers, to focus on skills of accommodation in speaking, defined as "a natural process of adapting speech habits to the constraints of the context and the perceived ability of the hearer to understand" (p. 2-6). Actually, it is possible to extract a range of communicative strategies from the ICAO implementation document (2010):

- Be aware of the challenges faced by speakers of English as a foreign language (EFL) and of an interlocutor's possible linguistic difficulties;
- develop an openness to accommodating different accents and dialects;
- replace high-risk (possibly unclear or ambiguous) features of the language to increase communicative efficiency;
- work on correcting or attenuating unintelligible features of their speech;

- avoid jargon, idioms, slang and colloquialisms whenever possible and be aware of the difficulty they may present;
- take particular care to be explicit, rather than indirect, in their communications;
- control intelligibility by moderating the rate of speech, limiting the number of pieces of information per utterance, and providing clear breaks between words and phrases;
- focus on keeping their intonation neutral and calm, admittedly difficult at busy control areas, but a good strategy to calm the language anxiety of an EFL speaker;
- aim at clarity, conciseness and correctness, both in the use of phraseology and plain language;
- acquire techniques for recognizing and negotiating communication breakdowns;
- ask for read backs and confirmation that their messages have been understood; and
- be aware of the threats presented by cross-cultural communications and attend more carefully to read backs.

On the one hand, this confirms ICAO's intention to consider NSs equally accountable for successful radiotelephony communication, while on the other hand, it reinforces the contradictory nature of the testing policy, which exempts NSs or expert speakers from being formally assessed on these strategies (Douglas, 2014; Farris, 2016).

Read and Knoch (2009) also emphasize that NSs of English communicating in ELF contexts "should eliminate idioms, cultural references and syntactic complexity from their speech in favor of a relatively plain form of language, in order to accommodate to the more limited proficiency of their L2 interlocutors" (p. 21.7). Nevertheless, this is not always the case in radiotelephony communications. As Kim (2013) pointed out, according to Korean pilots' and ATCOs' perceptions, "the tendency by some proficient speakers to deliberately use general English to show off their ability, often at the expense of efficient and effective communication" (p. 106) was considered problematic. Indeed, experts' feedback on six radiotelephony episodes confirmed NSs lack of accommodation skills: "Their lack of sensitivity in using general colloquial English when plain English was required was emphasized along with their unduly fast rate of speech and choice of words whose meanings were unlikely to be shared" (p. 107).

Of equal importance is the recognition in ELF research of the central role of communicative strategies associated with multilingual and multicultural communication (Baker, 2016). Those directly applicable to the context of AE include accommodation, preempting misunderstanding, linguistic awareness, cultural awareness, and the ability to adapt linguistic forms to the communicative needs at hand. When commenting on accommodation as a distinguishing aspect of ELF communication, Cogo and Dewey (2012) highlight the way Jenkins refers to this process in relation to both the speaker and the listener. The authors mention that "first, a speaker needs to be able to adapt and then move towards the audience of the moment; secondly, the listener needs to converge by developing greater tolerance of difference, which requires an adjustment to a listener's expectations of what is appropriate" (p. 103). The process of accommodation is also

mentioned by a number of other scholars in the field of ELF (Baker, 2012; Cogo, 2016; Jenkins, 2000; Seidlhofer, 2009; Sweeney & Zhu, 2010).

Although much is discussed about the need for NSs to accommodate to the less proficient interlocutors, NNSs should also develop and practice accommodation strategies, as they play the roles of both speakers and listeners in the interaction. In this respect, ICAO (2010) emphasizes that "native and non-native English-speaking pilots and controllers will benefit from an improved understanding of how language functions, with a focus on strategies that aid comprehension and clarity" (p. 5-4). Kim and Elder (2009) also join the call for the training of all pilots and ATCOs, whether native or non-native speakers, which should include communication strategies to facilitate successful and efficient communication. These strategies comprise "simplification of speech and avoidance of redundant information, paraphrasing of utterances when these are found to cause problems of comprehension, and more judicious deployment of available language resources, including the existing aviation phraseology repertoire" (p. 15). In addition, a deeper understanding of one's own communication style and the problems it could pose to ELF interactions would benefit not only NSs, as suggested by Sweeney and Zhu (2010), but NNSs as well. Similarly, Kim and Billington (2016) argue that the "LPRs for Pronunciation should include more specific mention of personnel having an awareness of the characteristics of their L1 phonology that may influence English pronunciation" (p. 19), which would ensure greater intelligibility and reduce the risk of miscommunication.

Furthermore, in her investigation of aspects of professional competence that contribute to effective RT communication, Kim (2018) reports on the role of background knowledge, sensitivity to each other's role and tasks, compliance with RT conventions,

ability or willingness to accommodate, and shared responsibility for communication failure. She argues that "the co-constructed nature of interactional competence is not at all reflected in the traditional linguistic-based ICAO rating scale. Interaction in the setting of air traffic control demands not just good language skills but also sufficient professional knowledge" (p. 420), a dual requirement for effective RT communication. Therefore, Kim (2018) proposes "an expanded construct of oral communication incorporating elements of professional knowledge and behaviour with a focus on interactional competence specific to this context" (p. 403).

3.5 Understanding the Interfaces: AE, ELF, ICA and IC

From the discussions in the previous sections, points of contact between conceptualizations and research in the fields of Aviation English, English as a Lingua Franca, Intercultural Awareness, and Interactional Competence were noted. Broad topics within these four domains are summarized and organized as a literature map (see Figure 3.3), so as to provide the reader with an overview of the literature that the present study builds on in relation to the first layer of architectural documentation in the test development process.

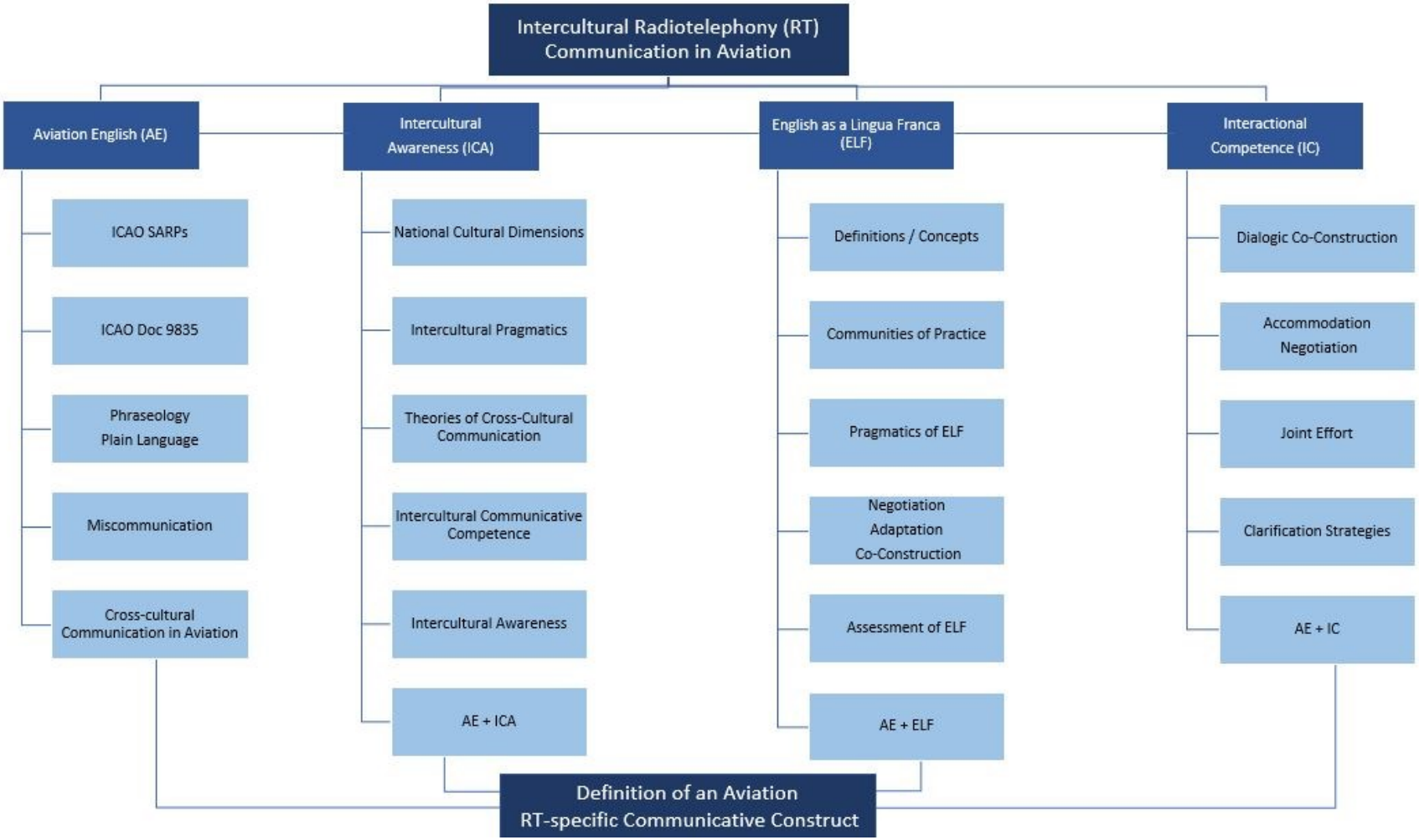


Figure 3.3. A literature map of the four domains of interest: AE, ICA, ELF, and IC

Noteworthy is the range of strategies required for successful intercultural communications using AE as a lingua franca, which go far beyond English proficiency. When referring to the diverse sociocultural settings of intercultural communication, Baker (2012) refers to the required strategies as 'the skills of multilingual communicators':

> These include the role of **accommodation** in adapting language to be closer to that of one's interlocutor in order to **aid understanding** and **solidarity**. **Negotiation** and **mediation** skills are also key, particularly between **different culturally based frames of reference**, which have the potential to cause misunderstanding or miscommunication. Such skills result in the ability of interlocutors to **adjust and align** themselves to different communicative systems and **cooperate** in communication. (p. 63, emphasis added)

Not surprisingly, some of the highlighted words in the above quotation are the same or similar to the ones mentioned in the definitions and features of interactional competence, which in turn also match the strategies required for pilots and ATCOs using AE as a lingua franca in multicultural contexts.

In this sense, the present study aims to build on conceptualizations and research in the fields of AE, ELF, ICA, and IC by disclosing core features that may contribute to effective communications in international radiotelephony. This will pave the way to the design of *Models* as a way to represent "the theoretical overview of what we understand by what it means to know and use a language" (Fulcher & Davidson, 2009, p. 126), and later to the ESP construct definition in this occupational context.

3.6 Existing Models of Communicative Competence

As the first (i.e., highest) layer of the test development process (see Figure 3.1) addresses *Models* of language use, it is important to highlight that a number of researchers have discussed conceptualizations of communicative language ability, communicative competence and related models which have heavily influenced L2 teaching and testing in the past decades (e.g., Fulcher & Davidson, 2007; Harding, 2014; McNamara, 1996). Fulcher and Davidson (2007) underscore the central role of models in considering the validity of test scores, as "from specific instances of language use we trace the meaning of the score back to its meaning in a framework (which includes contextual features), and then to a model of communicative competence" (p. 38).

Although it is beyond the scope of this study to discuss existing models of communicative competence at length, a brief summary of the most influential ones is provided in Table 3.1, for they demonstrate the evolution of what scholars considered as key in terms of communicating competently in a language use situation.

Table 3.1. Conceptualizations and models of communicative ability/competence

Author/Year	Models - Communicative language ability/communicative competence		
Chomsky (1965)	competence vs. performance ideal speaker-listener perfect competence homogeneous speech community independence of socio-cultural features		
Hymes (1972)	differential competence heterogeneous speech community socio-cultural factors		
	competence	knowledge ability for use	
Canale and Swain (1980)	communicative competence (knowledge)	grammatical competence sociolinguistic competence strategic competence	
	actual communication	the demonstration of knowledge in actual language performance	
Canale (1983)	communicative competence (knowledge and skill)	grammatical competence sociolinguistic competence strategic competence discourse competence	
	actual communication	instances of language use	
Bachman (1990)	communicative language ability	language competence strategic competence psychophysiological mechanisms knowledge structures	
	language competence	organizational competence	grammatical competence textual competence
		pragmatic competence	illocutionary competence sociolinguistic competence
Bachman and Palmer (1996)	language use	areas of language ability	language knowledge strategic competence or metacognitive strategies
		topical knowledge affective schemata characteristics of language use situation or test task	
	language knowledge	organizational knowledge	grammatical knowledge textual knowledge
		pragmatic knowledge	functional knowledge sociolinguistic knowledge
Celce-Murcia (1995)	communicative competence	discourse competence linguistic competence actional competence socio-cultural competence strategic competence	

However, in contexts where participants use English as a lingua franca in intercultural communication, emphasis on linguistic components and grammatical accuracy seems not to reflect the communicative demands of the target language use (TLU) domain, nor the appropriate assessment criteria to judge performance. As an example, Alptekin (2002) questions the model whose focus is on the notion of idealized native speaker-listener by arguing that "the conventional model of communicative competence, with its strict adherence to native speaker norms within the target language culture, would appear to be invalid in accounting for learning and using an international language in cross-cultural settings" (p. 63).

Moreover, Elder et al. (2017) observe that in spite of the recognition of theories of interactional competence (Kramsch, 1986), co-constructed nature of communication (Jacoby & Ochs, 1995), the social dimension of proficiency (McNamara & Roever, 2006), and other socio-cultural theories by the language testing field, there is still a need to move from the narrow linguistic criteria by which performance is currently judged. Added to that, Elder et al. (2017) state that theory construction related to communicative competence has not considered the perspectives of domain experts in defining the construct of communicative language ability.

These two limitations will be discussed in more detail in the next chapter (Chapter 4), which provides a review of the literature related to the second layer of the test development process – *Frameworks*.

Chapter 4 – Literature Review: Second Layer -- Frameworks

In Chapter 3, a discussion of theories, research, and practices related to the context of intercultural RT communications in aviation and the interfaces of Aviation English with other disciplines was presented. In this chapter, I will proceed with a review of the literature that highlights the links between previous research on construct definition and task design, and the current study, following the pyramid of layers and sub-layers of the test development process (see Fig. 3.1).

The second layer of architectural documentation, i.e., *Frameworks*, includes studies that inform the specification of a framework that maps the constructs to be measured, considered to be relevant to the target language use (TLU) domain and useful in the decisions that need to be made. In addition to the *Construct framework*, this layer also includes the sub-layers of *Evidence Models* and *Task Models*. Fulcher and Davidson (2006) synthesize these sub-layers in the following way: (a) *Construct framework*[12] – What are we testing?; (b) *Evidence Models* – What evidence do we need to test the construct(s)?; and (c) *Task Models* – How do we collect the evidence? (p.66-67). In line with that, when referring to a construct-centered approach to performance assessment, Messick (1994) describes a similar sequence:

> A construct-centered approach would begin by asking what complex of knowledge, skills, or other attributes should be assessed, presumably because they are tied to explicit or implicit objectives of instruction or are otherwise valued by society. Next, what behaviors or performances should reveal those

[12] In the Conceptual Assessment Framework within the ECD, this sub-layer is called *Student Model*, which describes the construct. However, Fulcher and Davidson (2009) explain that as "part of the notion of a *student model* is more associated with the practical activity of scoring and interpretation of test scores in use, rather than test design" (p. 128), they chose to use the term *construct framework*.

> constructs, and what tasks or situations should elicit those behaviors? Thus, the nature of the construct guides the selection or construction of relevant tasks as well as the rational development of construct-based scoring criteria and rubrics. (p. 16)

Additionally, Messick's (1996) notion of validity as a unified concept highlights the crucial role of the content aspect of construct validity, which details the "specification of the boundaries of the construct domain to be assessed – that is, determining the knowledge, skills, and other attributes to be revealed by the assessment tasks" (p. 10). Given that construct underrepresentation, i.e., when important dimensions of the construct are not included in a test, and construct irrelevant variance, i.e., when the test contains excess variance that is irrelevant (Messick, 1996), can both pose a serious threat to the validity of inferences drawn from tests, a thorough specification of the construct to be measured is of paramount importance.

A historical account on approaches to construct definition in language testing (Bachman, 2007) was also reviewed. Bachman's overview illustrates the "dialectic between what has been called 'trait/ability-focused' and 'task/context-focused' perspectives" (p. 42), ending with the 'interaction-focused' perspective and a discussion of some theoretical unresolved issues. The value of his article lies in his report of the implications for language assessment research and for practical assessment design, development and use. Having overlapping roles as a language testing researcher and practitioner, I agree with his view that in either role, "it is essential that we clearly define what it is we want to measure or what we want to investigate" (p. 66). Interestingly, he advocates for a pragmatist approach, both in relation to research, highlighting the need

to combine quantitative and qualitative methodologies, and also in relation to practice, stressing the need to address all three approaches (ability, task, interaction) in the design, development and use of language assessments.

McNamara and Roever (2006) discuss issues of validity and the social dimension of language tests, including value implications and social consequences to the candidates and to society. The authors emphasize that "both the construct (what we believe 'coping communicatively' in relevant settings means) and our view of the individual's standing are matters of belief and opinion, and each must be supported with reasoning and evidence before a defensible decision about the individual can be made" (p. 17). Therefore, in specific purpose language assessment, such as academic and occupational contexts, these decisions need to be aligned with the values of domain experts. As Elder, McNamara, Kim, Pill and Sato (2017) argue, "neglect of the perspective of lay (i.e., non-linguistic) judges on language and communication is a serious validity concern, since they are the ultimate arbiters of what matters for effective communication in the relevant context of language use" (p. 14). Thus, the next section expands this topic of relying on expert knowledge of domain specialists in defining constructs and assessment criteria by providing a review of studies in this field.

4.1 An ESP Perspective on Construct Definition

An ESP perspective on construct definition takes into account the 'indigenous' assessment criteria (Jacoby & McNamara, 1999), based on members of professional cultures' evaluation of the performance of novices or experienced peers, i.e., "an insider's view of the complex issues involved in communicating competently in some particular domain" (p. 214). A number of studies in different fields articulated the

construct of professional communicative competence based on 'indigenous' assessment criteria (e.g., Douglas & Myers, 2000; Elder & McNamara, 2016; Elder et al., 2017; Fox & Artemeva, 2017; Jacoby & McNamara, 1999; Knoch 2014; Pill, 2016; for a review see Knoch & Mcqueen, 2016), highlighting the importance of defining criteria that can be used to assess performance on a language test based on what domain experts consider as communicative effectiveness.

More specifically, a special issue of *Language Testing* focused on this topic, addressing questions of authenticity with particular reference to the *Occupational English Test* (OET) (Elder, 2016). Particularly useful to the present study is the paper by O'Hagan, Pill and Zhang (2016) who proposed an extension of the scope of the OET speaking assessment criteria by adding 'professionally relevant criteria' informed by the analysis of health professionals' feedback to trainees. Two new criteria were proposed: (a) Clinician engagement – "the ability to use language to demonstrate a professional manner towards and awareness of the patient" (p. 214), and (b) Management of interaction – "the ability to use language to manage the interaction and gather information from/give information to the patient effectively" (p. 214). In general terms, the authors' work offers "a promising approach to improving the alignment between specific-purpose language tests and the communicative demands of domains of interest" (p. 212).

In the specific context of the diagnostic assessment of engineering students, the development of a "disciplinary, ESP-based writing task and rubric to identify the needs of entering graduate engineering students" (Fox & Artemeva, 2017, p. 148) revealed the

central role of drawing on indigenous criteria, in this case, from the engineering community of practice.

Elder and McNamara (2016) explained that "these 'indigenous' criteria represent an articulation of the test construct and should therefore reflect what is germane to the particular domain of language use rather than general language-focused criteria familiar from other language tests" (p. 153). However, the authors underlined the difficulties involved in eliciting domain experts' opinions on what they considered relevant for successful workplace communication "in a way which emerges from contexts routinely and naturally occurring in the workplace setting" (p. 169). Their study on three different methods of investigating indigenous assessment criteria, i.e., two workplace settings with routinely occurring feedback and one artificial workshop with expert comments being elicited retrospectively, revealed that less authentic or less direct methods, such as the ones used by Douglas and Myers (2000) and Kim (2013), yielded richer insights on views of communicative success in the workplace.

In the field of aviation English, Kim (2012) reported that the 'indigenous' assessment criteria is essential to determine what really matters for communicative success, and that "linguistically oriented criteria alone cannot capture the key aspects of communication in this professional setting". These criteria also include a "thorough observance of the prescribed conventions, professional knowledge as embedded in efficient transmissions, appropriate speech rate, and strategic skills to deal with aviation personnel with different levels of expertise" (p. 229). In addition, Knock's (2014) study aimed to establish the indigenous assessment criteria that pilots use when assessing the language ability of peers and what level is sufficient as the operational level. The results

showed that the pilots focused on some but not all of the criteria on the ICAO rating scale, while at the same time focusing on the speakers' technical knowledge (p. 77). Commenting on her work, Douglas (2014) mentions that her findings also show "how difficult the pilots found it to separate language ability from aviation knowledge, thus reinforcing the view that language knowledge and background knowledge interact in communication" (p. 9). As a result, Kim and Elder (2015) mention that "questions of justice may arise when the construct espoused by a particular policy, and reflected in tests used to implement this policy, fails to reflect the real-life situation or to accord with the views of relevant stakeholders" (p. 2).

Douglas (2004) corroborates this view, emphasizing that "test developers need to learn more about the indigenous criteria experienced pilots and ATCOs use when evaluating the performance of their colleagues, so that these criteria can inform aviation English performance assessment" (p. 10). In line with that, the present study aims to follow an English for Specific Purpose (ESP) perspective on construct definition, drawing on input from different groups of stakeholders from the aviation context.

Yet, it is important to highlight some implications of adopting this approach, as discussed by Elder et al. (2017). First, it requires "the need to revise our understanding of the nature of communicative competence" (p. 19), not solely based on linguistic skills but on additional abilities that the individual uses in context. Second, the "already tenuous distinction between first- and second-language speakers … will be further reduced in scope" (p. 19), as what matters for effective communication in specific contexts will be the same, regardless of the individual's language background. Therefore, "the relevance of the native-speaker norm … and the justification for specific

tests for L2 speakers, when assessing readiness to manage the complex communicative demands of real-world encounters" (p.19) will need to be reconsidered.

In a similar fashion, when discussing emerging issues in the social dimension of language assessment, McNamara and Roever (2006) mention the status of the native speaker norm and the testing of English as a lingua franca as a topic that should be explored in future research. Given that in the aviation-specific domain there is a call for the assessment of both NSs and NNSs of English regarding the use of AE as a lingua franca (Douglas, 2014; Garcia, 2015; Kim, 2012, 2013; McNamara, 2012a), a review of studies has been carried out, both in terms of arguments for and implications of the assessment of ELF.

4.2 The Assessment of English as a Lingua Franca (ELF)

English as Lingua Franca (ELF) can be understood in at least four different ways, according to Elder and Davies (2006), including the use of English in an interaction in which:

- ELF 1 - at least some of the participants are non-native speakers (NNSs) of English;
- ELF 2 - all the participants are NNSs and do not share the same first language;
- ELF 3 - all the participants are NNSs and all share the same (or similar) first language; and
- ELF 4 - A (new) code used for interaction among NNSs, not standard English but based on standard English (SE). (p. 282-283)

The first three refer to uses of SE, relying on British or American English, while the latter does not, but includes the participation of native speakers (NSs) of English. The authors explain, citing Seidlhofer (2001), that ELF in sense (4) is influenced by "the lingua-cultural background of its speakers and the rules of use that characterize the domain within which the interventions take place (such as business, science, the media)" (p. 283). It is this definition that is of relevance for the present discussion which focuses on the assessment of pilots and ATCOs who use aviation English (AE) as a lingua franca, whose interactions are not limited to NNSs of English, but rather involve communications between NS-NS, NS-NNS and NNS-NNS of English.

Concerning the assessment of ELF and what a test of this nature would look like Elder and Davies (2006) propose two alternative models: (a) the first, based on definitions of ELF (1), (2), and (3), allows for some accommodations in the test delivery, making it accessible and fair for ELF users without changing the construct, still based on Standard English (SE) communication; and (b) the second, based on ELF definition (4), requires that ELF can be demonstrated to constitute a new code (rather than SE), and calls for a description of the language to be tested, without which assessment is not possible (p. 282).

Nevertheless, an ELF orientation to the testing of English has been the focus of debates among scholars in the field of language assessment and ELF research. On one side are those who argue for the assessment of ELF, or English as an International Language (EIL) or World Englishes (WE), supporting NNS varieties of English and additional skills for the testing of English (Canagarajah, 2006; Jenkins, 2006b; Jenkins & Leung, 2013; Harding, 2014; McNamara, 2011; 2012a; 2012b). On the other side are

those who take a more conservative position grounded in the dominant paradigms of testing and, although they appear to take a positive approach to ELF, are cautious about a quick move to the assessment of ELF before it has been properly described (Davies, 2009; Elder & Davis, 2006; Elder & Harding, 2008; Taylor, 2006).

Based on characteristics of test quality[13], although with different perspectives, both groups raise their arguments, either in favor of embracing the assessment of ELF, or emphasizing the constraints involved in so doing. It is worth mentioning that some authors, although listed in the more conservative group, put forward a number of arguments in favor of including the construct of ELF in the assessment of the English language, while others who advocate for the assessment of ELF recognize that it is a challenging endeavor for language testers.

4.2.1 Arguments for the assessment of ELF.

Firstly, it is important to underscore the changing global roles of English and, as a consequence, the changing communicative needs brought up by postmodern globalization which require language users to "shuttle between English varieties and speech communities" (Canagarajah, 2006, p. 233). McNamara (2011) recognizes that globalization has had a profound impact on the role of English, in the sense that "the emergence of English as a lingua franca as a key feature of a globalized world presents a powerful challenge to assumptions about the authority of the native speaker, an authority which is enshrined in test constructs" (p. 49). In addition, Jenkins, Cogo and Dewey (2011) describe ELF as a 'globalized and globalizing' communicative practice, in the sense that it is "simultaneously the consequence and the principal language medium of

[13] Bachman and Palmer (1996) describe five test qualities that complement one another and which good testing practice depends upon: test validity, reliability, authenticity, impact, and practicality.

globalizing processes" (p. 303). As ICAO (2010) recognizes, "most users of English will not be communicating with a native speaker of English but with another English-as-a-second-language speaker" (p. 2-6). More specifically, regarding pilot-ATCO communications, Kim and Elder (2009) state that "since English, plain or otherwise, is generally the language used and since the participants in the exchange are by no means all native speakers, it may be more helpful to think of aviation English as a lingua franca than as a restricted specific purpose code" (p. 23.14). Therefore, it does not seem to add to the validity of a test for pilots and ATCOs the use of assessment criteria based solely on NS norms (i.e., Standard English norms). As Elder and Davies (2006) asserted, using ELF norms as a basis for measurement instead of SE norms offers "more valid representations of target language use domains" (p. 296).

Moreover, Jenkins (2006b) argues that testing practice based on NS norms inhibits the teaching of accommodation strategies. She provides an example that is relevant to the present discussion, the use of idiomatic language. According to her, a speaker who employs an idiom will most likely be rewarded for his native-like production, regardless of the possible threats to the intelligibility needs of the interlocutors (p. 48). This is exactly what happens in the ICAO testing policy for expert speakers. Farris (2016) explains that the Level 6 descriptors of the rating scale not only disregard the accommodation strategies described in ICAO guidance material, but also include some contradictory behavior that could negatively affect the outcomes of pilot-ATCO interactions using AE as a lingua franca (p. 82). For example, in the ICAO (2010) rating scale, the 'Structure' descriptor mentions that "Both basic and *complex grammatical structures and sentence patterns* [emphasis added] are consistently well

controlled"; the 'Vocabulary' descriptor includes "Vocabulary is *idiomatic, nuanced, and sensitive to register* [emphasis added]"; and the descriptor for 'Fluency' states "Able to speak at length with a natural, effortless flow. Varies speech flow for *stylistic effect*…Uses appropriate *discourse markers and connectors spontaneously* [emphasis added]" (p. 4-10 to 4-12). Therefore, Jenkins and Leung (2013) conclude that testers need to move away from this narrow focus on native-like correctness and devise new approaches to assessing English, so that "they can assess whether ELF users' English is fit for ELF use, and the extent to which contingent uses of ELF in context have facilitated communication" (p. 1612).

A concern about the negative impact of the current testing policy (still based on NSs norms) on candidates and their life chances is also expressed by Jenkins and Leung (2013). Actually, Elder and Davies (2006) argue that using ELF norms as a basis for measurement would "have positive impact on test takers resulting in a reduction in anxiety on the part of ELF users, who would no longer feel pressured to adhere to norms of SE English" (p. 296). In addition, tests of ELF would also have positive *washback* on teaching, defined as "[t]he extent to which the test influences language teachers and learners to do things that they would not necessarily otherwise do" (Messick, 1996, p. 243). In relation to that, Elder and Davies further explain that "the syllabus would be designed around their [ELF users'] likely communicative needs rather than on unattainable native speaker norms" (p. 296). Nevertheless, Douglas (2004) is cautious about that and reminds us, referring to AE tests, that "the content and format of new tests will have a positive and beneficial impact on instructional programs, but only if they authentically reflect actual communication tasks in international aviation" (p. 251).

This means that if the construct of a test fails to address any important feature of the target language use (TLU) domain, it will be unlikely to foster a positive impact on teaching and learning. Thus, as mentioned by Cheng and Fox (2013), the principle of 'what is assessed becomes what is taught' appears as a way to provoke changes in this context. Regarding the current ICAO testing policy and the LPRs, Read and Knoch (2009) emphasize that "they place the onus on L2 speakers to improve their proficiency and by implication give native-speaking aviation personnel no incentive to develop their communicative competence in ELF terms" (p. 21.7).

Apart from the impact on test-takers and instructional programs, Jenkins and Leung (2013) also mention the impact of not welcoming the assessment of ELF on the English language itself, as testing based on the dominant paradigm is "preventing learners from exploiting the potential of the English language and their own resources as multilingual English speakers, and thus holding up English language change" (p. 1612). Indeed, McNamara (2011) highlights that "claims to the ownership of English and the privileging of native speaker identity inhibit the appropriation of English to the actual demands of communication in a globalized world" (p. 49). This holds true for aviation testing as well. So far, we have seen NNSs' movement towards increasing the level of their individual language proficiency, but not much effort in developing their interactional competence and intercultural awareness in order to communicate effectively in a multicultural context, not to mention the lack of effort in doing so from the majority of NSs. In this sense, Elder and Davies (2006) argue that the assessment of ELF may have a symbolic importance, in the sense that it gives greater authority and legitimacy to expanding and outer circle English speakers' voice, and also a practical

importance, as it is likely to enrich our understanding of effective intercultural communication (p. 282). Another point raised by Jenkins (2006b) is the concept of variation across Englishes, or interspeaker variation, and variation within Englishes, or intraspeaker variation. She explains that "whereas interspeaker variation concerns the influence of the wider EIL[14] social context, intraspeaker variation involves the context of the specific interaction and the way in which individuals adjust their speech to accommodate to the needs of their interlocutors" (p. 45). In a similar way, McNamara (2012b) highlights that what is required in communication in ELF is that both interactants are sensitive to the need to co-operate in the negotiation of understanding (p. 201). Canagarajah (2006) also advocates for negotiation skills, such as accommodation, and goes further in his proposal for a revision of assessment objectives and practices:

> The changing pedagogical priorities suggest that we have to move away from a reliance on discrete-item tests on formal grammatical competence and develop instruments that are sensitive to performance and pragmatics. In effect, assessment would focus on strategies of negotiation, situated performance, communicative repertoire, and language awareness. (p. 229)

Nevertheless, Canagarajah (2006) asserts that "the new paradigms of assessment are not only for multilingual speakers – they are also important to "native" speakers" (p. 241) who need to negotiate diverse varieties of English in the same way. This is in line with McNamara (2012b), who claims that "not all native speakers are equally good at ELF communication, which involves flexibility and accommodation, anticipation of communication difficulties and strategies for resolving them on the part of both

[14] EIL stands for English as an International Language.

interlocutors, regardless of their native speaker status" (p. 201). A number of studies in the field of international RT communications have shown the need for the development of accommodation strategies and interactional competence for effective communication, including the need for NSs to share the responsibility for lack of success while communicating with NNSs (Douglas, 2014; Garcia, 2015; Kim, 2012; 2013; Kim & Elder, 2009; 2015; Estival et al., 2016; Monteiro, 2016a, 2016b; Read & Knoch, 2009). As Estival et al. (2016) highlight, "the origins of miscommunication are too often attributed to pilots for whom English is not a native language (EL2 speakers)" and argue, based on results from an experimental study, that "pilots who are native English speakers commit, in some cases, as many communication errors as EL2 pilots" (p. 199). Therefore, Garcia (2015) corroborates that "native or native-like speakers of English need to be formally tested as they need to demonstrate their ability to communicate effectively on the radio. Assessing interactional competence, including awareness of cultural factors, is essential" (p. 57). Since those skills facilitate successful and efficient communication, accommodation strategies should be acquired by both NSs and NNSs of English. Thus, Estival et al. (2016) argue that they should be included in the assessment criteria for all levels of the rating scale (p. 201).

In addition, Canagarajah (2006) calls for a different attitude to proficiency, shifting the "emphasis from language as a system to language as social practice, from grammar to pragmatics, from competence to performance" (p. 234). He argues that there is no universal proficiency in English-language testing anymore, as "proficiency is the ability to use the English language effectively for specific purposes, functions and discourses in specific communities" (p. 235). Indeed, the concept of communities of

practice, from the work of Lave and Wenger (1991) on situated learning, is widely used in the context of ELF, where users of ELF do not form a single community but rather a range of communities with specific goals. Indeed, in addition to aviation, other specific domains where English is used for specific purposes (ESP) are also engaged in ELF research, including business, education (both school and university settings), tourism, politics, technology, and the media (Jenkins, Cogo & Dewey, 2011). As Hynnienem (2014) states, "the concept of community of practice thus provides us with a means to conceptualize communal practices and norms" (p. 298). Therefore, Canagarajah (2006) asserts that the current debate on which norm is to be used in assessing the English language becomes irrelevant and he is echoed by Elder and Harding (2008), who state that "rather than invoking EIL or SE as constructs we are arguing for a targeted and contextualized description of what we are attempting to measure" (p. 34.4). The authors further highlight that the definition of standards of communicative effectiveness depends on the test's purpose and that, in a particular domain, they need to be determined in consultation with the relevant stakeholders.

Furthermore, considering the two models for ELF assessment proposed by Elder and Davies (2006), the present discussion will focus on tests based on the second model as it is the one applicable to the ELF use in aviation. The authors speculate that those tests would probably include simulated interactions, integrated tasks and, for the sake of authenticity, NNSs interlocutors or paired- and group-speaking tasks where test-takers would include speakers at various levels of proficiency (p. 291). The authors argue that, differently from current tests, this would be considered essential for test validity/authenticity, because the ability to deal with differing levels of proficiency is

crucial to successful ELF interaction. As regards the criteria for assessing performance, they "would be concerned primarily with task fulfilment and with the participants' ability to accommodate to the other party and to self-repair or use other strategies to disambiguate in the event of misunderstanding" (p. 292).

4.2.2 Implications of the assessment of ELF.

From a different point of view, the second group of scholars in the field of language assessment appears to resist the move towards the assessment of ELF based on reasons that are "driven by tester's ethical responsibility to construct fair and useful tests … which impose necessary constraints to test design" (Elder & Harding, 2008, p. 34.2). As Taylor (2006) points out, testing is sometimes referred to as 'the art of the possible' (p. 58), requiring principled decisions regarding test quality and fairness. Elder & Harding (2008) stress the importance of a careful definition of the test construct, and argue that language testers have been reluctant to refer to ELF as the basis for test design due to the 'fluidity' of ELF norms (p. 34.3). Thus, they argue it would be unfair to test-takers "who would not know what kinds of language use were acceptable, what sources to draw upon in preparing themselves to pass a test, or indeed what standards were being applied by those judging their performance" (p. 34.3). In the field of AE testing, Farris (2016) comments on the implications of moving away from NSs norms and states that "ICAO's challenge of finding a viable alternative to the native speaker standard is one that is shared by the language testing research community" (p. 102).

Elder and Harding (2008) further explain, and are supported by Taylor (2006), that the emphasis in the last decades on communicative competence in language teaching has shifted the focus on knowledge and form towards a focus on function and

communication, and that language tests increasingly reflect the communicative approach (p. 52). Taylor (2006) notes that tests also rely on assessment criteria based on interactive communication and communicative effectiveness, concluding that "clearly the emphasis is on making oneself understood rather than on being 'native-like'" (p. 55). This holds true for ICAO testing policy, as we can see from a description of the scope and focus of the ICAO Rating Scale, from which two important points can be selected (ICAO, 2010):

- the ICAO Rating Scale has a distinct aeronautical radiotelephony focus: it addresses the use of language in a work-related aviation context, with voice-only communications, using strategic competences for safe communications in case of complications or unexpected turn of events, and emphasizing intelligibility in an international community of users; and
- ICAO Operational Level 4 does not target high degrees of grammatical correctness or native-like pronunciation. Grammar, syntax, vocabulary and pronunciation are judged primarily on the extent to which they do not interfere with effective oral communication. (p. 4-7)

However, what language testers are trying to argue is that as this shift is already under way, there is no urgency to change the current testing agenda.

Regarding the requirement of fairness, Elder and Harding (2008) mention the case of NNSs accents and the possible bias against language groups, due to the unfeasible task of sampling all of them to be fair with all test-takers. Therefore, testers' choice of standard varieties is based on SE neutrality. Yet, the authors recognize (based

on Harding's (2008) research findings) that this is an excessive conservative position, and that it is possible to propose approaches "to using speakers with L2 accents in listening assessment without compromising test designers' fundamental concerns for construct validity and fairness" (p. 34.7). Indeed, we find in the ICAO (2010) rating scale an explicit reference to different accents and speech varieties, as in the descriptor for 'Comprehension' at the Operational Level 4: "Comprehension is mostly accurate on common, concrete and work-related topics *when the accent or variety used is sufficiently intelligible for an international community of users* [emphasis added]"; and at the Extended Level 5: "Is able to comprehend *a range of speech varieties (dialect and/or accent) or registers* [emphasis added]" (p. 4-13). As a result, pilots and ATCOs need to be exposed to a variety of different accents.

Taylor (2006) makes her point regarding extra constraints facing test producers, explaining how they shape their response to English language variability. First, she mentions that "demands of validity and reliability mean that a test must include as wide and as appropriate a range of content as possible but without significantly disadvantaging any particular candidate group" (p. 57). Second, as regards the attitudes, needs and expectations of learners and teachers, most of the time they show a preference for 'native-like' varieties (British or American English), for learning and assessment (Elder & Davis, 2006; Elder & Harding, 2008; Taylor, 2006). Yet, Taylor (2006) further explains that "where regional varieties are perceived as useful by test stakeholders, a stronger case can be made for including them in assessment procedures; impact studies play a key role in enabling test producers to consult end-users about the types of English relevant to their needs" (p. 58). Nevertheless, this seems not to be the case in AE testing,

at least according to Kim's (2012) findings on Korean aviation experts' perspectives: "the policy unfairly targeted NNESs and overlooked the fact that NES members of the aeronautical community often do not adhere to the requirement to use prescribed phraseology and 'plain' English in routine and abnormal situations" (p. 221). In addition, when responding to Jenkins (2006b) about issues of washback and impact, Taylor (2006) emphasizes that it is more complex than just a direct linear relationship between teaching and testing, and states that "it may be naïve to hope that 'a substantial overhaul of English language testing' would bring about the desired changes in teaching/learning attitudes and practice" (p. 54). As a final comment, Taylor mentions that teaching and testing rely on well-described models of language use, but reminds us that ELF scholars acknowledge that a description of ELF communication is still in its early stages (p. 58).

Moreover, engagement in language testing research, as this second group argues, reflects language testers' awareness of sociolinguistic variation, its implications for test design, and provide a better understanding of ELF communication. According to Elder and Harding (2008) this research includes[15]: the use of NNS accents in listening test input, rater behavior across NS and NNS groups, assessment of intercultural communication skills in paired- and group-speaking tasks, and rating scale development including features of collaborative interaction (p. 34.6-34.8). Elder and Davies (2006) add to this list research endeavors[16] conducted on the assessment of individual competence when communication is co-constructed, as well as the use of indigenous assessment criteria in ESP tests and professional communication contexts. It appears that

[15] For a complete reference of the authors engaged in each field of research, see Elder and Harding (2008).
[16] For a complete reference of the authors engaged in each field of research, see Elder and Davies (2006).

findings from these contributions are already being put into practice in current tests and therefore would not require a change in the assessment of the English language. As the authors conclude, "the current quest by LSP researchers to bring testing practices into line with what really matters for participants in particular communicative contexts (Douglas, 2000) seems to be precisely what ELF researchers now seek" (p. 295).

Regarding tests based on the second model for ELF assessment, some implications are also considered. First, regarding scoring procedures, Elder and Davies (2006) state that raters would need to be NNSs familiar with the ELF code and with the pragmatic demands of the context. Ideally, they would be from different L1 backgrounds "to neutralize the possible effects of familiarity with candidate speech or communicative style on rating behavior" (p. 293). NSs could probably be used as raters, provided that they proved able to rate in a similar way to NNS ELF users. In AE tests, this diversity of raters could pose a challenge to test service providers, especially those who developed intra-national tests. Second, Elder and Davies' (2006) concern is what norms to apply and what counts as NNSs usage errors for NNSs. They state that "we are just not sure what norms to apply for ELF (4) and without that knowledge, assessment remains speculative" (p. 284). Based on this uncertainty, the authors go further to claim that in standard-setting exercises to judge task fulfilment, the standards applied by the professionals might differ considerably, "given the subjective nature of such judgements and the difficulty (already noted above) of specifying what constitutes an unacceptable departure from the ELF code" (p. 293). Third, Elder and Davies (2006) state that it is hard to imagine that ELF would become a stable undifferentiated generic code, according to its definition of ELF (4). If so, attempts to establish this uniformity would

"ignore the variation that exists within and between national boundaries and may end up reinforcing the reification of the NS and NNS categories and the stigmatization of the latter that the ELF movement attempts to counteract" (p. 295). Finally, the authors refer to the fact that ELF features in particular contexts could possibly be captured in ESP tests:

> Methodological and conceptual uncertainties still abound in the ELF literature and it may be more useful, until greater clarity is achieved, to conceive of ELF as a series of register varieties, each serving a highly specific communicative purpose. Such a context-specific formulation, we have argued, does not require a testing agenda radically different from what has already been adopted or at least entertained in other LSP contexts. (p. 296)

However, they signal some practicality considerations. ESP tests are restricted in scope and have limited generalizability, compared to current tests of SE which have greater prestige and wider acceptance (p. 295). In the special case of an AE test, this would not be considered a limitation, as pilots and ATCOs need to be assessed in their ability to use AE as a lingua franca for a very specific purpose, which is directly related to safety.

Finally, Elder and Davies (2006) express their concern about the politics of language, arguing that ELF norms, in the event of being codified and operationalized in ELF tests, could have the same power to discourage and 'oppress' non-proficient users of ELF as has been the case of SE tests, running the risk of becoming "a new monolithic standard with all the attendant consequences for those lacking the command of the new code" (p. 296). In addition, McNamara (2012a) believes that NS have a strong political

and social advantage, so requiring them to be tested in equal terms to NNS, and without a NS model, would probably lead to strong reactions.

4.2.3 Should the ELF construct be operationalized in aviation English test design?

From the above discussion, a call for a revision of the ICAO LPRs testing policy is necessary in order "to adequately reflect the realities of using English as a lingua franca" (Douglas, 2014, p. 10). Therefore, in order to increase the quality of tests developed to assess pilots' and ATCOs' language proficiency worldwide and the validity of inferences drawn from test scores, ICAO should set a group of experts and stakeholders to revise the Standards and Recommended Practices (SARPs) related to the LPRs (ICAO, 2010). As McNamara (2012b) argues, "standards can be seen as statements of test constructs, thereby playing a central role, in fact **the** central role, in determining the nature of the tests by means of which achievement is demonstrated" (p. 199). However, Estival et al. (2016) claim that there is still a lot to be done in order to define the construct of AE for training and testing purposes and that making changes to the rating scales "is not a simple matter and would, if deemed necessary, take time to resolve" (p. 202). Although I understand their concerns, which seem to be based on the issues of test quality and fairness discussed above, I prefer to follow their recommendation of, under the current LPRs, reconciling some of the contradictions in the ICAO testing policy by drawing on knowledge from ESP and ELF training and testing:

> The aviation language training and testing communities need not wait for ICAO to take the lead, and the intended role for native or expert-level speakers can

> begin to be operationalized now. Of course minimum requirements outlined in the ICAO Standards and Recommended Practices (ICAO, 2010) must be adhered to, but, in the interest of ensuring effective communication involving the use of Aviation English, test service providers can move beyond the requirements and include native or expert-level speakers in their language training and assessment programmes. Furthermore, the accommodation or adaptation strategies that are recommended for native or expert-level speakers can be included in training and assessment for speakers of all levels. (Estival et al., 2016, p. 201)

Noteworthy is the contribution of all areas of research mentioned in 4.1.2, which can undoubtedly inform language testers in designing AE tests including the ELF construct. Moreover, the emerging body of knowledge being produced by the increasing number of studies in the field of aviation English and ICAO language testing policy may also help in this endeavor (e.g., Aragão, 2018; Douglas, 2014; Emery, 2014; Estival, 2018; Estival et al., 2016; Garcia, 2015; Kim, 2012, 2013, 2018; Kim & Elder, 2009, 2015; Knoch, 2014; Monteiro, 2012, 2016a, 2016b; Read & Knoch, 2009).

Having said that, the discussion of the arguments for and the implications of the assessment of ELF, both in relation to a broader context of language assessment and specifically to the testing of pilots and ATCOs, which requires complex communicative abilities, reinforces the value of operationalizing the ESP construct in AE test design. In the interest of increased air traffic safety, Douglas (2014) also advocates for an ELF-enhanced test of AE, which would include, beyond language knowledge, Interactional Competence and background knowledge.

4.2.4 The ELF Construct: possible definitions and ways to operationalize it.

Harding (2014) proposes skills for an ELF test and suggests a focus on 'adaptability' to be added to the construct of Communicative Language Testing (CLT). He characterizes it as "how a candidate copes in a novel or challenging language situation in real time" (p. 192), including the assessment of the "test-taker's ability to deal with diverse, and potentially unfamiliar, varieties of English" (p. 194). The author discusses possible ways of how 'ability for use', "a component of Hymes's[17] original model, which took into account not just an individual's underlying ability for communication but also the capacity for using this across various contexts" (p. 191), might possibly be operationalized in communicative language testing.

Harding's (2015) proposed outline of an ELF construct includes a list of the following ELF competencies:

- The ability to tolerate and comprehend different varieties of English: different accents, different syntactic forms and different discourse styles;
- The ability to negotiate meaning when meaning is ambiguous;
- The ability to use those phonological features which are crucial for intelligibility across speakers of different L1 backgrounds;
- An awareness of appropriate pragmatics (e.g., awareness of politeness in cross-cultural situations);
- The ability to accommodate your interlocutor, to make yourself understandable to whomever you are speaking with; and

[17] Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 53–73). Harmondsworth, UK: Penguin.

- The ability to notice and repair breakdowns in communication. (p. 24)

Regarding the operationalization of the ELF construct, first Harding (2014) suggests that being paired with partners with different first languages or cultural backgrounds would also be a way to operationalize 'ability for use' in a testing context (p. 195). In this sense, interlocutors and test-takers' language variability in testing would resemble candidates' real-life intercultural communicative needs. Second, Harding (2015) reported on the piloting of a purpose-built ELF assessment task, both interactive and goal-oriented, in which "complications are built-in to the task to ensure that interlocutors need to negotiate both meaning and form" (p. 25). Although the pilot testing of the task indicated potential for its implementation, Harding and McNamara (2017) discussed challenges related to the scoring of an ELF task and called for further research in order to "explore the potential for more authentic tasks … and also to develop data-driven rating scales which capture the ELF-related strategic behaviour observed" (p. 579). The authors also mentioned issues related to the aviation language assessment policy and the difficulties in relation to the 'institutionalized conservatism' that comes from NSs of English in resisting changes, despite their contributions to communication breakdowns in ELF contexts:

> The shifting of focus towards accommodation and interactional strategies also calls into question the policy of exempting participants in ELF communication who have native-language proficiency from being tested at all, given what studies have revealed of the role of native-speaker behaviour in communicative failure in ELF, particularly in high-stakes contexts such as aviation and medicine. (p. 570)

To conclude, the summary of previous research and controversies in the field of ELF assessment highlights the links between what is known about the topic and the current study. It not only situates this study within this field but also shows where my contribution will be made: the definition of an expanded construct that better represents the communicative demands of pilots and ATCOs interacting in the occupation-specific domain of international RT communication, one that can inform test design and foster a positive washback on training practices, leading to greater communicative awareness and openness to accommodate difference.

4.3 Task Design

Douglas' (2000) seminal work on the assessment of languages for specific purposes made a profound contribution to the development of test tasks in occupational and academic contexts. Therefore, it is not only influential in the field of Language for Specific Purpose (LSP), but also extremely relevant to address the research problem presented in this study. The author discusses fundamental concepts in this field of language testing, and also provides examples to underpin the description of task characteristics and the movement from target language use (TLU) tasks to test tasks.

Considerations of "**authenticity of task** and the **interaction between language knowledge and specific purpose content knowledge**" (p. 2) are highlighted as key aspects of LSP testing. Another critical definition in LSP testing that Douglas (2000) provides is the one related to specific purpose language ability, which "results from the interaction between specific purpose background knowledge and language ability, by means of strategic competence engaged by specific purpose input in the form of test

method characteristics" (p. 40). In relation to test tasks, Douglas (2000) further argues that:

> It is not enough merely to give test-takers topics relevant to the field they are studying or working in: the material the test is based on must engage test-takers in a task in which both language ability and knowledge of the field interact with the test content in a way which is similar to the target language use situation. The test task, in other words, must be **authentic** for it to represent a specific purpose field in any measurable way. (p. 6)

In addition, the author's comprehensive explanation of a framework for analyzing TLU and test task characteristics (p. 50-71), comprising the characteristics of the rubric, the input, the expected response, the interaction between input and response, and of the assessment, proves very useful to inform the development of draft tasks to be included in a test for aviation professionals involved in international RT communications. This framework is in line with Bachman and Palmer's (1996) view that "the key to designing tests that will be useful for their intended purposes is to include, in the test, tasks whose distinguishing characteristics correspond to those of TLU tasks" (p. 45). The authors offer general strategies for developing test tasks, which include: (a) modifying TLU task types; or (b) creating original test tasks based on a set of test task specifications (p. 174-176).

With a direct relation to the present study, Douglas (2000) offers examples from different academic and occupational fields, including detailed features of the Proficiency Test in English Language for Air Traffic Controllers (PELA) listening and speaking tasks, while at the same time underscoring a big challenge in LSP testing:

> The *PELA* is a prototypical example of a classic narrowly focused specific purpose language test, dealing with a highly restricted linguistic register and situational context. It illustrates both the best qualities of LSP test development – the use of specialist informants and observations of the TLU domain – and one of the most fundamental problems of LSP testing – the inability to simulate the dynamic nature of the target language use in the test domain. (p. 188)

Later, Douglas (2004) also provides an insightful discussion on issues of validity and impact in the development of tests in response to the ICAO language proficiency requirements. In that occasion, he explained that in order to collect and present validity evidence in this specific context two things were necessary:

- a clear understanding of the nature of the entity being tested, aviation English; and
- a clear, complete, and unambiguous definition of the construct to be measured in relation to the purposes for which the measurement is being made. (p. 250)

Furthermore, Douglas (2014) advocates for an expanded construct to be measured in aviation English tests: one that includes Interactional Competence and components of ELF communication strategies, for both native and non-native speakers of English. The author goes further in suggesting ideas for potential ELF tasks to assess the required abilities, which underscores the relevance of his work to the present study.

Likewise, the work of Moder and Halleck (2009) also sheds light on the development of test tasks for this particular context. The authors investigated the performance of air traffic controllers in three different types of tasks from the Oral Proficiency Test for Aviation (OPTA): a work-related radiotelephony task, and two non-

specific English tasks on aviation topics, one called 'Common Occurrence' and the other 'Less Expected Occurrence'. Results demonstrated significant differences in the performance of test-takers across task types with respect to the minimum required proficiency level – Operational Level 4. Thus, the authors assert that "one-way tasks that require the test taker to listen to a general English prompt and provide a one-way response will not provide adequate opportunities to assess critical interactional competencies" (p.25.13), and conclude that "both phraseology and unexpected related contexts must be fully included in ICAO proficiency tests" (p. 25.14).

In addition, with an emphasis on the abilities of L2 learners, it is possible to mention other studies that address the measurement of Interactional Competence as part of the speaking construct, which add to the present discussion. In one of them, van Batenburg, Oostdam, van Gelderen and de Jong (2018) report on the use of "interactive speech tasks that engage candidates in achieving real-life communicative goals in a simulated setting, evoke functional language use and directly evoke the use of … interactional strategies in a standardized manner" (p. 78) through the use of a scripted format that standardizes the interlocutor's contributions. In another study, Roever and Kasper (2018) highlight the co-constructed nature of interactions, arguing that:

> measurements of interactional abilities expand the range of conclusions that can be drawn from speaking tests and support inferences as to test-takers' ability to engage in interactive talk with others, which is an ability that is currently not explicitly assessed but commonly and incorrectly assumed by users to be inferable from scores. (p. 348)

This is line with Weir (2005), who points out the need to include dialogic or 'reciprocity conditions' in tasks in order to assess spoken interaction, rather than just requiring candidates to answer questions in an interview: "So if we are interested in the candidate's capacity to take part in spoken interaction, there should be reciprocal exchanges where both interlocutor and candidate have to adjust vocabulary and message and take each other's contribution into account" (p. 72).

Also relevant to the present discussion is the work of Youn (2013) who investigated "the validity of task-based performance assessment of L2 pragmatics in interaction in an English for Academic Purposes (EAP) setting" (p. v) based on open role-play tasks that are relevant to stakeholders in the EAP context. He designed tasks to balance authenticity and standardization, and by collecting qualitative and quantitative evidence in his mixed methods study he was able to strengthen the validity argument.

When commenting on the complexities of newer models of communicative language ability, Harding (2014) highlights that "in translating complex models into workable testing blueprints, smaller-scale testing development teams might benefit from sharing knowledge about how conceptual frameworks have been interpreted and operationalized in different contexts" (p. 195) and also "how successful prototype assessment tasks have been in trial" (p. 195). In this respect, the present study will contribute to this end by reporting on the operationalization of the aviation RT construct as test tasks and on the results of their pilot testing, i.e., preliminary and unofficial trials.

Additionally, Harding (2014) suggests that discourse data yielded from tasks designed to assess those abilities "could be analyzed with a view to locating points at which these abilities are tapped in these interactions" (p. 195). Moreover, the author

proposes the following in relation to future research in the field, which the present study aims to address:

> Some of the skills and abilities just listed, which are not necessarily confined to intercultural communicative settings (e.g., ability to negotiate meaning, ability to accommodate, deployment of appropriate pragmatics), might also be assessed through carefully designed role-play tasks where an interlocutor is deliberately vague or inconsistent, or through integrated reading-to-write tasks where the initial input is deliberately challenging (e.g., responding to a letter of complaint where the complaint is unreasonable). Research into the types of strategic and pragmatic knowledge at play in tasks of this nature would be valuable. (p. 195)

Another important study by O'Sullivan, Weir and Saville (2002) addresses "the relatively neglected area of validating the match between intended and actual test-taker language with respect to a blueprint of language functions representing the construct of spoken language ability" (p. 33). The authors designed an observation checklist which "enables language samples elicited by the task to be scanned for these functions in real time, without resorting to the laborious and somewhat limited analysis of transcripts" (p. 33). They comment that an additional value of the use of observation checklists to validate speaking-test tasks lies in the fact that it provides a greater understanding of how tasks and task formats can be manipulated to result in specific language use" (p. 46). Their study will serve as a basis for the development and application of a checklist of communicative language functions associated with aviation during the pilot testing of the draft tasks.

As a contribution to the field of RT communications, this research study will inform the aviation industry of how an ESP construct framework that includes components of aviation English, English as a lingua franca, intercultural awareness and interactional competence can be operationalized in the context of the performance assessment of pilots and ATCOs. Added to that, it will provide insights of how prototype assessment tasks performed during pilot testing, and how they can be analyzed from the perspective of test-takers, interlocutors, raters and experienced pilots/ATCOs.

Having presented a review of the literature that addresses the sub-layers of *Construct Framework, Evidence Models* and *Task Models*, in the next chapter I will continue the literature review with an overview of studies that inform the third layer of architectural documentation, i.e., *Test Specifications.*

Chapter 5 – Literature Review: Third Layer -- Test Specifications

In the last chapter, I situated my study in relation to previous research in the fields of construct specification, the assessment of ELF and interactional strategies, and task design in LSP testing. In this chapter, I will present a review of studies that provide the foundation to the design of task specifications, included in the third layer of architectural documentation, i.e., *Test Specifications.* This is crucial because in the process of writing a task specification, the test developer has to reflect on and make decisions about a number of issues that may impact the usefulness of the task.

As shown in the representation of the test development process in Figure 3.1, at the layer of *Test Specifications* we find four sub-layers. The first, task/item specification provides details of test items or tasks that allow test developers to produce multiple equivalent forms (i.e., test forms or versions). They contain at least one sample of item or task and guiding language, which, at the level of tasks, "summarizes the relevant elements of the construct framework which the designers claim are being measured by a specific item or task type, and the evidence it is designed to elicit" (Fulcher & Davidson, 2009, p. 128). The following three sub-layers include additional information on presentation, assembly and delivery, drawn from Myslevy et al.'s (2003) models and described by Fulcher and Davidson (2009) as follows: (a) *Presentation Model* – how items and tasks are presented to test-takers; (b) *Assembly Model*[18] – how tasks and items should be combined to produce a test form; and (c) *Delivery Model* – how the actual test is delivered, including administration, security and timing.

[18] Although they are part of the test development process, the last two sub-layers, *Assembly Model* and *Delivery Model*, will not be detailed in the present study.

Fulcher and Davidson (2007) explain that understanding what needs to be tested and how tasks can be described makes it possible to create formalized test specifications that enable a controlled production of tasks, since "the specs function as an intellectual template, guiding the item writer, shaping the test and ensuring evidentiary clarity along the way" (p. 71). However, the authors observe that often, test creation "begins with an actual test question and infers the guiding language that drives it, such that equivalent items can be generated" (p. 57), an analytical process known as 'reverse engineering' (Davidson & Lynch, 2002). Another important attribute of 'specs', in the view of Fulcher and Davidson (2007) is that "both [their] form and content evolve in a creative, organic, consensus-driven, iterative process" (p. 61), which is corroborated by a number of test developers (e.g., Cheng & Fox, 2017; Davidson & Fulcher, 2012; Douglas, 2000).

Davidson and Fulcher (2012) further refer to the development of test specifications as an important stage in test design and explain how test specs can assume a more productive and pro-active role, in terms of increasing validity. The authors stress the benefits of seeking "feedback from as many informed parties as politically feasible" (p. 62), such as item writers, raters, etc., in an open and critical dialogue.

Likewise, Alderson, Clapham and Wall (1995) argue that the development of test specifications is crucial in the test construction and evaluation process and central to establish the construct validity of a test. They observe that a number of different stakeholders need test specifications, each for a particular purpose, such as test constructors, item writers, moderating/editing committees, test validators, test users

(e.g., teachers, candidates, admissions officers), and textbook publishers, highlighting the importance of having the audience in mind when writing a test specification.

Other representative studies that offer guidance to test developers in relation to the design of test task specifications include, for example, the work of Bachman and Palmer (1996) and Douglas (2000). The first authors emphasize the need to take into account the qualities of test usefulness (explained in more detail in the next section) in task design and suggest the following items to be included in a task specification: (a) the purpose of the test task; (b) the definition of the construct to be measured; (c) the characteristics of the setting of the test task; (d) time allotment; (e) instructions for responding to the task; (f) characteristics of input, response, and relationship between input and response; and (g) scoring method (p. 172-173). Douglas (2000), on the other hand, proposes a slightly different list of components, which "embodies the essentials for good LSP specifications" (p. 110): (a) the purpose(s) of the test; (b) the TLU situation and TLU tasks; (c) characteristics of the test-takers; (d) definition of the construct to be measured; (e) content of the test; (f) scoring criteria; (g) samples of task/items; and (h) a plan for evaluating the qualities of good testing practice (reliability, validity, authenticity, impact, and practicality). In sum, my attempt to produce a draft task specification for the proposed tasks to measure the specific purpose ability of pilots to communicate in international radiotelephony will draw on both templates of task specifications.

In the next section, a discussion of the qualities of test usefulness (Bachman & Palmer, 1996) and the qualities of good testing practice (Douglas, 2000) will be provided, for they are essential in all stages of test development.

5.1 Qualities of Test Usefulness

In order to answer the question "How useful is this particular test [or task] to its intended purpose(s)?" (p. 17), Bachman and Palmer (1996) propose a model of test usefulness that includes the qualities of reliability, construct validity, authenticity, interactiveness, impact, and practicality, along with three principles that they believe provide the foundation to operationalize the model in language test development and use. The three principles are specified as follows:

i) Principle 1 – It is the overall usefulness of the test that is to be maximized, rather than the individual qualities that affect usefulness;

ii) Principle 2 – The individual test qualities cannot be evaluated independently, but must be evaluated in terms of their combined effect on the overall usefulness of the test;

iii) Principle 3 – Test usefulness and the appropriate balance among the different qualities cannot be prescribed in general, but must be determined for each specific testing situation. (p.18)

The qualities of test usefulness, according to Bachman and Palmer (1996), can be broadly defined as shown in Table 5.1. In their book, the authors not only present a detailed explanation of each quality (p. 19-42) but also provide a comprehensive list of questions, organized in the form of a checklist (p. 150-155), to assist test developers in developing a plan for the evaluation of usefulness.

Table 5.1. Definitions of the qualities of test usefulness (Bachman &Palmer, 1996)

Qualities of test usefulness	Definition
Reliability	"A function of the consistency of scores from one set of tests and test tasks to another" (p.19).
Construct validity	"Refers to the extent to which we can interpret a given test score as an indicator of the ability (ies), or constructs, we wish to measure" (p. 21).
Authenticity	"The degree of correspondence of the characteristics of a given language test task to the features of a TLU task" (p. 23)
Interactiveness	"The degree to which the constructs we want to assess are critically involved in accomplishing the test task" (p. 39)
Impact	"The various ways in which test use affects society, an education system, and the individuals within these" (p. 39)
Practicality	"The relationship of the resources that will be required in the design, development, and use of the test, and the resources that will be available for these activities" (p. 39)

In the field of LSP testing, Douglas (2000) refers to these qualities as the 'qualities of good testing practice', to differentiate his list from that of Bachman and Palmer (1996). As a way to make these qualities more relevant to the context of specific purpose language testing, Douglas (2000) approached them a bit distinctly. Although Bachman and Palmer consider authenticity and interactiveness as separate qualities of usefulness, he considers authenticity as a single quality, residing in language users as they interact with texts and tasks, which has two separate aspects: situational and interactional. Douglas (2000) defines them as follows:

> Situational authenticity can be demonstrated by making the relationship between the test task characteristics and the features of tasks in the target language use situation explicit. The second aspect of authenticity, interactional…involves the

> interaction of the test taker's specific purpose language ability with the test task. The extent to which the test taker is engaged in the task, by responding to the features of the target language use situation embodied in the test task characteristics, is a measure of interactional authenticity. (p. 17)

As both authors stated, i.e., Bachman and Palmer (1996) and Douglas (2000), a consideration of the qualities of test usefulness is crucial in the whole process of test development, for there would be no point in developing a task or test not useful for its purpose(s). Therefore, the present study will draw on these qualities in the development of task specifications for an interactive speaking task to be included in the language assessment of pilots, as well as in the evaluation of how the draft tasks perform during the pilot testing.

5.2 From the Literature Review to the Specification of Study Objectives

In Chapters 3, 4 and 5 I situated my research study in the literature. This was accomplished by referring to different fields of inquiry and addressing theoretical, empirical and practical studies that proved relevant to achieve the objectives of the three layers of architectural documentation in the test development process (see Figure 3.1). An overview of what is still left unexplained or under-explored in relation to what is required for effective intercultural RT communications in aviation and for the LSP testing of pilots and ATCOs, led to a clearer understanding of the research problem. As stated at the beginning of Chapter 3, the literature review sought to provide a firm foundation for the research design and a rationale for the research questions.

As a result, responding to industry needs, this study extends previous research by addressing the following objectives, namely, to:

1. identify intercultural factors that can affect the way pilots and air traffic controllers (ATCOs) interact in the English language;
2. verify the extent to which experienced professionals perceive the potential threat of those factors to the safety of RT communications;
3. propose models of language use that account for the communicative demands of pilots and ATCOs ;
4. define the construct of international aviation RT communications in order to specify a framework to inform test development;
5. validate the matrix of construct specification by eliciting key stakeholders' perspectives from diverse linguistic and cultural backgrounds;
6. investigate how the construct of international aviation RT communications can be operationalized as test tasks; and
7. examine how prototype assessment tasks perform during pilot testing, taking into consideration the perspective of Aviation English Testing Experts (AETEs).

The overarching research question that guides this project, presented earlier in Chapter 1, is restated below, along with the specific research objectives that it aims to address:

> What are the communicative demands of pilots and ATCOs involved in intercultural RT communications that go beyond language proficiency (objectives 1, 2, and 3); how can they be specified within a construct framework and operationalized as test tasks (objectives 4, 5, 6, and 7)?

In the next chapter, the methodological approach and choice of design for this complex research study will be discussed, including methods for data collection, data analysis and

a procedural diagram with an overview of all design phases and more specific research questions that guide each of them.

Chapter 6 – Methodology

In Chapters 3, 4, and 5, I situated my study within what is already reported in the literature of relevance to the focus of the research, and specifically within the field of LSP testing in the occupational context of intercultural RT communications. In this Chapter I will focus on my choice of study design and methodological issues.

6.1 Study Design

Taking into account the complexity of the context under investigation and the comprehensiveness of the overarching research question guiding the study, arguably the best methodological approach to apply in this research study was Mixed Methods (MM) research. Mixed methods, which presupposes the integration or 'mixing' of qualitative (QUAL or qual, depending on the relative dominance of the method in the study) and quantitative (QUAN or quan) research approaches, has developed as a result of the need to address complex research problems, the legitimization of qualitative inquiry, and the need for more evidence in applied settings (Creswell & Plano Clark, 2011).

Therefore, in order to respond to the research question, three interrelated studies were conducted as part of the same program of research, along three phases. It began with an exploration of the intercultural RT communication context (Phase 1 – MM study) in order to identify intercultural factors that may affect the way pilots and ATCOs interact in the English language and to verify the extent to which those factors impact on safety, based on pilots' and ATCOs' perceptions. Then, based on a review of theory and research (Phase 2 – QUAL study), the aim was to propose models of language use relevant to the occupational domain of pilots and ATCOs, followed by the specification of the construct from the models to a framework, and giving voice to aviation

stakeholders to identify the key construct components. In order to verify the operationalization of what was considered relevant for inclusion in an aviation English test in terms of language and communication, draft tasks were designed which might be used in the assessment of pilots' English proficiency in this occupational context. Later, two tasks were pilot tested (Phase 3 – MM study) with Aviation English Testing Experts (AETEs), including interlocutors and raters with both language and operational backgrounds. This multiphase MM study was exploratory in nature; however, the QUAL and quan strands were conducted sequentially but also concurrently across phases and within phases, characterizing the range of alternatives and methodological mixes that the 'third research paradigm' affords.

Based on the overarching research question, three more specific and interconnected research questions guided my dissertation research, namely, two mixed methods questions (Phases 1 and 3) and one qualitative question (Phase 2). These were further elaborated in more detailed questions for each phase of the study. Table 6.1 provides an overview by phase of all the research questions that were necessary in order to address the comprehensiveness of the central question.

Table 6.1. Research questions specific to each phase of the study

Phase	Type of study	Research questions	
Multiphase MM study		**What are the communicative demands of pilots and ATCOs involved in intercultural RT communications; how can they be specified within a construct framework and operationalized as test tasks?**	
Phase 1	**Exploratory sequential MM study**	How do the quantitative results on pilots' and ATCOs' perceptions of the potential threat of intercultural factors to safety build on the qualitative exploration of those factors from RT communication case studies?	RQ 1.1) What intercultural factors arise from international pilot-ATCO communications that can affect the way they interact in the English language? RQ 1.2) To what extent do experienced pilots and ATCOs perceive the potential threats of intercultural factors to the safety of radiotelephony communications?
Phase 2	**Qualitative study**	What is the construct of international RT communication based on a review of current theory and research, and what components of this construct are validated by key aviation stakeholders?	RQ 2.1) What theoretical models of language use would account for the communicative needs of pilots' and ATCOs' occupational domain? RQ 2.2) How can this construct be articulated and specified from the models to a framework which better informs test development? RQ 2.3) What components of the construct are validated by key aviation stakeholders?
Phase 3	**Convergent parallel MM study**	To what extent do the quantitative results from the pilot testing of tasks agree with the interviews and focus group data reporting the views of test-takers, interlocutors and raters and with transcripts of task performances?	RQ 3.1) What behaviors or performances should reveal the proposed RT construct? RQ 3.2) What kind of test tasks can be developed to elicit those behaviors in a pilot's exam? RQ 3.3) What are the test-takers' insights on the draft pilot tasks, and, from their perspective, to what extent do the role-play tasks reflect pilots' communicative needs in RT communication? RQ 3.4) What are the interlocutors' insights on the draft pilot tasks, and, from their perspective, to what extent do they feel confident in following the role-play instructions and perceive the appropriateness of the task? RQ 3.5) What are the raters' insights on the draft pilot tasks? RQ 3.6) What language functions and behaviors indicative of effective communication arise during test-taker's performance? RQ 3.7) On the basis of expert judgment (raters), to what extent are: - The draft pilot tasks likely to elicit the desired language functions and behaviors indicative of effective communication? - The language functions and behaviors perceived as important?

The need to explore the construct of pilots and ATCOs' international RT communications and its specification and operationalization in test design called for multiple phases under a unique methodological framework, which multiphase MM designs provide. Therefore, considering the complex nature of the research problems identified, their intersection with various fields of inquiry, and the diverse context in which they are embedded, a multiphase MM investigation was conducted. According to Creswell and Plano Clark (2011), multiphase designs occur when a researcher "examines a problem or topic through an iteration of connected quantitative and qualitative studies …, with each new approach building on what was learned previously to address a central program objective" (p. 100). The authors state that this type of design is chosen when the researcher cannot achieve the program objective with only one MM study.

Some authors (e.g., Creswell & Plano Clark, 2011; Elliot & Lim, 2016) adopt the expression 'multiphase' to refer to the advanced mixed methods design that encompasses multiple QUAL and QUAN studies as part of the same program of research. Other authors (e.g., Ziegler & Kang, 2016; Galaczi & Khabbazbashi, 2016) use the expression 'multistage' for the same purpose. Throughout this dissertation, I use the expression 'multiphase' MM design to refer to my choice of MM design, which includes three interconnected phases (Phase 1, Phase 2 and Phase 3). However, in direct citations of the second group of authors, the terms 'multistage' and 'stages' are kept.

Ziegler and Kang (2016) further explain that through multiple iterations of data collection and analysis, "the multistage [multiphase] mixed methods design is able to address a larger research purpose by breaking it into multiple smaller, semiautonomous, studies" (p. 75). In respect to research projects that include test development processes,

the authors highlight that "inherent to multistage [multiphase] studies is an infinite range of possibilities for the application of mixed methods well suited to meet the complexities of test development" (p. 77). As examples of other multiphase studies in the field of language assessment, they cite Galaczi and Khabbazbashi's (2016) multiphase exploratory sequential rating scale development and Elliot and Lim's (2016) multiphase development and validation of a new Reading task in the *Cambridge English: Advanced* test, to mention a few.

In the present study, the three phases built on each other as an iterative process, moving in "a spiral-like manner, with subsequent stages drawing on the findings of previous ones and leading to more in-depth and meaningful insights" (Galaczi & Khabbazbashi, 2016, p. 211). The starting point was the literature review, based on existing theory and research, which informed all three phases through repeated iterations (See Figure 6.1), from the general to the specific. What goes on in the spiral after Phase 3 is the generation of meta-inferences, i.e., an overall understanding achieved by the integration of findings, which occurs within, between, and across all Phases of the study. This is the great advantage and strength of a mixed methods approach in addressing a complex research question, like the one considered in this dissertation. In my view, multiple vantage points potentially support the most useful, meaningful, and appropriate findings, by virtue of increased evidence of their validity (see Section 6.3 below for further discussion of meta-inferences).

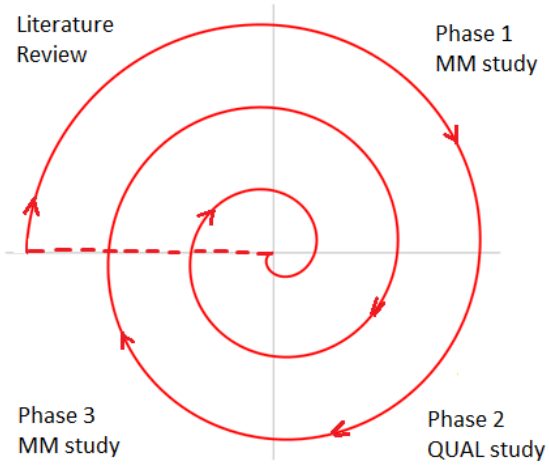


Figure 6.1. Multiphase mixed methods spiral-like design

However, the representation shown in Figure 6.1 does not convey the complexity of this multiphase MM study. It calls for the development of a procedural diagram to illustrate the stages of data collection and analysis (with blue and yellow boxes), mixing, merging or integrating results (with green ovals), while at the same time incorporating time frames in which each phase did occur. Figure 6.2 presents this procedural diagram, guiding the reader on the steps followed through the whole program of research and within each phase.

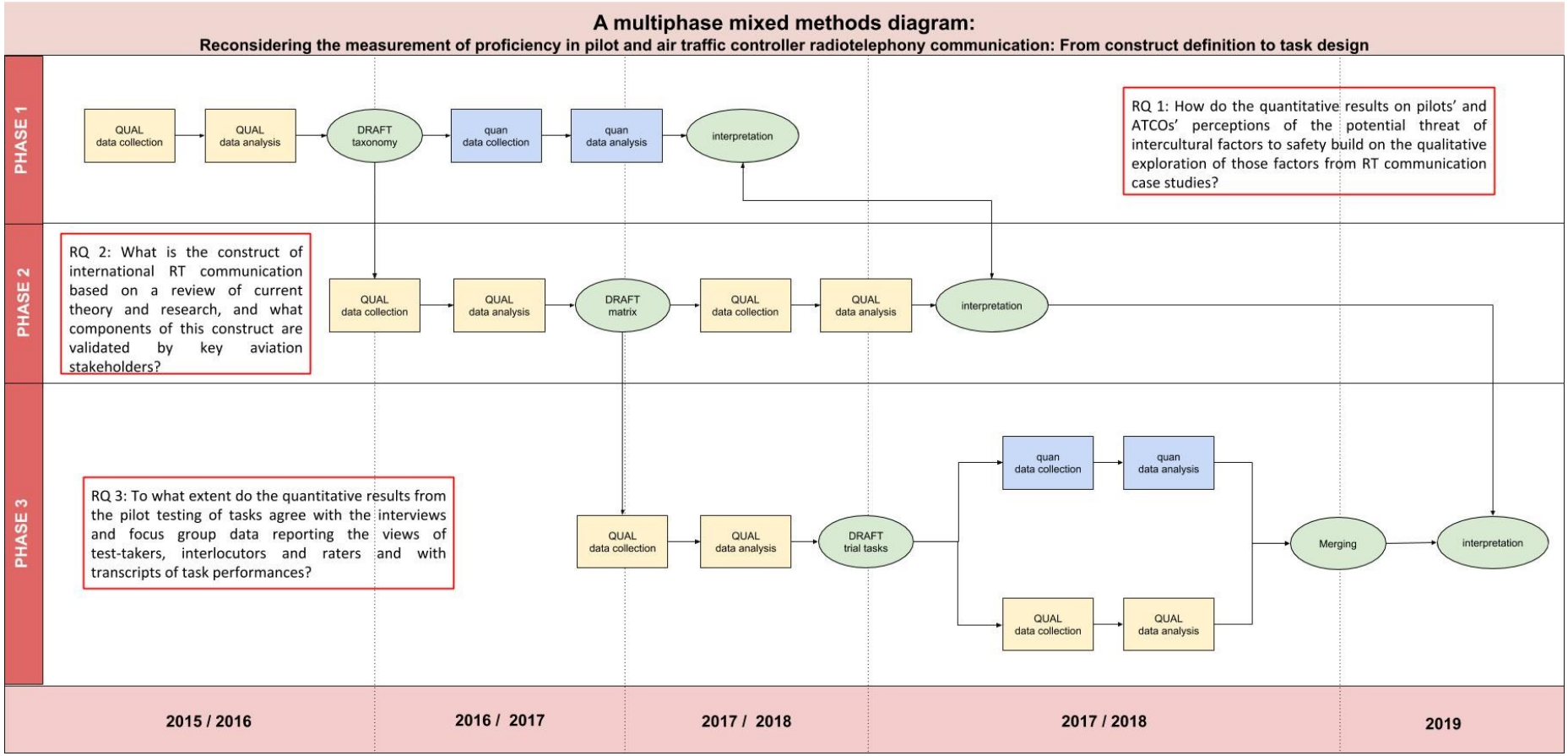


Figure 6.2. Procedural diagram for the multiphase mixed methods design

Having presented the choice of a specific MM design for this dissertation research, the next section will focus on the basic features of this methodological approach to underpin the argument for the use of MM as the most suitable for the present study.

6.2 Methodological Approach

Understanding the basics of MM as a research methodology is essential in order to recognize its value for conducting research in fields of inquiry that generally address complex research problems and questions.

According to Teddlie and Tashakkori (2009), mixed methods research has emerged as an alternative to the dichotomy of QUAL and QUAN research traditions during the past 20 years, and has been called the 'third methodological movement' (Teddlie & Tashakkori, 2003), the 'third research paradigm' (Johnson & Onwuegbuzie, 2004), and more recently the 'third research community' (Teddlie & Tashakkori, 2009).

Tashakkori and Teddlie (2010) put forward core characteristics of MM research (p. 273), to which I added two other features (see underlined text) highlighted by Johnson and Onwuegbuzie (2004), in an attempt to construct a definition of this research community. In my view, MM

- is an iterative, cyclical approach to research, based on methodological eclecticism and paradigm pluralism with a clear emphasis on continua rather than a set of dichotomies and on diversity at all levels of the research enterprise;

- focuses on the research question (or research problem) in determining the methods used within any given study and counts on a set of basic ‘‘signature’’ research designs and analytical processes;
- requires the integration of findings and inferences from both QUAN and QUAL approaches at some point of the study, besides relying on visual representations (e.g., figures, diagrams) and a common notational system; and,
- holds a tendency toward balance and compromise that is implicit within the “third methodological community” and takes an explicit value-oriented approach to research.

In an atempt to create a “more parsimonious and functional classification” of MM designs, Creswell and Plano Clark (2007) presented four major types of design, with variants within each type: triangulation, the embedded, the explanatory, and the exploratory design. Later, the authors updated their typology (Creswell & Plano Clark, 2011) and provided six types of MM design, as detailed in Table 6.2, according to their variants, timing, weight, mixing and notation. Nonethelesss, Creswell (2014) considers the three first types of design, convergent parallel, explanatory sequential and exploratory sequential as the three basic approaches, which can be used as a foundation to integrate the three advanced mixed methods designs: embedded, transformative and multiphase design.

Within this setting, the reasons for using MM and its applicability to the research problem under investigation are now described. First, because it is based on methodological eclecticism and paradigm pluralism, MM offers a way to get the best of

Table 6.2. Typology of MM designs (based on Creswell & Plano Clark, 2007, 2011)

Design Type	Variants	Timing	Weight	Mixing	Notation
Convergent parallel design	Parallel-databases Data-transformation Data-validation	Concurrent: QUAN and QUAL at the same time	Usually equal	Merge the data during interpretation or analysis (bring results together)	QUAN + QUAL
Explanatory sequential design	Follow-up explanations Participation-selection	Sequential: QUAN followed by qual	Usually QUAN	Connect the data between the two phases (one phase builds on the other)	QUAN → qual
Exploratory sequential design	Taxonomy/theory development Instrument development	Sequential: QUAL followed by quan	Usually QUAL	Connect the data between the two phases (one phase builds on the other)	QUAL → quan
Embedded design	Embedded-experiment Embedded instrument development and validation	Concurrent or sequential	Unequal	Embed one type of data within a larger design using other type of data (either building or bringing results together)	QUAN (qual) or QUAL (quan)
Transformative design	Feminist lens transformative variant Disability lens transformative variant Socioeconomic class lens	Convergent, Explanatory, Exploratory or Embedded Mixed Methods can be implemented within the transformative framework.			
Multiphase design	Large scale program development and evaluation Multilevel statewide study Single mixed methods studies that combine both concurrent and sequential phases	It is a combination of sequential and concurrent strategies used over time and it is most common in large funded or multiyear projects.			

QUAL and QUAN approaches by mixing them, resulting in an enhanced understanding of the research problem. Second, a unique research paradigm does not suffice to answer the complexity of research questions this context generates. As researchers are not limited to data collection procedures associated to either QUAN or QUAL approaches, MM provides more evidence for studying a research problem. Third, because MM is well suited for interdisciplinary research, it is suitable for inquiries related to the communicative needs of pilots and ATCOs and their operationalization into test tasks, which require expertise from a number of disciplines. For example, knowledge can be obtained from domain experts (experienced pilots and ATCOs), language teachers, applied linguists, scholars in the field of English as a lingua franca (ELF) and intercultural communication, and also language testers, especially the ones with expertise in performance assessment within the field of Languages for Specific Purposes (LSP). Finally, as regards knowledge production in MM studies, Johnson and Onwuegbuzie (2004) stress that QUAL and QUAN used together produce more complete knowledge necessary to inform theory and practice, and provide stronger evidence for a conclusion through convergence and corroboration of findings, such that MM can add insights and understanding that might be missed when only a single method is used.

Further, as this study focuses on language testing and assessment, it is worth mentioning that the combination of information from different sources "can often provide valuable insight into a deeper understanding of complex phenomena under study, most especially in the areas of validity and instrument development, classroom-based assessment, large-scale assessments, construct definition, and rater effects, to name a few" (Moeller, 2016, p. 11). Moeller (2016) also highlights that the complexity in

assessing task-based communication, the range of variables involved, and its inherent challenges of reliability, content validity and authenticity "underscore that one research method cannot fully capture the complexity of language skills" (p. 8). Moreover, Guetterman and Salamoura (2016) state that "conducting a rigorous mixed methods study can strengthen the investigation of test validity and the building of a validation argument in language assessment" (p. 153). Finally, Khalifa and Docherty (2016) report that the use of multiple data sources in MM research "provides reassurance to users of research results of the comprehensiveness of the findings" (p. 274).

When referring to MM research informing paradigms and their related worldviews, Riazi and Candlin (2014) state that "advocates of MMR methodology have recognized three paradigms, namely critical realism, critical theory or transformative learning, and pragmatism as potential foundations for MMR research" (p. 140). Concerning my philosophical worldview, which is one of the foundations upon which this research is built, and regarding my choice of MM research, I chose to frame my investigation on the paradigm of pragmatism, although during the qualitative parts of the study the constructivist perspective predominates, shifting into a more postpositivist worldview in the quantitative phases. Taking a pragmatic position, which Teddlie and Tashakkori (2009) considered "the philosophical partner of MM" (p. 73), offers an alternative way to facilitate communication among different paradigms and to address research problems. The pragmatist addresses the nature of reality (ontology) as both singular and multiple, collects data considering issues of practicality (epistemology), includes both biased and unbiased perspectives (axiology), combines inductive and deductive thinking (methodology), and may employ formal and informal styles of writing

(rhetorical) (Creswell & Plano Clark, 2011). Nonetheless, Riazi and Candlin (2014) point out that one of the challenges of MM research is that "claims need to be substantiated by showing how the researcher moves between the dataset and the theoretical framework underpinning the study to draw plausible and, above all, warrantable inferences" (p. 160).

An example of MM research in the field of Aviation English testing is Kim's (2012) study. Her aim was to "explore[s] the gap between the perception of important qualities of the construct of radiotelephony communication by Korean aviation personnel and the view of this construct as espoused by ICAO" as a way to further "explore the construct of pilot-controller communication and to better understand the resistance to the ICAO policy by Korean aviation personnel" (p. 62). Although there is no specific mention of MM design, she described her study as a combination of QUAL and QUAN research methods, using multiple sources and methods in a triangulation approach. She used three different sources (pilots, air traffic controllers and discourse samples) and five different methods (nonparticipant observation, discourse analysis of audio-recordings of pilot-controller communication, feedback analysis of the collected audio-recordings by aviation specialist informants, interviews and questionnaires) (p. 63). Nonetheless, she explains that the questionnaires (used for QUAN data collection) were built based on findings of the first phase, mainly from exploratory unstructured interviews (QUAL). Her goal was to gain a broader spectrum of views on communication issues in the aeronautical community which were, in turn, complemented by follow-up interviews conducted to probe these issues more deeply. Thus, it is possible to notice not only the exploratory nature of her design, but also the explanatory one, culminating with the integration of her findings to answer the research questions.

6.3 Generation of Meta-Inferences

Meta-inferences are defined by Tashakkori and Teddlie (2008) as "an overall conclusion, explanation, or understanding developed through an integration of the inferences obtained from the qualitative and quantitative strands of a mixed method study" (p. 101), which is, in the authors' opinion, the real source of the value supplemented by mixed methods. Ziegler and Kang (2016) further explain that the value of MM in multiphase designs "lies in the use of meta-inferences generated within (or between) stages to inform the direction of the following stages, as well as generating meta-inferences from all strands across stages [phases]" (p. 77). However, the validity of those meta-inferences depends on: (a) the purposeful integration of rich qualitative methods and rigorous quantitative methods (Moeller, 2016); and (b) the purposeful implementation of MM procedures at the design level, methods level and interpretation level of a study (Ziegler & Kang, 2016).

Therefore, based on the description and exemplification of MM strategies provided by Ziegler and Kang (2016)[19], a summary of those employed in this multiphase MM exploratory study in order to produce quality meta-inferences is provided in Table 6.3.

[19] Ziegler and Kang (2016) base their account of MM strategies mainly on the following authors:

a) Design level: Creswell and Plano Clark (2011);
b) Methods level: Fetters, Curry and Creswell (2013);
c) Interpretation level: Greene (2008); Onwuegbuzie and Teddlie (2003).

Table 6.3. Summary of MM strategies employed according to points of interface

Points of interface	Strategies
Design level	Multiple phases: 1) MM exploratory sequential design to investigate the context of aeronautical intercultural RT communication → 2) qualitative study to describe and validate the RT construct → 3) MM convergent parallel design to gather evidence of the functioning of test tasks
Method level	Building (analysis of one strand informs the data collection of the next) Merging (investigating parallel constructs in both strands to facilitate comparison)
Interpretation level	Data preparation: data reduction and data transformation Data analysis: Data comparison Data integration: Joint displays

The point at which those MM strategies were used and how they were employed in each phase of the study will be detailed in the next chapters.

6.4 Challenges of Multi-Phase Mixed Methods Research and Potential Validity Threats

Using the multiphase MM design presents special challenges which, according to Creswell and Plano Clark (2011), relate to its own "multifaceted nature and flexibility" (p. 103). The ones that I faced in this research project are the following:

1. The researcher needs sufficient resources, time and effort to successfully implement several phases over multiple years:

I overcame this challenge by: (a) counting on the assistance of other researchers who participated in the coding processes to inform inter-coder reliability, of those who provided additional equipment for data collection, and relying on the collaboration of Subject Matter Experts (pilots) in the design of the draft tasks; (b) receiving permission from two international conference organizers in the field of aeronautical communications

and from the National Civil Aviation Agency (ANAC – Brazil) to announce my study and recruit participants who are key stakeholders in the international context being investigated; (c) starting the first phase of my study in the early stages of my doctoral program; and by (d) being persistent, focused and systematic in undertaking the tasks of transcribing, coding, analyzing, interpreting, and reporting findings.

2. The researcher needs to consider how to meaningfully connect the individual studies in addition to mixing quantitative and qualitative strands within phases:

In order to achieve the 'synergetic effect'[20] (Ziegler & Kang, 2016) that MM produces, I had to take special care at the design level, method level and interpretation level, employing strategies (see Table 3.1) that would enable me to generate quality meta-inferences.

3. The researcher may need to submit new or modified protocols to the IRB (or Ethics Review Board) for each phase of the project:

In order to comply with the ethics requirements set by the Carleton University Research Ethics Board (CUREB), I had to submit an initial ethics application for Phase 1, a Change to Protocol to start data collection in Phase 2, and a completely new Protocol for Phase 3, which included different participants and instruments of data collection. Details related to Ethics considerations are discussed in the next section.

In respect to validity threats, Creswell and Plano Clark (2011) define them as potential problems that "might compromise the merging or connecting of the quantitative

[20] The synergism between quantitative and qualitative methods lies in the fact that the "combined effect (mixing methods) is greater than the sum of individual effects (qualitative or quantitative alone)" (Hall & Howard, 2008, p. 251, as cited in Ziegler & Kang, 2016).

and qualitative strands of the study and the conclusions drawn from their combination" (p. 239). The authors highlight that "researchers should actively use strategies to minimize the validity threats in their studies" (p. 243), which differ depending on the MM design being used. As the proposed multiphase MM study includes exploratory sequential and also convergent parallel designs, careful attention was taken regarding potential validity threats that might compromise data collection, data analysis and interpretation of findings in all three phases. For example, based on strategies for minimizing the threats in MM studies suggested by the authors (p. 241-243), Phase 1 (exploratory) required rigorous procedures to develop and validate the taxonomy, the use of relevant qualitative findings to follow up on quantitatively, and the need to enhance the reliability of the quantitative instrument. As for Phase 3 (convergent), it was necessary to draw QUAN and quan samples from the same population to make data comparable, make logical comparisons of the two results of analysis, and present them in an equal way, not favoring one set of results over the other. Considering the overall multiphase design, a validity consideration was to relate the stages or projects to each other meaningfully.

Furthermore, based on two independent frameworks to address MM validity concerns, one developed by Tashakkori and Teddlie (2008) and the other designed by Onwuegbuzie and Johnson (2006), Ziegler and Kang (2016) synthesized their constructs into a single table (p. 79). The main components include internal validity concerns, related to design quality and interpretive rigor, and external validity concerns, related to inference transferability. A discussion of how I have addressed these concerns will be

presented later in this dissertation, after the presentation and discussion of results from each phase of the study.

6.5 Ethical Considerations

As this research project involved humans as participants, it had to adhere to appropriate ethical standards as outlined in the Tri-Council Policy Statement: Ethical Conduct for Research Involving Human, 2nd edition, and the Carleton University Policies and Procedures for the Ethical Conduct of Research. Therefore, an initial application containing information on methodology, research participants, recruitment, research instruments, and consent documents was submitted to the Carleton University Research Ethics Board (CUREB) in 2015. The board granted ethics clearance for Phase 1 of this study on December 14, 2015 (see Appendix B), under Project Number 103859. In order to start data collection in Phase 2, an application for a Change to Protocol was submitted to CUREB-A and clearance was granted on April 20, 2017 (see Appendix C), followed by a renewal (see Appendix D), also issued on the same date. However, for Phase 3 of this project, a new application had to be prepared and submitted, incorporating details of the new group of participants, different instruments of data collection and consent documents. The board granted ethics clearance for Phase 3 of this study on September 21, 2017 (see Appendix E), under Project Number 107816, followed by a renewal (see Appendix F) issued on September 24, 2018. It is important to mention that, after each year, the Ethics Board requires a report of the project in order to grant its renewal.

Regarding other ethical considerations applicable to the present research, it is worth mentioning the following:

1. As the researcher works for ANAC – Brazil, a pre-existing relationship with Brazilian pilots is a possibility. No publicly available recordings or transcripts involving Brazilian pilots have been selected in the qualitative part of Phase 1. Regarding the questionnaire responses, participants were given the right to withdraw from the study if they perceived this kind of relationship would be an issue.
2. As the researcher has been to many international conferences related to pilots' and ATCOs' language proficiency and communication problems and as a Board member of the International Civil Aviation English Association (ICAEA), there was a possibility of a previous relationship with some of the participants from other countries. They were given the right to withdraw from the study while they were responding to the survey (Phase 1), or at any time up to a month after the focus group or interview took place (Phases 2 and 3).
3. In phase 3, participants were all Aviation English Testing Experts (AETEs), who are employees at the Brazilian Civil Aviation National Agency (ANAC) and from ANAC accredited institutions. The Coordinator of ANAC's Language Proficiency Group assisted in the recruitment of participants, and granted explicit authorization to pilot test the draft tasks at ANAC, in the collection of audio, video and written data (Appendix G).
4. Participants in phase 3 were granted the right to end their participation in the study at any point during the task administration and interview/focus group discussions, if they perceived the previous relationship with the researcher an

issue. However, all data that they provided up to the point of withdrawal would be retained for analysis.

5. In the case of focus groups, although I safeguarded the confidentiality of the discussions to the best of my ability, the nature of focus groups prevents me from guaranteeing that other members of the group will do so. Participants were requested to respect the confidentiality of the other members of the group by not repeating what was said in the focus group to others, and to be aware that other members of the group may not respect their confidentiality.
6. In all phases of this MM research, informed consent (Appendices H, I, J, K, and L) was gained before actual data collection, and identity of participants was protected throughout the study and in publications/presentations.

In the sections above, I presented the study design as a multiphase exploratory MM research, stated the research questions, provided the reader with reasons for adopting MM as the methodological approach, discussed strategies to generate meta-inferences, and considered how to overcome challenges and validity threats posed by this choice of design. Having also discussed ethical considerations, in the next chapters I will describe each phase of this multiphase MM study separately, including participants, instruments, procedures and analyses, as well as present and discuss the results according to the research questions related to each phase.

Chapter 7 Phase 1: Exploring Intercultural Factors in International Pilot-ATCO Communications: Validating a Taxonomy Using Mixed Methods Research

This chapter describes the first phase of the broader multiphase MM study, which is in itself an exploratory MM investigation, including its purpose, research questions, and details of the qualitative and the quantitative strands, ending with a presentation of results and discussions.

7.1 Purpose and Research Questions

The purpose of this initial phase of the study was to explore the multicultural context of aeronautical radiotelephony (RT) communications. Its aims were to: (a) identify intercultural factors that can affect the way pilots and air traffic controllers (ATCOs) interact in the English language, such as differences in communication styles, power distance, reluctance to declare emergency, face saving, and impoliteness, among others; and (b) verify the extent to which experienced professionals perceive the potential threat of those factors to the safety of radiotelephony communications.

In order to address this research problem, the MM exploratory sequential design was used, by expanding the scope of the qualitative strand and further investigating the problem quantitatively with a larger sample. The theory-development variant (Creswell & Plano Clark, 2011) applied "to develop an emergent theory or a taxonomy or classification system, and … [which] examines the prevalence of the findings and/or tests the theory with a larger sample" (p. 90) was chosen, with the priority placed on the initial qualitative phase of the study. Six scenarios were analyzed, consisting of transcripts of pilot-ATCO communications extracts publicly available on aviation-related websites. By thematically coding emerging categories (Saldaña, 2009), a taxonomy of factors was

proposed and used to inform the subsequent quantitative phase, namely, the development of an online survey, aiming to give voice to pilots' and ATCOs' perceptions on the potential threats of intercultural factors to the safety of RT communications. Thus, in this phase, the methods level building strategy (Ziegler & Kang, 2016) was applied, which presupposes the analysis of one strand, in this case QUAL, informing the data collection of the other, the quan strand. By giving voice to study participants to respond to my initial qualitative findings, my aim was to develop what Matusov (2007) called the 'dialogic truth of the research' (p. 328).

Specifically, the goal of Phase 1 was to answer the following two research questions:

> RQ 1.1) What intercultural factors arise from international pilot-ATCO communications that can affect the way they interact in the English language?
>
> RQ 1.2) To what extent do experienced pilots and ATCOs perceive the potential threats of intercultural factors to the safety of radiotelephony communications?

In the two sections which follow, I explain the methods (participants, instruments, procedures, and analysis) used in the QUAL and QUAN strands of Phase 1. Having presented the methods, I then discuss the results of the two strands in Section 7.4.

7.2 Qualitative Strand

7.2.1 Instruments.

The first step of this MM study was a qualitative exploration of culturally influenced categories that arise from pilot-ATCO interactions, through the analysis of audio recordings and transcripts of radiotelephony communication extracts that are publicly available on aviation-related websites. From www.liveatc.net, extracts of live recordings that are considered relevant to be transcribed are usually made available in the format of YouTube videos at www.youtube.com. From www.planecrashinfo.com, we can find transcripts from previous aviation incidents and accidents. A purposive sample was drawn of six scenarios involving native speakers (NSs) and non-native speakers (NNSs) of English, in which 'something' unexpected happened that affected their communicative behavior. As will be demonstrated throughout data analysis in the QUAL and quan strands, none of the scenarios were considered irrelevant or unrealistic by aviation stakeholders.

As no background information related to the first language of pilots and ATCOs from the selected transcripts was available at the aviation-related websites, it was not possible to define their status as either NS or NNS of English. Therefore, I considered the country of the airline companies and location of ATC facilities involved in each communication extract as a guide to classify the interaction type (see Table 7.1). This distinction is necessary because ICAO testing policy treats NSs and NNSs of English differently, as explained before. Thus, I tried to select examples of radiotelephony communications involving NS-NS, NS-NNS and NNS-NNS as data sources, in an

attempt to capture how cultural differences may impact these distinct types of interactions.

Table 7.1. Selection of scenarios according to type of interaction

Scenarios	Country - Airline companies	ATC facility	Type of interaction considered
Scenario 1	England	Dublin Tower, Ireland	NS-NS
Scenario 2	Singapore	Kennedy Ground, USA	NNS*-NS
Scenario 3	Colombia	New York Approach, USA	NNS-NS
Scenario 4	United Arab Emirates	Kennedy Ground, USA	NNS-NNS*
Scenario 5	France	Ezeiza Approach, Argentina	NNS-NNS
Scenario 6	USA	John Wayne Tower, USA	NS-NS

Note. *proficient or very proficient in English, but with a foreign accent to a native English speaker's ear.

The transcriptions of all six scenarios are provided in Appendices M to R with a more precise reference for each, and they consist of transcriptions available on the aviation-related websites, such as www.youtube.com, where we can find extracts of recordings from www.liveatc.net with subtitles included, and www.planecrashinfo.com, where we can find transcripts from previous incidents and accidents. The names of the airline companies involved have been removed. Extracts of those transcriptions were selected for inter-coder reliability and incorporated into the 'Results and discussions' section, as a way to illustrate my comments.

7.2.2 Procedures.

Selection of scenarios from aviation-related websites took place in November 2015, by listening to a number of audio recording and reading the transcriptions of pilot-ATCO international radiotelephony communications. It was a purposive sampling, in the sense that I was looking for six scenarios in which culturally related factors, beyond language proficiency, could have an impact on the outcome of the communication. Thus,

it was important to also verify if the selection of scenarios covered the range of interaction types necessary to address the research problem, i.e., a mix of language backgrounds. Following that, familiarization with the dataset was crucial, by reading through the transcripts and organizing the data (written transcripts) into tables in order to proceed with coding. Data were coded in a systematic way, including first cycle coding and second cycle coding (see definitions in the Glossary). Then a training package was prepared and sent to two other coders for independent coding, providing an overview of the purpose of the study, research questions and an example of my own coding from the first cycle. The codes assigned by Coders 2 and 3 were organized according to similarities with my own codes. Grouping of similar codes and verification of the number of occurrences for each coder was undertaken. Correlations among coders and inter-coder reliability were calculated. Codes were organized into sub-categories and categories, generating an initial taxonomy of intercultural factors.

7.2.3 Analysis.

The analysis of qualitative data followed Saldaña's (2009) methods of coding, aiming to interpret what was said explicitly but also implicitly (e.g., intentions, emotions, perceptions, beliefs). For the First Cycle Coding, I explored the data by breaking it down into pilot's and ATCO's *utterances*, my unit of analysis, which are limited by the change of speaking subjects and represent a link in the chain of speech communication, in relation to both previous utterances and to subsequent ones (Bakhtin, 1986). Scollon and Scollon (2001) put it this way: "as people communicate with each other, the process of interpretation moves through real time from utterance to utterance in an ongoing process of interpretation" (p. 83). By employing Process Coding (-ing words), I realized that I

was using the communicative language functions associated with aeronautical RT communications (e.g., giving order, requesting confirmation, giving authorization, etc.) already compiled in a list of more than 100 functions, published by ICAO (2004). Indeed, this first step was Provisional Coding, a top-down process, i.e., the use of a predetermined set of codes generated from the literature review and/or from previous research findings.

However, to answer my research questions I needed to go deeper into the participants' emotions, values, conflicts, and judgements which highlighted the need to employ Values Coding, from a bottom-up perspective, i.e., constructed during the coding of data. According to Saldaña (2009), this affective coding method is appropriate to "explore cultural values and intrapersonal and interpersonal participant experiences and actions in case studies" (p. 90). Nonetheless, at the very early stages of this second step, I noticed that the richness and complexity of the interactions required, most of the time, the assignment of two or more codes to a single utterance, which also led to the use of Simultaneous Coding. Table 7.2 shows an example of First Cycle Coding.

Table 7.2. Example of First Cycle Coding

	Scenario 3	**First Cycle Coding**	
	Transcript	**Process Coding / Provisional Coding**	**Values Coding / Simultaneous Coding**
23 ATCO	[ ] zero five two heavy, turn left, heading zero seven zero.	Giving order	Directness
24 PILOT	Heading zero seven zero, [ ] zero five two heavy.	Reading back	Compliance with order
25 ATCO	And [ ] zero five two heavy, ah, I'm going to bring you about fifteen miles northeast, and then turn you back onto the approach, is that fine with you and your fuel?	Stating intentions / Requesting confirmation	Concern for efficiency/Concern for potential complications
26 PILOT	I guess so, thank you very much.	Giving confirmation	Failure to declare emergency
		Thanking	Deferential - avoiding disagreement
			Deferential - using excessive politeness in emergencies

During the Second Cycle Coding, Pattern Coding was employed to organize my First Cycle codes into sub-categories and categories according to similarity (Appendix S). At this time, it was also necessary to draw on Maxwell and Miller's (2008) connecting strategies, by considering a more holistic dimension in the interpretation of data and approaching my data analysis as an iterative process. Looking for antecedents and consequences, I benefited from considering any unexpected relationships among the 14 sub-categories that had been identified and their contextual connections.

As previously discussed in Chapter 3, the literature review, this analysis was informed by theories and concepts associated with intercultural communication/awareness and pragmatics. That is, theories of cross-cultural communication (Gudykunst, Lee, Nishida, & Ogawa, 2005), including face-negotiation theory, conversational constraints theory, expectancy violation theory, anxiety/uncertainty management theory, and communication accommodation theory. Moreover, Hofstede's cultural dimensions, i.e., individualism-collectivism, power distance, masculinity-femininity and uncertainty avoidance were also drawn on to inform the analysis. The same applies to notions of face, i.e., "notions of being embarrassed or humiliated … that can be lost, maintained, or enhanced" (Brown & Levinson, 1987, p. 63), and impoliteness strategies (Culpeper, 1996), such as being unsympathetic, seeking disagreement, making the other feel uncomfortable, and associating the other with a negative aspect. Those theories and conceptualizations enabled me to assign "Values" codes and to thematically organize them into a provisional taxonomy of intercultural factors that may affect pilot-ATCO international communications (see Table 7.11).

In order to increase the reliability and validity of my research findings, I asked two other coders to independently code excerpts of the selected pilot-ATCO interactions (see Tables 7.5 to 7.10) that I had already coded. One of the coders (Coder 2) was an NS of English and experienced qualitative researcher, while the other (Coder 3) was an NNS of English and qualitative researcher with an aviation background. They coded only for Values, as these were the codes most relevant to the present analysis in order to answer the research questions. The codes assigned by Coder 2 and Coder 3 were organized according to similarities with my own codes (Coder 1). The process of grouping similar codes was followed by the verification of their occurrences for each coder (see example in Table 7.3), across the 14 sub-categories identified before, in order to calculate the correlation among the three coders and the inter-coder reliability. The correlation was run in the SPSS Software, version 23, which gave the following results (Pearson's r): between coders 1 and 2, r = .78; between coders 2 and 3, r = .81; and between coders 1 and 3, r = .80. This indicates a significant correlation between the coders. In addition, the reliability analysis was run using the intra-class correlation coefficient and the absolute comparison option. The test was based on a 95% confidence level. Cronbach's Alpha provided a reliability measure of .92. The results of the intra-class correlation coefficient analysis provided information on the reliability and predictability of the variance: both single measures (.80) and average intra-class correlation measures (.92) were above .70, indicating a positive and significant correlation in the data. The significance level of $p < .0005$ indicated a high level of reliability that the positive correlation among the coders was not due to random chance over 95% of the time.

Table 7.3. Example of Second Cycle Coding – grouping of codes with occurrences for Coders 1, 2 and 3

Category	Sub-category	Coder 1	Coder 2	Coder 3
Non-collaborative behavior	Unprofessional attitude	Unwillingness to help (1)	Unhelpful (2)	Unprofessional attitude – did not do what the pilot asked him to do (1) / Unprofessional attitude: the pilot needed help and the ATC is there to help them, no matter how busy it is (1)
		Buying time (2)	Stalling (1)	Attempt to gain time (1)
		No time to support x time to reprimand (1)		
		Non-accommodating to interlocutor's needs (3)		Native speaker: Not trying to accommodate (1)
		Failure to question severity of problem (1)	Unconcerned (1)	Lack of concern for safety/efficiency (4)
			Unprofessional (8)	Lack of professionalism (5) / Unprofessionalism (why did he clear the pilot to ILS 35 if it is out of service?) (1)
		Impoliteness: seeking disagreement (1)	Gratuitous (1) / Unnecessarily nasty (1)	Provocativeness (1)
		Impoliteness: Making the other feel uncomfortable (1)		
		Impoliteness: accusing/blaming the other (5)		Accusing the other (3)
		Threatening (3)	Threatening (3)	Threatening (3)
		Total: (18)	Total: (17)	Total: (21)

The emerging sub-categories from the taxonomy informed the development of a quantitative survey to be administered to a larger sample of participants. This was the purpose of the quantitative strand of Phase 1, detailed in the next section.

7.3 Quantitative Strand

7.3.1 Participants.

Participants in this strand were the key stakeholders in international radiotelephony communications, pilots and ATCOs, both NS and NNS of English, males and females, experienced in international operations.

Fully completed questionnaires were received from 38 professionals: 23 pilots and 15 air traffic control officers. In terms of language background, seven participants speak English as a first language while 31 do not. Their nationalities are represented in the following way: American (5), Argentinian (2), Australian (1), Brazilian (20), British (2), Cape Verdean (1), Croatian (1), French (1), Irish (1), Japanese, (1), Nigerian (1), Portuguese (1), and Spanish (1). Figure 7.1 shows the percentage of pilots and ATCOs, males and females, NSs and NNSs of English in the sample and participants' distribution according to their experience either as a pilot or an ATCO.

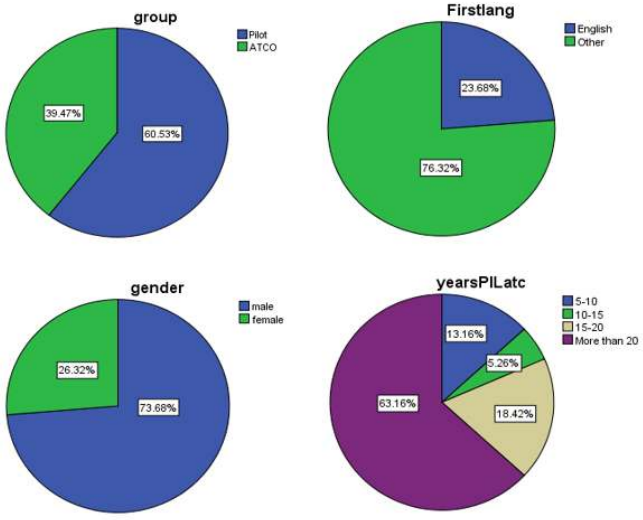


Figure 7.1. Percentage of survey participants according to group, first language, gender and years of experience (N= 38)

7.3.2 Instruments.

A questionnaire (Appendix T) was designed to investigate the constructs highlighted in the proposed taxonomy. This was accomplished by transforming the codes that originated within each sub-category into questions as a building strategy. Rigorous development of a quantitative instrument requires a number of steps. Thus, the first versions of the questionnaire were reviewed by an experienced questionnaire designer and also by the two groups of stakeholders, first by ATCOs, and secondly by pilots. Their feedback generated some amendments that led to the final version of the questionnaire.

The questionnaire contains, in its first section, nine items eliciting background/demographic information. In section II, items 10-17 elicit responses regarding participants' perceptions of expected practices in international communications, which refer to more positive attitudes or behavior. These items are sub-divided into three, the first two being Likert-scale questions (scale 1 to 6): (1) How often do you encounter pilots/ATCO who…?; (2) In your view, how important is this?; and (3) Please comment (a qualitative component asking participants to comment on the same issue). In Section III, however, items 18-34 elicit responses on the participants' perceptions of not so desired attitudes and behaviors in radiotelephony communication, including their importance as a potential threat to safety. These items are also sub-divided into three, the first two being Likert-scale questions (scale 1 to 6): (1) How often do you encounter pilots/ATCO who…?; (2) How important, in your view, were these events as potential threats to safety?; and (3) Please comment (a qualitative component asking participants to comment on the same issue). Item 35 is a final open-ended question regarding additional comments on other issues that may affect the safety of pilot-ATC

communications. Appendix T clarifies how the questionnaire was structured and provides a complete list of the questions.

In order to investigate the quality of this quantitative instrument in terms of internal consistency, which may impact on the quality and reliability of the data collected, a reliability analysis was carried out on Sections 2 and 3 together, comprising 50 items. SPSS software, version 23, was used, and Cronbach's alpha showed the questionnaire to reach acceptable reliability, $\alpha = 0.881$.

The Likert-scale items in Sections II and III of the questionnaire are operational definitions of each of the sub-categories, or constructs, identified in the QUAL strand. Thus, Table 7.4 was designed to portray the use of integration strategies, in the form of a joint display (Creswell and Plano Clark, 2011), by aligning the qualitative sub-categories with the questions in the quantitative instrument.

Table 7.4. Alignment of the sub-categories with the online survey questions

Theme	**Categories**	**Sub-Categories**	**Operational Definitions in the Questionnaire**
Intercultural factors in international pilot-ATCO communications	Power Distance	Power relations	Q18, Q19
		Deferential role	Q20, Q21
	Face-work strategies	Self-face concern	Q22
		Mutual-face concern	Q12
	Conflict management	Conflictual direction	Q23, Q24
		Neutral direction	Q10, Q11
		Expectancy violations	Q25
	Communication styles	Directness	Q13
		Indirectness	Q26, Q27
	Non-collaborative behavior	Unprofessional tone	Q28, Q29
		Unprofessional attitude	Q30, Q31, Q32
		Non-compliance with rules	Q33,Q34
	Collaborative behavior	Professional attitude	Q14, Q15, Q16
		Supportiveness	Q17

7.3.3 Procedures.

In this phase of the project, recruitment of participants occurred online, via email. Contact information for the first invitations were obtained from pilots and ATCOs' business cards exchanged with the researcher during previous national/international events that she had participated in or organized, or from the lists of participants of those events. In a snowballing process, the researcher asked those who accepted to participate if they would be willing to pass information about the study to other potential eligible participants over whom they were not in a position of authority.

Participation in the quantitative strand entailed responding to an online questionnaire. Participants were informed about the details of the project and actions to preserve their anonymity and confidentiality of research data when they received the email invitation, but also before starting the online survey. By submitting their survey responses participants provided their informed consent. The study was considered minimal risk research and this phase began only after receiving approval from the Carleton University Research Ethics Board (see Appendix B).

7.3.4 Analysis.

Quantitative data from participants' responses were inserted into the SPSS software, version 23 and analyzed using descriptive statistics and frequency distribution (Vogt, 2007; Larson-Hall, 2016). The purpose was to extract mean values and frequencies, as a strategy of data reduction, by reducing the quantitative findings into manageable chunks of information. The categorical variables (e.g., group, first language,

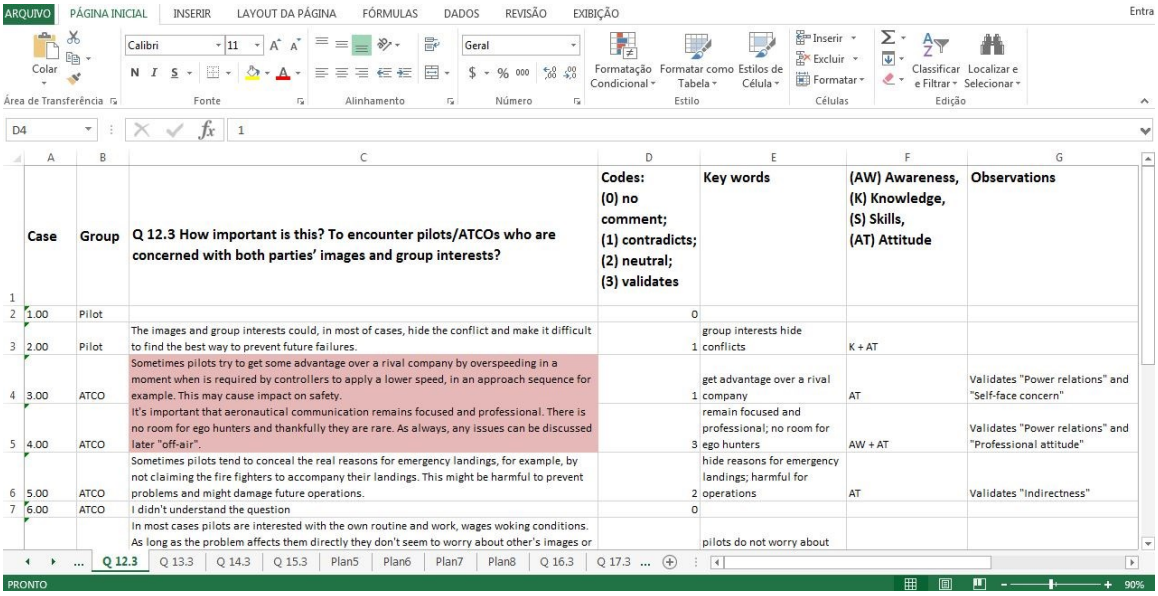
gender) helped the researcher to describe the characteristics of the sample in terms of respondents, as detailed in Section 7.3.1. As for the continuous variables, they enabled the researcher to have an overview of the responses, with the aim of selecting the most striking and relevant features for the present discussion according to the research questions.

Analysis of the data considered all responses together, but also compared perspectives from different groups of participants, i.e., pilots vs. ATCOs, NS vs. NNS of English, and male vs. female. Their perceptions on the frequency of occurrence of the situations presented in the questions, as well as their opinions on the importance of those issues to the safety of RT communications, were the focus of the analysis.

For triangulation purposes, open-ended responses were thematically coded with the final goal of validating the questionnaire, validating the provisional taxonomy and addressing the research questions. Following Saldaña (2009) coding included four steps:

1. Magnitude Coding, in the following way: (0) no comment provided; (1) comment contradicts the sub-category; (2) comment is neutral; (3) comment validates the sub-category;
2. Initial Coding: selecting key-words to summarize each comment;
3. Provisional Coding I: classifying each comment according to the dimensions of awareness (AW), knowledge (K), skills (S) and/or attitudes (AT), which are the core features of the matrix of the construct specification built in Phase 2 of this multiphase MM research study;
4. Provisional Coding II: identifying if, within a comment, reference to other sub-categories from the provisional taxonomy was present.

By doing so, not only data reduction was accomplished, but also data transformation, i.e., the conversion of one dataset into the other (Onwuegbuzie & Teddlie, 2003), by assessing the quality of the open-ended comments and applying magnitude codes to them (0 to 3). The purpose was to facilitate the fusion of both datasets in order to conduct data comparison of qualitative and quantitative findings which ultimately would assist in the validation of the provisional taxonomy. A snapshot of how this was carried out is shown in Figure 7.2.



	Case	Group	Q 12.3 How important is this? To encounter pilots/ATCOs who are concerned with both parties' images and group interests?	Codes: (0) no comment; (1) contradicts; (2) neutral; (3) validates	Key words	(AW) Awareness, (K) Knowledge, (S) Skills, (AT) Attitude	Observations
2	1.00	Pilot		0			
3	2.00	Pilot	The images and group interests could, in most of cases, hide the conflict and make it difficult to find the best way to prevent future failures.	1	group interests hide conflicts	K + AT	
4	3.00	ATCO	Sometimes pilots try to get some advantage over a rival company by overspeeding in a moment when is required by controllers to apply a lower speed, in an approach sequence for example. This may cause impact on safety.	1	get advantage over a rival company	AT	Validates "Power relations" and "Self-face concern"
5	4.00	ATCO	It's important that aeronautical communication remains focused and professional. There is no room for ego hunters and thankfully they are rare. As always, any issues can be discussed later "off-air".	3	remain focused and professional; no room for ego hunters	AW + AT	Validates "Power relations" and "Professional attitude"
6	5.00	ATCO	Sometimes pilots tend to conceal the real reasons for emergency landings, for example, by not claiming the fire fighters to accompany their landings. This might be harmful to prevent problems and might damage future operations.	2	hide reasons for emergency landings; harmful for operations	AT	Validates "Indirectness"
7	6.00	ATCO	I didn't understand the question	0			
			In most cases pilots are interested with the own routine and work, wages woking conditions. As long as the problem affects them directly they don't seem to worry about other's images or		pilots do not worry about		

Figure 7.2. Coding of survey open-ended responses

7.4 Results and Discussion

The results from the analysis are presented and discussed in this section in relation to each research question.

RQ 1.1) What intercultural factors arise from international pilot-ATCO communications that can affect the way they interact in the English language?

Evidence to answer this question was collected mainly from the qualitative strand of the study; however, findings from the quantitative strand and from the analysis of survey open-ended responses also substantiate my discussions.

First, the interpretation of the scenarios introduced in Section 7.2.1 suggests how the intercultural factors identified may affect pilots' and ATCOs' discourse in different ways.

Scenario 1: NS-NS. The first example is an interaction between two very proficient speakers of English, presumably from different nationalities. The extract analyzed, shown in Table 7.5, involves a male pilot who does not comply with the female ATCO's orders and states readiness, when in fact he was not ready for departure, causing trouble for the ATCO and other aircraft in a busy airport.

Table 7.5. Excerpt from Scenario 1

		Scenario 1 – Transcript (Available at https://www.youtube.com/watch?v=uWg7IpphPc8)
5	ATCO	[ ] 845 proceed onto runway 28 now and vacate right onto runway 34, there is traffic behind you waiting to depart.
6	PILOT	Yep, we've just got a phone call standby…
7	PILOT	And [ ] 845, we are actually fully ready.
8	ATCO	[ ] 845, line up runway 28 and wait.
9	PILOT	Line up and wait 28, [ ] 845.
10	ATCO	And for future reference [ ] 845, err, I suggest you advise the, err, ground controller…that you are unable to take departure yet and you shouldn't really be taxiing out to E1 when you're not ready for departure…because there is traffic behind, waiting, that is ready.
11	PILOT	Standby.
12	ATCO	[ ] 845, are you fully ready for departure?
13	PILOT	Affirm [ ] 845, we're just doing the checks whilst err, whilst you keep talking over us, standby.
14	ATCO	Ok, negative! Turn right please onto runway 34. That's the third time I've asked you to vacate onto runway 34 if you are not ready, turn right onto runway 34.
15	PILOT	Madam, we are fully ready, we're just trying to complete the checklists, but err, you just keep interrupting our checks, standby.
16	ATCO	Negative! Turn right onto runway 34, I've asked you three times now! Turn right, to vacate onto runway 34.

In this scenario, the pilot's concern with his own interests and resultant non-compliance with the ATCO's order triggers a competing conflict style and unprofessional

tone, while at the same time reveals relations of power. This interaction demonstrates how the female ATCO, concerned with efficiency, resorts to her own authority in an effort to solve the situation.

Scenario 2: NS-NNS. In this interaction involving a NS air traffic controller and a NNS of English pilot, after four previous contacts requesting the ATCO to confirm information, the dialogue transcribed in Table 7.6 shows the way the ATCO reacted to a new request for clarification from the same pilot.

Table 7.6. Excerpt from Scenario 2

		Scenario 2 (Available at https://www.youtube.com/watch?v=2t_NT7aUrE0)
33	PILOT	And…[ ] 7997, uh…just want to confirm you the point before Hartford, could you give me the name again, please?
34	ATCO	[ ] you gonna kill me, what do you want now?
35	PILOT	Okay, ground, (…) checking (…) our routing just the point before Hartford and Partham, could you give me the point again?
36	ATCO	Now sir, you've been given a change of frequency, you'd be talking to the same guy all night long, see? You're going back for a million questions, but let's go over it: MERIT intersection, that's spelled: Mike Echo Romeo India Tango; direct Hartford, that is Hotel Foxtrot Delta; direct Partham, that is Papa Uniform Tango, and then as filed. Do you have any further questions about your route, your taxi route, the route you gotta fly, anything else?
37	PILOT	Not for now, sir, thanks.
38	ATCO	…now. I'm sure in 30 seconds you'll have another one, but continue to the runway.
39	PILOT	Okay.

Scenario 2 contrasts the ATCO's use of non-standard phraseology and condescending language, combined with an aggressive conflict style, impatient/sarcastic tone, and judgmental attitude, with the more accommodating conflict style of a less-confident pilot, who constantly seeks confirmation and attempts to restore face loss.

Scenario 3: NS-NNS. This interaction between a NS air traffic controller and a NNS of English pilot, as shown in Table 7.7, is part of the transcript of a fatal accident. A few utterances from the cockpit interaction involving the Captain, the First Officer (FO) and the Flight Engineer (FE) are included to add understanding of the situation (for the

complete transcript follow the link provided on Table 7.7). One of the contributing factors to this accident was the pilot's (First Officer) deferential and submissive communicative style, which may have prevented him from declaring an emergency.

Table 7.7. Excerpt from Scenario 3

		Scenario 3 (Available at http://www.planecrashinfo.com/cvr900125.htm)
14	PILOT	Approach, [ ] zero five, ah, two heavy, we just missed a missed approach, and ah, we're maintaining two thousand and five on the...
15	APPR	[ ] zero five two heavy, [ ], good evening, climb and maintain three thousand.
16	Captain	*(Advise him we don't have fuel.)*
17	PILOT	Climb and maintain three thousand, and ah, we're running out of fuel, sir.
18	APPR	Okay, fly heading zero eight zero.
19	PILOT	Flying heading zero eight zero, climb to three thousand.
20	Captain	*(Did you already advise that we don't have fuel?)*
21	FO	*(Yes sir, I already advise him, hundred and eighty on the heading. We are going to maintain three thousand feet, and he's going to get us back.)*
22		*(Some time later…)*
23	APPR	[ ] zero five two heavy, turn left, heading zero seven zero.
24	PILOT	Heading zero seven zero, [ ] zero five two heavy.
25	APPR	And [ ] zero five two heavy, ah, I'm going to bring you about fifteen miles northeast, and then turn you back onto the approach, is that fine with you and your fuel?
26	PILOT:	I guess so, thank you very much.
27	Captain	*(What did he say?)*
28	FE	*(The guy is angry.)*
29		(Some time later…)
30	PILOT	Ah, can you give us a final now? [ ] zero five two heavy.
31	APPR	[ ] zero five two, affirmative sir, turn left, heading zero four zero.
32	PILOT	[ ] zero five two heavy, left turn two five zero, and ah, we're cleared for ILS.
33	APPR	[ ] fifty two, climb and maintain three thousand.
34	PILOT	Ah, negative sir. We just running out of fuel. We okay three thousand. Now okay.
35	APPR	Okay, turn left, heading three one zero sir.
36	PILOT	Three one zero, [ ] zero five two.
37		(Some time later…)
38	APPR	[ ] fifty two, fly heading of three six zero please.
39	PILOT	Okay, we'll maintain three six zero now.
40	APPR	Okay, and you're number two for the approach. I just have to give you enough room so you can make it without, ah, having to come out again.
41	PILOT	Okay, we're number two and flying three six zero now.
42	APPR	Thank you sir.
43	APPR	[ ] zero five two heavy, turn left, heading three three zero.
44	PILOT	Three three zero on the heading, [ ] zero five two.
45	APPR	[ ] zero five two, turn left, heading two five zero, intercept the localizer.
46	PILOT	Roger.
47		(Some time later…)
48	PILOT	[ ] zero five two, we just, ah, lost two engines and, ah, we need priority, please.
49	ATCO	[ ] zero five two, turn left, heading two five zero, intercept the localizer.
50	PILOT	Roger.

In scenario 3, the deferential style of the First Officer (FO) communicating with the ATCO (and also in relation to his Captain), was reflected by his agreement to act despite operational limitations and by his failure to declare an emergency. Issues of power distance, involving compliance and acceptance of authority and power rankings, seemingly inhibited the FO from approaching the ATCO directly and assertively stating his needs in a busy and native-speaking environment, which in turn prevented the ATCO from acting appropriately to the seriousness of the situation.

Scenario 4: NS-NNS. This example represents a conversation between a NNS of English pilot and a native-like ATCO who asks many questions in sequence of the pilot but without accommodating to the pilot's communicative needs. Differences in expectations give rise to issues of impoliteness and power relations (see Table 7.8).

Table 7.8. Excerpt from Scenario 4

		Scenario 4 - Transcript (Available at https://www.youtube.com/watch?v=ZWOOKQlEe5s)
1	ATCO	[ ] 503, where you park?
2	PILOT	Bravo 28, Sir.
3	ATCO	Not taxiway, the LETTER!
4	PILOT	Oh negative sir, we are on 22R holding short of Foxtrot.
5	ATCO	What taxiway do you enter the ramp?
	PILOT	Okay, so we just exit the runway and we're holding short of Foxtrot on 22R.
6	ATCO	You are not listening to what I'm asking you. What taxiway do you enter the ramp?
7	PILOT	I'm not on the ramp yet, sir.
8	ATCO	What taxiway do you enter the ramp. Tell me. What letter?
9	PILOT	Okay we can enter at KILO for [ ] 503.
10	ATCO	That's what I need get out of you. We talked like 6 times. Straight ahead and hold short of HOTEL, sir.
11	PILOT	Straight ahead, hold short of HOTEL, roger.
		(Some time later...)
12	ATCO	[…] 503 follow […] 222, hold short Juliette on the runway.
13	PILOT	Yes, we'll follow the […], and next time I would like you to be polite with me. Thank you.
14	ATCO	Okay, but if I got to talk to you 6 times, and I got all other people I got to talk to, and you don't know what I'm saying.
15	PILOT	(…)…nice day, polite with me. All right?
16	ATCO	Are you impolite with me?
17	PILOT	I'll make a report.
18	ATCO	Go ahead!

This situation reveals an impatient and aggressive ATCO who fails to use standard phraseology or to accommodate to the pilot's communicative needs, but instead blames the pilot for his lack of comprehension. The perceived lack of collaboration and conflicting style of the ATCO may also have raised the pilot's levels of anxiety, who, despite having his face under attack, attempts to restore neutral communication. However, the pilot's reaction to the ATCO's disrespectful and impolite behavior triggers an even more aggressive response.

Scenario 5: NNS-NNS. This scenario involves a very proficient NNS of English pilot interacting with the ATCO with limited English proficiency, as shown in Table 7.9. The ATCO had previously authorized the pilot to land via the Instrument Landing System (ILS) of runway 35, which in fact had been out of service for years. This generated an operational limitation for the aircraft with regards to fuel reserves and a far from effective interaction between the pilot and the ATCO.

Table 7.9. Excerpt from Scenario 5

		Scenario 5 - Transcript (Available at https://www.youtube.com/watch?v=Lpxc0B-t5tM)
1	PILOT	Could you please confirm we are cleared to ILS 35?
2	ATCO	Authorized ILS 35.
3	PILOT	[ ] 417, could you confirm the ILS is serviceable, we are not receiving it.
4	ATCO	ILS to 35….out of service, 417
5	PILOT	417, Roger. We are unable to continue this approach. I inquired about the ILS on runway 35 before. I want you to know that I asked you many times if the ILS 35 was operative and you said it was. Tell me, how can it not function anymore?
6	ATCO	([ATCO] does not respond) (Some time later…)
7	PILOT	I am declaring a low fuel.
8	ATCO	([ATCO] does not respond)
9	PILOT	[ATCO], [ ] 417, did you copy?
10	ATCO	Affirmative, Sir. Report stablished on localizer…
11	PILOT	OK, we are taking heading 310 and I would like to see the authorities on the ground.
12	ATCO	([ATCO] does not respond)
13	PILOT	Did you copy my request about seeing the people on the ground?
14	ATCO	([ATCO] does not respond)

15	PILOT	Still low fuel. So I want priority for landing.
16	ATCO	([ATCO] does not respond)
17	PILOT	Did you copy, [ ] 417?
18	ATCO	Affirmative, 417.
19	PILOT	I will be filling a report about this charade because it's quite amazing what happened.
20	ATCO	Can you repeat, please?
21	PILOT	I'll be filling a report against you, what happened is truly, truly amazing.

In Scenario 5, the utterances show the ATCO's less powerful role due to language limitations, his fear of losing face and avoiding conflict style, as opposed to the dominant pilot's directness, concern for clarity and efficiency, and his resort to higher authority after getting involved in an operational complication.

Scenario 6: NS-NS. The last scenario involves the interaction of two NS of English, which shows how aviation professionals may react when their expectations are violated (see Table 7.10).

Table 7.10. Excerpt from Scenario 6

		Scenario 6 - Transcript (Available at https://www.youtube.com/watch?v=JYROSTV_KVG)
1	PILOT	And [ ], we are VFR under the clouds right now. And if you could give me a (inaudible).
2	ATCO	You're not familiar with this airspace?
3	PILOT	Yes sir, I'm very familiar with this airspace. But just coming through the clouds now it would be easier if you just give me my heading for a moment.
4	ATCO	What kind of NAV equipment do you have on board?
5	PILOT	Slant Uniform, VOR sir.
6	ATCO	[ ] fly heading 150. Vectors Mile Square Park.
7	PILOT	Okay, we are currently 150 sir. Thank you sir, just wanted a little help. Thank you.
8	ATCO	Well, let me give you some advice. We are really busy. We've got one controller working all the airspace and a lot of inbounds coming in, the last airliners coming into John Wayne. I probably don't always have time to hold your hand. Sorry to say that, but that's the truth.
9	PILOT	25 years I have been flying this airspace sir. I've never had a controller talk to me like that.
10	ATCO	Well, you are welcome to call me on the phone.
11	PILOT	Love to!

Finally, scenario 6 displays how an ATCO, unwilling to cooperate and unnecessarily conflictual, violates the pilot's expectations and triggers a reaction which is far from the expected professional tone.

Apart from the interpretation of the six scenarios which highlighted the presence of intercultural factors throughout the interactions, a second comment needs to be made based on the theories of cultural variability (Gudykunst et al., 2005) discussed in Chapter 3. These theories helped to recognize cultural features generally associated with individualistic cultures, i.e., societies in which the interests of the individual predominates over the interests of the group (Hofstede, 1991), such as directness, concern for efficiency, self-face concern, aggressive conflict style, dominant style of communication, unwillingness to help or to accommodate to interlocutor's needs, and aggressive response to expectancy violations. Furthermore, characteristics generally linked to collectivistic cultures, i.e., societies in which the interests of the group predominates over the interests of the individual (Hofstede, 1991) were also detected, such as indirectness, concern for avoiding negative evaluation by the hearer, mutual-face concern, avoiding or accommodating conflict style, deferential style of communication, supportiveness, and response to expectancy violations with withdrawal. These features appeared in both pilots' and ATCOs' utterances but in most cases, they were associated with the communicative behavior of very proficient speakers of English and of less proficient ones, respectively. This association indicates a pattern observed in the sample analyzed, but further investigations are necessary to confirm it as a general rule of thumb. I am not arguing here that individualistic features are positive and collectivistic are negative, or the opposite. Instead, they should be considered as different characteristics

which may positively or negatively affect the way pilots and ATCOs communicate, depending on the specific moment of interaction and whom they are communicating with.

Moreover, it is important to mention how the cultural dimensions of power distance and inequality may affect the outcome of pilot-ATCO interactions. These may be due to perceived superiority vs. inferiority (e.g., Scenario 3), differences regarding participants' professions (all Scenarios: pilot vs. ATCO), gender (e.g., Scenario 1), differing levels of language proficiency (e.g., Scenario 5), or to a combination of all these features. Scollon and Scollon (2001) remind us that among different cultures relationships of face politeness are also treated differently, leading to the use of different rhetorical strategies. Scollon and Scollon (2001) give as an example of rhetorical strategies the use of deductive and inductive strategies for introducing topics. They state that "there is a tendency for Asians to show deference or respect in interactions with non-intimates, in contrast to westerners, who tend to emphasize egalitarian interpersonal relationships" (p. 24). The authors further explain that these differences in rhetorical strategies can lead to the development of differences in interpersonal power, thus creating sources of "power disparities in discourse" (p. 24). In some of the scenarios analyzed, it was possible to note the relation between the more powerful participant and the freedom to use impoliteness strategies (e.g., Scenario 2 and Scenario 4) and, at the other extreme, how a deferential and submissive style led to a fatal accident (e.g., Scenario 3). Although pilots and ATCOs belong to different professions, instances of pride or other hidden motives should not lead to impoliteness, unprofessional tone nor to unwillingness to cooperate, as observed in some of the scenarios described above. In regards to the

organizational level, the safety culture of each organization can also shape certain attitudes, beliefs and values. Tolerance for non-compliance with orders or rules and for the use of non-standard phraseology are some of the negative examples.

In addition, violations to the expected flow of communication or to expected attitudes may increase the levels of anxiety and uncertainty of pilots and ATCOs. As a result, conflicts may arise, as noted in some of the transcripts analyzed (e.g., Scenario 4 and Scenario 6). However, the way participants managed conflict situations, the face-work strategies they used, and how they continued communicating after a face-threating act, varied according to their own cultures. This may be because pilots and ATCOs have their own set of expectancies according to their national, professional and organizational cultures, but also according to their individual values and assumptions. As Scollon and Scollon (2001) state, problems arise in international communication in English "when participants in a conversation hold different assumptions because of membership in different groups … The problem is, of course, exacerbated when communication is across more than one group boundary such as culture and gender" (p. 83) (e.g., Scenario 1). Hofstede and Hofstede (2005) also argue that "as almost everyone belongs to a number of different groups and categories at the same time, we unavoidably carry several layers of mental programming within ourselves, corresponding to different levels of culture" (p. 10-11), such as national, regional, gender, generation, social class, and organizational or corporate levels. However, the authors further state that "conflicting mental programs within people make it difficult to anticipate their behavior in a new situation" (p. 11). Added to that, in terms of processes of interpretation, members of different groups "will find it more difficult to draw inferences about what the other person means … When our

inferences about what others mean are wrong, we can form a negative opinion of the other, which might increase the likelihood of a misunderstanding" (Scollon & Scollon, 2001, p. 22). Therefore, the combination of these cultural identities, or the conflict among them, may influence what people say, how they say it, the responses they expect, and how they react to previous utterances.

As stated previously, the outcome of the qualitative strand was a preliminary taxonomy of intercultural factors that may affect pilot-ATCO international communications. Table 7.11 presents how it was organized into six categories and fourteen sub-categories. A more detailed presentation of the codes that originated each sub-category is included in Appendix S.

Table 7.11. Preliminary taxonomy of intercultural factors that may affect pilot-ATCO international communications

Theme	Categories	Sub-Categories
Intercultural factors in international pilot-ATCO communications	Power Distance	Power relations
		Deferential role
	Face-work strategies	Self-face concern
		Mutual-face concern
	Conflict management	Conflictual direction
		Neutral direction
		Expectancy violations
	Communication styles	Directness
		Indirectness
	Non-collaborative behavior	Unprofessional tone
		Unprofessional attitude
		Non-compliance with rules
	Collaborative behavior	Professional attitude
		Supportiveness

This preliminary taxonomy is not an exhaustive list of possible culturally influenced factors that may impact aviation safety, but it does contain a useful inventory of elements against which intercultural communications in aviation can be analyzed.

Considering the limited number of scenarios analyzed, combined with the fact that they were not randomly selected, one might say that these were just localized examples, or even extreme instances that do not reflect the reality of daily practices in aviation around the world. Therefore, this set of concepts or constructs still needed to be validated or recognized by the professionals directly involved in RT communications, both on the basis of the frequency of their occurrences and also on their importance and potential threats to aviation safety. This is the focus of the discussion which follows.

Regarding the quantitative strand, responses to the questions 'How often do you encounter pilots/ATCOs who…?' corroborate to answer the first research question. Descriptive statistics of these questionnaire questions, in terms of means (*M*) and standard deviations (*SD*), revealed the situations that were perceived as the most and least frequent in the participants' opinion, for both the expected practices (Section II – questions 10.1 to 17.1) and the potential threats to safety (Section III – questions 18.1 to 34.1) (see Appendix T for questionnaire questions).

Considering all responses, a variation of $M = 1.89$ (questions 32.1 and 31.1) to $M = 5.38$ (question 15.1) in a scale from 1 to 6, confirms that all situations do occur in RT communications, based on participants' opinions.

The two least frequent situations are related to the sub-category of unprofessional attitude: Q32.1 (How often do you encounter pilots/ATCOs who are unprofessional and/or unwilling to help?) and Q31.1 (How often do you encounter pilots/ATCOs who seek disagreement and/or make the other feel uncomfortable?). They correspond, respectively, to the notions of lack of cooperation/collaboration and to impoliteness strategies, not a frequent occurrence, but still perceived as present in the context of pilot-

ATCO communications. On the other hand, the situation rated as most frequent refers to the sub-category of professional attitude: Q15.1 (How often do you encounter pilots who comply with ATCOs' orders/ATCOs who comply with rules?). Indeed, this is a critical attitude for the safe and expeditious management of aircraft operations. It was anticipated that responses from Section II questions (e.g., Q15.1), would be rated higher as they presented expected practices in international communications, while Section III questions (e.g., Q32.1) would probably be rated as less frequent because they included situations that may pose a threat to safety. However, if we consider the sections separately, we find the results shown in Table 7.12, which details all participants' perceptions in terms of the least and the most frequent situations, per section.

Table 7.12. All respondents' perceptions per section – Frequency of occurrence

All respondents	**Least frequent**	**Most frequent**
Section II – 10.1 to 17.1	Q 12.1 (*M* =3.00, *SD* =1.41)	Q 15.1 (*M* =5.38, *SD* =0.79)
Section III – 18.1 to 34.1	Q 32.1 (*M* =1.89, *SD* =1.06)	Q 33.1 (*M* =3.60, *SD* =1.35)

Question Q12.1 (How often do you encounter pilots/ATCOs who are concerned with both parties' images and group interests?), which operationally defines the sub-category of mutual-face concern, was rated as the least frequent occurrence from Section II. This indicates that some participants perceive their interlocutors as more concerned with their own images and interests, which, in aviation communications, may be a warning signal for lack of teamwork and cooperation. However, a greater concern arises from the fact that Q33.1 (How often do you encounter pilots/ATCOs who use non-standard phraseology?), which represents the sub-category of non-compliance with rules, was rated as the most frequent from Section III. This means that many respondents from the sample analyzed are encountering pilots and ATCOs who are not complying with

radiotelephony standard expressions, which may be a serious threat to safety, especially due to the increasing number of NNSs of English involved in RT communications around the world.

Apart from displaying results from all respondents (N=38), comparing different groups' opinions on how frequently they encounter certain situations may give us useful insights into similarities and differences in perception of the impact of culture on RT communications. To this end, results were first compared between pilots and ATCOs' responses. Table 7.13 details pilots' and ATCOs' perceptions in terms of the least and the most frequent situations.

Table 7.13. Pilots and ATCOs' perceptions per section – Frequency of occurrence

Pilots' perceptions	**Least frequent**	**Most frequent**
Section II – 10.1 to 17.1	Q 12.1 (*M* =2.95, *SD* =1.65)	Q 15.1 (*M* =5.27, *SD* =0.88)
Section III – 18.1 to 34.1	Q 32.1 (*M* =1.69, *SD* =0.93)	Q 33.1 (*M* =3.56, *SD* =1.50)
ATCOs' perceptions	**Least frequent**	**Most frequent**
Section II – 10.1 to 17.1	Q 12.1 (*M* =3.07, *SD* =0.99)	Q 15.1 (*M* =5.53, *SD* =0.64)
Section III – 18.1 to 34.1	Q 24.1 (*M* =1.67, *SD* =0.62)	Q 33.1 (*M* =3.67, *SD* =1.13)

Interestingly, pilots and ATCOs agreed on the most frequent situations related to the impact of culture on their communications, despite a small difference in mean values. Question 15.1 (How often do you encounter pilots who comply with ATCOs' orders/ATCOs who comply with rules?) in Section II, is one of them, and operationally defines the sub-category of professional attitude. Similarly, Q33.1 (How often do you encounter pilots/ATCOs who use non-standard phraseology?) was rated as the most frequent in Section III, and operationalized the sub-category of non-compliance with rules. Paradoxically, it seems that participants did not consider the use of non-standard phraseology as a way of not complying with rules, which in fact it is, according to ICAO

regulations. Figures 7.3 and 7.4 show the ratings given by pilots and ATCOs to questions Q15.1 and Q33.1, respectively.

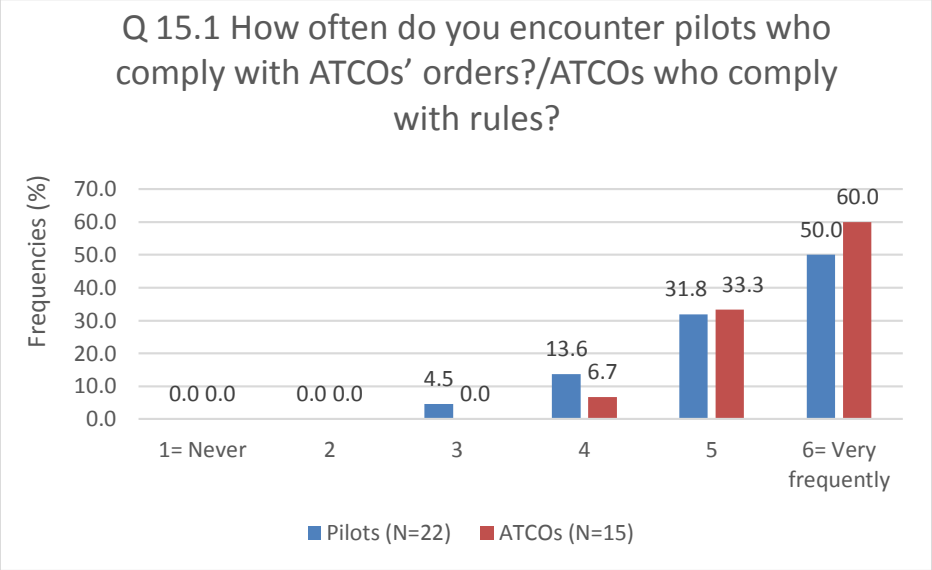


*Figure 7.3.*Pilots' vs. ATCOs' perceptions (Q15.1)

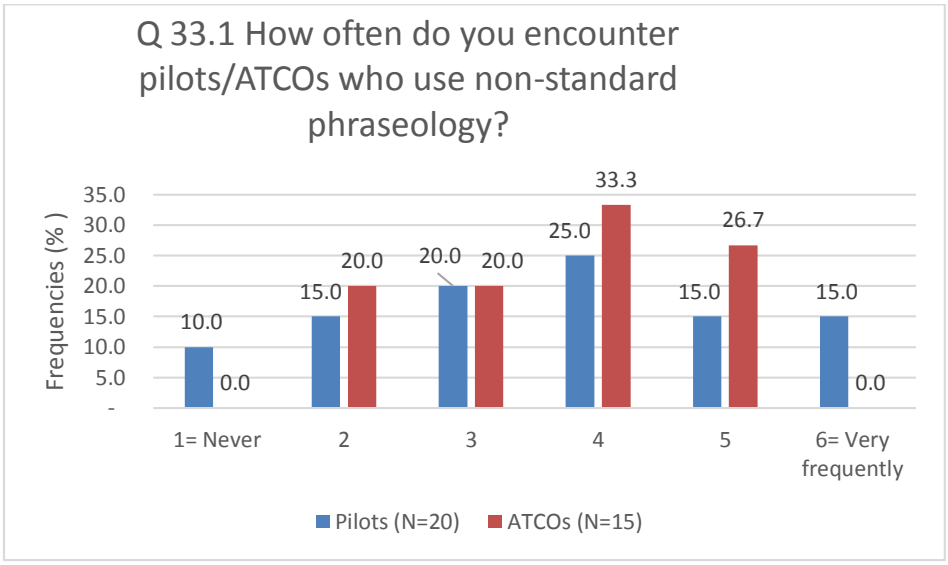


Figure 7.4. Pilots' vs. ATCOs' perceptions (Q 33.1)

However, these two groups of respondents had a slightly different opinion on the least frequent situations. From Section II, Q12.1 (How often do you encounter pilots/ATCOs who are concerned with both parties' images and group interests?) was

ranked as the least frequent by both groups and operationally defines the sub-category of mutual-face concern. However, pilots and ATCO did not agree on the least frequent instance from Section III. For example, Q32.1 (How often do you encounter pilots/ATCOs who are unprofessional and/or unwilling to help?), which operationalizes the sub-category of unprofessional attitude, was rated as the least common situation by pilots, whereas ATCOs considered it a bit more frequent (M =2.20, SD =1.21). On the other hand, ATCOs ranked Q24.1 (How often do you encounter pilots/ATCOs who engage in upfront and aggressive conflicts?), which operationally defines the sub-category of conflictual direction, as the least frequent while from the pilots' point of view it was a bit more common (M =2.22, SD =1.24). Figure 7.5 highlights the percentage of ATCOs who very frequently encounter pilots who are unprofessional and /or unwilling to help, whereas in Figure 7.6, we can see that 20% of pilots do encounter ATCOs who engage in upfront conflicts on the radio.

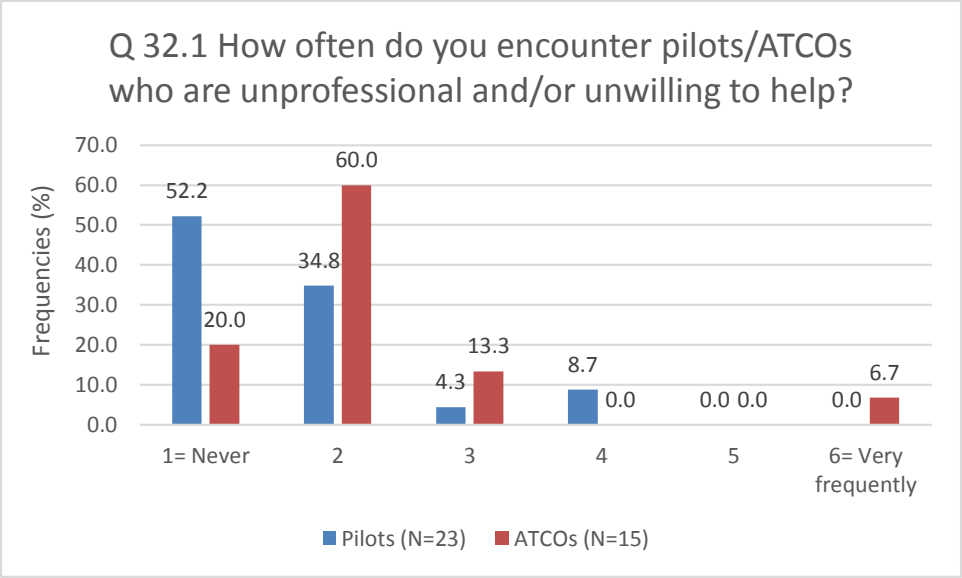


*Figure 7.5.*Pilots' vs. ATCOs' perceptions (Q32.1)

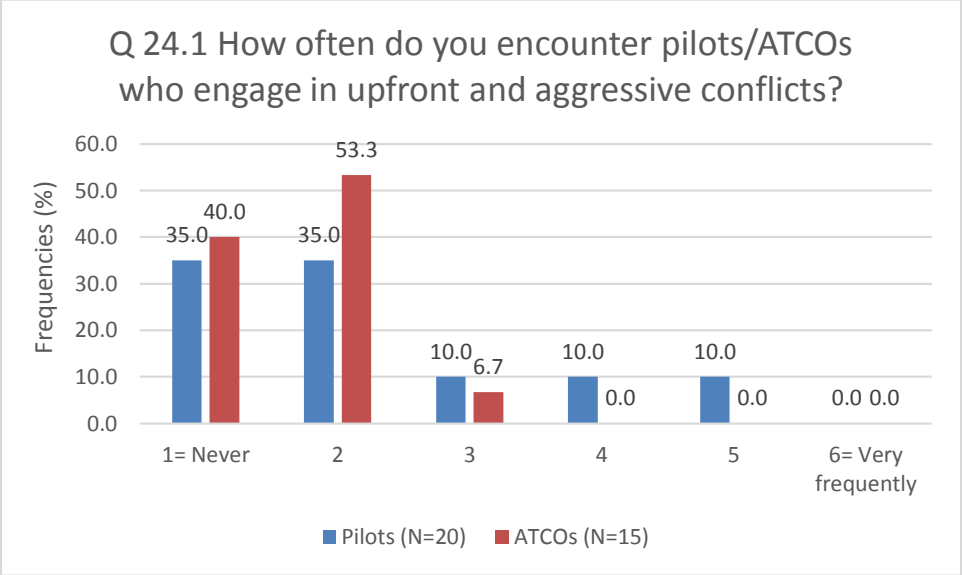


Figure 7.6. Pilots' vs. ATCOs' perceptions (Q24.1)

Secondly, the perceptions of NSs and NNSs of English on cultural issues in terms of frequency of occurrence were also compared. Table 7.14 details NSs and NNSs' perceptions in terms of the least and the most frequent situations.

Table 7.14. NSs and NNSs' perceptions per section – Frequency of occurrence

NSs' perceptions	**Least frequent**	**Most frequent**
Section II – 10.1 to 17.1	Q 12.1 (*M* =2.25, *SD* =0.89)	Q 15.1 (*M* =5.75, *SD* =0.46)
Section III – 18.1 to 34.1	Q 27.1 (*M* =1.78, *SD* =0.67)	Q 33.1 (*M* =3.55, *SD* =1.13)
NNSs' perceptions	**Least frequent**	**Most frequent**
Section II – 10.1 to 17.1	Q 12.1 (*M* =3.21, *SD* =1.47)	Q 15.1 (*M* =5.28, *SD* =0.84)
Section III – 18.1 to 34.1	Q 31.1 (*M* =1.90, *SD* =1.37)	Q 33.1 (*M* =3.62, *SD* =1.42)
	Q 32.1 (*M* =1.90, *SD* =1.11)	

Similar to the pilots and ATCOs' perceptions, NSs and NNSs of English agreed on the most frequent situations related to the impact of culture on their communications. That is, Q15.1 (How often do you encounter pilots who comply with ATCOs' orders/ATCOs who comply with rules?) in Section II, and Q33.1 (How often do you

encounter pilots/ATCOs who use non-standard phraseology?) in Section III. In addition, they also agreed on the least frequent situation in Section II, Q12.1 (How often do you encounter pilots/ATCOs who are concerned with both parties' images and group interests?), but disagreed on the least frequent instances from Section III. In this respect, the NNSs' opinions were analogous to those of pilots: Q32.1 (How often do you encounter pilots/ATCOs who are unprofessional and/or unwilling to help?), which operationalizes the sub-category of unprofessional attitude, was rated as the least frequent. Added to that, Q31.1 (How often do you encounter pilots/ATCOs who seek disagreement and/or make the other feel uncomfortable?) had the same mean value (M =1.90) of Q32.1, being equally rated by NNSs as the least frequent. It operationally defines the sub-category of unprofessional attitude as well, but addresses specific issues related to impoliteness strategies. Not surprisingly, the NNSs of English are the ones who go through this situation more often. On the other side, NSs ranked Q27.1 (How often do you encounter pilots/ATCOs who are reluctant to share critical information about a fact/state?), which operationalizes the sub-category of indirectness, as the least frequent situation in Section III. Although at a low rate, NNSs of English still perceive this as an actual issue, which may reduce the levels of situational awareness of all involved. Figures 7.7 and 7.8 show, respectively, the perceptions of NSs and NNSs of English in relation to Q31.1 and Q27.1.

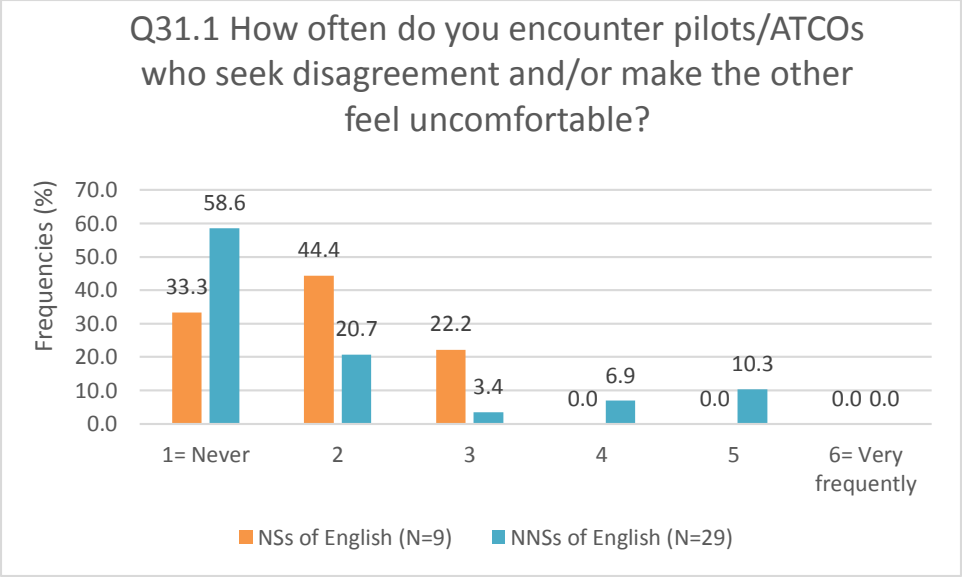


Figure 7.7. NSs' vs. NNSs' perceptions (Q31.1)

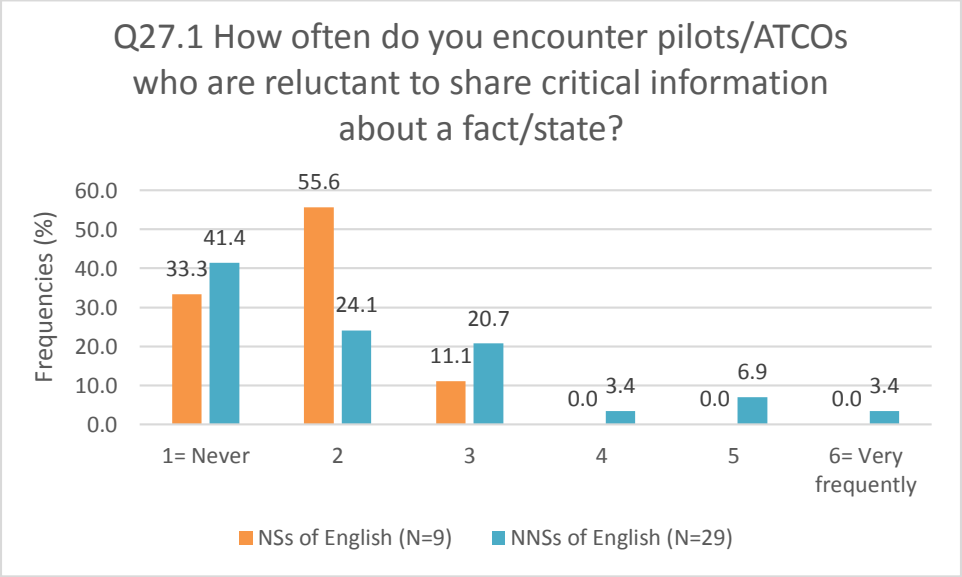


Figure 7.8. NSs' vs. NNSs' perceptions (Q27.1)

Of great importance to the present study is the discussion about the use of accommodation strategies in intercultural communications. Hence, it is worth comparing the perceptions of NSs and NNSs of English in relation to Q30.1 (How often do you encounter pilots/ATCOs who do not accommodate to less proficient speakers' needs?), which operationally defines the sub-category of unprofessional attitude. Figure 7.9 shows

more details of the participants' responses to Q30.1, revealing that NNSs encounter more frequently pilots/ATCOs who do not accommodate to less proficient speakers' needs than do NSs of English.

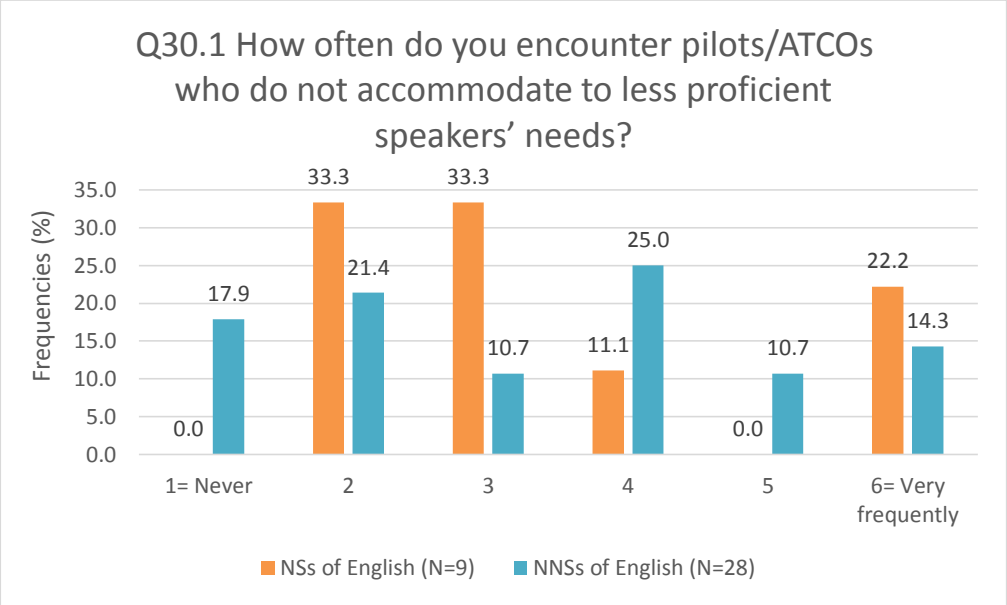


Figure 7.9. NSs' vs. NNSs' perceptions (Q30.1)

Another question that deserves some attention in this context, which is related to one of the contributing factors of the fatal accident described in Scenario 3 (QUAL strand), is Q21.1 (How often do you encounter pilots/ATCOs who avoid any kind of disagreement or demand in their speech?). This question operationalizes the sub-category of deferential role. Based on the sample analyzed, NNSs of English encounter more frequently pilots/ATCOs who tend to not demand or disagree with their interlocutors on the radio than do NSs, as shown in Figure 7.10.

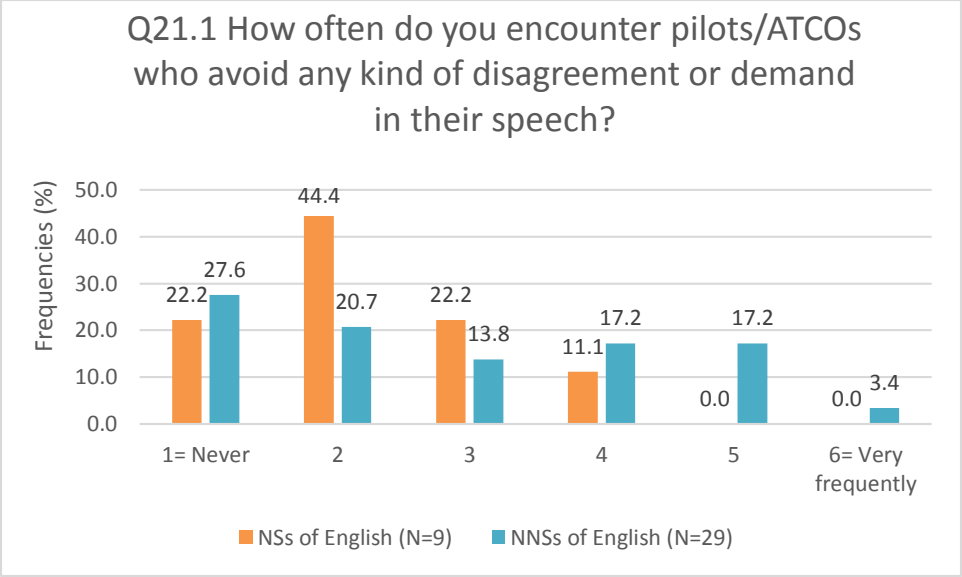


Figure 7.10. NSs' vs. NNSs' perceptions (Q21.1)

An additional and noteworthy comparison is the one related to males and females' perceptions of intercultural factors that can affect pilot-ATCOs communications. Table 7.15 specifies these two gender groups' opinions, including the responses from all pilots and ATCOs, in terms of the least and the most frequent situations.

Table 7.15. Males and females' perceptions per section – Frequency of occurrence

Males' perceptions	**Least frequent**	**Most frequent**
Section II – 10.1 to 17.1	Q 12.1 (*M* =2.93, *SD* =1.46)	Q15.1 (*M* = 5.28, *SD* =0.81)
Section III – 18.1 to 34.1	Q 32.1 (*M* =1.71, *SD* =0.85)	Q 33.1 (*M* =3.53, *SD* =1.45)

Females' perceptions	**Least frequent**	**Most frequent**
Section II – 10.1 to 17.1	Q 12.1 (*M* =3.25, *SD* =1.28)	Q15.1 (*M* = 5.67, *SD* =0.71)
Section III – 18.1 to 34.1	Q 24.1 (*M* =1.80, *SD* =0.63)	Q 33.1 (*M* =3.80, *SD* =1.03)

As we can see, males and females also agreed on the most frequent situations related to the impact of culture on their communications. That is, Q15.1 (How often do you encounter pilots who comply with ATCOs' orders/ATCOs who comply with rules?) in Section II, and Q33.1 (How often do you encounter pilots/ATCOs who use non-standard phraseology?) in Section III. Moreover, they also agreed on the least frequent

situation in Section II, Q12.1 (How often do you encounter pilots/ATCOs who are concerned with both parties' images and group interests?), but disagreed on the least frequent one from Section III. Interestingly, this pattern has repeated for all three types of comparisons/contrasts presented so far, for pilots vs. ATCOs, NSs vs. NNSs, and males vs. females. Similar to pilots' opinions, males considered Q32.1 (How often do you encounter pilots/ATCOs who are unprofessional and/or unwilling to help?), which operationalizes the sub-category of unprofessional attitude, as the least frequent situation, whereas females considered it a bit more frequent (M =2.40, SD =1.43). On the other hand, females ranked Q 24.1 (How often do you encounter pilots/ATCOs who engage in upfront and aggressive conflicts?), which operationally defines the sub-category of conflictual direction, as the least frequent, while from the males' perspectives, it was considered a bit more common (M =2.07, SD =1.18). These findings suggest that males may experience more conflictual situations on the radio than females, while females feel less cooperation from their interlocutors. Figures 7.11 and 7.12 present the perceptions of males and females in relation to Q32.1 and Q24.1.

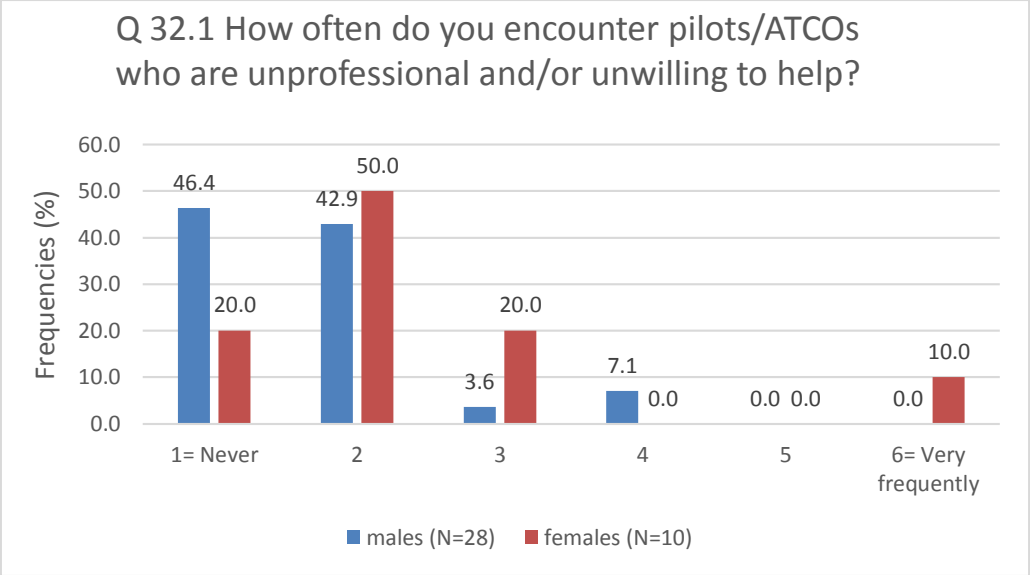


Figure 7.11. Males' vs. females' perceptions (Q32.1)

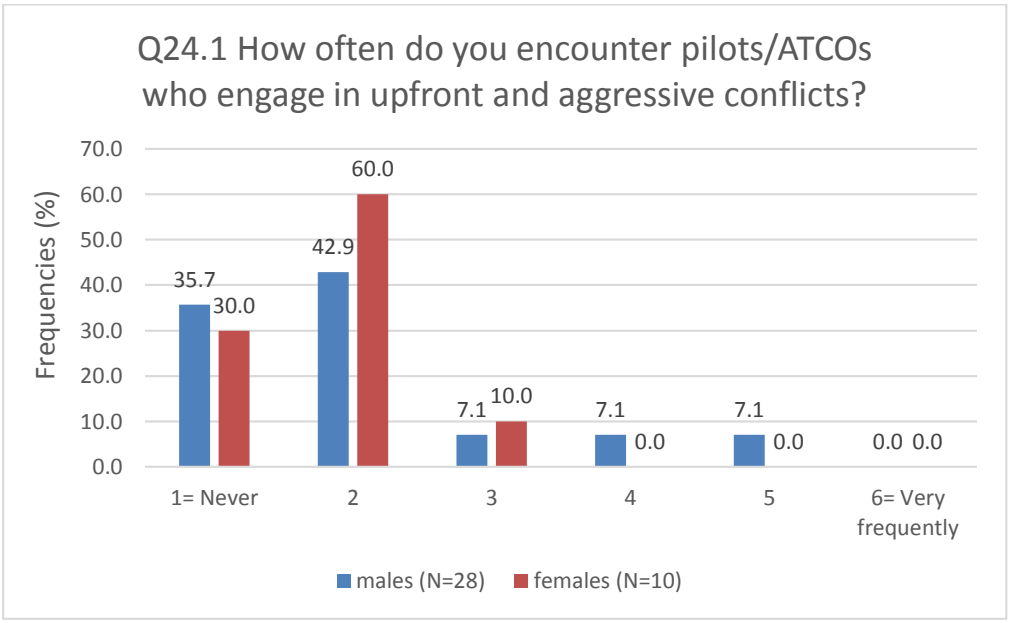


Figure 7.12. Males' vs. females' perceptions (Q24.1)

As power relations, perceptions of authority and superiority, and deferential roles might be related to or impacted by differences in gender, it is important to compare the participants' views on this topic in the specific context of pilot-ATCO radiotelephony communications. To this end, responses from males and females to Q18.1 (How often do

you encounter pilots/ATCOs who demonstrate excessive authority or superiority in their speech?), which operationally defines the sub-category of power relations, and Q20.1 (How often do you encounter pilots/ATCOs who respond in a deferential/submissive style and use excessive politeness?), operationalizing the sub-category of deferential role, are presented and discussed.

Figures 7.13 and 7.14 show that females perceive in their interlocutors' speech an excess of authority and superiority more frequently, as compared to their male counterparts. On the other hand, females perceive as more common the use of excessive politeness and deferential style in the responses of their interlocutors over the radio than do males.

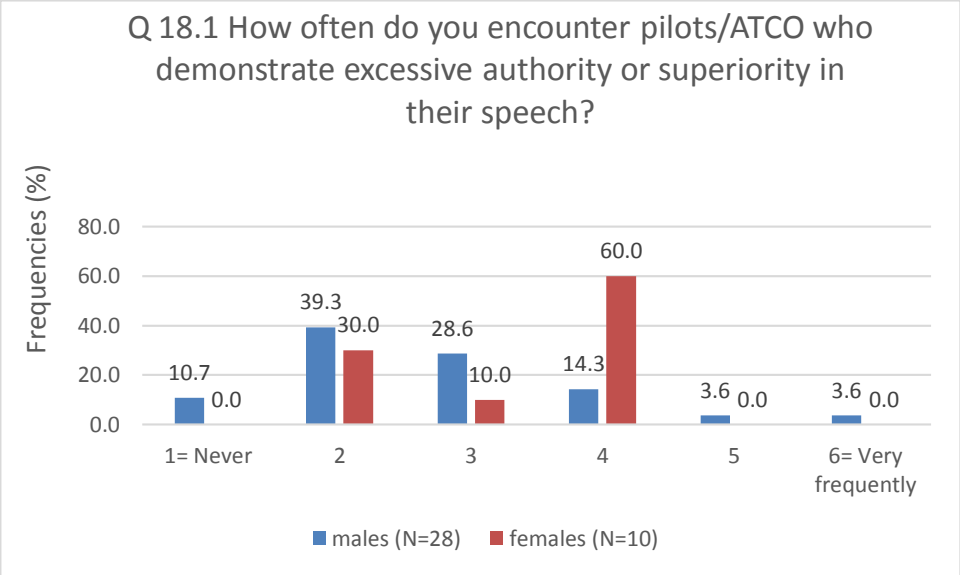


Figure 7.13. Males' vs. females' perceptions (Q18.1)

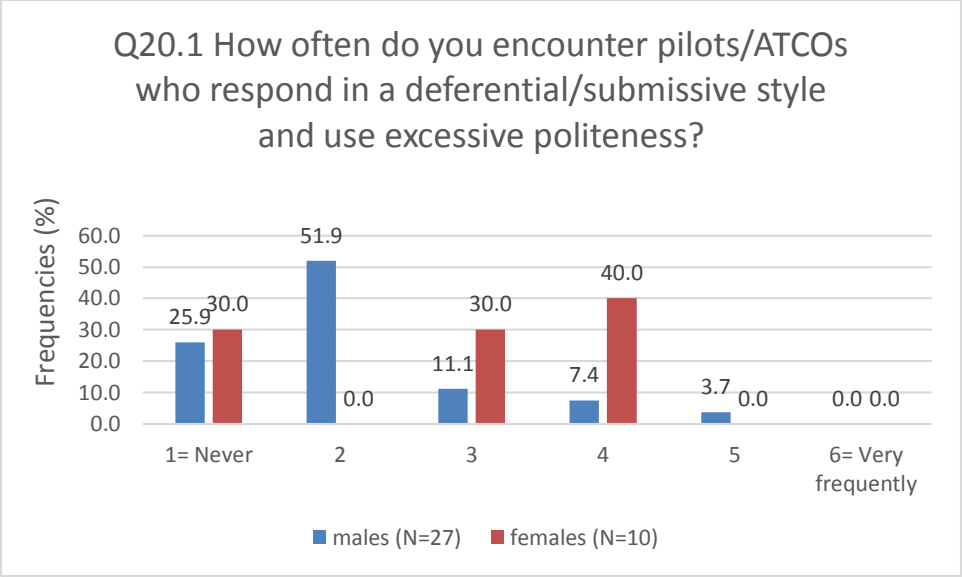


Figure 7.14. Males' vs. females' perceptions (Q20.1)

It was also possible to find evidence from the survey's open-ended responses to support the validation of the sub-categories from the provisional taxonomy, and thus to answer the first research question. Figure 7.15 presents the number of valid comments for each sub-category organized according to the Magnitude Codes attributed, i.e., if the comment contradicted the sub-category, was neutral or validated it. As can be seen, all sub-categories received more comments that validated the situations presented in the questions, except for the sub-category of mutual-face concern. Question 12 is the only one which operationally defines this sub-category, which refers to being concerned with group's images and interests. In the aviation context this would be related to a face orientation towards the collective (see Ting-Toomey (2005), as discussed in Chapter 3, Section 3.2.1), to teamwork and cooperation, and to fulfilling the goal of the group, not of the individual. Yet, unexpectedly, this question was considered the least frequent by all groups of participants: pilots, ATCOs, NSs, NNSs of English, males and females, although the lowest mean value was of M =2.25 (SD =0.89), according to the perspective

of NSs of English. Additionally, six contradictory comments were provided in response to this question, which might indicate a certain difficulty on the part of the participants in understanding the question. Still, this question received five neutral comments and four comments that validated the sub-category. On top of that, in six comments from other questions, reference was made to the construct operationalized in Q12, i.e., this question received six extra comments that contributed to its validation.

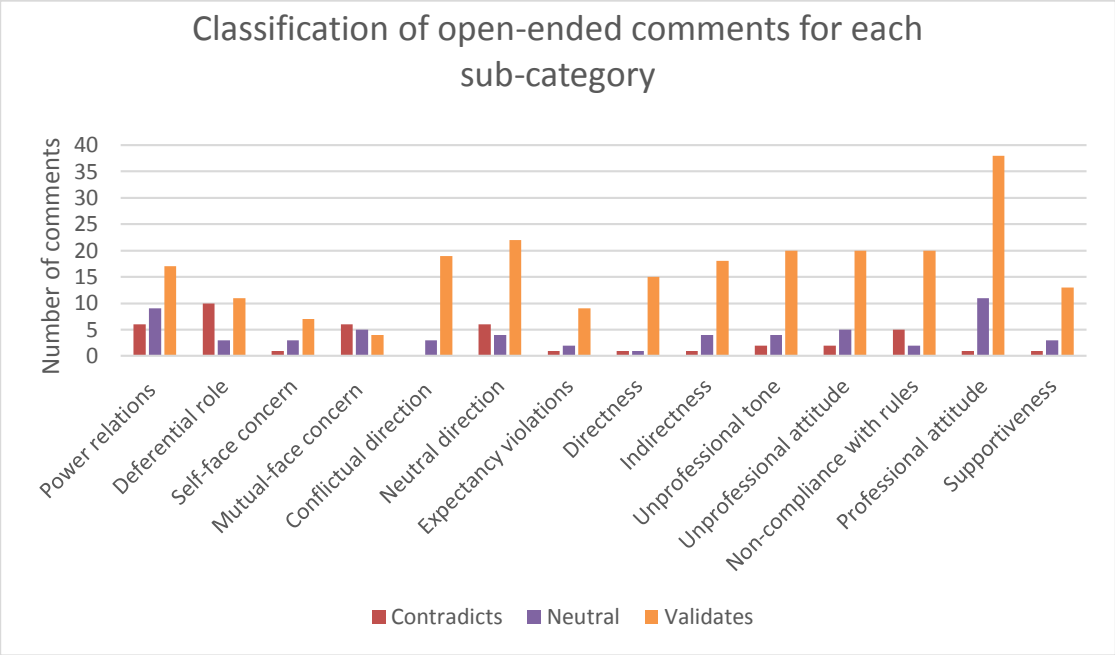


Figure 7.15. Summary of Magnitude Coding (Contradicts, Neutral, Validates) of survey open-ended responses

In order to illustrate how participants' open-ended responses also support the validation of the sub-categories, some comments have been selected and are presented below as anonymous quotes. However, for the purpose of providing a better understanding of the participant's perspective portrayed in the quote, the group (pilot vs. ATCO), gender (male vs. female) and language background (English as L1 vs. English as

L2) of the respondent is provided. It is worth noting that these responses reveal, most of the times an inter-relation among the intercultural factors identified.

- Question 15 (How often do you encounter pilots who comply with ATCOs' orders/ATCOs who comply with rules?):

> Most controllers attempt to follow the rules. The pressure that they work under is mostly self- inflicted. If they choose not to follow a rule, you would have to peal the onion back, and get to the root causes. If you do, you would probably find; than[21] an earlier mistake or error, took place. Which, if addressed properly at the time, could have avoided the problem which occurred later. There is peer pressure among the controllers. How much traffic can you handle? When are you going to say NO more traffic in my sector? When are you going to ask for help? Controllers do not want to "lose face" with their fellow controllers by not being able to keep up with the traffic. (Pilot 17, male, English as L2)

Interestingly, this pilot confirms that most controllers attempt to follow the rules, but highlights the existence of peer pressure among this group and the controllers' fear of losing face by saying "NO" to more traffic. Consequently, they frequently feel overloaded and may get impatient or annoyed easily (e.g., Scenario 2, Scenario 6).

- Question 18 (How often do you encounter pilots/ATCOs who demonstrate excessive authority or superiority in their speech?): "Those controllers who demonstrate excessive authority can, in some moments, in some ways, to

[21] The comments have been copied *ipsis litteris* from the actual online texts produced by the participants, and have not been corrected for any typos or grammatical/lexical mistakes.

diminush the pilot innitiative to clarify misunderstanding" (Pilot 2, male, English as L2). Demonstrations of power and asymmetric interactions lead to pilots, and also ATCOs, being inhibited or prevented from clarifying information in RT (e.g., Scenario 3, Scenario 4, and Scenario 5). Reactions to that may vary according to the cultural background of interlocutors.

- Question 21(How often do you encounter pilots/ATCOs who avoid any kind of disagreement or demand in their speech?): "Yes. Sometimes pilots contest our order specially because I'm a woman and this makes my frequency busy" (ATCO 29, female, English as L2). This comment highlights an important perception of a female ATCO related to the way pilots react to females' orders, by not complying with them directly (e.g., Scenario 1).
- Question 23 (How often do you encounter pilots/ATCOs who dominate or compete during an argument?): "There is no room for such persons in aviation. Such behaviour can be very destabilising and threatening to a controller. The basis is generally cultural which thankfully is slowly changing (re CRM/TRM[22] and inter-cultural behaviour programmes)" (ATCO 4, male, English as L1). Arguments and conflicts on the radio frequency can trigger less than efficient communications and negative outcomes (e.g., Scenario 1, Scenario 4, and Scenario 6). This ATCO pointed out that the basis of this behavior lies on cultural factors, which directly answers the first research question.
- Question 26 (How often do you encounter pilots/ATCOs who speak in a confusing and unclear way?):

[22] CRM stands for Crew Resource Management and TRM for Team Resource Management.

> Same comment as in previous question - very common in the USA. Once I've got, for instance the following question: “What can you give me”. He was intending to know how fast could I fly. However, this is not the way they should address a pilot, moreover a foreigner. (Pilot 10, male, English as L2)

We see here a clear confirmation that some speakers of English as L1 do not comply with the RT triad “be clear, concise and unambiguous” (e.g., Scenario 4), and seem not to be aware of how this may affect NNSs’ understanding.

- Question 28 (How often do you encounter pilots/ATCOs who show impatience and/or sarcasm in their speech?): “By not showing respect one is not being professional. Impatience has been identified as one of the causes of fatal accidents” (ATCO 7, male, English as L2). According to this ATCO, instances of impatience not only happen in RT communications, (e.g., Scenario 2, Scenario 4) but also were considered as contributing factors to accidents. Being respectful is the professional tone desired.
- Question 29 (show annoyance and/or arrogance in their speech): “We all get tired, irritable, frustrated but try and remain professional. Let your behaviour set the tone of the communication” (ATCO 4, male, English as L1). Despite other human factors that are inherent to stressful activities, we get from this comment that remaining professional is the key to setting the tone of communications. However, this is not what happens all the time (e.g., Scenario 2 and Scenario 4).
- Question 30 (How often do you encounter pilots/ATCOs who do not accommodate to less proficient speakers’ needs?):

> It's easy to find this kind of behavior in regions OR countries where the mother language is English, the air controllers ONLY speak English and NEVER had the need to know how to speak another language, besides the English language (lack of empathy in communications with foreigners). For example, it happens sometimes with United States air controllers. Sometimes, it looks like that they don't care if they are being understood, and if the pilot ask "say again", they repeat the same words and they are not able to say the same information using another words (lack of the ability to paraphrase). (Pilot 8, male, English as L2)

This sounds like this NNS of English pilot is releasing his frustration about how some NSs of English lack communicative strategies to interact with less proficient speakers (e.g., Scenario 2 and Scenario 4). Words like “it’s easy to find” and “it happens sometimes” confirm this as a real issue and “lack of empathy”, as a cultural factor.

- Question 33 (How often do you encounter pilots/ATCOs who use non-standard phraseology?): “Far too often these days. ‘Now that I have a level 5 (or level 6, even worse) I want to show you how well I can speak English.’ OK perhaps a slight exaggeration, but this is a trait that is giving concern these days” (ATCO 4, male, English as L1). Another participant states the following: “It happens ALL time everywhere. Pilots are also guilty” (Pilot 17, male, English as L2). Both pilots and ATCOs perceive the use of non-standard phraseology as a frequent issue in RT communications (e.g. Scenario 2 and Scenario 4), and as a global safety concern.

To conclude, regarding the sample analyzed, the intercultural factors arising from international pilot-ATCO communications that can affect the way they interact in the English language are the ones displayed in the provisional taxonomy, as validated by survey participants.

Research question 2:

To what extent do experienced pilots and ATCOs perceive the potential threats of intercultural factors to the safety of radiotelephony communications?

Evidence to answer this question comes from the quantitative phase, including the participants' open-ended comments. Responses to the questions 'How important is this?', in Section II, and 'How important were these events as potential threats to safety?', in Section III, provide the answer to this research question. Descriptive statistics of the questions on the survey revealed the situations that were considered the most and least important in the participants' opinion.

> A variation of M = 3.45 (Q20.2) to M =5.97 (Q14.2), in a scale from 1 to 6, indicates that all situations were considered important and related to the safety of RT communications, based on participants' opinions.

The situation that was considered the least important as a potential threat to safety was the one presented in Q20.2 (How important, in your view, were these events as potential threats to safety? To encounter pilots/ATCOs who respond in a deferential/submissive style and use excessive politeness?). This question is from Section III and operationally defines the sub-category of deferential role. Although politeness per se may not be a risk to safety, its excess might compromise effective communication and, above all, a deferential style coming from issues of power distance may inhibit a pilot or

ATCO to assertively state his/her operational needs, as was the circumstance in the fatal accident reported in Scenario 3 (QUAL strand).

On the other hand, the situation rated as most important by all respondents, Q 14.2 (How important is this? To encounter pilots/ATCOs who are concerned with safety and potential complications?) from Section II, operationalizes the sub-category of professional attitude. No doubt, this should be the driving force of all involved in aeronautical communications and operations: to strive for safety and be aware of and prepared to deal with unexpected situations.

Added to that, a frequency count of references to "safety/safe/safely/unsafe" or "incident/accident" was conducted in open-ended comments related to questions Q10.3 to Q17.3 in Section II. They confirmed that the constructs operationalized in all these questions were considered important, specifically in relation to safety. Question 14.3 (Please comment: How important is this? To encounter pilots/ATCOs who are concerned with safety and potential complications?) received the greatest number of references to safety, followed by Q15.3 (Please comment: How important is this? To encounter pilots who comply with ATCOs' orders/ATCOs who comply with rules?). Both questions operationally define the sub-category of professional attitude.

Participants' perceptions in terms of the least and the most important situations, per section, are detailed in Table 7.16, which also informs the means (*M*) and standard deviations (*SD*) for the highlighted questions.

Table 7.16. All respondents' perceptions per section – Importance to safety

All respondents	**Least important**	**Most important**
Section II – 10.2 to 17.2	Q 12.2 (*M* =4.48, *SD* =1.29)	Q 14.2 (*M* =5.97, *SD* =0.16)
Section III – 18.2 to 34.2	Q 20.2 (*M* =3.45, *SD* =1.54)	Q 26.2 (*M* =5.56, *SD* =0.82)

Question Q12.2 (How important is this? To encounter pilots/ATCOs who are concerned with both parties' images and group interests?), which operationally defines the sub-category of mutual-face concern, was not only rated as the least frequent, as discussed before, but also was ranked the least important from Section II, though with a mean value above 4.00 (*M* =4.48, *SD* =1.29). This might suggest that participants may not perceive a face orientation towards the collective as so important to their radiotelephony communications, or, alternatively, did not understand what it really entails. In contrast, we have Q 26.2 (How important, in your view, were these events as potential threats to safety? To encounter pilots/ATCOs who speak in a confusing and unclear way?) rated as the most important from Section III, which operationalizes the sub-category of indirectness. In ordinary conversation this style is not desirable, much less in safety-critical communications such as radiotelephony, in which being clear, concise and unambiguous is the norm. However, certain cultures are known to be more direct, objective, and right to the point, whereas others not so much.

Following what was done in comparing different groups' opinions on how frequently they encounter certain situations, responses to questions related to the degree of importance participants attributed to each instance were also compared across groups in an attempt to uncover any meaningful connections or dissimilarities. Table 7.17 details pilots' and ATCOs' perceptions in terms of the least and the most important situations.

Table 7.17. Pilots' and ATCOs' perceptions per section – Importance to safety

Pilots' perceptions	**Least important**	**Most important**
Section II – 10.2 to 17.2	Q 12.2 (*M* =4.59, *SD* =1.37)	Q 14.2 (*M* =6.00, *SD* =0.00)
Section III – 18.2 to 34.2	Q 20.2 (*M* =3.44, *SD* =1.26)	Q 26.2 (*M* =5.65, *SD* =0.83)

ATCOs' perceptions	**Least important**	**Most important**
Section II – 10.2 to 17.2	Q 12.2 (*M* =4.36, *SD* =1.21)	Q 14.2, Q15.2 (*M* = 5.93, *SD* =0.26)
Section III – 18.2 to 34.2	Q 20.2 (*M* =3.47, *SD* =1.85)	Q 34.2 (*M* =5.71, *SD* =0.61)

This time, pilots' and ATCOs' perceptions were very much alike, with slight differences in means, for the least important situations in both sections and the most important one in Section II. Agreement was found for Q12.2 (How important is this? To encounter pilots/ATCOs who are concerned with both parties' images and group interests?), as the least important in Section II, and for Q20.2 (How important, in your view, were these events as potential threats to safety? To encounter pilots/ATCOs who respond in a deferential/submissive style and use excessive politeness?), as the least important in Section III. The same happened to Q14.2 (How important is this? To encounter pilots/ATCOs who are concerned with safety and potential complications?), considered as the most important in Section II. All of them have already been identified and discussed in previous paragraphs. From the ATCOs' point of view, Q15.2 (How important is this? To encounter pilots who comply with ATCOs' orders/ATCOs who comply with rules?) was also rated as one of the most important in Section II and it operationalizes the sub-category of professional attitude. However, differences in perceptions appeared in the most important situations as a potential threat to safety, from Section III. Pilots rated Q26.2 (How important were these events as potential threats to safety? To encounter pilots/ATCO who speak in a confusing and unclear way?), which operationally defines the sub-category of indirectness as the most important one, whereas ATCOs rated Q 34.2 (How important were these events as potential threats to safety? To encounter pilots who do not comply with ATCO's orders/ATCOs who do not comply with rules?), as the most critical one. Figures 7.16 and 7.17 present the perceptions of pilots and ATCOs in relation to Q26.2 and Q34.2, respectively.

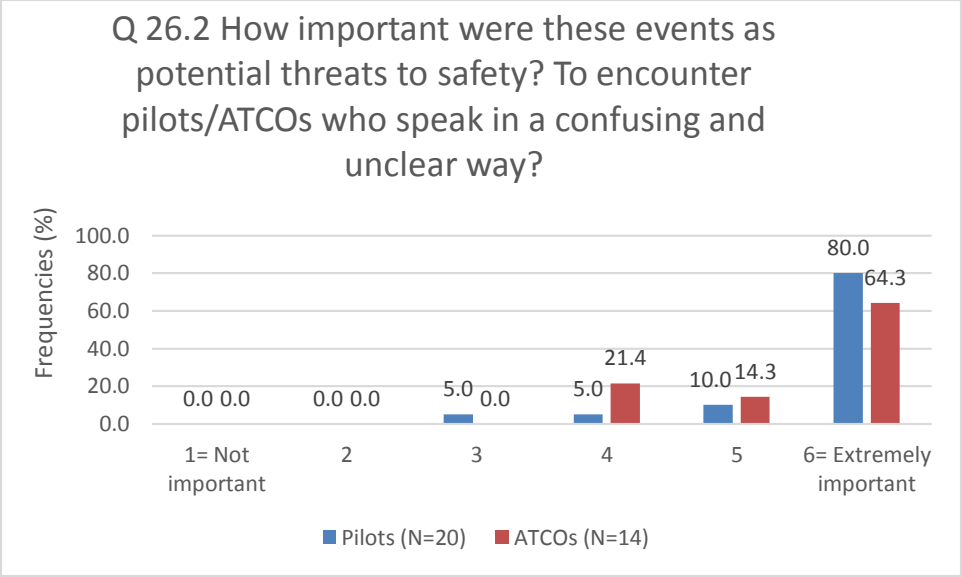


Figure 7.16. Pilots' vs. ATCOs' perceptions (Q26.2)

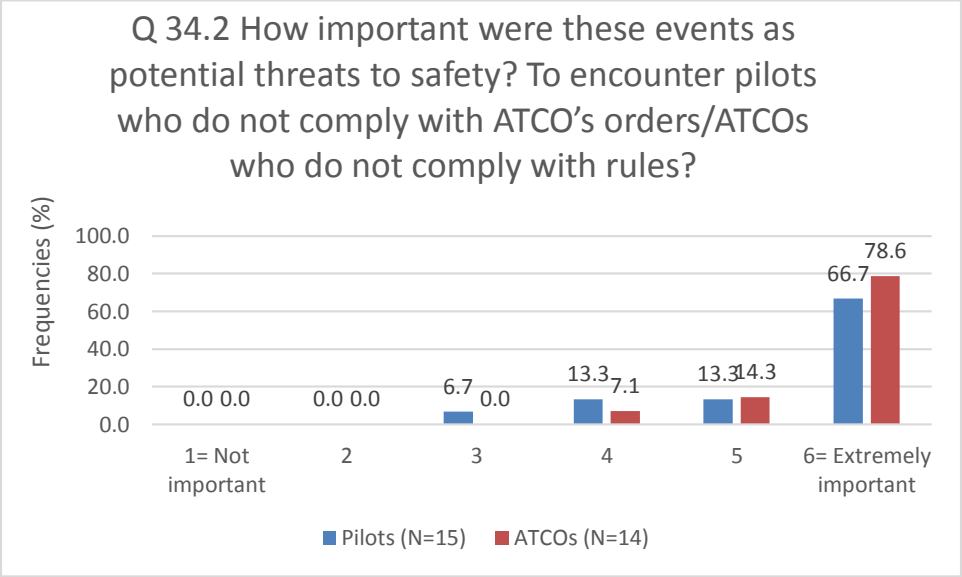


Figure 7.17. Pilots' vs. ATCOs' perceptions (Q34.2)

In addition, the perceptions of NSs and NNSs of English on cultural issues in terms of their importance to safety were also compared. Table 7.18 details NSs and NNSs' perceptions in terms of the least and the most important situations.

Table 7.18. NSs' and NNSs' perceptions per section – Importance to safety

NSs' perceptions	**Least important**	**Most important**
Section II – 10.2 to 17.2	Q 12.2 (*M* =4.00, *SD* =1.67)	Q 15.2 (*M* =6.00, *SD* =0.00)
Section III – 18.2 to 34.2	Q 20.2 (*M* =3.28, *SD* = 1.38)	Q 26.2 (*M* =5.22, *SD* =1.20)
NNSs' perceptions	**Least important**	**Most important**
Section II – 10.2 to 17.2	Q 12.2 (*M* =4.60, *SD* =2.19)	Q 14.2 (*M* =6.00, *SD* =0.00)
Section III – 18.2 to 34.2	Q 20.2 (*M* =3.50, *SD* = 1.61)	Q 34.2 (*M* =5.80, *SD* =0.41)

As noted in the previous discussion about pilots' and ATCOs' perceptions, NSs and NNSs of English also agreed on the same least important situations in Section II (Q12.2) and Section III (Q20.2). Although there were differences between NSs' and NNSs' perspectives on the most important issues to safety, their responses to Section III questions were similar to those given by pilots and ATCOs, i.e., Q26.2 (How important were these events as potential threats to safety? To encounter pilots/ATCO who speak in a confusing and unclear way?), was ranked by pilots and NSs as the most important, whereas Q34.2 (How important were these events as potential threats to safety? To encounter pilots who do not comply with ATCO's orders/ATCOs who do not comply with rules?), was ranked by ATCOs and NNSs as having greater importance as a potential threat to safety. Interestingly, the most frequent instances in Section II received different opinions. NSs rated Q15.2 (How important is this? To encounter pilots who comply with ATCOs' orders/ATCOs who comply with rules?) as the most important situation, while NNSs rated Q14.2 (How important is this? To encounter pilots/ATCOs who are concerned with safety and potential complications?) as the most crucial one. For these two questions, which operationalize the sub-category of professional attitude, consensus was achieved (*M* =6.0) within each group.

With regards to the opinions of males and females in relation to the importance of the situations presented in the survey, a comparison is found in Table 7.19 with a summary of their responses per section.

Table 7.19. Males' and females' perceptions per section – Importance to safety

Males' perceptions	**Least important**	**Most important**
Section II – 10.2 to 17.2	Q 12.2 (*M* =4.54, *SD* =1.33)	Q 14.2 (*M* =5.96, *SD* =0.19)
Section III – 18.2 to 34.2	Q 20.2 (*M* =3.23, *SD* =1.27)	Q 34.2 (*M* =5.53, *SD* =0.90)
Females' perceptions	**Least important**	**Most important**
Section II – 10.2 to 17.2	Q 12.2 (*M* =4.37, *SD* =1.30)	Q 16.2, Q15.2 and Q14.2 (*M* =6.00, *SD* =0)
Section III – 18.2 to 34.2	Q 22.2 (*M* =3.56, *SD* =1.42)	Q 26.2 (*M* =5.80, *SD* =0.42)

It is noticeable that agreement was mainly achieved in terms of the least and the most important situations in relation to Section II. That is, Q12.2 (How important is this? To encounter pilots/ATCOs who are concerned with both parties' images and group interests?) was ranked as the least important, for both males and females, whereas Q14.2 (How important is this? To encounter pilots/ATCOs who are concerned with safety and potential complications?), as the most important one. In this last case, females rated as equally important responses to Q16.2 (How important is this? To encounter pilots/ATCOs who speak in a professional tone?) and to Q15.2 (How important is this? To encounter pilots who comply with ATCOs' orders/ATCOs who comply with rules?), both of which operationally define the sub-category of professional attitude.

For Section III, however, males' and females' opinions diverged. Males rated Q20.2 (How important, in your view, were these events as potential threats to safety? To encounter pilots/ATCOs who respond in a deferential/submissive style and use excessive politeness?), which operationalizes the sub-category of deferential role, as the least important situation, whereas females considered it a bit more significant (*M* =4.00, *SD* =2.06). On the other hand, in the females' opinions, Q22.2 (How important, in your

view, were these events as potential threats to safety? To encounter pilots/ATCOs who are concerned with preserving their own images and interests?), which operationally defines the sub-category of self-face concern, was regarded as the least important of all, while males rated it as slightly more important (M =3.96, SD =1.54). Figures 7.18 and 7.19 present the perceptions of males and females in relation to Q20.2 and Q22.2, respectively.

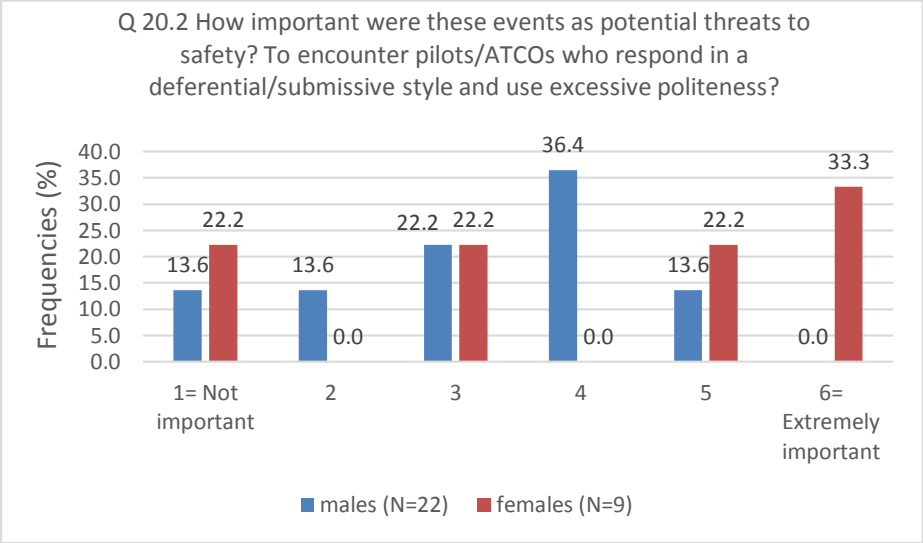


Figure 7.18. Males' vs. females' perceptions (Q20.2)

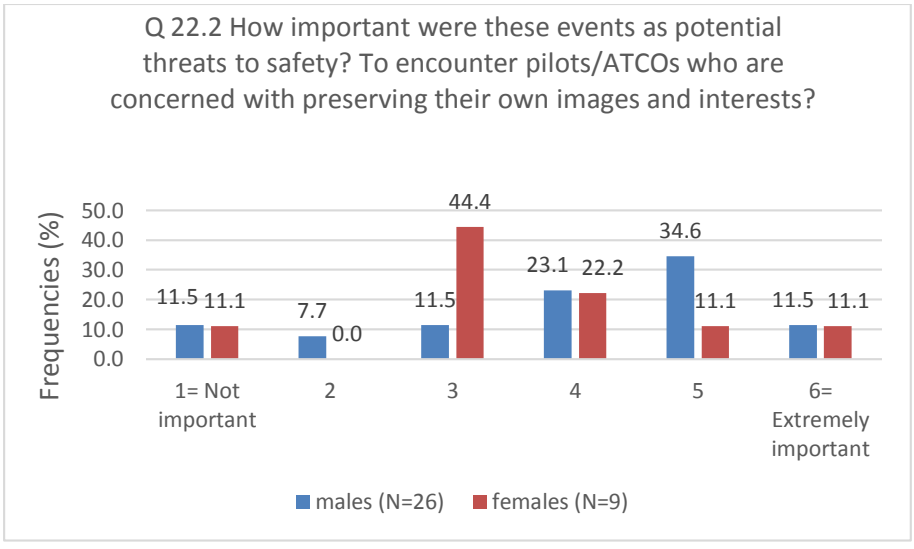


Figure 7.19. Males' vs. females' perceptions (Q22.2)

In terms of the most important instances for Section III, differences were also noticed between males' and females' perceptions. The most important situation for males, Q34.2 (How important were these events as potential threats to safety? To encounter pilots who do not comply with ATCO's orders/ATCOs who do not comply with rules?), matches the opinions of NNSs of English, whereas the most important one for females, Q26.2 (How important were these events as potential threats to safety? To encounter pilots/ATCO who speak in a confusing and unclear way?), matches the perceptions of NSs of English.

Participants' open-ended responses well illustrate their perceptions of the importance of intercultural factors to the safety of pilot-ATCO communications using the English language as potential triggers of incidents and accidents. Consequently, they corroborate in the answers to research question 2.

- Question 13 (How important is this? To encounter pilots/ATCOs who are concerned with clarity and efficiency?): "Pilots will always need and want clarity and efficiency from the controllers, but sometimes they themselves don't practice this. Both parties should always be concerned with clarity and efficiency in the name of safety" (ATCO 3, male, English as L2). Interestingly, this comment not only emphasizes the importance of being clear and efficient to safety, but also mentions the shared responsibilities of all participants in achieving this goal. (e.g., Scenario 1 and Scenario 5).
- Question 14 (How important is this? To encounter pilots/ATCOs who are concerned with safety and potential complications?): "Safety should be a teamwork practice and concern. Aviation involves the coordinated work of

several teams both in flight and on the ground, the primary concern of all should be with safety" (Pilot 9, male, English as L2). In a similar way, concepts of teamwork and coordination are directly linked to aviation safety. For example, in Scenario 5, the ATCO did not have the tools to act collaboratively and to be aware of potential complications.

- Question 16 (How important is this? To encounter pilots/ATCOs who speak in a professional tone?): "The adherence to standard phraseology is the most professional tone that can be used in radiotelephony. This is crucial for a better understanding and for safety" (ATCO 3, male, English as L2). This ATCO linked the notion of professional tone to adherence to standard phraseology as a way to point out that compliance with prescribed rules is a crucial attitude for safety in radiotelephony (e.g., Scenario 2 and Scenario 4).
- Question 18 (How important, in your view, were these events as potential threats to safety? To encounter pilots/ATCOs who demonstrate excessive authority or superiority in their speech?):

> When this occurs, it icauses a somber pressage that something bad may happen. The environment becomes gloomier, the effective communication may be hindered by some possible unnecessary debate over trivial matters. In a worst scenario, this may lead to disobedience and consequently to an incident. (ATCO 3, male, English as L2)

From this comment, we can infer that the respondent perceives the potential threat of excessive authority and radio debates to the safety in RT communications (e.g., Scenario 1 and Scenario 6), which can escalate to incidents.

- Question 20 (How important, in your view, were these events as potential threats to safety? To encounter pilots/ATCOs who respond in a deferential/submissive style and use excessive politeness?): “Too much is also not too good. High ‘power distance’ behaviour for both sides are not good for safety. A well grounded posture and positioning should do better for safety” (Pilot 9, male, English as L2). A confirmation of the influence of power distance to participants’ behavior is provided above, as one that compromises safety (e.g., Scenario 3). According to this pilot, achieving a balance in terms of politeness and behavior should be the goal.
- Question 21 (How important, in your view, were these events as potential threats to safety? To encounter pilots/ATCOs who avoid any kind of disagreement or demand in their speech?): “Whenever necessary pilots disagree and demand. This is also very important for safety because they have to be aware of their aircraft performance and air traffic regulations” (ATCO 5, male, English as L2). Again, achieving the correct balance in terms of assertiveness should be the goal, according to regulations and aircraft performance. Unfortunately, this was not how the First Officer from Scenario 3 acted in an emergency.
- Question 25 (How important, in your view, were these events as potential threats to safety? To encounter pilots/ATCOs who violate your expectations of a standard flow of communication?):

    > Sometimes first timers costumers or inexperienced pilots who are not used to the procedures do "violate" our expectation of a standard flow of communication and that will always mean extra work, a high level of

> stress and the concern that safety might be afected. (ATCO 13, female, English as L2)

Whenever our expectations are violated, levels of anxiety and stress may increase. Safety may be compromised by an increase in workload or unnecessary message exchanges (e.g., Scenario 2, Scenario 4, Scenario 5 and Scenario 6).

- Question 27 (How important, in your view, were these events as potential threats to safety? To encounter pilots/ATCOs who are reluctant to share critical information about a fact/state?): “It still happens specially when the issue is declaring emergency and that adds workload for the controller that works under the suspition that he is not holding all the cards and that can really affect safety and efficiency” (ATCO 13, female, English as L2). Not being direct and withholding information is critical for aviation safety. This ATCO confirms that it frequently happens and can create or exacerbate emergencies (e.g., Scenario 3).
- Question 29 (How important, in your view, were these events as potential threats to safety? To encounter pilots/ATCOs who show annoyance and/or arrogance in their speech?):

  > Controllers and pilots must remember that this is a high-stake environment, safety is the motto. So, if the pilot or the controller are not in good mood, at least they should be professional, adhere to the standard phraseology, comply with the instructions, issue correct instructions and be polite with each other. (ATCO 3, male, English as L2)

This ATCO summarizes what is expected from pilots and controllers in terms of professional communication and attitudes (e.g., Scenario 2 and Scenario 4), specifically in this high-stakes safety environment.

- Question 32 (How important, in your view, were these events as potential threats to safety? To encounter pilots/ATCOs who are unprofessional and/or unwilling to help?):

> Sometimes it happens. These pilots often think only about themselves and don't care about the possible problems that this attitude may cause. They want all the advantages to themselves. Controllers must be alert to this type of pilots so that they cannot interfere with his job and jeopardize safety. (ATCO 3, male, English as L2)

In addition to referring to the notion of lack of collaboration, this comment also highlights the issue of self-face concern, in terms of not being aware of the needs of the others and the collective (e.g., Scenario 1 and Scenario 6).

In sum, to answer research question 2, it is possible to affirm that pilots and ATCOs in this sample perceived, to a great extent, the potential threats of intercultural factors to the safety of radiotelephony communications. Although to different degrees, participants considered that all intercultural factors operationalized in the questions are important to safety.

A last figure (Figure 7.20), comparing the means for frequency and importance per question, discloses the fact that the frequency of occurrence of intercultural factors that can affect pilot-ATCOs communications was generally lower than their perceived importance as a potential threat to safety. However, as those situations do happen in RT,

confirmed by the survey respondents, and in some scenarios a combination of more than one instance at a time, interlocutors should be aware of their impact and develop skills on how to accommodate differences and be effective intercultural communicators.

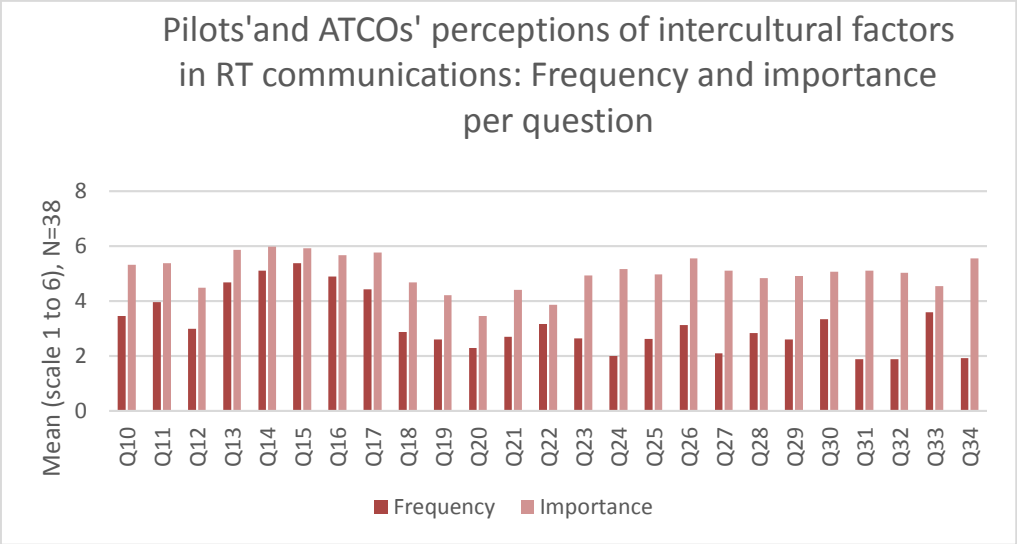


Figure 7.20. Comparison of means for frequency and importance per question

Finally, it is crucial to conduct a final interpretation of data and a discussion of the value added by MM to this phase of the study. To reach this goal, a joint display integrating the reduced findings from the QUAL and the quan strands in a single visual representation was designed (see Table 7.20). It not only assists the researcher and the reader in understanding the existing relationships between the two strands, but also in the generation of meta-inferences. As more evidence was yielded throughout the study, this exploratory sequential MM design increased confidence in the findings and added further insights into the complex phenomena of multicultural RT communications in aviation.

Table 7.20. Phase 1 - MM exploratory sequential joint display

QUAL ANALYSIS		QUAN ANALYSIS (N=38)				OPEN-ENDED COMMENTS ANALYSIS			
Categories	**Sub-Categories**	**Questions**	**Mean/SD**	**Question**	**Mean/SD**	**Contradicts**	**Neutral**	**Validates (Valid %)**	**Extra comments**
Power Distance	Power relations	Q18.1	*M*=2.87, *SD*=1.14	Q18.2	*M*=4.66, *SD*=1.14	2	5	13 (65.0%)	9
		Q19.1	*M*=2.60, *SD*=1.65	Q19.2	*M*=4.22, *SD*=1.45	4	4	4 (33.3%)	
	Deferential role	Q20.1	*M*=2.30, *SD*=1.13	Q20.2	*M*=3.45, *SD*=1.54	5	0	5 (50.0%)	1
		Q21.1	*M*=2.71, *SD*=1.49	Q21.2	*M*=4.40, *SD*=1.57	5	3	6 (42.8%)	
Face-work strategies	Self-face concern	Q22.1	*M*=3.16, *SD*= 1.67	Q22.2	*M*=3.86, *SD*=1.50	1	3	7 (63.6%)	15
	Mutual-face concern	Q12.1	*M*=3.00, *SD*=1.41	Q12.2	*M*=4.48, *SD*=1.29	6	5	4 (26.6%)	6
Conflict management	Conflictual direction	Q23.1	*M*=2.65, *SD*=1.29	Q23.2	*M*=4.94, *SD*=1.28	0	2	10 (83.3%)	3
		Q24.1	*M*=2.00, *SD*=1.06	Q24.2	*M*=5.16, *SD*=1.29	0	1	9 (90.0%)	
	Neutral direction	Q10.1	*M*=3.45, *SD*=1.46	Q10.2	*M*=5.31, *SD*=1.06	3	2	9 (64.3%)	2
		Q11.1	*M*=3.97, *SD*=1.55	Q11.2	*M*=5.37, *SD*=1.13	3	2	13 (72.2%)	
	Expectancy violations	Q25.1	*M*=2.63, *SD*= 1.34	Q25.2	*M*=4.97, *SD*=1.29	1	2	(75.0%)	1
Communication styles	Directness	Q13.1	*M*=4.67, *SD*=1.41	Q13.2	*M*=5.86, *SD*=0.35	1	1	15 (88.2%)	10
	Indirectness	Q26.1	*M*=3.13, *SD*=1.19	Q26.2	*M*=5.56, *SD*=0.82	0	1	13 (92.8%)	2
		Q27.1	*M*=2.10, *SD*=1.27	Q27.2	*M*=5.11, *SD*=1.15	1	3	5 (55.5%)	
Non-collaborative behavior	Unprofessional tone	Q28.1	*M*=2.84, *SD*=1.15	Q28.2	*M*=4.83, *SD*=1.22	1	1	13 (86.7%)	3
		Q29.1	*M*=2.60, *SD*=1.13	Q29.2	*M*=4.91, *SD*=1.15	1	3	7(63.6%)	

	Unprofessional attitude	Q30.1	*M*=3.35, *SD*=1.65	Q30.2	*M*=5.06, *SD*=1.32	0	2	10 (83.3%)	16
		Q31.1	*M*=1.89, *SD*=1.25	Q31.2	*M*=5.11, *SD*=1.24	1	2	5 (62.5%)	
		Q32.1	*M*=1.89, *SD*=1.06	Q32.2	*M*=5.03, *SD*=1.32	1	1	5 (71.4%)	
	Non-compliance with rules	Q33.1	*M*=3.60, *SD*=1.35	Q33.2	*M*=4.54, *SD*=1.92	3	1	14 (77.8%)	15
		Q34.1	*M*=1.92, *SD*=0.77	Q34.2	*M*=5.55, *SD*= 0.82	2	1	6 (66.7%)	
Collaborative behavior	Professional attitude	Q14.1	*M*=5.10, *SD*=1.06	Q14.2	*M*=5.97, *SD*=0.16	1	2	14 (82.3%)	23
		Q15.1	*M*=5.38, *SD*=0.79	Q15.2	*M*=5.92, *SD*=0.28	0	5	11 (68.7%)	
		Q16.1	*M*=4.89, *SD*=0.92	Q16.2	*M*=5.67, *SD*=0,62	0	4	13 (76.5%)	
	Supportiveness	Q17.1	*M*=4.42, *SD*=1.00	Q17.2	*M*=5.76, *SD*=0.49	1	3	13 (76.5%)	11

Thus, it can be said that the intercultural factors identified in the qualitative analysis of the Scenarios, organized into categories and sub-categories (first two columns of Table 7.20), were confirmed by the key stakeholders of international RT communications. Their responses to the Likert scale questions in the quantitative instrument (columns 3 to 6) and also to the qualitative component added by the open-ended questions (columns 7 to 10) corroborate in the validation of the provisional taxonomy. Nevertheless, beyond confirmation, integrated findings from both strands suggest a possible expansion of the taxonomy, including issues related to gender roles, perceptions and expectations. This possible expansion will be considered after the second phase of the study, based on other sources of data generated from a bigger number of participants.

In conclusion, integration between Phases 1 and 2 of the larger multiphase MM study can be explained by the fact that results from Phase 1 informed the selection of theoretical and empirical studies for Phase 2, while Phase 2 findings provided validation and expansion of the taxonomy of intercultural factors. In the next chapter, details of data collection, analyses and results of Phase 2 of this research study are presented and discussed.

Chapter 8 Phase 2: Construct Specification in the Testing of Pilot-ATCO Communication: The Contribution of Domain Experts

In the last chapter, I showed how a MM study was used to identify the intercultural factors that can affect the way pilots and ATCOs interact in the English language, through a qualitative analysis of authentic scenarios of RT communications, and the confirmation of those factors by international pilots and ATCOs, through a quantitative analysis of their responses to an online survey.

In this chapter, I further investigate the ESP construct of international aviation RT communications. This is crucial in order to specify a framework to inform test development. Therefore, following Fulcher and Davidson's (2009) three layers of the test development process, in this Phase of this multiphase mixed methods study, the first layer (related to *Models)* and the second layer (related to *Frameworks*) are addressed. I begin by explaining how Phase 2 is linked to Phase 1, then I state the purpose and research questions specific to Phase 2, present the method used and, finally, I provide a discussion of results.

8.1 Purpose and Research Questions

Phase 2 is a qualitative study that builds on Phase 1 in a number of ways. First, the interfaces of Aviation English and intercultural communications highlighted in Phase 1 and confirmed by the taxonomy of intercultural factors suggested points of contact with other disciplines. Thus, Phase 1 served as a basis to guide the selection of studies to be included as part of a systematic review of theoretical and empirical research (see Figure 3.3), from which models of language use that account for the communicative demands of international RT communication were proposed. Second, in order to move from these

models to the specification of a framework that maps the constructs considered to be relevant to the target language use (TLU) domain of pilot and ATCO interactions, the same structure of the Intercultural Communicative Competence models that underpinned data analysis in Phase 1 (e.g., Byram, 1997; Fantini, 2000) was applied, i.e., one that addresses the dimensions of *awareness* and *attitudes* along with *knowledge* and *skills.* Further, for the validation of the matrix of construct specification, the six scenarios of RT communication analyzed in Phase 1 were used in Phase 2 to elicit the perceptions of language testing scholars and aviation stakeholders, regarding the communicative needs of pilots and ATCOs in the multicultural context of international radiotelephony. Finally, data from the focus group discussions with aviation stakeholders were also used to validate the taxonomy of intercultural factors, which was the outcome of Phase 1.

The goals of Phase 2 were to: (a) investigate the ESP construct of international aviation RT communications in order to specify a framework to inform test development, by eliciting key stakeholders' perspectives from diverse linguistic and cultural backgrounds; (b) stress the critical role that domain experts play in construct definition within LSP contexts; and (c) validate the taxonomy of intercultural factors proposed in Phase 1, based on aviation stakeholders' perceptions.

In terms of design, this qualitative study was conducted in three sequential steps, as shown in Figure 8.1, with each step aiming to answer a specific research question.

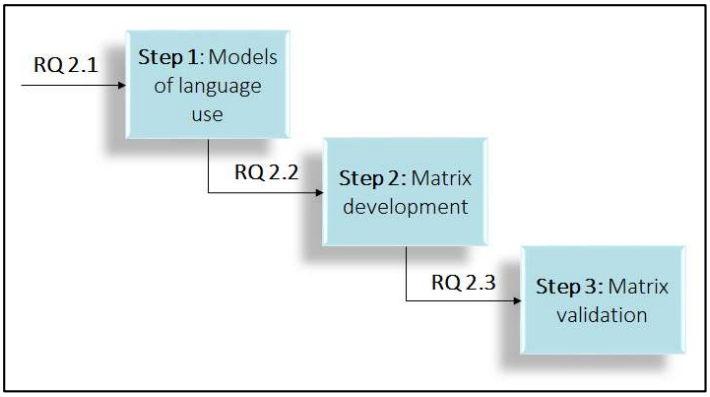


Figure 8.1. Diagram of steps within the qualitative study (Phase 2)

The research questions that guided each step within Phase 2 are as follows:

RQ 2.1) What theoretical models of language use would account for the communicative needs of pilots' and ATCOs' occupational domain?

RQ 2.2) How can this construct be articulated and specified from the models to a framework which better informs test development?

RQ 2.3) What components of the construct are validated by key aviation stakeholders?

Additionally, as a validity strategy, Phase 2 also aimed to validate the taxonomy of intercultural factors proposed in Phase 1, by triangulating data from different sources. Therefore, data from the focus group discussions with aviation stakeholders were also analyzed in order to confirm the taxonomy sub-categories. Figure 8.2 illustrates the existing interconnection between Phases 1, 2, and 3 of this multiphase MM research study and the contribution of all three in the generation of meta-inferences.

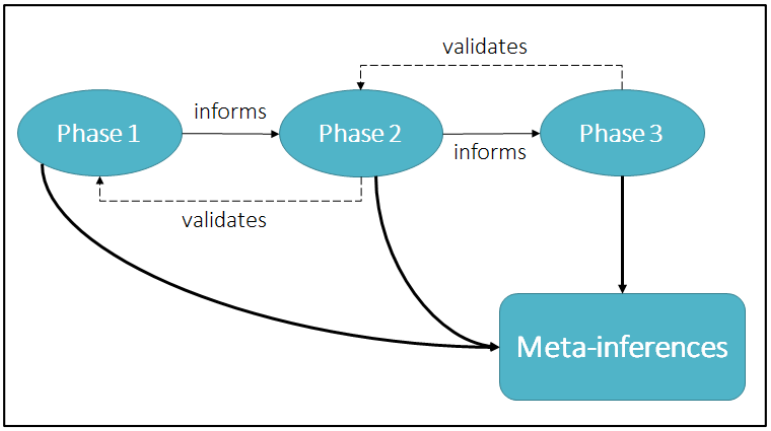


Figure 8.2. Interconnection between Phases 1, 2, and 3

In the sections which follow, I explain the method (participants, instruments, procedures, and analysis) used in this qualitative study, which represents Phase 2 of the larger multiphase mixed methods research. Having presented the method, I then discuss the results in Section 8.3.

8.2 Method

8.2.1 Participants.

Steps 1 and 2 of this qualitative study did not require interaction with human participants for data collection. However, in Step 3 participants were recruited in order to discuss, in small groups, their perceptions of the communicative needs of pilots and ATCOs in the multicultural context of aviation RT. A purposeful sampling approach was used, in order to include both native and non-native speakers of English, ideally from different linguacultural backgrounds, who were pilots, ATCOs, aviation English teachers/material developers, aviation English examiners/test developers, researchers in the field of aviation communications/aviation English teaching and testing, or regulators. This was possible because participants were recruited in two international conferences/workshops related to aviation communications and aviation English

teaching/testing, where the researcher announced the study[23]. The first took place in Dubrovnik, Croatia, in April 2017, and the second in São José dos Campos, Brazil, in September 2017. Conference organizers were contacted in advance of the events and granted approval. Participants were made aware of the study through an oral script used for recruitment (see Appendix U) and the researcher also advertised the study using a poster (see Appendix V).

A total of 128 participants agreed to take part in this Phase of the study. In terms of language background, there were 20 native speakers of English and 108 non-native speakers of English, and in relation to gender, 52 were male and 76 were female. Figure 8.3 details their main occupation in the specific domain of aviation.

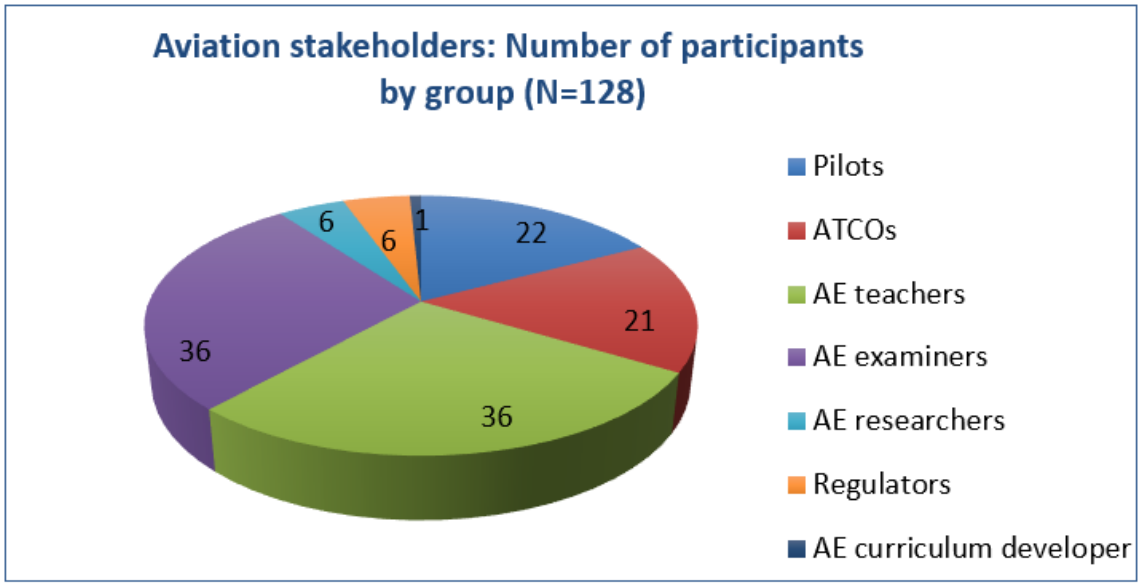


Figure 8.3. Number of participants by group of stakeholders

The participants were organized into 26 different groups, of which 13 were multilingual groups (i.e., a mix of NSs and NNSs of English, from different L1

[23] Ethics approval to conduct this part of the study was received from Carleton University Research Ethics Board (CUREB) on April 20, 2017 (see Appendices C and D).

backgrounds) and 13 monolingual groups (i.e., NNSs of English sharing the same L1 – Portuguese). Most importantly, pilots and ATCOs, the ones who actually communicate over the radio, account for 43 (33.6%) of the sample. Comments from all participants provided evidence in support of the validation of the matrix of construct specification developed in Step 2.

8.2.2 Instruments.

For Steps 1 and 2, that is, the design of models of language use and the matrix development, no other instruments of data collection were necessary. Both steps considered the information extracted from the literature review, as described in Chapter 3, which drew on theoretical and empirical studies.

In Step 3, each group of aviation stakeholders that participated in the focus group discussions, which I called Focus Group Type 1, received the following material:

1. the Letter of Introduction containing details of the study (see Appendix W);
2. consent forms to all participants (see Appendix I);
3. a handout including: (a) a question related to the number of participants in the group according to their main role in aviation and to their language background; (b) the written transcript of one of the six scenarios of authentic radiotelephony communication analyzed in Phase 1; and (c) a set of six questions to guide participants' discussions and enough space for them to provide some notes related to their perspectives (see Appendix X); and
4. a recording device.

8.2.3 Procedures.

8.2.3.1 Step 1: Models of language use.

A review of conceptual (theory), empirical (research) and practical (policies/practice) studies representative of the field of intercultural encounters in aviation and relevant to answer this study's research questions was conducted in order to propose models of language use that account for the communicative demands of pilots and ATCOs involved in RT communication. This literature review is detailed in Chapter 3 and presented as a literature map (see Figure 3.3), i.e., a visual summary of articles, books and documents in the field of aviation radiotelephony communications, and research conducted by scholars in a number of interconnected fields, both theoretical and empirical.

All these readings, informed by findings from Phase 1, made it possible to build different representations of the specific occupational domain of international communications between pilots and ATCOs. First, from Doc 9835 (ICAO, 2010), four key areas of interest were identified, and studies were reviewed and structured thematically according to the major domains of Aviation English, English as a lingua franca, intercultural awareness, and interactional competence. Second, relevant features of each domain that apply to the context of RT communications, and/or that could somehow have an impact on their outcomes, were carefully chosen according to their importance to the context and suitability to build theoretical models. As a result, these representations or models, convey: (a) what is required for effective communication in the intercultural and highly specific context of RT; (b) what affects the interaction between pilots and ATCOs in terms of fixed cultural frames of reference and emergent

features; (c) what needs to be included in a test to identify if a pilot or ATCO is ready to communicate successfully in intercultural RT communications.

8.2.3.2 Step 2: Matrix development.

After building the models of language use that account for the communicative demands of international RT communication, the next step was to develop a framework that specifies the constructs considered to be relevant to this specific context. Initially, the structure of the matrix was defined, specifically in what relates to the four key domains to be included, i.e., four specified sub-areas of activity or knowledge, namely, Aviation English, English as a lingua franca, Intercultural awareness, and Interactional competence, and the aspects that would constitute the dimensions of interest, both drawn from the proposed models. Second, a synthetic organization (Li & Wang, 2018) of recurring themes and patterns emerging from the studies was conducted, followed by a categorization of components of the construct, i.e., relevant features of the RT context that pilots and ATCOs should be aware of, know, use appropriately, and display as attitude for successful intercultural encounters over the radio. Finally, these components were organized according to their best fit to each domain and dimension intersection, generating the draft matrix of construct specification.

8.2.3.3 Step 3: Matrix validation.

In order to validate the matrix of construct specification, focus group discussions were chosen as the method of data collection. The discussions started after a brief presentation explaining how I got to the provisional taxonomy of intercultural factors in Phase 1 (including only the level of categories), and after describing the Model of RT communications in intercultural contexts (see Fig. 8.8) as well as the AE, ELF, ICA and

IC overlap model (see Fig. 8.9). Participants were divided in small groups, and group size ranged from three to seven participants, each including ideally at least one operational participant (pilot or ATCO), one aviation English teacher/material developer, one examiner/test developer, one researcher in the field of aviation communications/aviation English teaching and testing, and one regulator. Each group received the material described in Section 8.2.2, and data collection started only after participants signed the consent form (see Appendix I). The aviation stakeholders were invited to analyze the transcript of one scenario of authentic RT communication and to answer six questions, within 30 minutes. Subsequently, an inter-group discussion took place, moderated by the researcher, in which a member of each group presented participants' main perceptions related to the scenario analyzed and the questions proposed. Intra-group and inter-group discussions were audio recorded. The researcher kept a handout from each group containing participants' background information and notes from their discussions. The same procedures were used in both conferences. However, in Croatia two sessions were conducted with a total of 10 focus groups, whereas in Brazil three sessions took place, totalizing 16 focus groups. Audio files from the 26 focus groups were fully transcribed with the aid of *Express Scribe Transcription Software* version 6.10. Each focus group transcription received a sequence number and was also classified according to the scenario analyzed by the group (e.g., FG 01 of 26 Scenario 1). Essential information related to sites of data collection, time frame, data format and number, participants characteristics, and the context of each scenario were included at the beginning of the each data set, as a form of Attribute Coding (Saldaña, 2009). Data were coded in a systematic way, including first cycle and second cycle coding. For reliability purposes, a

second coder also coded a sample of units previously coded by the researcher and inter-coder reliability was calculated.

8.2.4 Analysis.

8.2.4.1 Step 1: Models of language use.

Fulcher and Davidson (2009) define *Models* as the first layer of architectural documentation that articulate test design decisions (see Figure 3.1), more specifically, "a theoretical overview of what we understand by what it means to know and use a language" (p. 126). Models of language use can be represented in different ways, depending on the purpose the researcher wants to achieve by focusing on different aspects of the communicative context.

Therefore, informed by the review of theoretical, empirical and practical studies discussed in Chapter 3, three models were proposed in this study. They were developed, and will be discussed, from the general, i.e., a broader theoretical view of language use in intercultural communications, to a more specific model for the occupational purpose of international RT communications.

The criteria that guided the design of the models are based on the overall purpose of Phase 2 and its related research questions, comprising the following:

1. Comprehensiveness – a model that would expand the existing notion of communicative competence (e.g., Canale & Swain, 1980; Bachman and Palmer, 1996; Celce-Murcia, 1995) to include other competencies required for international communications in the globalized world of aviation;
2. Interpretability – a model that would depict the dynamic and co-constructed nature of the interactions between two individuals with

distinct cultural frames of reference, using Aviation English as a lingua franca and relying on specific background knowledge in the context of RT communications; and

3. Usefulness to support test development – a model that would represent the ESP construct in international RT communications, required to identify and recognize professionals with language and communication skills for the aviation RT workplace.

The three models and how they evolved based on the criteria above are discussed in Section 8.3.

8.2.4.2 Step 2: Matrix development.

In order to represent the construct, we need to specify it. As Fulcher and Davidson (2009) point out: ". . . the constructs to be tested, [are] selected from models, because they are shown to be *relevant* [emphasis added] to the specific context in question, and *useful* [emphasis added] in the decisions that need to be made" (p.127). Therefore, the construct framework, or the matrix of construct specification, was built by breaking down the construct of interest into four sub-areas or domains, which define features of the construct that "reflect the realities of language use in the target situation" (Douglas, 2000, p. 69), in relation to what was available in the literature and what was available theoretically. Thus, the domains of Aviation English (AE), English as a lingua franca (ELF), Intercultural Awareness (ICA), and Interactional Competence (IC) were the key ones that structured the matrix (see literature map in Figure 3.3), which were also represented in the three proposed models of language use for the RT domain. In addition, the matrix covers four dimensions, i.e., knowledge, skills, attitudes, and awareness,

referred to as important components in discussions of Intercultural Communicative Competence (e.g., Byram, 1997; Fantini, 2000) and Intercultural Awareness (e.g. Baker, 2011), which, notably, were included in the first model. As Messick (1994) states, when referring to performance assessments, "a construct-centered approach would begin by asking what complex of knowledge, skills, or other attributes should be assessed" (p. 16), and those were set as the columns of the matrix. Table 8.1 illustrates the overarching structure of the matrix of construct specification.

Table 8.1. Overarching structure of the construct framework

	Awareness	Knowledge	Skills	Attitudes
Aviation English				
English as a lingua franca				
Intercultural awareness/competence				
Interactional Competence				

In order to populate the matrix of construct specification, a synthesis of the key and recurring features relevant to the target language use (TLU) domain of intercultural pilot and ATCO interactions was carried out, informed by what scholars highlighted as important in the theoretical, empirical and practical studies included in the literature review (see Chapter 3).

The components of the construct were selected and organized in the matrix according to the intersection of the cells, i.e., considering first what was related to the domain of Aviation English and the awareness dimension (e.g., situational awareness), then, AE and the knowledge dimension (e.g., background knowledge: rules and procedures), AE and skills (e.g., CRM - Crew/Corporate Resource Management), and finally, AE and attitudes (e.g., professional tone and attitude). The same process was

carried out for the domains of English as a lingua franca, Intercultural awareness/competence, and Interactional Competence.

8.2.4.3 Step 3: Matrix validation.

In preparation for the validation of the matrix of construct specification, I triangulated my initial findings from the literature review with a relevant group of stakeholders (e.g., language testers, ESL teachers) to whom I showed the structure of the matrix and had a brief discussion on the possible key components of the cells. Field notes were collected and a cross-check was undertaken in each cell of the matrix in order to verify if these components were already part of the draft matrix proposed in Step 2 or if they were new components that emerged from this discussion.

Then, I moved to the analysis of the focus group discussions with aviation stakeholders. As this phase of the study involved a large amount of text-based qualitative data, data analysis was conducted using *NVivo Version 12 Plus for Windows*, a qualitative data analysis computer software package produced by QSR International, which makes it possible to organize, store, categorize data, and visualize results in different formats whenever necessary. Focus group files were imported into Nvivo and classified according to the language background of participants as "monolingual" or "multilingual".

In order to validate the matrix of construct specification, in the First Cycle coding I explored the data using Provisional Coding (Saldaña, 2009), which begins with a 'start list' of codes generated from "literature reviews related to the study, the study's conceptual framework and research questions, previous research findings, pilot study fieldwork, the researcher's previous knowledge and experiences (experiential data), and researcher-

formulated hypotheses or hunches" (p. 120). Therefore, the nodes used for this coding were the four dimensions of interest used to build the matrix in Step 2: awareness, knowledge, skills, and attitudes. Each focus group file was coded in the following way: the selected text, focusing on crucial aspects of RT communication, was highlighted and dragged to the respective node folder. Figure 8.4 portrays how this cycle of coding looks in Nvivo, with an example from "FG 08 of 26 Scenario 4". We can see on the left the highlighted parts of the text that have been coded and on the right the coding density and a color-coding scheme for knowledge, skills, attitude, and awareness.

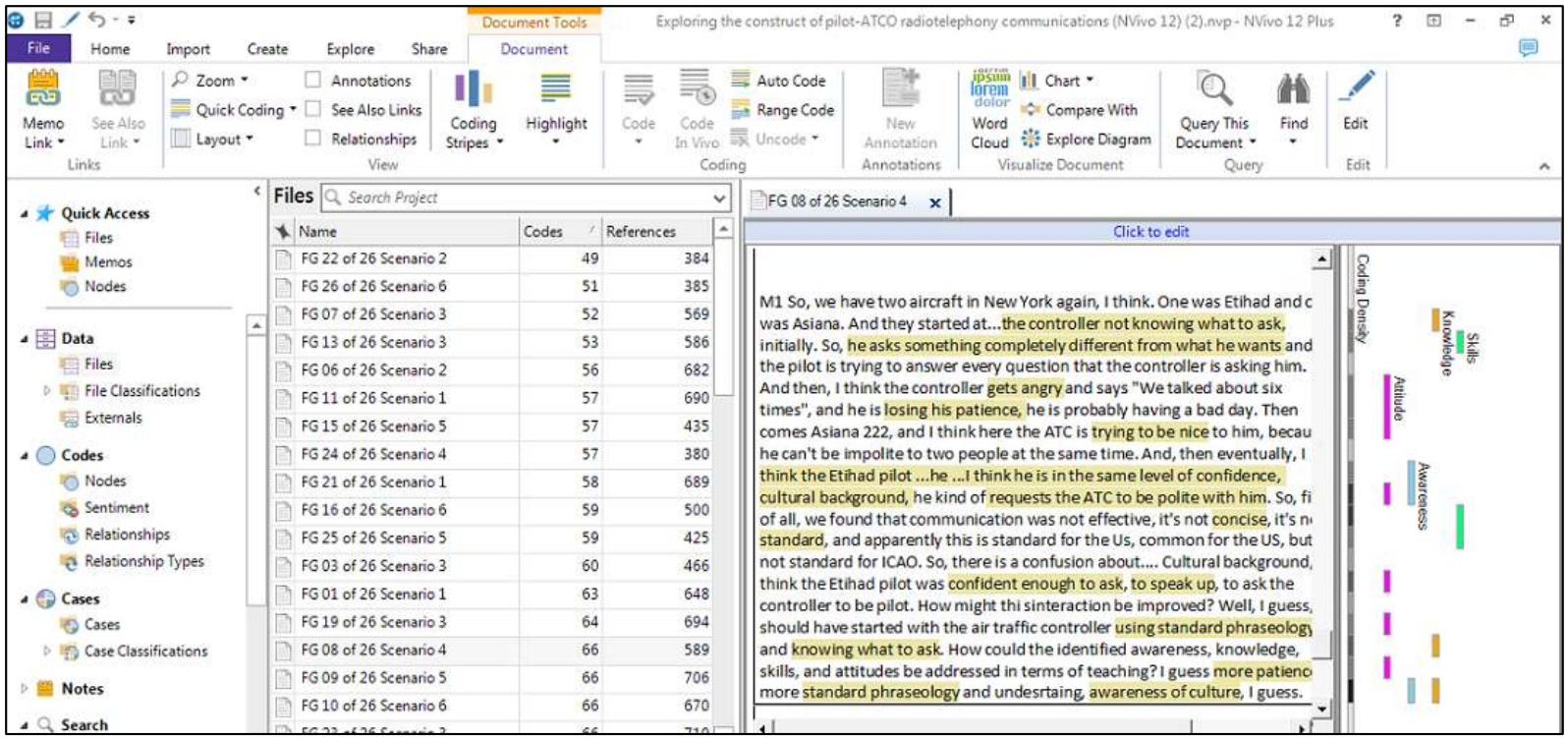


Figure 8.4. Example of First Cycle Coding using NVivo

After completion of the First Cycle coding, inter-coder reliability analysis was undertaken to establish the reliability of the coding process. Two out of the 26 focus group discussions were selected, one from Scenario 1 and the other from Scenario 4. Then, 99 units of coding were coded by a second coder, a NS of English and experienced qualitative researcher but with no background in aviation. A Cohen's Kappa coefficient was calculated in SPSS, version 23, and the result was significant at .81, which represents a very good agreement[24]. Sensitivity analysis of agreement, i.e., the consistency of Coder 2 classification against the 'gold standard' from Coder 1 (Pallant, 2007), was also determined for each of the four dimensions: attitude (85%), awareness (95%), skills (95%), and knowledge (70%).

At this point of data analysis it was possible to have an overview of the number of references coded for attitude, awareness, knowledge, and skills in each focus group file, by using the Crosstab function of NVivo (i.e., crosstab codes against cases). However, in order to answer RQ 2.3, data analysis had to continue into a Second Cycle of coding. The objective was to verify which components of the construct in the draft matrix would be validated, the weight of each component (i.e., the number of coding references) and whether new components would emerge from the focus group discussions. This was achieved by doing the following in NVivo:

- Creating new nodes for each domain (i.e., Aviation English, English as a lingua franca, Intercultural awareness and Interactional competence), sub-nodes within each node for each dimension (i.e., awareness, knowledge, skills, and attitudes), and sub-sub-nodes within each sub-node for each

[24] Pallant (2007) explains that "a value of .5 for Kappa represents moderate agreement, above .7 represents good agreement, and above .8 represents very good agreement (p. 220).

component of the construct in the draft matrix (see example in Figure 8.5 of ‘standardized phraseology’ for the dimension of ‘knowledge’, within the domain of ‘Aviation English’); and

- Coding all the references previously coded in the First Cycle coding into the new nodes/sub-nodes/sub-sub-nodes.

Again, for the Second Cycle coding, the coding method applied for the validation of the matrix of construct specification was Provisional Coding, using the construct components from the draft matrix as the start list.

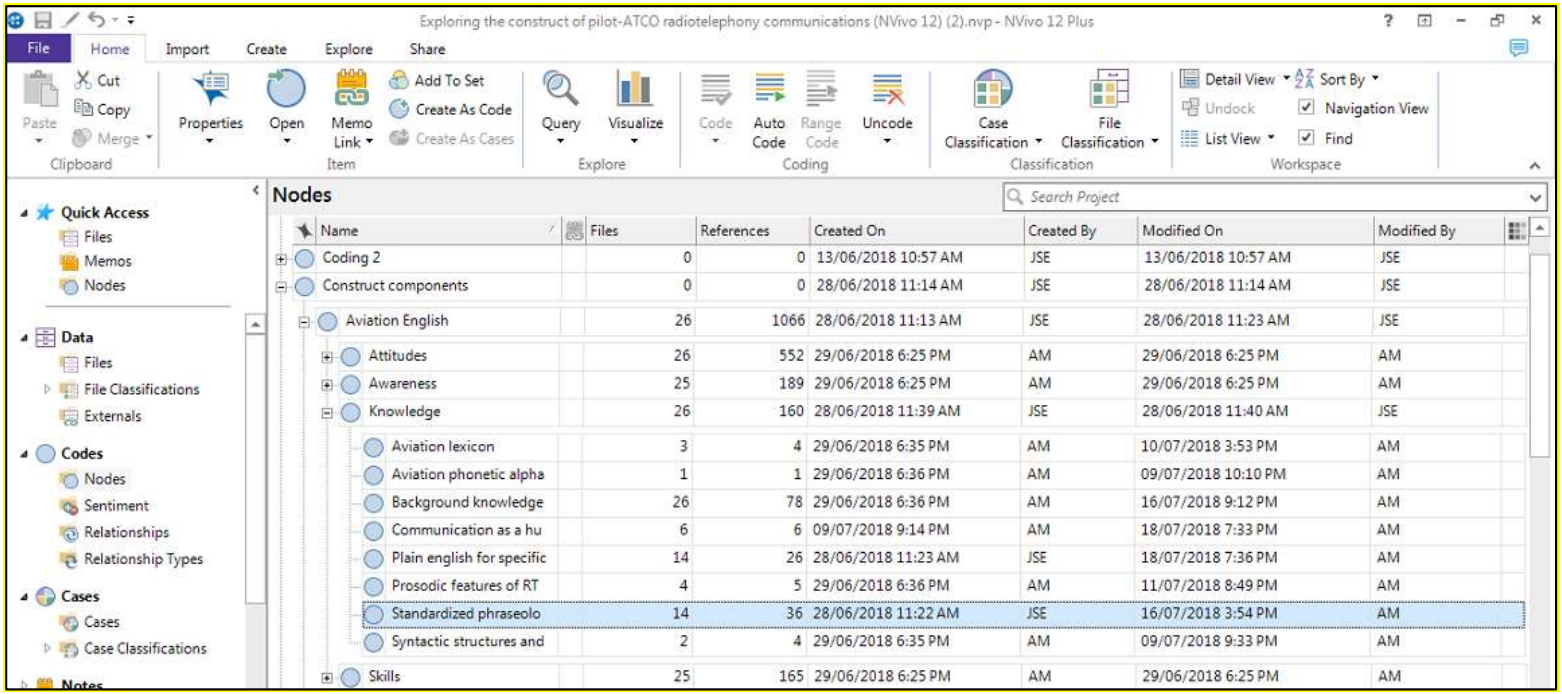


Figure 8.5. NVivo example of sub-nodes and sub-sub-nodes within the node of Aviation English

Notwithstanding, data from the focus groups with aviation stakeholders also served another purpose: to validate the taxonomy of intercultural factors proposed in Phase 1 by using a different data source. To this end, coding focused on finding culture-related factors that could impact pilot-ATCO communications on the radio. The chosen method of coding was Elaborative Coding, given that the study in Phase 2 builds on the MM study in Phase 1. Saldaña (2009) explains that:

> Elaborative coding is appropriate for qualitative studies that build on or corroborate previous research and investigations. Basically, the second study elaborates on the major theoretical findings of the first, even if there are slight differences between the two studies' research concerns and conceptual frameworks. Different participants or populations can also be used for the second study. This method can support, strengthen, modify, or disconfirm the findings from previous research. (p. 168)

Following Saldaña (2009), the sub-categories of the taxonomy of intercultural factors (i.e., the major theoretical findings from Phase 1) were used as the nodes for the coding process. The objective was to verify if all existing subcategories would be validated, if new sub-categories would emerge, and also verify the number of references for each sub-category to get the idea of their importance. Figure 8.6 shows a snapshot of this coding process in NVivo.

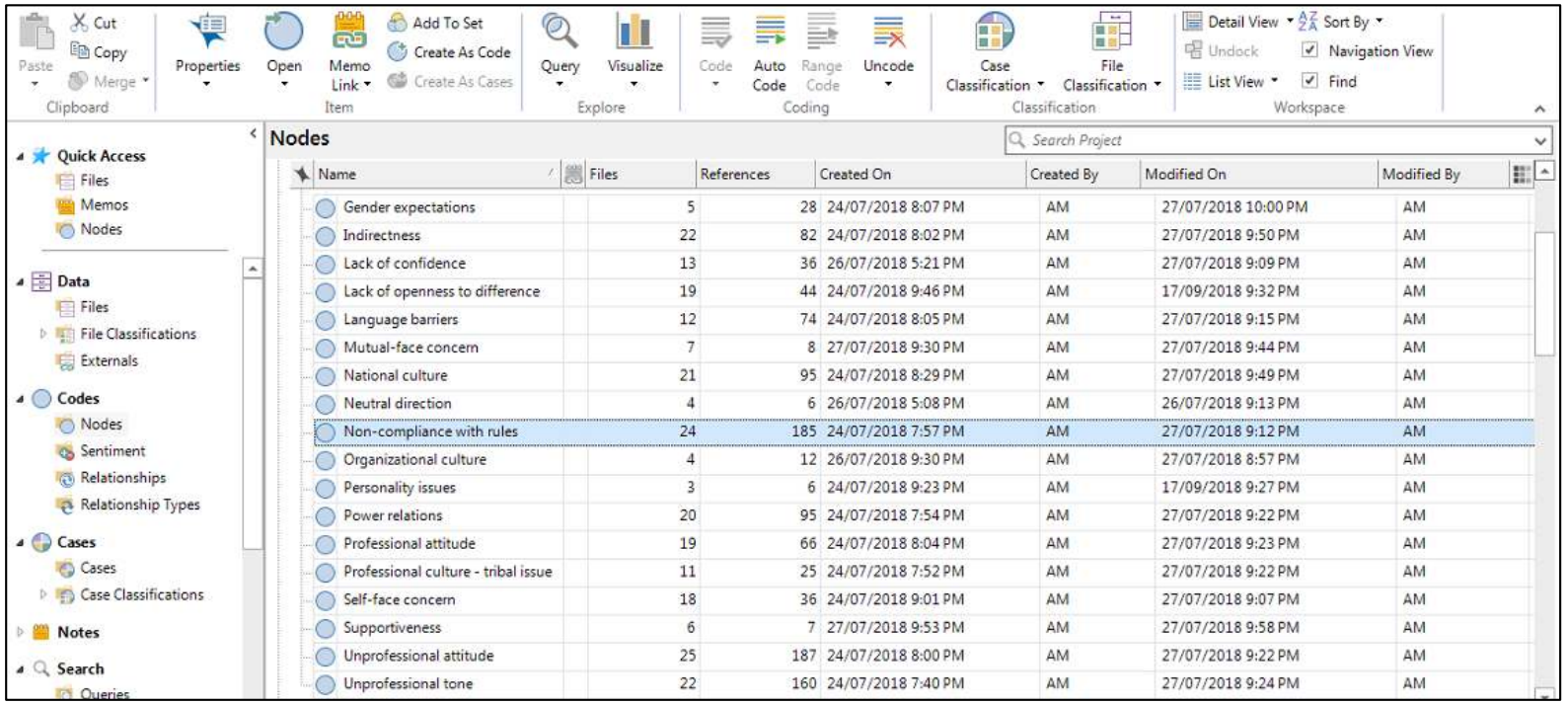


Figure 8.6. NVivo example of nodes used to validate the taxonomy of intercultural factors

8.3 Results and Discussion.

Results from Phase 2 are presented and discussed in this section in relation to each research question.

RQ 2.1) What theoretical models of language use would account for the communicative needs of pilots' and ATCOs' occupational domain?

Considering the characteristics of this complex and intercultural context, as a result of the literature review and combined with findings from Phase 1, it is possible to say that a more comprehensive notion of communicative competence is necessary to cope with pilots' and air traffic controllers' communicative needs, whether native or non-native speakers of English. Alptekin (2002) questions models defined by the notion of idealized native speaker-listeners: "The conventional model of communicative competence, with its strict adherence to native speaker norms within the target language culture, would appear to be invalid in accounting for learning and using an international language in cross-cultural settings" (p. 63). Likewise, Byram (1997) privileges the notion of 'intercultural speaker' and the inclusion of the intercultural dimension to the traditional models of communicative competence. His proposed model of Intercultural Communicative Competence (ICC) encompasses five components or *savoirs*: *savoir être* (intercultural attitudes), *savoirs* (knowledge), *savoir comprendre* (skills of interpreting and relating), *savoir aprendre/faire* (skills of discovery and interaction) and *savoir s'engager* (critical cultural awareness).

Byram's model proves useful in informing an initial attempt to account for pilots' and ATCOs' communicative demands, both for native and non-native speakers of English. Therefore, first I began by incorporating the intercultural dimension within the

more traditional communicative competence framework, guided by the way it was defined in ICAO Doc 9835 (2010, Section 2.3.2) for the aviation context. It resulted in an Intercultural Communicative Competence model for radiotelephony communications which maps the territory where RT communications take place, represented by the radar screen, as shown in Figure 8.7.

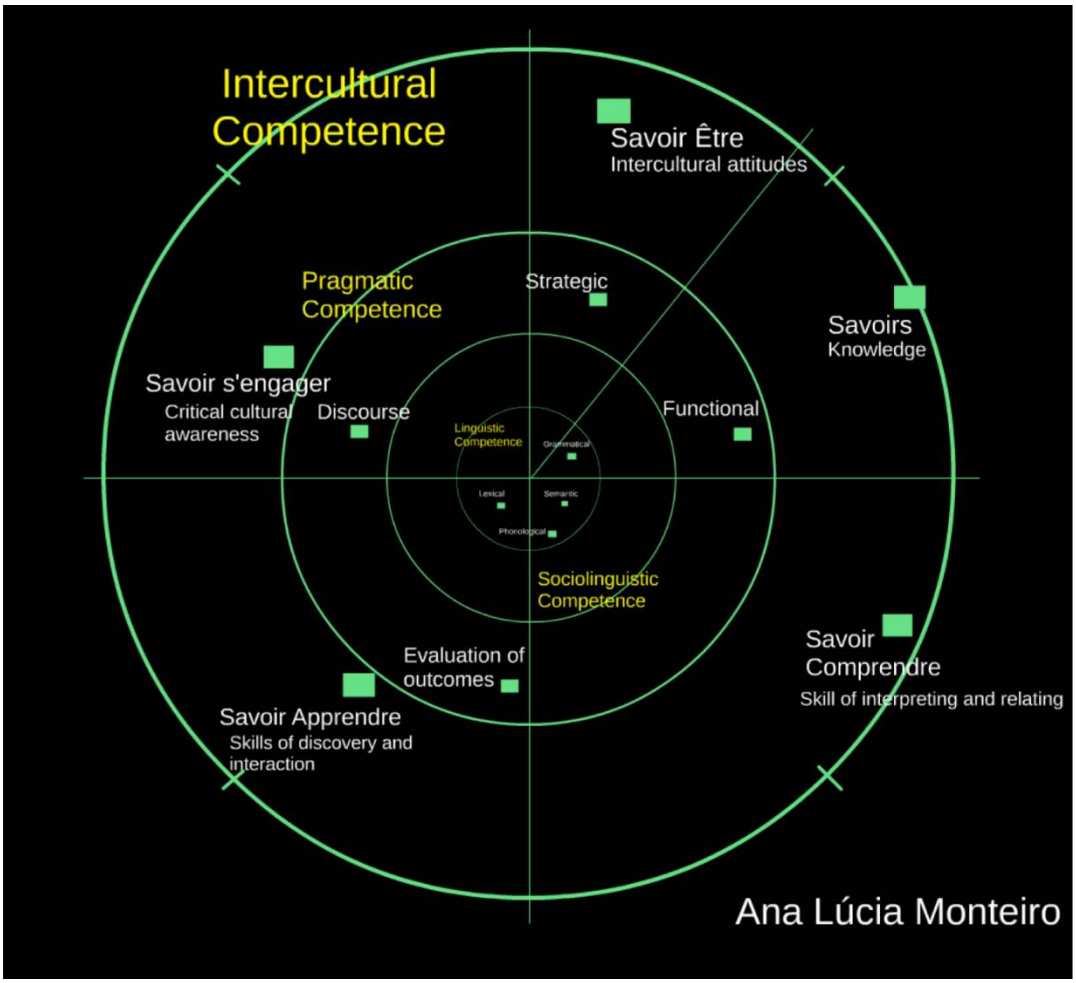


Figure 8.7. First Model of the discursive space of RT communications (Monteiro, 2016a)

This first model comprises, from the inner layer to the outer ones: (a) Linguistic Competence, defined as "the knowledge and meaningful use of the linguistic features of a given language or languages" (ICAO, 2010, p. 2-2), which can be split into lexical, grammatical, semantic, and phonological subskills; (b) Sociolinguistic Competence,

which "involves understanding the social (including occupational) context in which language is used …[and] being able to make appropriate use of markers of social relations, politeness conventions, register differences, dialect and accent" (p. 2-2); and (c) Pragmatic Competence, referring to "a number of skills used to make or give meaning to language in a given situation or context"(p. 2-2). According to Doc 9835 (ICAO, 2010), Pragmatic competence further includes: (a) strategic competence, i.e., "how language users mobilize or balance their resources to activate skills and procedures, in order to fulfil the demands of communication in context and successfully complete the task in question in the most comprehensive or most economical way feasible" (p. 2-2); (b) discourse competence, which "refers to the ability to combine sentences or utterances to make coherent, whole texts" (p. 2-2); (c) functional competence, related to "the awareness of and ability to make use of the rules governing the way in which language structures are interpreted … in a given context — 'language functions' — and the ways in which these functions are commonly sequenced to establish conversational structures" (p. 2-3); and (d) evaluation of "outcomes of the use of language in the real world, for example, impacts on safety or impacts on efficiency" (p. 2-3).

Finally, this first model is expanded in the outer layer with the inclusion of the dimensions of Intercultural Competence, i.e., attitudes, knowledge, skills and awareness. As Sussex and Curtis (2018) explain, "it is now possible to propose that we are moving, not into a post-communicative framework for language education, but into one enhanced by a major focus on intercultural communication" (p. 4), with a special emphasis on intercultural communicative competence.

The second model corresponds to a narrower conceptualization of the communicative demands of pilots and ATCOs as they exist in this particular workplace context. This model shows the interaction of several layers of culture (Hofstede, 1991), or discourse systems (Scollon & Scollon, 2001), i.e., an individual's cultural frames of reference, with the dialogic nature of the utterances (Bakhtin, 1986) in an interaction between two individuals using AE as a lingua franca and governed by the rules of the air traffic control system (see Figure 8.8).

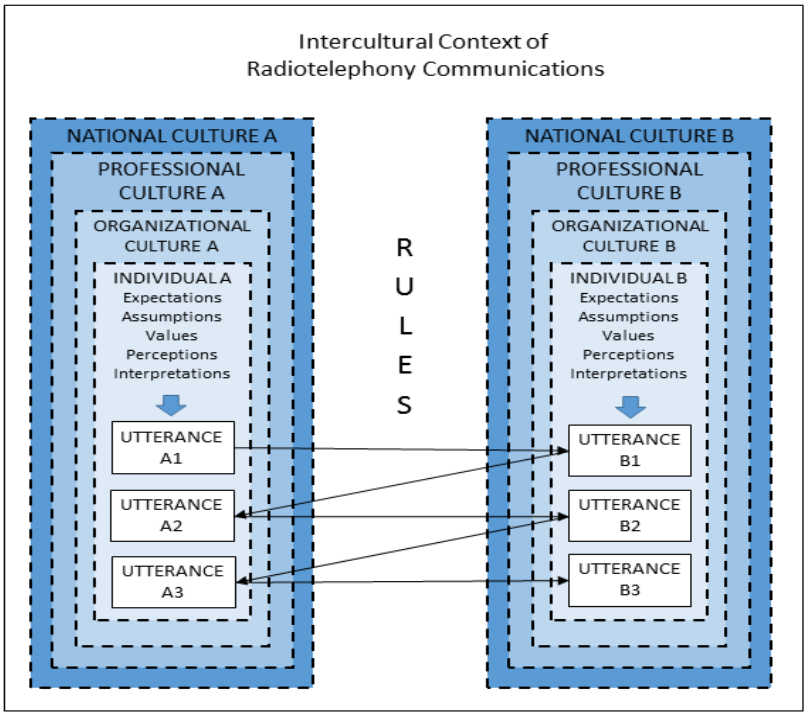


Figure 8.8. Second Model of radiotelephony communications in intercultural contexts (Monteiro, 2016b)

Byram, Gribkova and Starkey's (2002) definition of intercultural competence as the "ability to ensure a shared understanding by people of different social identities, and their ability to interact with people as complex human beings with multiple identities and their own individuality" (p. 10), seems to explain the complexity of factors portrayed in

the blue rectangles. Moreover, the combination of the interlocutors' cultural identities, or the conflict among them, may influence what they say, how they say it, the responses they expect, and how they react to previous utterances. As Snow (2018) explains, in intercultural encounters the interpretation process is fundamental, and he echoes Bakhtin's (1986) communication chain and dialogism stating that "in the chain of communication moves, most moves are impacted by one's interpretation of whatever move came before" (p. 60). As a result, to account for each individual's – pilot or ATCO – own set of expectations, assumptions, values, perceptions and interpretations, according to the various cultural groups they are inserted in, a model which also includes a more dialogic, dynamic and emergent interaction of culture, language and communication is necessary. As Kesckes (2014) highlights, culture has *a priori* elements, i.e., ethnic or cultural marking in communicative behavior (see blue rectangles in the model, portrayed in Figure 8.8) and emergent features, co-constructed in the moment of interaction (see utterances A1, A2, A3 and their responsive reactions in B1, B2 and B3), which should be combined to approach culture in a dialectical and dynamic way. In addition, this second model, which presupposes the use of AE as a lingua franca, takes into consideration Baker's (2011) notion of intercultural awareness (ICA) as an expanded and dynamic framework for intercultural competence. His definition of ICA as, "a conscious understanding of the role culturally based forms, practices and frames of reference can have in intercultural communication, and an ability to put these conceptions into practice in a flexible and context specific manner in real time communications" (p. 202), includes the two main ideas the model purports to convey.

Although it is an interesting model, it would be very hard for a test developer to move from this second model to an actual test specification without breaking down the theoretical concepts into smaller components that are relevant to a particular testing situation. As the goal of this study is to address the operationalization of the aviation RT-specific construct in test design, a third model is necessary, one that illustrates clearly each one of the critical constructs that interact in this intercultural workplace context, how and where they overlap (see Figure 8.9).

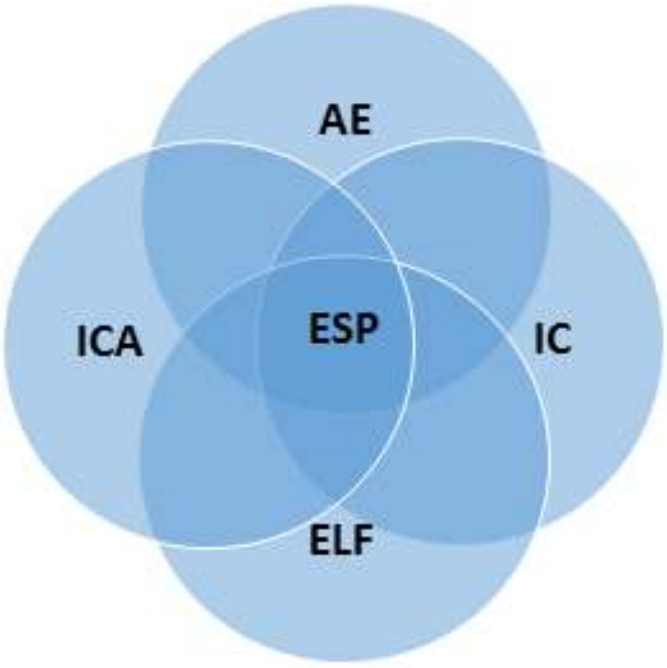


Figure 8.9. Third Model of ESP in RT communications = AE, ELF, ICA, and IC overlap (Monteiro, 2016c)

As explained in the discussion of the second model, the focus of the analysis is on the interaction and negotiation (IC) between two individuals using Aviation English (AE) as a lingua franca (ELF), governed by the rules and procedures of the air traffic control system (AE), shaped by their cultural frames of reference, but in a flexible and context-specific manner (ICA). And in fact, the third model illustrates these four constructs in the

figure's circles and the ESP overlap, where exactly the test should be situated, that is, the core features that must operate smoothly and simultaneously for effective communication in this aviation workplace.

What this third model does not give us are the details on what constitutes each one of these overlapping circles, nor does it give us a sense of how important they are. Thus, it was necessary to move from these models to the specification of a framework that better maps the constructs of interest, namely Aviation English (AE), English as a lingua franca (ELF), Intercultural awareness/competence (ICA), and Interactional competence (IC). This is the topic of the discussion that follows, based on results that address my research questions.

RQ 2.2) How can this construct be articulated and specified from the models to a framework which better informs test development?

The framework presented in this section maps the construct of international RT communication in the aviation context and addresses research question 2.2. It is worth noting that the results of this study suggest that all three proposed models convey, in different ways, the message that background knowledge is relevant for performance in the aeronautical RT occupational context. This is corroborated by what Douglas (2000) stated in relation to specific purpose language ability, that "the construct contains, by definition, specific purpose background knowledge … [which] will necessarily be called upon in the interpretation of the communicative situation and in the formulation of a response" (p. 39). This aspect of the construct is considered in the matrix within the domain of Aviation English across the four columns which represent the four complementary dimensions.

Not surprisingly, the overlap of the domains (i.e., AE, ELF, ICA, IC) is apparent in the matrix, which makes it difficult sometimes to define where a particular component of the construct, e.g., accommodation skills, would best fit or whether it should appear in more than one domain. The same happens to the four dimensions (i.e., awareness, knowledge, skills, and attitudes), which sometimes appear to suit the same component of the construct. For example, standard phraseology appears in AE knowledge but it is also extremely important as a component of AE attitudes when it is a matter of compliance with rules and procedures.

Although the components of the construct that populated the draft matrix were drawn from the models of language use and from theoretical and empirical studies addressing the communicative needs of pilots and ATCOs, it was necessary, as well, to give voice to other domain experts in order to confirm such components as relevant to the specific context of RT communications. This is, in fact, an English for Specific Purposes (ESP) perspective on construct definition, which takes into account the TLU's 'indigenous' assessment criteria (Douglas & Myers, 2000; Elder & McNamara, 2016; Elder et al., 2017; Fox & Artemeva, 2017; Jacoby & McNamara, 1999; Knoch 2014; Pill, 2016). Within international RT communication, these criteria should inform evaluation of the language proficiency requirements applied to this professional/workplace context. Jacoby & McNamara (1999) note the importance of "an insider's view" and point out that such a view is essential in identifying (and addressing) ". . . the complex issues involved in communicating competently" (p. 214) in a TLU domain. Thus, as described in Section 8.2.4.3, an initial group of relevant stakeholders contributed to the specification of the matrix. Their perceptions of what components should be included in the construct

framework are highlighted in Table 8.2: in bold, the ones that were already part of the draft matrix proposed in Step 2, and as underlined text, new components suggested by language testers and ESL teachers.

Table 8.2. Preliminary matrix of construct specification

Construct definition within the aviation radiotelephony domain				
	Awareness	**Knowledge**	**Skills**	**Attitudes**
Aviation English	- rules of use that characterize the domain - safety-critical requirements for **intelligibility[a],** directness, appropriacy, **non-ambiguity** and concision - **threats** presented by cross-cultural communications - **impact o**f communication on safety and efficiency - social and occupational context in which AE is used	- standard phraseology - plain English for the specific purpose of aeronautical RT communications - syntactic structures and language functions used in RT - aviation lexicon - aviation phonetic alphabet and pronunciation of numbers - prosodic features of RT - background knowledge	- apply speech transmitting techniques - use the linguistic features of AE meaningfully - communicate effectively in routine and in highly unpredictable situations - use strategic skills to deal with aviation personnel with different levels of expertise	- compliance with prescribed rules and procedures (e.g. use of phraseology, read back/hear back) - discipline - **professional tone** and attitude - clarity, conciseness and correctness
English as a lingua franca	- **different varieties of English** and speech communities - challenges faced by speakers of EFL and interlocutors' possible linguistic difficulties - difficulty presented by the use of jargon, idioms, slang and colloquialisms - the need to speak English as a lingua franca[b] - language use and language processing	- language as a social practice - different pragmatic norms for different contexts - one's own communicative style and the problems it could pose to ELF interactions - characteristics of one's L1 phonology that may influence English pronunciation - exposure to different international accents	- mediate and negotiate meaning - **accommodate** different accents and dialects - adapt linguistic forms to the communicative needs at hand - adjust and align to different communicative systems (new patterns of phonology, syntax, discourse styles) - self-repair, **rephrase**, paraphrase, and clarify - **notice and repair breakdowns in communication** - preempt misunderstanding - ascertain and deploy appropriate pragmatics - eliminate ambiguous expressions and sentence patterns - adapt speed and rate of speech - use auditory skills to perceive a wide variety of Englishes	- collaborative behavior - patience - **tolerance** - flexibility - openness and humility to negotiate differences - avoidance of any kind of superiority of one variety over another

Intercultural Awareness/ Competence	- culture as having *a priori* elements (ethnic or cultural marking in communicative behavior) and emergent features (co-constructed in the moment of interaction) - impact of the cultural background of participants on the complex and dialogic nature of their communications - individuals with multiple membership in various cultural groups - importance of being a multilingual communicator - critical cultural awareness - <u>tone as a potential cause of cultural misinterpretation</u>	- theories of cross-cultural communication - how social groups and identities function - different cultural frames of reference (communication style, conflict management, face-work strategies, etc) - what is involved in intercultural interaction - causes and processes of misunderstanding between members of different cultures - <u>potential threats posed by intercultural communications</u>	- adjust (cultural) ways of speaking - apply and refine one's own cultural schemata - engage with and negotiate sociocultural differences - **accommodate** to difference and to multilingual aspects of intercultural communication - engage with politeness conventions - act as mediator between people of different cultural origins - analyze, interpret, and relate - acquire new knowledge of cultural practices and operate it in interaction - move beyond cultural stereotypes and generalizations	-willingness to cooperate - respect - flexibility - openness - curiosity - readiness to suspend disbelief about other cultures and belief about one's own - willingness to relativize one's own values, beliefs, behaviors
Interactional Competence	- shared responsibility for successful communication - communication as 'a two-way negotiative effort' - discourse as co-constructed among participants	- rhetorical scripts - register specific to the practice - patterns of turn-taking - topical organization - an appropriate participation framework - signaling of boundaries between practices - the processes we go through to solve communication issues	- build a 'sphere of inter-subjectivity' through collaborative efforts - **accommodate** to the constraints of the context and perceived ability of the hearer - eliminate idioms, cultural references and syntactic complexity from speech - **deal adequately with apparent misunderstandings, by checking, confirming and clarifying** -attenuate unintelligible features of one's own speech	- cooperation - openness - flexibility - tolerance

Note. [a] In **bold**, components of the construct confirmed by language testers/ESL teachers.
[b] As <u>underlined text</u>, additional components of the construct suggested by language testers/ESL teachers.

At this point in the study, it was not possible to have an idea of how important each construct component is, which is crucial information for LSP test design. Therefore, validation of the matrix with key aviation stakeholders was conducted. These two aspects are considered in the discussion of results relating to RQ 2.3.

RQ 2.3) What components of the construct are validated by key aviation stakeholders?

It was only possible to answer this research question after the Second Cycle of coding. However, it is important to present and discuss results from the First Cycle coding, as they suggest the extent to which participants of the 26 focus groups accounted for the importance of aspects related to the four dimensions of awareness (AW), knowledge (K), skills (S) and attitudes (AT). For ease of comparison, Table 8.3 shows the number of initial coding references, in absolute numbers and percentages, by grouping together focus groups that analyzed the same scenario and by specifying the type of group as either multilingual or monolingual, the latter shaded in light grey color.

Table 8.3. Coding references for AW, K, S and AT according to scenario and type of group

Scenario	Type of group	Focus Group number	Awareness	Knowledge	Skills	Attitude	Total number of references
1	Multilingual	FG 01	19(30.6%)	13(21.0%)	10(16.1%)	20(32.3%)	62(100%)
		FG 05	13(15.8%)	4(4.9%)	20(24.4%)	45(54.9%)	82(100%)
	Monolingual	FG 11	13(16.7%)	6(7.7%)	20(25.6%)	39(50.0%)	78(100%)
		FG 17	26(25.7%)	3(3.0%)	11(10.9%)	61(60.4%)	101(100%)
		FG 21	18(25.3%)	5(7.1%)	4(5.6%)	44(62.0%)	71(100%)
2	Multilingual	FG 02	5(9.6%)	7(13.5%)	8(15.4%)	32(61.5%)	52(100%)
		FG 06	6(8.5%)	4(5.6%)	29(40.8%)	32(45.1%)	71(100%)
	Monolingual	FG 12	15(20.5%)	6(8.3%)	26(35.6%)	26(35.6%)	73(100%)
		FG 18	24(24.5%)	6(6.1%)	28(28.6%)	40(40.8%)	98(100%)
		FG 22	1(2.7%)	6(16.2%)	11(29.7%)	19(51.3%)	37(100%)
3	Multilingual	FG 03	13(28.9%)	6(13.3%)	10(22.2%)	16(35.6%)	45(100%)
		FG 19	15(19.0%)	13(16.5%)	31(39.2%)	20(25.3%)	79(100%)
	Monolingual	FG 07	11(17.7%)	9(14.5%)	19(30.7%)	23(37.1%)	62(100%)
		FG 13	12(18.2%)	12(18.2%)	12(18.2%)	30(45.4%)	66(100%)
		FG 23	21(26.3%)	10(12.5%)	19(23.7%)	30(37.5%)	80(100%)
4	Multilingual	FG 04	13(27.1%)	4(8.3%)	9(18.8%)	22(45.8%)	48(100%)
		FG 08	12(18.8%)	7(10.9%)	10(15.6%)	35(54.7%)	64(100%)
		FG 24	6(14.6%)	2(4.9%)	7(17.1%)	26(63.4%)	41(100%)
	Monolingual	FG 14	20(25.0%)	7(8.7%)	26(32.5%)	27(33.8%)	80(100%)
		FG 20	19(31.7%)	5(8.3%)	14(23.3%)	22(36.7%)	60(100%)
5	Multilingual	FG 09	8(10.4%)	8(10.4%)	20(26.0%)	41(53.2%)	77(100%)
		FG 25	8(17.4%)	4(8.7%)	10(21.7%)	24(52.2%)	46(100%)
	Monolingual	FG 15	3(6.1%)	11(22.5%)	16(32.6%)	19(38.8%)	49(100%)
6	Multilingual	FG 10	15(19.7%)	4(5.3%)	12(15.8%)	45(59.2%)	76(100%)
		FG 26	12(27.3%)	1(2.3%)	7(15.9%)	24(54.5%)	44(100%)
	Monolingual	FG 16	4(7.3%)	6(10.9%)	7(12.7%)	38(69.1%)	55(100%)
Total number of references			**332(19.6%)**	**169(9.9%)**	**396(23.3%)**	**800(47.2%)**	**1697(100%)**

Coding frequency provides a more objective measure of the prevalence of a dimension between and within groups. Percentages for each dimension make it possible to compare the variation in the coding references across the focus groups. For example, FG 16, Scenario 6 had the highest percentage of references for attitude: 69.1%, whereas FG 26, Scenario 6 had the lowest percentage for knowledge: 2.3%. By comparing absolute numbers of total references, we notice a range of 37 to 101, considering all focus

groups. In order to check if group size was a factor in the number of total references, the number of participants in FG 17 and FG 18, which had the two highest number of references (i.e., 101 and 98), and in FG 22 and FG 24, which had the two lowest number of references (i.e., 37 and 41), was verified. Actually, FG 17 and 18 were composed of only 4 participants, whereas FG 22 and 24 had both 7 participants, which leads to the conclusion that group size was not a factor in the number of coding references. A different view of this data is provided through the analysis of descriptive statistics for the number of coding references related to awareness, knowledge, skills and attitudes, calculated in SPSS version 23. Table 8.4 shows the mean, median, mode, standard deviation, and range for each of the four dimensions and for the total number of references. It reveals that the dimension of attitude had the widest range (45) of all four, almost four times higher than the lowest range (12), for the dimension of knowledge.

Table 8.4. Descriptive statistics of coding references across the four dimensions

	Awareness	Knowledge	Skills	Attitude	Total number of references
N Valid	26	26	26	26	26
Missing	0	0	0	0	0
Mean	12,8	6,5	15,2	30,8	65,2
Median	13	6	12	28,5	64,5
Mode	13	6	10	19,00[a]	48,00[a]
SD	6,5	3,2	7,8	10,8	17,1
Range	25	12	27	45	64
Minimum	1	1	4	16	37
Maximum	26	13	31	61	101

Note. [a] Multiple modes exist. The smallest value is shown.

The process of coding during the Second Cycle disclosed that most of the components of the construct (sub-sub-nodes) in the draft matrix were confirmed by the aviation stakeholders, i.e., appeared in their discussions of the scenarios. They are highlighted in yellow in Table 8.5. However, some components did not receive any

coding reference, i.e., they were not explicitly mentioned by the stakeholders during the focus group discussions. For example, awareness of the social context in which AE is used (AE); knowledge of different pragmatic norms for different contexts, exposure to different international accents, and use of auditory skills to perceive a wide variety of Englishes (ELF); awareness of the importance of being a multilingual communicator, knowledge of theories of cross-cultural communication, skills to apply and refine one's own cultural schemata, to act as mediator between people of different cultural origins, to analyze, interpret and relate, to acquire new knowledge of cultural practices and operate it in interaction (ICA); knowledge of rhetorical scripts, patterns of turn-taking, topical organization and of signaling of boundaries between practices (IC). Although these components were extracted from relevant sources in the literature, sampled participants in this study did not confirm them in their discussions of the six scenarios of RT communications.

On the other hand, other components not included in the draft matrix were considered relevant and discussed by participants. These emerging components were included in the matrix cells according to their best fit to the domains (ie., AE, ELF, ICA and IC) and dimensions (i.e., awareness, knowledge, skills, and attitudes). They include, for example, awareness of group identities and authority gradients in aviation and conflict management skills (AE), knowledge of nuances of the language (ELF), politeness as attitude (ICA), and skills to declare non-understanding (IC), among others. Table 8.5 highlights in blue all emerging components.

Table 8.5. Matrix of construct specification

Construct definition within the aviation radiotelephony domain				
	Awareness	**Knowledge**	**Skills**	**Attitudes**
Aviation English	- rules of use that characterize the domain[a] - safety-critical requirements for **intelligibility**[b]**,** directness, appropriacy, **non-ambiguity** and concision - **threats** presented by cross-cultural communications - **impact o**f communication on safety and efficiency - social and occupational context in which AE is used - group identities and authority gradients in aviation[c] - situational awareness	- standard phraseology - plain English for the specific purpose of aeronautical RT communications - syntactic structures and language functions used in RT - aviation lexicon - aviation phonetic alphabet and pronunciation of numbers - prosodic features of RT - background knowledge (rules and procedures) - communication as a Human Factor	- apply speech transmitting techniques - use the linguistic features of AE meaningfully - communicate effectively in routine and in highly unpredictable situations - use strategic skills to deal with aviation personnel with different levels of expertise - language proficiency (ability to use the language) - conflict management - Crew Resource Management (CRM)	- compliance with prescribed rules and procedures (e.g. use of phraseology, read back/hear back) - discipline - **professional tone** and attitude - clarity, conciseness and correctness - assertiveness - self-confidence
English as a lingua franca	- **different varieties of English** and speech communities - challenges faced by speakers of EFL and interlocutors' possible linguistic difficulties - difficulty presented by the use of jargon, idioms, slang and colloquialisms - the need to speak English as a lingua franca[d] - language use and language processing	- language as a social practice - different pragmatic norms for different contexts - one's own communicative style and the problems it could pose to ELF interactions - characteristics of one's L1 phonology that may influence English pronunciation - exposure to different international accents - nuances of the language	- mediate and negotiate meaning - **accommodate** different accents and dialects - adapt linguistic forms to the communicative needs at hand - adjust and align to different communicative systems (new patters of phonology, syntax, discourse styles) - self-repair, **rephrase**, paraphrase, and clarify - **notice and repair breakdowns in communication** - preempt misunderstanding - ascertain and deploy appropriate pragmatics - eliminate ambiguous expressions and sentence patterns - adapt speed and rate of speech - auditory skills to perceive a wide variety of Englishes	- collaborative behavior - patience - **tolerance** - flexibility - openness and humility to negotiate differences - avoidance of any kind of superiority of one variety over another

Intercultural Awareness/ Competence	- culture as having *a priori* elements (ethnic or cultural marking in communicative behavior) and emergent features (co-constructed in the moment of interaction) - how the cultural background of participants can impact the complex and dialogic nature of their communications - individuals with multiple membership in various cultural groups - importance of being a multilingual communicator - critical cultural awareness - tone as a potential cause of cultural misinterpretation - power distance - gender expectations - face concern	- theories of cross-cultural communication - how social groups and identities function - different cultural frames of reference (communication style, conflict management, face-work strategies, etc) - what is involved in intercultural interaction - causes and processes of misunderstanding between members of different cultures - potential threats posed by intercultural communications	- adjust (cultural) ways of speaking - apply and refine one's own cultural schemata - engage with and negotiate sociocultural differences - **accommodate** to difference and to multilingual aspects of intercultural communication - engage with politeness conventions - act as mediator between people of different cultural origins - analyze, interpret, and relate - acquire new knowledge of cultural practices and operate it in interaction - move beyond cultural stereotypes and generalizations	- willingness to cooperate - respect - flexibility - openness - curiosity - readiness to suspend disbelief about other cultures and belief about one's own - willingness to relativize one's own values, beliefs, behaviors - avoidance of judgments
Interactional Competence	- shared responsibility for successful communication - communication as 'a two-way negotiative effort' - discourse as co-constructed among participants	- rhetorical scripts - register specific to the practice - patterns of turn-taking - topical organization - an appropriate participation framework - signaling boundaries between practices - the processes we go through to solve communication issues	- build a 'sphere of inter-subjectivity' through collaborative efforts - **accommodate** to the constraints of the context and perceived ability of the hearer - eliminate idioms, cultural references and syntactic complexity from speech - **deal adequately with apparent misunderstandings, by checking, confirming and clarifying** - attenuate unintelligible features of one's own speech - declare non-understanding - use communicative/interactional skills	- cooperation - openness - flexibility - tolerance - avoidance of intimidation and threatening behavior

Note: [a]In yellow, components of the construct confirmed by aviation stakeholders.
[b]In **bold**, components of the construct confirmed by language testers/EFL teachers.
[c]In blue, additional components of the construct suggested by aviation stakeholders.
[d]As underlined text, additional components of the construct suggested by language testers/EFL teachers.

Analysis also revealed the number of coding references for each component of the matrix, and as a consequence, for the intersection of each dimension with the four domains of interest. This information is crucial to inform test development. As it indicates the degree of importance or the weight of each cell in the matrix, it ultimately guides the test developer in the test assembly model to produce test forms, in such a way as to consider the "mix of items or tasks on the test that must be included in order to represent the domain adequately" (Fulcher & Davidson, 2007, p. 67). Table 8.6 provides the weighting of construct components based on the number of coding references. It is important to mention that the number of total references for each dimension is not the same as shown in Table 8.3. The reason lies in the fact that some of the references were coded in more than one dimension or domain, indicating the overlap among them. For example, the total number of coding references in the First Cycle was 1697, whereas in the Second Cycle it increased to 1986.

Table 8.6. Weighting of construct components based on coding references

	AW	**K**	**S**	**AT**	**Total**
AE	189	160	165	552	1066
ELF	82	14	105	178	379
ICA	143	37	26	159	365
IC	9	14	123	30	176
Total	423[a]	225[a]	419[a]	919[a]	1986[a]

Note: [a] Overlap counted.

As can be noted, the number of coding references for each domain is included in the last column of Table 8.6 and decreases as it moves down from AE to IC. Not surprisingly, AE holds the greatest weight in the matrix, and this is so because it comprises not only knowledge of standard phraseology and plain English, but also background knowledge, compliance with rules and procedures, professional tone and

attitude, just to name a few. In terms of the four dimensions, we can find the number of coding references for each of them in the last line of the table. It reveals that attitude weighs way more than the other dimensions, followed by awareness, skills and knowledge.

As explained before, while some authors consider awareness as being at the core of all four dimensions (e.g. Fantini, 2000), attitude may also be understood as putting one's awareness, skills and knowledge into practice.

By considering the number of coding references for each component of the construct separately, i.e., for each sub-sub-node, it was possible to organize them within each cell from the highest to the lowest number of references. As a result, the four most cited components of each cell of the matrix were identified and included in the final matrix of construct specification (see Table 8.7). This does not mean that the other components are not relevant to the context of RT communications, but may indicate a better understanding and awareness of the ones that were kept in the final matrix, in the view of participants.

Table 8.7. Final matrix of construct specification

Construct definition within the aviation radiotelephony domain				
	Awareness	**Knowledge**	**Skills**	**Attitudes**
Aviation English	- situational awareness (67) - group identities and authority gradients in aviation[c] (50) rules of use that characterize the domain[a] (27) - threats presented by cross-cultural communications (19)	- background knowledge (rules and procedures) (78) - standard phraseology (36) - plain English for the specific purpose of aeronautical RT communications (26) - communication as a Human Factor(6)	- Crew Resource Management (CRM) (55) - language proficiency (ability to use the language) (45) - communicate effectively in routine and in highly unpredictable situations (39) - conflict management (12)	- **professional tone** and attitude (195) - compliance with prescribed rules and procedures (e.g. use of phraseology, read back/hear back) (193) - assertiveness (87) - clarity, conciseness and correctness (37)
English as a lingua franca	- challenges faced by speakers of EFL and interlocutors' possible linguistic difficulties (34) - difficulty presented by the use of jargon, idioms, slang and colloquialisms (17) - the need to speak English as a lingua franca[d] (17) - **different varieties of English** and speech communities (9)	- nuances of the language (5) - language as a social practice (4) - one's own communicative style and the problems it could pose to ELF interactions (3) - characteristics of one's L1 phonology that may influence English pronunciation (2)	- adjust and align to different communicative systems (new patters of phonology, syntax, discourse styles) (23) - eliminate ambiguous expressions and sentence patterns (21) - adapt linguistic forms to the communicative needs at hand (20) - self-repair, **rephrase**, paraphrase, and clarify (13)	- patience (68) - collaborative behavior (45) - avoidance of any kind of superiority of one variety over another (39) - **tolerance** (12) - openness and humility to negotiate differences (12)
Intercultural Awareness/ Competence	- how the cultural background of participants can impact the complex and dialogic nature of their communications (58) - power distance (27) - gender expectations (17) - face concern (12)	- what is involved in intercultural interaction (11) - potential threats posed by intercultural communications (11) - different cultural frames of reference (communication style, conflict management, face-work strategies, etc) (10) - how social groups and identities function (3)	- move beyond cultural stereotypes and generalizations (11) - engage with and negotiate sociocultural differences (5) - engage with politeness conventions (5) - **accommodate** to difference and to multilingual aspects of intercultural communication (4)	- politeness (90) - willingness to cooperate (25) - respect (20) - readiness to suspend disbelief about other cultures and belief about one's own (9) - willingness to relativize one's own values, beliefs, behaviors (9)
Interactional Competence	- shared responsibility for successful communication (5) - discourse as co-constructed among participants (3) - communication as 'a two-way negotiative effort' (1)	- register specific to the practice (10) - an appropriate participation framework (3) - the processes we go through to solve communication issues (1)	- **deal adequately with apparent misunderstandings, by checking, confirming and clarifying (44)** - use of communicative/interactional skills (36) - **accommodate** to the constraints of the context and perceived ability of the hearer (20) - declare non-understanding (9)	- avoidance of intimidation and threatening behavior(10) - cooperation(9) - tolerance (6) - flexibility (4)

Note: [a]In yellow, components of the construct confirmed by aviation stakeholders.
[b]In **bold**, components of the construct confirmed by language testers/EFL teachers.
[c]In blue, additional components of the construct suggested by aviation stakeholders.
[d]As underlined text, additional components of the construct suggested by language testers/EFL teachers.

In order to verify if the number of coding references for these components would be a good representation of the total coding references reported in Table 8.6, a new table was constructed (see Table 8.8) including the sum of coding references for the selected components in the final matrix.

Table 8.8. Weighting of construct components included in the final matrix

	AW	K	S	AT	Total
AE	163	146	151	512	972
ELF	77	14	77	176	344
ICA	114	35	25	153	327
IC	9	14	109	29	161
Total	363	209	362	870	1804

The numbers in Table 8.8 coupled with the ones in Table 8.6 reveal that 85.8% of the total coding references for awareness were kept in the final matrix, 92.9% for knowledge, 86.4% for skills, and 94.7% for attitudes. Calculating the percentages for each domain, 91.2% of the total coding references for AE were kept in the final matrix, 90.7% for ELF, 89.6% for ICA, and 91.5 % for IC. This suggests that the final matrix is a very good representation of the international aviation RT construct, based on the perceptions of key stakeholders in the sample analyzed.

Interestingly, across the four domains, the weighting of the four dimensions displays some differences. For both the domains of Aviation English and English as a Lingua Franca, the number of coding references follow the decreasing pattern of AT, AW, S, K; for Intercultural Awareness, a difference is noted only in the last two dimensions – AT, AW, K, S; whereas for Interactional Competence, the pattern is totally different – S, AT, K, AW.

In contrast to the above discussion which focused on the total number of coding references for each component of the construct, it is also interesting to note the number of focus groups in which a certain component was mentioned. This information gives us

another perspective on the data. However, the fact that some groups did not mention a particular component may be related to group dynamics and the way their discussion developed rather than an indicator of decreased importance. Table 8.9 displays a list of the 26 construct components that were mentioned by the highest number of focus groups; it applies a specific color to each of the four domains for ease of contrast and comparison: green for AE, blue for ELF, orange for ICA and pink for IC.

Table 8.9. Construct components mentioned by the highest number of focus groups

Construct component	Domain/ dimension	Number of focus groups
Background knowledge	AE/K	26
Professional tone and attitude	AE/AT	25
Compliance with prescribed rules and procedures (e.g., use of phraseology, read back/hear back)	AE/AT	25
Politeness	ICA/AT	19
Situational awareness	AE/AW	19
Impact of the cultural background of participants on the complex and dialogic nature of their communications	ICA/AW	19
Patience	ELF/AT	18
Assertiveness	AE/AT	17
Communicate effectively in routine and in highly unpredictable situations	AE/S	17
Clarity, conciseness and correctness	AE/AT	17
Collaborative behavior	ELF/AT	15
Avoidance of any kind of superiority of one variety over another	ELF/AT	15
Standard phraseology	AE/K	14
Self-confidence	AE/AT	14
Rules of use that characterize the domain	AE/AW	14
Plain English for the specific purpose of aeronautical RT communications	AE/K	14
Language proficiency (ability to use the language)	AE/S	13
Use of communicative/interactional skills	IC/S	13
Crew Resource Management (CRM)	AE/S	12
Group identities and authority gradients in aviation	AE/AW	12
Challenges faced by speakers of ELF and interlocutors' possible linguistic difficulties	ELF/AW	12
Willingness to cooperate	ICA/AT	12
Eliminate ambiguous expressions and sentence patterns	ELF/S	12
Deal adequately with apparent misunderstandings, by checking, confirming and clarifying	IC/S	11
Adapt linguistic forms to the communicative needs at hand	ELF/S	11
Threats presented by cross-cultural communications	AE/AW	11

Note. Aviation English; English as a Lingua Franca; Intercultural Awareness; Interactional Competence

Only one construct component appeared in all 26 focus groups – 'background knowledge', followed by 'professional tone and attitude' and 'compliance with prescribed rules and procedures (e.g., use of phraseology, read back/hear back)', both mentioned by 25 focus groups. This suggests that participants, while analyzing the six scenarios, perceived the relevance of these three AE components to effective communications in international RT. These components are all related to knowledge and attitudes specific to this occupational domain, and as Douglas (2000) stated, specific purpose content knowledge "is a necessary, integral part of the concept of specific purpose language ability" (p. 2). However, these components are not taken into consideration in the assessment of pilots and ATCOs according to the ICAO LPRs, nor do they seem to be the focus of continuous surveillance linked to actual corrective actions or sanctions within ICAO Contracting States. Other components from the ICA domain (e.g., politeness and the impact of the cultural background of participants on the complex and dialogic nature of their communications) and from the ELF domain (e.g., patience and collaborative behavior) appeared in the discussions of a great number of focus groups, but yet are not part of the ICAO testing policy.

Having examined the most important components of the construct and the unexpected ones that emerged during the focus group discussions, representative quotes are presented below in order to give readers a sense of the kind of statements that were made in support of particular construct components. Participants' comments are identified by the number of focus group and scenario analyzed, with an "M" or "F" indicating whether it was said by a male or female. They are organized by domain (i.e.,

AE, ELF, ICA and IC) and dimension (i.e., AW, K, S and AT), and, due to limitations of space, only one quote is provided for each component of the matrix.

8.3.1 Aviation English.

8.3.1.1 Awareness.

Within this dimension, *situational awareness* was the component with the highest number of coding references. It is a common expression in the aviation field related to being conscious of what is happening in a situation, i.e., the comprehension of "present system and environmental conditions" and anticipation of future changes (ICAO, 1998, p. 2-2-17). In relation to communication, one participant mentioned an example in which *"*the situational awareness of this people in the cockpit, maybe the situational awareness of the controller himself, not seeing maybe, not noticing nervousness from pilot's speech, all of this contributed to this accident"[25] (M – FG 23 of 26 Scenario 3). Added to that, having an understanding of *group identities and authority gradients in aviation* was cited as relevant and contextualized in the following quote: "I think he was rude when he used the phrase...she was all the time, and also because the hierarchy, like, he is a pilot, she is a controller. So, the power relationship might be...because she was trying to be polite, because the first time she asked, she even said please, you need to enter runway 28...." (F – FG 17 of 26 Scenario 1). As a highly regulated occupational domain, safe RT communications also require the realization of *rules of use that characterize the domain*: "Well, the ATCO should be aware that he should follow the procedures, which has to be from ICAO. Safety issues, any danger..." (F – FG 09 of 26 Scenario 5). Further, being

[25]The comments have been transcribed *ipsis litteris* from the actual oral discourse of participants, and have not been corrected for grammatical or lexical mistakes.

conscious of *threats presented by cross-cultural communications* may help pilots and ATCOs to communicate more effectively, as one participant explained:

> I think we need to aware them about what they might face, they might experience when flying to a certain place, certain country, you know. We should aware them. It can happen if they interact in this way. I do that with my students. I tell them that Americans have problems to paraphrase, they are impatient, and they have to keep calm. (F – FG 18 of 26 Scenario 2)

8.3.1.2 Knowledge.

Background knowledge was mentioned by all focus groups and considered crucial for successful communications between pilots and ATCOs. When discussing one scenario, a participant mentioned that "I think the ATCO should have used his background in terms of phraseology and in terms of knowing the airport, because he knew the airport and the pilot probably didn't. So, he should use that language and phraseology to try to explain the taxiways, routes, and try to share the knowledge with the pilot". (F- FG 12 of 26 Scenario 2). Similarly, knowledge of *standard phraseology* was deemed paramount in this specific field, but sometimes not considered the reality around the world: "When we get pilots from Anglophone countries, that's quite often the case, they are perfect in plain language but they are ....they don't know standard phraseology. (F – FG 2 of 26 Scenario 2). On the other side, knowledge of *plain English for the specific purpose of aeronautical RT communications* was confirmed as essential, and its absence perceived as an issue: "Communication was not effective, lack of assertiveness, lack of English knowledge" (M – FG 13 of 26 Scenario 3). Knowledge of

communication as a Human Factor[26] includes the notion that communication takes place between individuals, i.e., the human element, defined as "the most flexible, adaptable and valuable part of the aviation system, but … also the most vulnerable to influences which can adversely affect its performance" (ICAO, 1998, p. 1-1-2). As one participant stated, "human actors, human errors" (F – FG 20 of 26 Scenario 4), referring to human shortcomings can occur while communicating with each other.

8.3.1.3 Skills.

Within the Aviation English domain, *Crew Resource Management,* known as "a widely implemented strategy in the aviation community that acts as a training countermeasure to human error" (ICAO, 1998, p. 1-5-26), has been highlighted as a necessary skill, not only for pilots but also for ATCOs: "There are issues of CRM here, of CRM across the frequency, not just on the flight deck, but on the frequency and of course both of those could be addressed in training" (M – FG 5 of 26 Scenario 1). In addition, *language proficiency* (or lack thereof), regarded as the ability to use the language to speak and understand, i.e., both productive and receptive skills, was mentioned as a concern, which was confirmed by the following comment: "So, the two main problems in this interaction are: the first, the air traffic controller English was not enough to understand what the pilot was asking and the pilot was using a language that was too complex" (M – FG 15 of 26 Scenario 5). The skills needed to *communicate effectively in*

[26] According to ICAO (1998), the task of Human Factors training includes "the explanation of common communication problems as well as the reinforcement of a standard of language to ensure the error-free transmission of a message and its correct interpretation. Ambiguous, misleading, inappropriate or poorly constructed communication, combined with expectancy, have been listed as elements of many accidents, the most notorious one being the double B747 disaster in Tenerife (March 1977)" (p. 1-1-15).

routine and in highly unpredictable situations are central to successful radiotelephony exchanges, and could have prevented a fatal accident from happening:

> He was confirming and saying "I am just running out of fuel". Everyone is always...you deplete your fuel at all times. I'm running out fuel, of course it means, it's more than you would like to consume but, you have to declare a low fuel situation in that specific scenario. If you are that close of running out of fuel, you should actually say that you are on the verge of a fuel starvation scenario, you are actually in a mayday situation. (M – FG 19 of 26 Scenario 3)

Conflict management is also a skill that aviation professionals should develop in order to deal with tensions while interacting over the radio and to avoid potential communication clashes: "But it can also be conflict management, because in the second case, the second pilot is trying to solve this conflict in a polite way" (F – FG 08 of 26 Scenario 4).

8.3.1.4 Attitudes.

A professional stance in aviation communications includes displaying a *professional tone and attitude*, which was noted as crucial by study participants in almost all groups: "Then, the attitude, obviously reduce attitudes from both, or a more professional attitude, if you want a more positive spin on that. Non-confrontational" (M – FG 5 of 26 Scenario 1). Likewise, *compliance with prescribed rules and procedures* (e.g., use of phraseology, read back/hear back, etc.) was a recurring topic and deemed crucial also, or mainly, for native speakers of English: "Yes, I think what you said is ok, because they speak the same language, they are both native speakers, so I think they didn't care about the regulations, I don't know....phraseology" (M – FG 11 of 26 Scenario 1). In terms of *assertiveness*, defined by ICAO (1998) as a behavior when a pilot or ATCO

"queries others especially during ambiguous situations to clarify actions to be taken [and] constructively asserts views and contributes to overall team effectiveness" (p. 2-2-20), it was deemed crucial for expeditious and efficient communications: "But in aviation everything happens so fast. The controller needs to be assertive, the pilot needs to be assertive in order to make things flow. You see, during this two minutes that there was this bad, poor interaction, two aircraft had to go around. This is serious" (M – FG 17 of 26 Scenario 1). By the same token, *clarity, conciseness and correctness* were confirmed as relevant to the safety of this highly specialized context, as exemplified in the comment below:

> But in a nutshell, the pilot and the controller, they had no interaction at all, because there was no concise and clear information, in relation to what the pilot should do. The controller accepted his ideas, 'I am not ready', 'I am going to stay here on the runway', and he didn't follow the instructions. We thought the controller should have been more informative, concise and direct, and assertive, to make the pilot leave the runway or stay there, if necessary. But there was no comprehension from either sides. (M – FG 11 o 26 Scenario 1)

8.3.2 English as a lingua franca.

8.3.2.1 Awareness.

Within the domain of ELF, being conscious of the *challenges faced by speakers of ELF* was considered important for effective communications, as cited by one of the participants:

> Yes, they take for granted and they have, they need to have this awareness, that it's not just... they have to be involved in the whole process. They have to be

> involved not only in speaking, but also in receiving and understanding and trying to accommodate the necessity of specific communication that is being held in the ATCO-pilot situation. They need to know that on the other side they have a non-native speaker. They need to be aware that they can't just throw out their speech... (M – FG 23 of 26 Scenario 3)

Additionally, having an understanding of the *difficulty presented by the use of jargon, idioms, slang and colloquialism* was confirmed as a key component for the successful outcome of RT exchanges: "I think, once again, the blame is on the ATCO. He is a native speaker and used to that... those slangs" (FG 12 of 26 Scenario 2). Likewise, *the need to speak English as a lingua franca* was understood as a way to avoid the notion of the 'native speaker norm', still present in some English-speaking countries: "Americans have this idea that when you are there you must speak like native" (M – FG 18 of 26 Scenario 2). Furthermore, the realization of *different varieties of English and speech communities* was also confirmed among focus group participants as necessary for effective ELF interactions, highlighting existing differences, such as: "*Ok, so British and Irish...* "(M – FG 21 of 26 Scenario 1).

8.3.2.2 Knowledge.

Knowing the *nuances of the language* was considered relevant in the context of Aviation English as a lingua franca, in order to inform language choice: "Knowing the nuances of the language. They need to be aware of the nuances of the language, and the choice of language" (M – FG 5 of 26 Scenario 1). Knowledge of *language as a social practice*, i.e., co-constructed by participants at the moment of interaction, was also deemed important to emphasize the dialogic nature of the utterances as 'responsive

reverberations' of previous utterances (Bakhtin, 1986), which may not always be complied with for a number of reasons: "*The ATC. This is the first day, or, the first day that he comes face to face and that's why he cannot answer, or she doesn't know, or he doesn't know how to react in this situation. That's my ....he goes away from answering the questions"* (F – FG 09 of 26 Scenario 5). Added to that, knowing *one's own communicative style and the problems it could pose to ELF interactions* was also confirmed as a construct component, meaning, for example, that a more aggressive style may raise the anxiety of interlocutors, inhibit the use of clarification strategies, etc, as explained in the example below:

> ....this was what I was trying to say, it's kind of archaic. I remember ...anyway, regardless of that. Let's assume he did say "where you park". That's pretty much North American, that's ok. Ah....his response is fine, the pilot's. And yet we know he hasn't stopped. "Not taxiway, the letter!" So, really it is the ATC giving the aggression from the start. (M – FG 4 of 26 Scenario 4)

Knowledge of the *characteristics of one's L1 phonology that may influence English pronunciation* may also contribute to the production of more intelligible messages, which is key to aid understanding: "Probably his pronunciation was not very good, he is French. Probably..." (F – FG 15 of 26 Scenario 5).

8.3.2.3 Skills.

Among the skills identified as part of the ELF construct, *adjust and align to different communicative systems (new patterns of phonology, syntax, discourse styles)* appeared as a central one, as mentioned in relation to a less than effective interaction: "The controller jumped out of normal phraseology and gave a lot of long, wordy

directives which the pilot not...and not even listen to anyway. So, the controller should have just said 'Expedite, get off the runway'....in my opinion" (F – FG 1 of 26 Scenario 1). Added to that, the ability to *eliminate ambiguous expressions and sentence patterns* is a critical one in radiotelephony communications, but not always the case: "Very confusing, because the controller is using a lot of fillers, and he is not precise. He is kind of confusing" (M- FG 18 of 26 Scenario 2). In addition, having the skill to *adapt linguistic forms to the communicative needs at hand* was highlighted as a way to reach a positive outcome of the interaction, as explained in the following comment: "You have to adapt, so who had the skills was the ATCO. With his language skills, he could have adapted the language to reach the ..." (F – FG 12 of 26 Scenario 2). Furthermore, the *ability to self-repair, rephrase, paraphrase, and clarify* whenever necessary to repair communication failures is illustrated in another quote: "....as you already said, that repetition thing, like he is not even paraphrasing, he is just saying again and again. And then he complains that he had to say it again" (F – FG 04 of 26 Scenario 4).

8.3.2.4 Attitudes.

Concerning attitudes related to the ELF domain, the first most cited was *patience*, and its absence was reported as an issue for smooth interactions over the radio: "There were traces of a dispute in the interaction, lack of patience, something like that" (F – FG 11 of 26 Scenario 1). Very close to that one is *collaborative behavior*, bearing in mind that both pilots and ATCOs need to perform as a coordinated team in the context of radiotelephony: "It was not a matter of being proficient in terms of linguistic knowledge, but it was mainly being disrespectful, non-collaborative and this kind of thing" (M – FG 17 of 26 Scenario 1). Moreover, an ELF context requires the *avoidance of any kind of*

superiority of one variety over another while communicating, especially when safety is at risk. But a comment was made about the behavior of native speakers of English in that matter: “I think it's a bit about an economic superiority, financial, linguistic, cultural, and they believe they are the best” (F – FG 14 of 26 Scenario 4). Therefore, *tolerance,* from both sides, emerges as a central attitude in intercultural settings: “Koreans and middle-eastern people demand respect. They won't tolerate that from a New Yorker” (M – FG 24 of 26 Scenario 4). Thus, displaying *openness and humility to negotiate differences* was confirmed as part of the ELF construct: “Good will, patience, politeness, to be sensible to the different....empathy. That's it...empathy” (F – FG 20 of 26 Scenario 4).

8.3.3 Intercultural awareness/competence.

8.3.3.1 Awareness.

Being conscious of *the impact of the cultural background of participants on the complex and dialogic nature of their communications* was largely discussed in Phase 1 but also appeared in the focus group discussions, as highlighted in this example: “So, we thought it was interesting culturally, because obviously JFK controllers, traditionally are more straightforward, they are ....they can be rude and aggressive at times, then we had a middle-eastern and Korean guy that really demand respect” (M – FG 24 of 26 Scenario 4). Moreover, issues of *power distance* revealed how unequal distribution of power, due to a number of factors, may result in asymmetric communications. In the example that follows, the imbalance was due to different levels of language proficiency: “Well, I think the power distance was based on the language factor” (F – FG 25 of 26 Scenario 5). On top of that, *gender expectations* became apparent in some discussions, also considered to be culturally related and a trigger to issues of superiority or more dominant behavior:

"Perhaps the pilot being male and being an airline captain, makes him feel superior to the female air traffic controller. I think it's a cultural issue (M – FG 17 of 26 Scenario 1). Besides, having an understanding of *face concern* and how different people react differently when they feel threatened or humiliated, may reduce communication clashes: *"*Yeah, language choice, awareness of...people often say Asian cultures are more aware of face" (M – FG 5 of 26 Scenario 1).

8.3.3.2 Knowledge.

In order to participate in international RT communications, it is essential to know *what is involved in intercultural interaction,* and participants discussed issues related to the several layers of culture that affect the way an individual communicates, including professional culture, related to the concept of communities of practice : "There may be gender issues, male and female, and much more likely a tribal issue, ATC tribal needs versus the pilot's community needs...some big issues there" (M – FG 1 of 26 Scenario 1). Further, another quote highlights the *potential threats posed by intercultural communications,* addressing both sides of the interaction:

> And we discussed that there might be, among others, a cultural reason for this flawed communication between the ATC and pilot 1. I apologize if there are other Asians, I have told her I have no experience with Asian people, just recounting what I heard, but apparently they have two sides, they either don't want to lose their face, and so they don't ask if they are not sure, or they do the opposite, they want to do it as much as possible, to make sure they are doing the things correctly, which is basically a good thing. And on the other hand, we have this ATCO who comes from a culture where people are quite self-confident and are used to take

> the decisions on their own, so that is already a potential conflict if we have a pilot from a culture where they are used to ask what to do and the ATCO used to the opposite, and is expecting that the pilot will also behave that way ....and decide by himself, and talk about the routes, etc, etc., not asking all these questions. (F – FG 2 of 26 Scenario 2)

Knowing that individuals possess *different cultural frames of reference (communication style, conflict management, face-work strategies, etc)* may prevent potential problems while at the same time increase respect and openness to difference, as suggested in the following quote: "If we look at the other ones, like...conflict management, communication styles, Korean is probably going to be more..." (F – FG 08 of 26 Scenario 4). Likewise, knowing in advance *how social groups and identities function* can also add to the effectiveness of RT communications: "Maybe it's the work culture in JFK...the controllers become normal arrogant. It's such a big airport" (M – FG 10 of 26 Scenario 6).

8.3.3.3 Skills.

The ability to *move beyond cultural stereotypes and generalizations* reinforces the notion that each individual is different, despite belonging to a certain nation or cultural group, and performs differently according to the constraints of the situation, as exemplified in the following comment: "So, we've got an international pilot landing in a very busy international airport, with an ATC obviously ...who clearly had a busy day and is not happy" (M – FG 4 of 26 Scenario 4). Also, the need to *engage with and negotiate sociocultural differences* also emerged as an issue in a situation where the ATCO did not show the skills to do so: "I think it might have something to do also with skills. If the

controller showed social skills... He would be using polite... it would be different (M – FG 10 of 26 Scenario 6). Further, having the skills to *engage with politeness conventions* in this intercultural workplace context was also confirmed by the quote: "So, I think the cultural background helped the pilot to, you know, helped the UAE pilot to tell the controller, or to remind him politely, to be polite with him" (M – FG 8 of 26 Scenario 4). In addition, the ability to *accommodate difference and to multilingual aspects of intercultural communication* was mentioned as an essential component of the construct to all involved in international operations:

> … because culture in my opinion, is not sufficiently addressed, cultural differences. It is mentioned only at Level 6 ... "cultural subtleties", but I think it is a general problem, and not just a Level 6 problem. And I think even a private pilot because he could also fly somewhere in the States, in Asia, or wherever, he goes everywhere, even a private pilot should know at least a minimum about important cultural differences in communication. And I think this is missing in the testing... (FG 02 of 26 Scenario 2)

8.3.3.4 Attitudes.

The behavior and posture that pilots and ATCOs display while interacting over the radio can lead to professional and effective intercultural communications or, on the contrary, trigger less than desired reactions by either interlocutor. In one scenario, a participant noted a concern related to *politeness*: "Talking about language, I can see an issue, because there is a time that the pilot uses "Madam", and I think basically it is related to politeness, it is something like ironic" (FG 21 of 26 Scenario 1), whereas another one highlighted the need to address issues related to attitude and *politeness*:

"Well, clearly they need to have, the ATCO needs to be upskilled, if it wasn't a transmission problem, attitudes need to be addressed, a discussion on attitudes and politeness need to be addressed, and that's more on the pilot rather than on the ATCO" (F – FG 9 of 26 Scenario 5). Further, showing *willingness to cooperate* emerged as an important attitude in pilot-ATCO interactions: "Yes, but he is asking for help. He didn't say any PAN PAN or MAYDAY, but he is asking for help. And the controller doesn't seem to be very willing to help. 'Oh, I'm not going to be able to hold your hand'" (F – FG 16 of 26 Scenario 6). On top of that, lack of *respect* was also identified as an attitude problem: "He was ironic, he patronized her, he was disrespectful. So, the only problem I see clearly here is attitude" (M – FG 17 of 26 Scenario 1). Moreover, the need to demonstrate *readiness to suspend disbelief about other cultures and belief about one's own* was captured in an instance related to differences in the linguacultural background of interlocutors: "The ATCO is playing the role of the native speaker so, he is saying the same thing over and over, and we can see that he is not putting himself in the shoes of the non-native speaker, possibly a pilot who does not know the airport structure, the taxiways, and so on" (F – FG 12 of 26 Scenario 2). Likewise, displaying *willingness to relativize one's own values, beliefs, behaviors* was confirmed as a component of the construct, and is exemplified by a quote in which the focus group participant highlights the impact of not acting in this way: "Yes, but Americans don't. They are really impatient regarding the communication. They want you to understand the first time they say something, they want you to understand" (F – FG 18 of 26 Scenario 2).

8.3.4 Interactional Competence

8.3.4.1 Awareness

Being conscious of the *shared responsibility for successful communication* was mentioned by one participant as central in pilot-ATCO international RT communications: "Responsibility. Both should be responsible for the communicative success" (F – FG 17 of 26 Scenario 1). Likewise, having an understanding of *discourse as co-constructed among participants* disclosed the role of both interlocutors in finding the best way to get their messages across: "He did not use any clarification strategy, he started being nervous and got irritated. Both are there to reach an understanding, and not a misunderstanding (F – FG 14 of 26 Scenario 4). On top of that, being aware *of communication as a two-way negotiative effort* can be explained by the following quote:

> Yes, they take for granted and they have, they need to have this awareness, that it's not just... they have to be involved in the whole process. They have to be involved not only in speaking, but also in receiving and understanding and trying to accommodate the necessity of specific communication that is being held in the ATCO-pilot situation. They need to know that on the other side they have a non-native speaker. They need to be aware that they can't just throw out their speech... (M – FG 23 of 26 Scenario 3)

8.3.4.2 Knowledge

In order to participate in effective aviation RT communications, pilots and ATCOs should know and comply with the *register specific to the practice,* yet, this is not always the case as explained in the following quote: "What my impression is ... there is too much too bla-bla-bla, unnecessary things, no use of standard phraseology in routine

situations, and so it takes much longer and it's much more confusing than if they would just say what is necessary to be said" (F – FG 2 of 26 Scenario 2). Apart from that, knowing the roles each participant plays in a particular interaction through the use of an *appropriate participation framework* is also central to this intercultural occupational context. However, some focus groups' participants expressed their confusion about these roles, as shown in the extract below:

M1 The captain was flying and the co-pilot was talking.

F1 Ah....I understand.

F2 The captain was...

F1 But the co-pilot, I can't understand that, it was his role.

M1 Yes, to communicate (FG 13 of 26 Scenario 3).

Moreover, being knowledgeable about *the processes we go through to solve communication issues* can help pilots and ATCOs to act in a timely and efficient manner when problems arise. One participant highlighted the potential of a specific scenario to address this and other issues: "So, we think there is plenty of opportunity to use this as a case study for learning really … It brings into areas like lack of situational awareness, ah...generally what you do when there is a breakdown of effective communication and certainly not really following ICAO guidelines in a way" (FG 01of 26 Scenario 1).

8.3.4.3 Skills

The ability to *deal adequately with apparent misunderstandings, by checking, confirming and clarifying* is crucial in the management of the dialogue and, thus, for safe RT communications. It was highlighted by one participant who gave some examples of these strategies: "Try to use some clarification strategies. Although he is not, he is a non-

native speaker, he should use clarification strategies, like 'Repeat', 'Say again', 'Speak slower'…" (F – FG 12 of 26 Scenario 2); and by another, who mentioned the need to include this skill in all levels of the ICAO rating scale: "There is the checking, confirming and clarifying thing, but only at Level 4. So, if you consider that the communication should...all interlocutors should be responsible for reaching the communicative goal, then it should be spread out throughout the rating scale, from Level 1 to 6" (F – FG 17 of 26 Scenario 1). In terms of the *use of communicative/interactional skills,* some participants confirmed that they are relevant, but did not break them down into smaller elements:

> F1 There's no problem with their skills, except communication skills.
>
> M3 Or interactional.
>
> M2 The communication skills, their communication should be more concise.
>
> F1 It's communication skills, definitely (FG 5 of 26 Scenario 1).

The need to *accommodate to the constraints of the context and perceived ability of the hearer* was also highlighted as a central skill in the international RT context, as the following example illustrates: "The end of the story was that we realized there was lack of accommodation on both parts, because the ATCO, who was the native speaker, could have accommodated, the pilot did not try to use any strategy to clarify or try to negotiate, because he could not understand, perhaps" (FG 14 of 26 Scenario 4). Also, being able to *declare non-understanding* when necessary is a quality of skillful interlocutors and should be practiced by all: "The pilot did not use other strategies as well. He ended up being aggressive. He could have said 'I didn't understand what you said, please repeat'" (FG 14 of 26 Scenario 4).

8.3.4.4 Attitudes

Certain attitudes are required of a person to be considered interactionally competent, such as *avoidance of intimidation and threatening behavior.* The lack thereof may raise the levels of anxiety of interlocutors and trigger less than effective communications, especially in a high-stake context:

> And he is not being supportive as she said, he is not having this collaborative behavior, he is not accommodating his language, he is being too direct, I think. In some ways, in the way he said things, is ... straight, it's intimidating, the way you talk to a person intimidates, creates more nervousness on the part of the pilot. (F – FG 14 of 26 Scenario 4)

On the other hand, *cooperation* appears as a requirement in RT interactions: "Compromised in that situation. Cooperative..." (M – FG 24 of 26 Scenario 4), as well as tolerance, as the ability to tolerate opinions or behaviors that one does not agree with: "Koreans and middle-eastern people demand respect. They won't tolerate that from a New Yorker" (M – FG 24 of 26 Scenario 4). Finally, *flexibility* was also confirmed as a relevant component of the construct, illustrated by the following quote: "Certainly, if the controller was oriented to, next time, behave more politely, more flexible, be aware of cultural differences and also about standard phraseology..." (M – FG 14 of 26 Scenario 4).

In conclusion, participants' comments in the examples above illustrate what aviation stakeholders perceived as necessary to improve the effectiveness of RT communication in terms of awareness, knowledge, skills and attitudes, based on their discussions of real scenarios. It was noted that construct components overlap across the

domains and dimensions, but more critically, a problem with one of them can be, many times, exacerbated by other issues specified in different cells of the matrix. The analysis confirms the narrow view of proficiency defined by the current ICAO LPRs, that is, the current language proficiency testing underrepresents the international RT communication construct, leading to questions regarding the validity of inferences drawn from current testing practices. Kim (2018) corroborates this finding, when she states that "the co-constructed nature of interactional competence is not at all reflected in the traditional linguistic-based ICAO rating scale. Interaction in the setting of air traffic control demands not just good language skills but also sufficient professional knowledge" (p. 420). Consonant with this view, Douglas (2000) argues that "when test content is highly specialized, and is based on complex concepts which are familiar to only a limited group of language users, good language proficiency alone will no longer be sufficient for effective performance" (p. 34). Instead, results indicate that intercultural communications in aviation require a broader view of communicative competence, including 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). As Snow (2018) notes, emphasizing the growing role of English as a lingua franca, "building effective intercultural communication skills is at least as important as building linguistic accuracy, if not more so" (p. 69). Therefore, findings from this and other research studies provide evidence that both first language (L1) speakers of English and those who speak English as a second (L2) or additional

language should develop the range of competencies outlined in the matrix of construct specification. Exempting native speakers of English from being tested in their specific purpose language ability to communicate in international radiotelephony seems to go against the safety requirements of aviation.

8.4 Member-checking

According to Creswell (2014), member-checking is used "to determine the accuracy of the qualitative findings through taking the final report or specific descriptions or themes back to participants and determining whether these participants feel that they are accurate" (p. 251). In this respect, I took the final matrix of construct specification (Table 8.7) back to three participants: one non-native speaker of English and AE test developer + AE rater (who I called M1); one native speaker and AE teacher + AE test developer + AE rater (who I called M2); and one native speaker of English and former ATCO + current AE teacher + AE test developer + AE rater (who I called M3), and gave them the opportunity to comment on the findings. I asked if, in their opinion, there was anything missing in the matrix, or anything that surprised them. Generally speaking, their comments were positive in regards to the final matrix, as for example: "This work is very important to our field and future research. It should help a lot with the revision of the ICAO language proficiency requirements and the development of valid tests" (M1). Other important observations were made during member-checking and are listed below:

- M3 pointed out that findings support the relevance of background knowledge in this context of language use, but the question remains "whether it is [should be] tested by default or IS actually implicitly or explicitly tested".

- M2 and M3 observed that some components in the matrix are broad and overlapping (e.g., use of communicative/interactional skills – Interactional competence/Skills; professional tone and attitude – Aviation English/Attitude), while some other inter-related ones are broken down into smaller elements (e.g., deal adequately with apparent misunderstandings, by checking, confirming and clarifying – Interactional competence/Skills; patience, politeness and tolerance – Aviation English/Attitude). This may be explained by the fact that some components were more clearly spelled out in the literature of, for example, ELF or IC. However, the broader term "use of communicative/interactional skills" was used a lot by the participants in their comments, so I decided to keep it in the final matrix. It certainly does not address the smaller components, which may be indicative of a lack of participants' knowledge of what they would be. This issue also discloses how challenging it was to code participants' comments into one domain/dimension or the other, especially in relation to ELF and ICA, two interconnected disciplines. For example: openness and humility to negotiate differences; and difficulty presented by the use of jargon, idioms, slang and colloquialisms could be coded into one or the other.
- M3 mentioned that pragmatic competence should be included in the matrix, referring to "being able to use language to achieve certain goals … to perform a function or express an intention clearly…[in a way] appropriate to the social context". My response to this is that the draft matrix drawn from the literature (see Table 8.2) includes knowledge of "language functions used in RT" (AE) and of "different pragmatic norms for different contexts" (ELF). However, the first did not rank high in the focus group discussions and the last received no comments at all, though they did not appear in

the final matrix. Nevertheless, the use of language functions associated with RT communications will be addressed in Phase 3 of this study.

- M2 raised the question if it would be possible to differentiate the components of the construct in relation to "skills/elements of a speaker [in relation] to another (e.g., accommodation/adaptation skills) that affect the communication as opposed to internal speaker skills not so linked to the interaction (e.g., language proficiency/face-concern, etc.)". The response to this question is not a straightforward one (nor within the scope of this study), as all these elements are intertwined and come into play at the moment of interaction.
- M3 pointed out that the domain of Aviation English "has been the focus of so many of the responses compared to the other three parts of the matrix… It certainly shows again that maybe more needs to be done to make one aware of all the multiple factors that affect communication in this domain". In fact, the objective was to validate the initial matrix drawn from the literature review with evidence from aviation stakeholders commenting on real cases of RT communications. Although their responses DO confirm the role of the four domains and dimensions in effective pilot-ATCO exchanges, findings suggest that a greater awareness of the role the last three domains play in this ESP context still needs to be achieved among those involved in RT communications.

8.5 Validation and Validity Strategies

It is important to mention that, in order to overcome internal and external validity threats, the study presented in this chapter applied rigorous qualitative methods (Creswell, 2014; Creswell & Miller, 2000; Morse, 2015), such as:

1. triangulation with different data sources: in Step 1, several theoretical and empirical studies were selected in order to inform the development of theoretical models of language use appropriate to the context of intercultural RT communications; in Step 2, the matrix of construct specification was developed from three theoretical models that offered different perspectives on the many factors involved in RT communications; in Step 3, six different scenarios of authentic pilot-ATCO communications were selected to trigger the focus group discussions, which provided evidence to "build a coherent justification for themes" (Creswell, 2014, p. 251) included in the preliminary matrix of construct specification;
2. appropriate sampling of participants, which included several groups of key aviation stakeholders with different language backgrounds and experiences (see Section 8.2.1);
3. development of a coding system and reporting of inter-coder reliability (see Section 8.2.4.3);
4. clarification of my research stance or standpoint in the study (see Section 1.2);
5. presentation of contradictory evidence, in order to provide a more realistic account of the research problem, which resulted in the exclusion of a number of components, not discussed in the focus groups, from the initial matrix of construct specification (see Section 8.3, discussion of RQ 2.3);

6. a rich account of the phenomenon under study, by presenting multiple perspectives about the topic. The inclusion of quotes from focus groups' participants in the discussion, provided richer and more realistic results (see Section 8.3.1 to 8.3.4); and
7. member-checking, by taking qualitative findings back to participants to verify if they feel they are accurate (see Section 8.4).

As mentioned in the beginning of the chapter, Phase 2 also aimed to validate the taxonomy of intercultural factors proposed in Phase 1 by triangulating data from the focus group discussions with aviation stakeholders as a validation strategy. The objective was to verify if all existing subcategories would be validated, if new sub-categories would emerge, and would also confirm the number of references for each sub-category. Table 8.10 displays the results from this analysis, including the existing sub-categories in black and the emerging ones in red, with their respective number of coding references in parentheses.

Table 8.10. Taxonomy of intercultural factors after Phase 2 analysis

THEME	CATEGORIES	SUB-CATEGORIES
Intercultural factors in international pilot-ATCO communications	Power Distance[a]	Power relations (95)[b] Deferential role (34)
	Face-work strategies	Self-face concern (36) Mutual-face concern (8)
	Conflict management	Conflictual direction (30) Neutral direction (6) Expectancy violations (7)
	Communication styles	Directness (38) Indirectness (82)
	Non-collaborative behavior	Unprofessional tone (160) Unprofessional attitude (187) Non-compliance with rules (185)
	Collaborative behavior	Professional attitude (66) Supportiveness (7)
	Layers of culture	National[c] (95) Professional (25) Organizational (12) Gender expectations (28)
	Individual traits	Personality issues (6) Lack of confidence (36) Language barriers (74) Lack of openness to difference (44)

Note. [a] In black, existing sub-categories
[b] In (parentheses), number of coding references
[c] In red, emerging sub-categories

As can be noted, participants' comments from the 26 focus group discussions also confirm and validate all the 14 existing sub-categories in the taxonomy of intercultural factors. Moreover, the number of coding references for each sub-category highlights their perceived importance according to this group of aviation stakeholders. In particular, issues related to non-collaborative behavior, comprising unprofessional tone and attitude and non-compliance with rules, which includes the lack of adherence to standard phraseology and other prescribed procedures, were substantially mentioned during the discussions. The number of coding references to these sub-categories is much higher than

the others, indicating their observed weight as an information to be considered in test design decisions. Further, the sub-categories relative importance can also be discussed in terms of the number of focus groups in which they were mentioned, as shown in Table 8.11.

Table 8.11. Number of focus groups that mentioned each sub-category

Sub-category	Number of Focus Groups
Unprofessional attitude[a]	25
Non-compliance with rules	24
Unprofessional tone / Indirectness	22
National culture[b]	21
Power relations	20
Professional attitude / Lack of openness to difference	19
Self-face concern	18
Conflictual direction / Directness / Lack of confidence	13
Language barriers	12
Professional culture	11
Deferential role / Mutual-face concern	7
Suportiveness	6
Gender expectations	5
Expectancy violation / Neutral direction /Organizational culture	4
Personality issues	3

Note. [a] In black, existing sub-categories
[b] In red, emerging sub-categories

Again, the top three sub-categories, i.e., the ones that were mentioned in a greater number of focus groups, are those belonging to the category of non-collaborative behavior, with unprofessional attitude ranking the highest. Figures from Table 8.11 also suggest that 'Language barriers', an emerging sub-category that comprises comments related to lack of language proficiency, did not appear as a relevant issue in more than half of the focus group discussions.

It is worth pointing out that the emerging sub-categories (see Tables 8.10 and 8.11 in red) were created in order to best categorize some of the participants' comments. These emerging sub-categories suggest connectivity to the previous ones, but so far there is not enough evidence to confirm their exact location within the existing categories. Therefore, they have been placed at the end of the taxonomy and organized into two main categories: layers of culture and individual traits, each comprising four sub-categories. Notwithstanding, these emerging sub-categories proved very useful to confirm the 'Model of the communicative demands of the RT occupational context', presented in Figure 8.8. The sub-categories and the model account for several layers of culture and also for individual traits or characteristics that may impact the interactions of pilots and ATCOs using AE as a lingua franca. In a way, this model advances what Wu and Li (2018) described as the need to "move beyond the fragmentary perspectives to a more holistic view of culture that includes not only nation-state culture, but organizational and technological culture, which will lead to a more multi-layered conception of culture" (p. 228).

In conclusion, having proposed models of language use that account for the communicative demands of pilots and ATCOs in intercultural RT communications, and specified a construct framework to inform test development, the validation of this framework was possible by giving voice to key stakeholders from diverse linguistic and cultural backgrounds. Within LSP contexts, domain experts play a critical role in construct definition, as indicated in this study. In addition, the taxonomy of intercultural factors proposed in Phase 1 was also validated, based on aviation stakeholders' perceptions.

As a next step in Phase 3, information regarding the relative weighting of the components of the construct will be incorporated in the design of new test tasks and the development of their specifications. This will be discussed in the next chapter.

Chapter 9 Phase 3: Operationalizing the Construct of International RT Communications as Test Tasks: Using Mixed Methods Research to Address Stakeholders' Perspectives

This chapter details the third phase of the multiphase mixed methods design that this dissertation describes. Having addressed in Phase 2 the first and second layers of the test development process (see Figure 3.1 re. design of *models* and specification of construct *framework*), Phase 3 of this study moves forward to the operationalization of the specified construct in test tasks that can elicit the behaviors indicative of effective RT communication. To that end, draft task specifications were developed, referred to by Fulcher and Davidson (2007) as the "generative explanatory document[s] for the creation of test tasks" (p. 52), and two draft tasks were designed. As a following step, the tasks were subjected to pilot testing, viewed as an initial and less formal pretesting than the main trials (Alderson, Clapham & Wall, 1995) with a group of Brazilian aviation stakeholders, who participated in a role-play enactment of the task.

9.1 Purpose and Research Questions

At this point, it may be useful to review how Phase 3 relates to the other Phases and to the overall conceptualization of the thesis (see Figure 1.1, presented in Chapter 1, for an overview of the study from the point of view of Phase purpose and an explanation of how all three Phases fit together).

Phase 3 builds on Phases 1 and 2, with the objective of developing and pilot testing test tasks that generate the evidence needed to make inferences from test performance to the constructs of interest. If this were a live testing project rather than exploratory research, at this point, one would begin to look at assessment criteria.

However, as the goal of this study was not to develop a whole new test, but to explore the inclusion of new tasks in order to more adequately represent the construct, Phase 3 investigated whether or not it would be possible to develop tasks that could reflect the construct as identified in previous Phases. This was done with the participation of aviation stakeholders to verify how viable the proposed tasks would be, based on their opinion and data from the initial pilot testing of the tasks. This initial step was essential, prior to the development of the assessment criteria and main trials with real test-takers. Evidence from this initial step also serves the purpose of informing policy makers about how essential construct components in the LSP testing of pilots and ATCOs could be operationalized as test tasks.

A first qualitative step was taken in order to move from the construct framework to the description of the evidence needed to test the construct (i.e., the sub-layer of *Evidence Models*) and to the design of tasks that generate the evidence needed (i.e., the sub-layer of *Task Models*). The third layer in the test development process, *Test Specifications*, is also addressed in Phase 3 through the development of task specifications for a role-play task to assess pilots' abilities to communicate in international radiotelephony contexts.

Specifically, the goal of Phase 3 was to answer the research questions outlined in Table 9.1.

Table 9.1. Research questions for each strand of Phase 3

	qual	QUAL	quan
Research questions	RQ 3.1) What behaviors or performances should reveal the proposed RT construct?	RQ 3.3) What are the test-takers' insights on the draft pilot tasks, and, from their perspective, to what extent do the role-play tasks reflect pilots' communicative needs in RT communication?	RQ 3.7) On the basis of expert judgment (raters), to what extent are: - the draft pilot tasks likely to elicit the desired language functions and behaviors indicative of effective communication? - the language functions and behaviors perceived as important?
	RQ 3.2) What kind of test tasks can be developed to elicit those behaviors in a pilot's exam?	RQ 3.4) What are the interlocutors' insights on the draft pilot tasks, and, from their perspective, to what extent do they feel confident in following the role-play instructions and perceive the appropriateness of the task?	
		RQ 3.5) What are the raters' insights on the draft pilot tasks?	
		RQ 3.6) What language functions and behaviors indicative of effective communication arise during test-taker's performance?	

In order to answer RQ 3.3 to RQ 3.7, a convergent parallel MM design was applied during the pilot testing of two draft test tasks. According to Creswell (2014), this type of MM design is used when "a researcher collects both quantitative and qualitative data, analyzes them separately, and then compares the results to see if the findings confirm or disconfirm each other" (p. 269). It was chosen due to the fact that evidence drawn from both qualitative and quantitative data and their evaluation through merging the results "facilitates a comparison between the two strands, significantly enhancing the study by generating a more comprehensive understanding of the phenomenon of interest" (Ziegler & Kang, 2016, p. 61). Therefore, individual interviews were conducted with test-takers and interlocutors right after task administration (QUAL). Their perceptions on the

issues involved in the task administration, appropriateness of the task, impact on test-taker preparation and confidence of interlocutors were also collected (QUAL). In addition, the degree of agreement among raters using an *Observation Checklist of language functions* in real time and an *Indicator Checklist of skills/behaviors indicative of effective communication* (quan) was compared to the raters' comments during focus group discussions and to the functions and behaviors identified in the transcripts of data generated as a result of task performances (QUAL). Figure 9.1 portrays the procedural diagram specific for Phase 3, which includes an initial qualitative component and a concurrent parallel MM design.

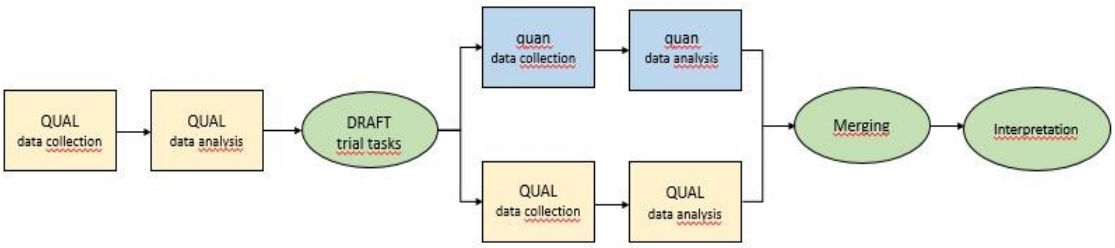


Figure 9.1. Phase 3 procedural diagram

In the sections which follow, I first explain the qualitative step that led to the design of two draft pilot tasks. Then, I describe the method (participants, instruments, procedures, and analysis) used in the QUAL and quan strands of the convergent parallel MM design. Having presented the method, I then discuss the results of the two strands in Section 9.5.

9.2 Initial Step: The QUAL Strand

9.2.1 What evidence do we need?

In order to conduct the pilot testing of the draft tasks, it was first necessary to move from the construct framework (outcome of Phase 2) to a list of observable

behaviors or performances that might provide evidence about the construct we wish to measure. Therefore, drawing on the tool that O'Hagan, Pill and Zhang (2016) used to extend the scope of the Occupational English Test (OET) speaking assessment criteria, the *Indicator Checklist of behaviors indicative of effective communication* in pilot-ATCO interactions (see Appendix Y) was developed. As a result of the complexity of this multi-phase MM design and the sequencing of events in the study timeline, this *Indicator Checklist of behaviors* was constructed based on the preliminary matrix of construct specification (see Table 8.2), immediately after the initial triangulation with language testers and ESL teachers. A set of 21 indicators was drawn from the matrix, consisting of five indicators of Aviation English (AE) competence, five indicators of English as a lingua franca (ELF) competence, five indicators of Intercultural awareness/competence (ICA), and six indicators of Interactional competence (IC). The indicators comprise a combination of components of the construct extracted from all four dimensions of interest: awareness, knowledge, skills and attitudes. Although the construct is well represented in the *Indicator Checklist*, it does not perfectly reflect the weighting of the final matrix of construct specification, which was the outcome of Phase 2.

In addition, as part of the evidence about the construct of interest, an *Observation Checklist of language functions* used by pilots and ATCOs in aeronautical radiotelephony communications was developed (see Appendix Z). The development of this checklist was informed by O'Sullivan, Weir and Saville (2002), who built a checklist of language functions to validate speaking tests within the University of Cambridge Local Examinations Syndicate (UCLES) examinations, " designed to be complementary to the use of transcriptions and to provide an additional source of validation evidence" (p. 39).

The language functions used in radiotelephony are organized into four broad categories: (a) Directed towards triggering action; (b) Sharing information; 3) Management of the pilot-controller relation; and (d) Management of the dialogue. The complete list of "Communicative language functions associated with aviation" can be found in the *Manual on the Implementation of ICAO Language Proficiency Requirements* (ICAO, 2010, p. B-1-B-4). They were compiled based on research by Mell and Godmet (1997).

At this point, it is possible to address the first research question in Phase 3, although it will be further discussed at the end of this Chapter.

RQ 3.1) What behaviors or performances should reveal the proposed RT construct?

The behaviors or performances that may provide evidence about the construct we wish to measure are the ones included in the *Indicator Checklist of skills/behaviors indicative of effective communication* in pilot-ATCO interactions (see Appendix Y), which comprises the indicators of Aviation English competence, English as a lingua franca competence, Intercultural awareness/competence and Interactional competence. In addition, the language functions associated with aviation, included in the *Observation Checklist of language functions* used by pilots and ATCOs in aeronautical radiotelephony communications (see Appendix Z) are also part of the evidence needed to make inferences from test performance to constructs.

9.2.2 Task design

In sequence, it was crucial to design tasks or situations that should elicit the behaviors we need as evidence, and the underlying task features. This was accomplished by drawing on Douglas' (2000) framework for the analysis of specific purpose test task

characteristics, which includes characteristics of the: (a) rubric, (b) input, (c) expected response, (d) interaction between input and response, and (e) assessment (p. 50-71).

Two draft tasks were designed by the researcher with the assistance of two airplane pilots, with different levels of flight experience. The proposed tasks represent a type of interactive task between a pilot (test-taker) and an air traffic controller (interlocutor) in the form of a role-play that elicits the language used for international radiotelephony communications, including sequential phases of a flight. The objective was to address the progression of contexts and communicative tasks pilots are used to perform in the TLU domain, with their corresponding content. Task 1 addresses the initial phases of a flight, and Task 2, on the other hand, addresses the final ones.

A response to the second research question that this initial qualitative strand addresses is provided below:

RQ 3.2) What kind of test tasks can be developed to elicit those behaviors in a pilot's exam?

In the proposed tasks, the test-taker is required to respond orally to an interlocutor, who plays the role of several air traffic controllers in sequential phases of a flight, during 12-14 minutes. Task 1 begins with a listening component of an Automatic Terminal Information Service (ATIS), defined as "the automatic provision of current, routine information, to arriving and departing aircraft" (ICAO, 2007, p. 1-6), which pilots normally listen to before initial contact with ATCOs. As in real life, it gives the test-taker a lot of contextual information about the airport conditions in the flight planning phase. The test-taker listens to it using headsets and takes notes. Then, he starts interacting according to the contextual information in his role-play card in five sequential mini role-

plays. He contacts clearance delivery (pre-flight), ground controller, (push-back and taxi), tower controller (take-off), and departure controller (climb). In Task 2, conversely, the test-taker communicates with area controller (en-route), listens to ATIS information, contacts approach controller (approach), tower controller (final and landing), and ground controller (taxi). However, as it is not possible to include all potential pieces of communication within these phases of flight, it was necessary to include some directions in the prompt to guide test-takers to the following phase. As Douglas (2010) explains, "even though we may not be able to reproduce target tasks exactly in our language tests, we can try to incorporate as many of the task and language characteristics as possible in our test tasks" (p. 53), so that test-takers can demonstrate the best of their language ability and we can interpret their performances in a "fair, accurate and meaningful way" (p. 54).

In both tasks, throughout the phases of the flight, the pilot is expected to elicit and respond to information provided by the ATCO, as well as to give information and state intentions using appropriate language functions and standard expressions. Unexpected situations and linguistic complications are intentionally included in the tasks, which require the use of plain language and communicative strategies to manage the interaction, negotiate meaning, and accommodate his/her interlocutor. The characteristics of the proposed role-play task for pilots, specifically of Task 1, are described in Table 9. 2 below, including the features outlined in Douglas' framework.

Table 9.2 Characteristics of the proposed role-play task for pilots (Task 1)

Rubric		
Objective	To demonstrate the ability to speak and understand the language used for radiotelephony communications	
Procedures for responding	Speak to interlocutor, interaction will be video-recorded	
Structure		
Number of tasks	5 small role-plays	
Relative importance	Role-plays of equal importance	
Distinction between tasks	Quite distinct – new phases are announced	
Time allotment	1-minute preparation; 1 minute to listen to ATIS; 2 minutes for each role-play	
Evaluation		
Criteria	ICAO rating scale: pronunciation, structure, vocabulary, fluency, comprehension, interactions; Test-takers are aware of the *Indicator Checklist of skills/behaviors of effective communication.*	
Procedures	Two raters (1 ELE and 1 SME) use a standard form to score performance independently after the test; third rater in case of disagreement between final levels 3 and 4; all categories scored on a scale of 1 – 6.	
Input		
Prompt		
LSP context	Test-taker's card	Interlocutor's card
Setting	Aeronautical RT communication	Aeronautical RT communication
Participants	Pilot	ATCO
Purposes	Communicate with ATCO from pre-flight (Airport A) to climb (heading to Airport B) Explicit: Listen to ATIS information. Contact ATCO and reply to his messages in an appropriate way, taking into account the phase of the flight and the information given (in both routine and non-routine situations) Implicit: manage the dialogue and the pilot-ATCO relation	Explicit: Interact with the pilot according to the phase of the flight and the scenario given. Explicit: Produce an intentionally ambiguous utterance/Indicate failure to comprehend/Request clarification/check understanding/request repetition
Form/Content	Oral interaction with no visual contact; Implicit: routine information exchange in each phase of the flight Explicit: Non-routine situations to handle	Oral interaction with no visual contact; Implicit: routine information exchange in each phase of the flight Explicit: Non-routine situations to handle
Tone	Professional manner	Professional manner
Language	ICAO standardized phraseology and plain English for aeronautical communication; Communicative language functions for aviation	Same as for candidate
Norms	Implicit: Pilot/ATCO interaction over the radio; elements of radio transmitting techniques	Same as for candidate
Genres	International aeronautical radiotelephony communications	Same as for candidate
Identification of problems	Deal with: medical emergency during taxi/runway incursion at take-off/depressurization during climb Negotiate meaning	Provide information, directions and assistance to pilot Simulate non-understanding/ produce intentionally ambiguous utterance/request repetition
Input data		
Format	Aural and visual	

Vehicle	Aural - Taped: Genuine recording of (or simulated) ATIS information Aural - Live: oral input from interlocutor Visual: three pictures portraying unexpected situations
Length	Prompt: 20 lines Recording: 0.5 minutes Oral input: 4-5 minutes
Level of authenticity	
Situational	Fairly high: common type of situations to pilots; pictures portray real scenarios
Interactional	High to moderate: some interlocutors (SME-pilots) are not highly trained in their role as ATCO
Expected response	
Format	Oral
Type	Extended
Response content	
Language	ICAO standardized phraseology and plain English for aeronautical communication; language functions associated with aviation; strategies related to Aviation English competence, ELF competence, Intercultural awareness/competence and Interactional competence.
Background knowledge	Aviation knowledge, RT communication rules and procedures
Level of authenticity	
Situational	High situational authenticity: The task shares many features of TLU situation
Interactional	High to moderate: The task engages an appropriate discourse domain in test-takers (aeronautical RT communications)
Interaction between input and response	
Reactivity	Highly reciprocal: adaptation on both sides as necessary for mutual comprehension
Scope	Moderately broad: must process information on card
Directness	Fairly indirect: must use background knowledge
Assessment	
Construct definition	Refer to the matrix of construct specification, which details the dimensions of interest (awareness, knowledge, skills and attitudes) across the domains of Aviation English, English as a lingua franca, Intercultural awareness/competence, and Interactional competence
Criteria for correctness[a]	ICAO Rating Scale descriptors: pronunciation, structure, vocabulary, fluency, comprehension, interactions. Expanded assessment criteria which values strategies related to Aviation English competence, ELF competence, Intercultural awareness/competence and Interactional competence.
Rating procedures	Two raters (1 ELE + 1 SME) rate the sample independently after the test terminates, by listening to the audio/video recording. A third rater is used only in case of disagreement between final levels 3 and 4.

Note. Based on the framework of specific purpose test task characteristics (Douglas, 2000)

[a] Assessment criteria needs to be expanded based on trialing data accumulated over time as part of the task validation process.

As Task 2 features are very similar to those of Task 1, the role-play task characteristics specific for Task 2 are detailed in Appendix AA – *Characteristics of the proposed role-play task for pilots (Task 2)*.

9.2.3 Task specifications

Moving to the third layer of architecture documentation, we arrive at 'Test Specifications', in which emphasis will be given to the sub-layer of 'Task specifications'. The design of task specifications is required as a 'blueprint' for the subsequent generations of equivalent test tasks. According to Davidson and Fulcher (2012), typically "a spec has two elements: guiding language and at least one sample item. The guiding language sets out the rationale, background, description of resources, and other details to justify and explain the particular items [tasks] being created" (p. 59). In addition, qualities of good language testing practice, as proposed by Douglas (2000) for LSP testing, were also incorporated to the task specifications, which can be found in Appendix BB – *Draft task specifications*. The task specifications include an excerpt of role-play Task 1 (one phase of the flight only), comprising a sample of the test-taker's card and the corresponding interlocutor's card, used in the pilot testing (see Appendix BB - Attachment A).

In the next section, I describe the pilot testing of two draft test tasks, conducted in the context of Brazilian language proficiency assessment of pilots, using a convergent parallel MM design in order to answer my research questions.

9.3 Convergent Parallel MM Design: QUAL Strand

In this convergent parallel MM study, the QUAL strand is the dominant one. This strand comprises multiple qualitative methods nested within it.

9.3.1 Participants.

During the pilot testing of the two draft tasks, participants were Aviation English Testing Experts (AETEs) who, at the time of data collection, were employees at the

Brazilian Civil Aviation National Agency (ANAC) and Aviation English examiners from ANAC accredited institutions. The pilot testing of the tasks required AETEs:

- with a minimum of 2 years' experience in applying/rating Aviation English proficiency tests;
- who were English language experts (ELEs) and/or subject matter experts (SMEs), pilots or ATCOs;
- who spoke English fluently[27].

Task 1 and Task 2 were both pilot tested with two different groups of participants, leading to four task administrations. Therefore, four AETEs, who were pilots, played the role of test-takers, and four AETEs, two pilots and two ATCOs, played the role of interlocutors. Their background information is described in Table 9.3.

Table 9.3. Background information of AETEs who acted as test-takers and interlocutors

Group	Role	Gender	Age range	Years as AE examiner	Expertise	L1	Highest level of education
1 Morning	Test-taker 1a	Male	> 45	2-4 years	SME (pilot)	Portuguese	Master's degree
	Test-taker 2a	Male	> 45	2-4 years	SME (pilot)	Portuguese	Master's degree
	Interlocutor 1a	Male	>45	6-8 years	SME (ATCO)	Portuguese	College
	Interlocutor 2a	Male	>45	2-4 years	SME (pilot)	Portuguese	Master's degree
2 Afternoon	Test-taker 1b	Male	> 45	2-4 years	SME (pilot)	Portuguese	High school
	Test-taker 2b	Male	36-45	2-4 years	SME (pilot)	Portuguese	University
	Interlocutor 1b	Male	>45	> 8 years	SME (ATCO) + ELE	Portuguese	University
	Interlocutor 2b	Male	36-45	4-6 years	SME (pilot)	Portuguese	University

In addition, live tasks performances, both in the first group (morning) and in the second group (afternoon) were observed by AETEs who were experienced Aviation

[27] Fluency was determined based on the requirements to be an ANAC language proficiency examiner : (a) for SMEs, pilots and ATCOs: hold a pilot or ATCO license with a valid language proficiency endorsement at Level 5 or a Level 6 endorsement; and (b) for ELEs, hold one of the following language proficiency certificates: CPE (minimum grade C), CAE (minimum grade B), IELTS (minimum score 7,0) – Cambridge University; TOEFL (minimum score: paper-based: 560; computer-based: 220; internet-based: 110) – Educational Testing Service; ECPE (minimum score: low pass) – Michigan University) (ANAC, 2018).

English raters, both ELEs and SMEs. They acted as observers of test-takers' performances and of interlocutors' behaviors in order to evaluate the task. A third group of raters, who were not able to join the first two groups, observed a video recording of one of the tasks later at ANAC, so that they could also provide their comments and feedback on the task. The groups were organized as detailed in Table 9.4, totaling 35 observers. Each column displays the number of observers according to gender, age range, years as an AE examiner, expertise in years as an ELE or SME, L1, and highest level of education.

Table 9.4. Background information of AETEs who acted as observers across groups

Group	Gender	Age range	Years as AE examiner	Number of AETEs by Expertise	L1	Highest level of education
1 Morning	Male: 09	25-35: 01	2-4: 06	ELE: 09	Portuguese: 14	University: 08
	Female: 08	36-45: 05	4-6: 02	SME (pilot): 06	English: 02	Master's degree: 06
		>45: 11	6-8: 04	SME (ATCO): 02	Other (Spanish): 01	Other:03
			>8: 05			
2 Afternoon	Male: 08	25-35: 02	2-4: 05	ELE: 07	Portuguese: 14	High school: 01
	Female: 07	36-45: 03	4-6: 01	SME (pilot): 06	English: 01	College: 02
		>45: 10	6-8: 04	SME (ATCO): 02		University: 11
			>8: 05			Other:01
3 ANAC	Male: 01	25-35: 02	2-4: 02	ELE: 03	Portuguese: 03	University: 01
	Female: 02	36-45: 01	>8: 01			Master's degree: 01
						Other:01

In order not to cause any harm to participants who would play the role of test-takers, pilots with a lower level of English proficiency, who could potentially feel uncomfortable or at risk to perform the task in front of a group of AE examiners, were not chosen for the pilot testing. Rather, the ones who performed this role were members of the same group of AETEs, to who were assured that their individual performance in the task would not be evaluated. Rather, the focus was to evaluate the task itself.

The Coordinator of ANAC's Language Proficiency Group assisted in the recruitment of participants and granted explicit authorization to pilot test the draft tasks at ANAC, through the collection of audio, video and written data (see Appendix G). Specific consent forms were developed for each group of participants (see Appendices J, K and L) and the Carleton University Research Ethics Board (CUREB) granted ethics clearance for Phase 3 of this study on September 21, 2017 (see Appendix E), under Project Number: 107816.

9.3.2 Instruments.

In order to pilot test the tasks, a range of instruments of data collection was used. First, I will mention the ones that were used by the participants who acted as test-takers and interlocutors, comprising the following:

1. Two test-takers' role-play cards, one for Task 1 and one for Task 2 (see an excerpt in Appendix BB - Attachment A;
2. Two interlocutors' role-play cards, one for Task 1 and one for Task 2 (see an excerpt in Appendix BB - Attachment A;
3. A set of pictures as visual input, one for Task 1 and one for Task 2;
4. Two simulated recordings of Automatic Terminal Information Service (ATIS), one for Airport A, used in Task 1, and one for Airport B, used in Task 2;
5. A set of six questions asked to AETEs who played the role of test-takers, during a face-to-face semi-structured interview (see Appendix CC);

6. A set of eight questions asked to AETEs who played the role of interlocutors, during a face-to-face semi-structured interview (see Appendix CC);
7. A feedback form with 12 Likert scale questions for AETEs who played the role of test-takers (Appendix DD), with a background information form (Appendix FF);
8. A feedback form with 12 Likert scale questions for AETEs who played the role of interlocutors (Appendix EE), with a background information form (Appendix FF); and
9. Consent forms specifically designed for test-takers (Appendix K) and interlocutors (Appendix J).

In addition, qualitative data were also collected through focus group discussions with AETEs, i.e., ELEs and SMEs raters, who acted as observers. To this end, a specific set of nine questions was developed to guide their discussions (Appendix GG), to which they responded after signing a consent form (Appendix L).

Finally, all four task performances in the pilot testing were video-recorded and transcribed, as discourse-level data would also be used for qualitative analysis.

9.3.3 Procedures.

Before the actual piloting of the tasks, the researcher explained to each group of participants (morning and afternoon) that the objective of the session was to pilot test two draft tasks developed to assess pilots' ability to communicate with air traffic controllers over the radio within the high-stakes context of international aviation radiotelephony using the English language. Then, the researcher made clear what their participation

would entail identifying the ones who volunteered to act as test-takers, interlocutors and observers. They signed a consent form and filled in another form with their demographic information.

While interlocutors for Task 1 and Task 2 were familiarizing themselves with the content and guidance to follow during the interaction, included in the interlocutors' role-play card, the room was being set up for the piloting of the tasks. A rectangular table, with a visual barrier in the middle, was positioned in a place where everybody in the room could see both the test-taker and the interlocutor. The side of the test-taker had a headset, blank sheets of paper and pens for note-taking, and the test-taker's role-play card. The side of the interlocutor contained a computer linked to the test-taker's headset to play the ATIS recording, the interlocutor's role-play card and the set of pictures. Figure 9.2 portrays the room setting, showing the test-taker on the left side and the interlocutor on the right side[28]. This was the perspective from which the observers watched the role-plays.

[28] Participants consented in having their images used in publications as long as their identity would be protected by obscuring any identifiable features from the video (see Appendices J and K).

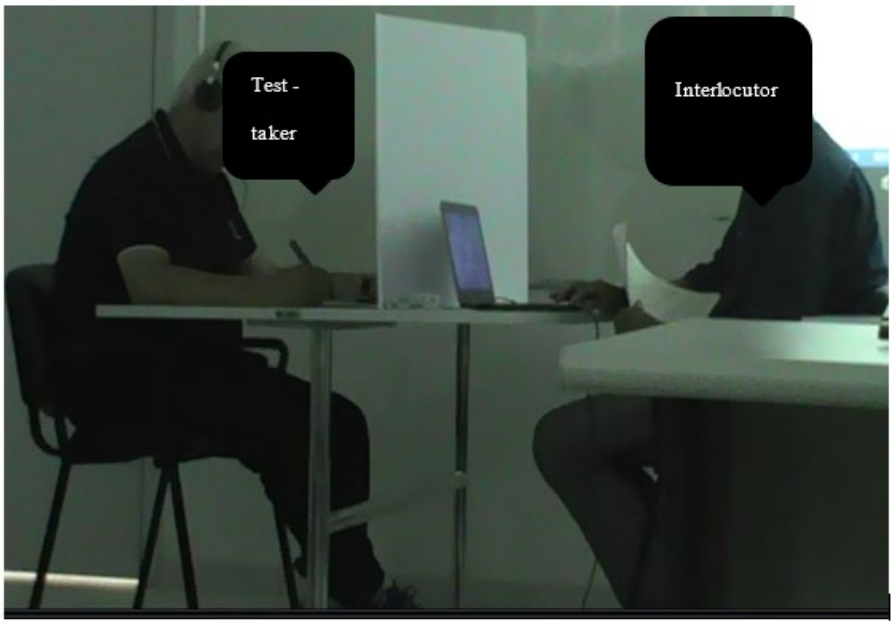


Figure 9.2. Room setting for task pilot testing

Immediately after the role-play Task 1 performance, the researcher left the room with the test-taker 1a and interlocutor 1a. One at a time, in a separate room, test-taker 1a and interlocutor 1a were interviewed by the researcher, who audio-taped their responses. After the interviews, they were also given a feedback form with 12 Likert scale questions to answer (see Appendices DD and EE). Back in the main room, Task 2 was piloted following the same steps, with test-taker 2a and interlocutor 2a, who were also interviewed orally and responded to 12 Likert scale questions. Then the researcher moderated and audio-recorded a focus group discussion with observers, including the test-takers and interlocutors who role-played Tasks 1 and 2. The same procedures were followed with the second group of participants (in the afternoon), who also role-played Task 1 and Task 2.

A third group of participants (N=3), who did not have the opportunity to watch the live piloting of the tasks, had the chance to contribute to the study at a later date, acting as observers. The researcher provided the same initial explanations to these

participants and played the video recording of Task 1. After that, a focus group discussion was moderated and audio-recorded by the researcher, applying the same set of nine questions used previously in the other groups.

Video files from four task performances, and audio files from four interviews with test-takers, four interviews with interlocutors, and three focus group discussions with observers were fully transcribed with the aid of *Express Scribe Transcription Software* version 6.10. Interview and focus group data were imported into *NVivo Version 12 Plus for Windows* in order to be coded in a systematic way, whereas discourse data from task performances were coded manually, with the objective of finding evidence of the language functions produced by the test-takers as well as skills and behaviors indicative of effective RT communications.

9.3.4 Analysis.

9.3.4.1 Semi-Structured interviews with test-takers and interlocutors.

After importing data from the eight interview transcripts into *NVivo Version 12 Plus for Windows,* four with different test-takers and four with different interlocutors, the coding process started with reading and re-reading the data in order to have initial thoughts on codes. Therefore, a First Cycle coding was conducted applying the method of Descriptive coding, which, according to Saldaña (2009), "summarizes in a word or short phrase – most often as a noun – the basic topic of a passage of qualitative data" (p. 70). Thus, as participants' answers reflected the main topics elicited by each interview question, nodes were created based on key words from these questions. For test-takers, they included: general feeling, difficulties, task appropriateness, engagement, addition/removal of something, suggestions/comments. For interlocutors, the nodes were

slightly different, comprising: general feeling, difficulties, task appropriateness, script information, addition/removal of something, suggestions/comments, and interlocutors' training.

A Second Cycle coding was carried out in sequence using a feature of *NVivo Version 12 Plus for Windows* called 'Sentiment analysis', which allows the researcher to "grasp the range of sentiment across . . . content by categorising over four streams: very positive, moderately positive, moderately negative or negative" (NVivo, https://www.qsrinternational.com/nvivo/enabling-research/sentiment-analysis, accessed on May 27, 2019). This is, in fact, what Saldaña (2009) calls Magnitude coding, which "consists of and adds a supplemental alphanumeric or symbolic code or sub-code to an existing coded datum or category to indicate its intensity, frequency, direction, presence, or evaluative content" (Saldaña, 2009, p. 58). In this case, numeric values were attributed to the responses of test-takers and interlocutors using the same four streams from *NVivo*. The following codes were used: 0= Very negative; 1= Moderately negative; 2= Moderately positive; and 3= Very positive. As Saldanã (2009) explains, Magnitude Coding can be used as a way of 'quantitizing'[29] a phenomenon's evaluative content, which, combined with text, can "work in concert to compose a richer answer and corroborate each other" (p. 58).

9.3.4.2 Feedback forms from test-takers and interlocutors.

The feedback forms that test-takers (Appendix DD) and interlocutors (Appendix EE) responded to each contained 6-point Likert scale questions, so that participants could rate the degree to which they agreed or disagreed with 12 topics related to the pilot

[29] 'Quantitizing' is an expression used by Saldaña (2009) and also by Ziegler and Kang (2016) when referring to the quantification of qualitative data, as a method of data transformation and analysis.

testing of the tasks. Their responses were inserted into *IBM SPSS Statistics version 24,* and descriptive statistics, frequencies and percentage analyses were carried out. Despite the quantitative nature of Likert scale instruments, the sample size of respondents, i.e., test-takers (N=4) and interlocutors (N=4), did not allow for a quantitative analysis, that is, to make inferences from the sample to the population. Therefore, both the interview transcripts and the Likert scale responses are better interpreted as converging multiple qualitative methods, with a 'quantitized' component.

9.3.4.3 Focus group discussions with raters acting as observers.

Focus groups data were imported into *NVivo Version 12 Plus for Windows* in order to be coded in a systematic way. A bottom-up approach was privileged and, using Descriptive coding, a mind map was built in *NVivo* to depict preliminary interpretations of data and potential nodes. Soon, the amount of detail provided by participants required an expansion of the map to include "child ideas", or sub-nodes, and also "sibling ideas", i.e., sub-nodes from the same "parent" node. Figure 9.3 displays the mind map with the coding hierarchy that accounts for raters' perceptions on the draft tasks. Then, this mind map was transformed into nodes in *NVivo*, so that the transcriptions of the three focus groups could be coded.

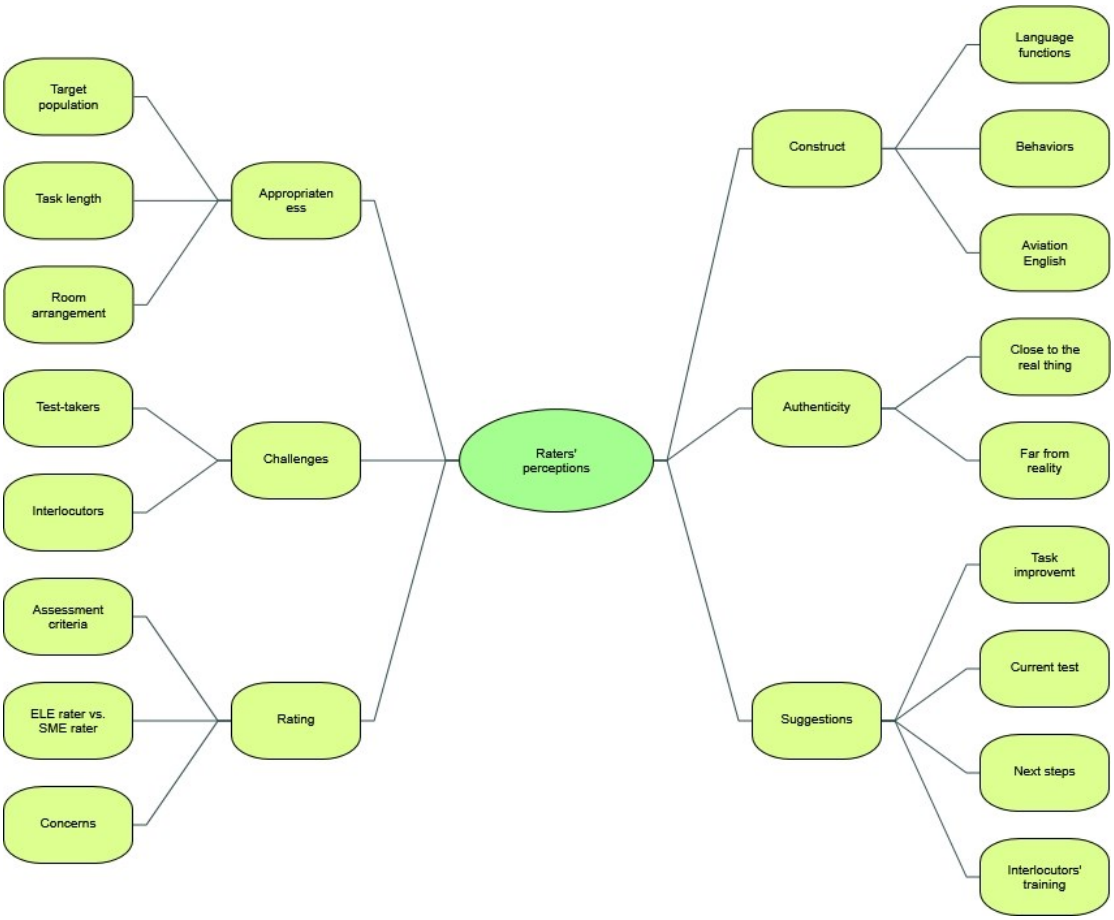


Figure 9.3. Mind map of raters'/observers' perceptions, with nodes and sub-nodes

At this point, inter-coder reliability was calculated. In order to increase the reliability and validity of my research findings, I asked another coder (Coder 2) to independently code the transcript of the focus group discussion conducted with Group 1. This second coder, a NNS of English graduate student in Applied Linguistics and Discourse Studies, applied Provisional coding as the coding method, using the mind map as the starting list. The codes assigned by Coder 2 were organized according to similarities with my own codes (Coder 1). The process of grouping similar codes was followed by the verification of their occurrences for each coder across the 17 sub-categories identified in the mind map in order to calculate the correlation among the coders and the inter-coder reliability. The correlation was run in the SPSS Software,

version 24, which gave the following result: Pearson's r between coders 1 and 2, r = .87. This indicates a significant correlation between the coders. In addition, the reliability analysis was run using the intra-class correlation coefficient and the absolute comparison option. The test was based on a 95% confidence level. Cronbach's Alpha provided a reliability measure of .92. The results of the intra-class correlation coefficient analysis provided information on the reliability and predictability of the variance: both single measures (.83) and average intra-class correlation measures (.91) were above .70, indicating a positive and significant correlation in the data. The significance level of p <.0005 indicated a high level of reliability that the positive correlation among the coders was not due to random chance over 95% of the time.

Then, a Second Cycle coding was undertaken in the same way as was done with the interview data: 'Sentiment analysis' in *NVivo* and Magnitude coding – creation of nodes in order to rate participants' comments according to their evaluative content, i.e., from very negative to very positive. The following codes were used: 0= Very negative; 1= Moderately negative; 2= Moderately positive; and 3= Very positive.

9.3.4.4 Analysis of transcripts of task performances.

In order to proceed with the analysis of the interactions between test-takers and interlocutors during the role-plays, I first explored the data by breaking it down into pilot's and ATCO's utterances, which are defined by Bakhtin (1986) as "a unit of speech communication" (p. 73), giving rise to "responsive reactions and dialogic reverberations" (p. 94). The author explains that utterances possess three main features: a) boundaries delimited by the change of speaking subjects, b) the finalized wholeness of the utterance, guaranteeing the possibility of a response or of responsive understanding, and c) the

relation of the utterance to the speaker himself and to the "other" participants. This notion of the "other" is explained by the existing relations among whole utterances in a speech communication, such as question and answer, command and execution, which are expected only among utterances of different interlocutors. In radiotelephony communications, these relations are well defined by the communicative language functions performed by pilots and air traffic controllers, who are the two speaking subjects of this particular speech genre and who alternate the roles of author and addressee. Each language function presupposes a response from the "other": e.g., give an order/announce compliance with an order, request information/give information, check understanding/acknowledge, and so forth. Therefore, as highlighted by Bakhtin (1986), addressivity is "a constitutive feature of the utterance" (p. 99) which helps to define its genre, in the sense that each utterance has an author and is directed to someone, the addressee, whose specific characteristics vary according to each area of speech communication.

Then, in the First Cycle coding, the coding method of Provisional Coding was applied in order to identify the language functions produced by both the test-taker (role of pilot) and the interlocutor (role of ATCO), through the use of the existing list of communicative language functions associated with aviation (ICAO, 2010). It was a manual coding, in which the actual words produced by the pilot and ATCO during the task performance related to a language function were included inside square brackets [ ], and their respective coding was identified immediately after and included inside parentheses ( ) and highlighted in red. An example of language functions analysis from task performance in Task 1, morning group, is shown in Figure 9.4. The same procedures

were followed in the identification and coding of language functions in the four task administrations.

P: [Clearance delivery] (name addressee), this is ANAC 123, [ready to copy the clearance] (announce readiness).

I: [ANAC 123] (name addressee), [cleared to 'Airport b'; via upper alpha 317, 'VOR Yankee', Whisky 8; flight level 320; 'rom... (intentionally unintelligible) departure', 'BRAVO' transition; after airborne maintain runway heading; squawk 3252.] (give permission)

P: [Say again the clearance delivery] (request repetition), please.

I: [ANAC 123] (name addressee), [I say again, 'ROMEO departure'] (give repetition). [ANAC 123] (name addressee), [cleared to 'Airport b'; upper alpha 317, 'VOR Yankee', Whisky 8; flight level 320; 'ROMEO departure', 'BRAVO' transition; after airborne maintain runway heading; squawk 3252.] (give permission)

P: [Roger] (acknowledge), eh...[ANAC 123 is cleared to 'airport b' via upper alpha 317, 'VOR Yankee', then Whisky 8; flight level 320. after take-off maintain runway heading; squawk 3252. The departure is expected to be ROMEO and transition is VICTOR.] (Read back)

I: [ANAC123] (name addressee), [I say again, the transition is BRAVO transition.] (correct a misunderstanding)

P: [BRAVO transition,] (read back) [copied that.] (acknowledge)

I: [ANAC 123] (name addressee), [read back is correct] (give confirmation). [CONTACT 'Airport A' ground on 121.35.] (give an order) [ANAC 123] (name addressee).

P: [121.35] (read back), [good day] (take leave).

I: [Good day] (respond to leave-taking).

Note. Red indicates coding for language functions.

Figure 9.4. Example of language functions analysis from actual task performance – Task 1 (morning)

In order to answer the research questions, the transcripts of task performances were also subjected to a Second Cycle coding, aiming to identify the evidence related to the behaviors of effective communication in radiotelephony in terms of AE competence, ELF competence, Intercultural awareness/competence, and Interactional Competence. Again, it was a Provisional Coding, but this time using the *Indicator Checklist of skills and behaviors indicative of effective communication.* Nonetheless, at the very early stages of this analysis, I noticed that in a single utterance it was possible to find evidence of a range of behaviors, or components of the construct, requiring the assignment of two or more codes simultaneously. This explains the use of Simultaneous Coding for this part of the qualitative analysis. Figure 9.5 illustrates an example of how the number of each

indicator in the checklist was used as a code, e.g., number 3 – communicating effectively in routine and in unpredictable situations (indicative of AE competence); number 9 – avoiding the use of jargon, idioms, slang and colloquialisms (indicative of ELF competence), etc. Through manual coding, the actual words produced by the pilot and ATCO during the task performance were kept inside square brackets [ ], and their respective coding was identified immediately after and included inside parentheses ( ) in red color. The words highlighted in yellow correspond to instances of plain language, either replacing or adding to correct phraseology.

P: [Airport B approach, ANAC 123, we have information Charlie] (indicators 1, 2, 3, 5, 8, 9, 16, 17, 20).

I: [ANAC123, Roger. Descend to flight FL 060, fly direct ZULU VOR, altimeter setting… xrixri (unintelligible), expect YANKEE ILS approach to runway 20, report 20 miles out] (indicators 1, 5, 9, 16, 20)

P: [ANAC 123, cleared to descend to FL 060, ah….VOR … Can you say again the information] (indicator 1, 2, 3, 5, 8, 9, 10, 16, 17, 19, 20), [please?] (indicators 4,12)

I: [ANAC123, Roger. Descend to flight FL 060, fly direct ZULU VOR, altimeter setting… xrixri (unintelligible), expect YANKEE ILS approach to runway 20, report 20 miles out] (indicators 1, 5, 9, 16, 20)

P: [Okay] (indicator 5), [ANAC 123, cleared to descend to FL 060, direct to ZULU VOR, ILS Yankee, for runway 20, I will call you 20 miles out. Can you confirm the altimeter setting] (indicators 1, 2, 3, 5, 8, 9, 16, 17, 20), [please?](indicators 4, 12)

I: [ANAC 123, I say again, altimeter setting 1019] (indicators 1, 2, 3, 5, 8, 9, 16, 17, 20).

P: [1019, ANAC 123] (indicators 1, 2, 3, 5, 8, 9, 16, 17, 20).

Note. Red indicates coding for behaviors of effective communication.
Instances of plain language are highlighted in yellow.

Figure 9.5. Example of behaviors analysis from actual task performance – Task 2 (afternoon)

Having described the method (participants, instruments, procedures, and analysis) used in the QUAL strand, in the next section I provide details about the quan strand.

9.4 Convergent Parallel MM Design: Quan Strand

9.4.1 Participants.

Participants who took part in the quantitative strand of this MM study were the same AETEs, i.e., experienced Aviation English raters who volunteered to participate in the qualitative strand. As described in Section 9.3.1, they were either employees at ANAC or Aviation English examiners from ANAC accredited institutions. In the study, however, they only acted as observers of the task in order to evaluate it and not as raters of test-takers` performances. It is important to mention that, in the context of pilots` language proficiency testing in Brazil, the population of raters at the time of data collection consisted of 40 ELEs and SMEs (pilots and ATCOs). That said, the sample of AETEs who acted as observers, i.e., N=35, characterizes a very good representation of the population. Figure 9.6 provides a visual summary of their background information, in terms of gender, age, group, first language, years as aviation English examiner, and highest level of education.

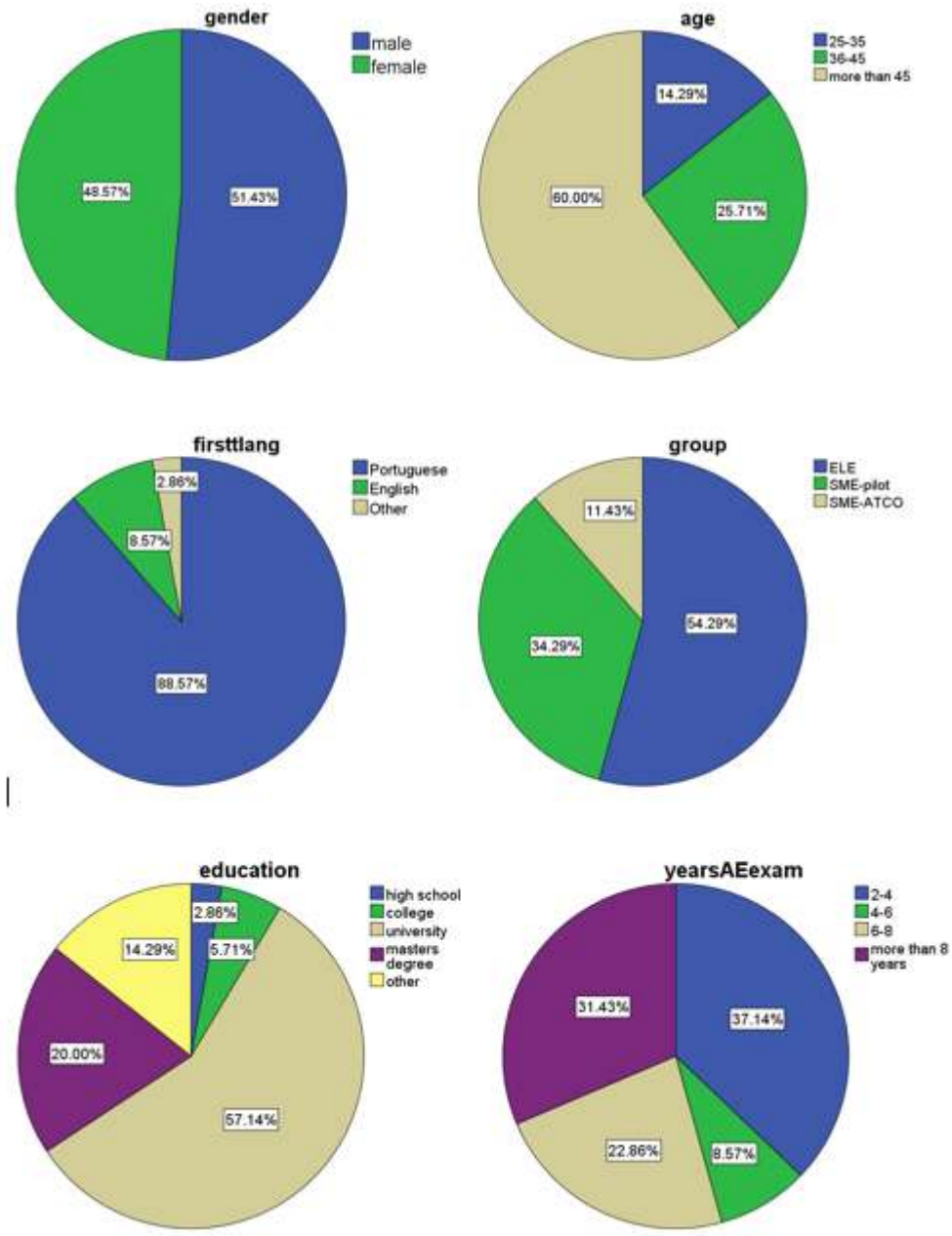


Figure 9.6. Percentage of participants in the quan strand according to gender, age, group, first language, years as Aviation English examiner and education (N= 35)

9.4.2 Instruments.

In the quantitative strand, two instruments were used to collect data:

1. the *Indicator Checklist of skills/behaviors indicative of effective communication* in aviation radiotelephony (Appendix Y) – It contains five indicators of Aviation English competence, five indicators of ELF competence, five indicators of Intercultural Awareness/competence, and six indicators of Interactional competence. The checklist was designed with three columns. In reference to the first two, observers were told to select the option that best characterized each behavior in Task 1 and in Task 2, in terms of being demonstrated or evidenced during task performance: (Yes) demonstrated; (No) not demonstrated; or (N/A) not applicable. This last option would be used when participants believed a particular task did not require the demonstration of certain skills/behaviors. The third column had a different purpose: participants would rate the level of importance of each behavior, ranging from 1 (not important) to 4 (very important);
2. the *Observation Checklist of language functions* (Appendix Z) – The list of communicative language functions associated with aviation (ICAO, 2010) was included in the *Observation Checklist* in its entirety, keeping the four main categories and sub-categories that organize them into smaller groups. It was designed with the same three columns included in the *Indicator Checklist of skills/behaviors*, so that observers could select if

the language functions were demonstrated or not during task performances, as well as rate their perceived level of importance.

9.4.3 Procedures.

In both the morning and afternoon groups, the first thing observers did, before the role-plays started, was to rate the level of importance of the language functions listed in the *Observation Checklist* (Appendix Z). The idea was to ease participants' understanding of the language functions, get a sense of their perceived relevance to RT communications and facilitate their later identification during the application of the *Observation Checklist*.

Then, the tasks were pilot tested and video-recorded, and observers were instructed to apply the checklist of language functions (Appendix Z) during task performances for both Task 1 and Task 2. Immediately after task administration, observers also filled in the *Indicator Checklist of skills and behaviors indicative of effective communication* (Appendix Y) for both tasks, i.e., first and second columns. After that, they rated the level of importance of each behavior indicative of AE competence, ELF competence, Intercultural awareness/competence, and Interactional competence, in the third column.

However, the third group that convened later at ANAC did not have the chance to observe a live task administration, due to time constraints, logistics and lack of AETEs, i.e., pilot and ATCO to perform the task as test-taker and interlocutor. Therefore, this group of participants was given the *Indicator Checklist of skills and behaviors indicative of effective communication* and rated only the level of importance of each indicator.

9.4.4 Analysis.

Each participant received a case number. Data from the demographic information form and from both checklists, i.e., the *Observation Checklist of language functions* and the *Indicator Checklist of skills and behaviors indicative of effective communication* were inserted into SPSS software, version 24. Variables were created for indicators 1 to 21 in Task 1 (e.g., beh1task1), for indicators 1 to 21 in Task 2 (e.g., beh21task2), as well as for the level of importance of each indicator (e.g., beh16imp). The same was done for the language functions in Task 1and 2 (e.g., LF1.1.1task1) and for their level of importance (e.g., LF4.1imp). Data were analyzed using descriptive statistics and frequency distribution (Vogt, 2007; Larson-Hall, 2016), with the objective of extracting mean values and frequencies as a strategy of data reduction of quantitative findings into manageable chunks of information.

The categorical variables (e.g., age, group, first language, gender, etc) helped the researcher to describe the characteristics of the sample in terms of participants' background, as detailed in 9.4.1. The continuous variables, on the other hand, enabled the researcher to have an overview of the responses so as to select the most striking and relevant issues for discussion, according to the research questions.

In this MM convergent parallel study, data from the QUAL and quan strands were collected and analyzed during the same time frame but kept separately. Then, integration of the two data sets was accomplished through merging and combining results. This integration is presented in the next section, followed by an interpretation and discussion of results, aiming at producing a better understanding of the issues involved in the pilot testing of the tasks and detecting points of convergence or divergence.

9.5 Results and Discussion

Results from QUAL and quan analysis within the convergent parallel MM design were merged and compared, so as to answer the overarching MM research question of Phase 3:

> To what extent do the quantitative results from the pilot testing of tasks agree with the interviews and focus group data reporting the views of test-takers, interlocutors and raters and with transcripts of task performances?

Nonetheless, before doing so, results are presented and discussed to answer the more specific research questions. Having answered in Sections 9.2.1 and 9.2.2 research questions RQ 3.1 and RQ 3.2, respectively, now I move on to discuss results and answer research questions from RQ 3.3 to RQ 3.7.

9.5.1 Research question 3.3

RQ 3.3) What are the test-takers' insights on the draft pilot tasks, and, from their perspective, to what extent do the role-play tasks reflect pilots' communicative needs in RT communication?

Evidence to answer this question was collected from the multiple methods used in the qualitative strand of the study. However, employing data transformation strategies, such as 'quantitizing' qualitative data through the use of Magnitude coding and also percentage analyses of Likert scale responses, facilitated the process of comparing and contrasting concurrent sources of evidence, which provided a greater insight into test-takers' perspectives. Thus, to generate meta-inferences and conduct an integrated data interpretation, findings were organized around common constructs in a joint display (see Table 9.5). Following Ziegler and Kang's (2016) example, the constructs of interest, i.e.,

the main themes discussed with test-takers, were entered along the vertical axis and reduced findings from the two sources of data were included along the horizontal axis.

Table 9.5. Test-takers' perceptions from individual interviews and Likert scale questions

Themes	Face-to-face semi-structured interviews				Likert scale questions
	Task 1 – Morning	**Task 2 - Morning**	**Task 1 - Afternoon**	**Task 2 - Afternoon**	
General Feeling	"Very, very nice. It was an excellent exercise". (3)	"I felt comfortable". (2)	"I felt fine". (2)	"Good, well, I thought it was very well created, I think".(3)	**25% strongly agree/50% agree/25% somewhat agree** - I was confident during task administration (IMP)
Difficulties	"The transition from one situation to another. It should be made clear to the student". (1)	"But in my opinion, although the pilot has the script to follow, yes, it would be better if the interlocutor said the changing situations. Because just with the.....now taxi". (1)	"Only at times I found that the prompts, or the script, was a little confusing about when to start and when not to start". (2)	"A little bit, you know". (2)	**25% strongly agree/50% agree/25% somewhat disagree** - The task instructions were clear to me (R)
		"Yes, that's what I said, if you follow this card, guidelines...." (1)	"Yes, I had some difficulty because of the zig-zag between pilot and ATCO, going back to pilot. And sometimes, right after you get a prompt as a pilot, you have another one as a pilot. You are supposed to interact right afterwards. But it is not very clear where and when". (1)	"I think the guidelines could be a little bit better, more ...concise, I think. I could understand it, but I think it could be a little bit better".(2)	**25% strongly agree/50% agree/25% disagree** - The information received was enough to guide me through the interaction (R)
Task appropriateness	"Sure". (3)	"Yes, yes, for sure".(3)	"Yes, I believe so, because the language was not a problem. Phraseology and plain English used in the test that I did, they are concise with what we do on a day-by-day basis".(3)	"Yes, I think so". (3)	**75% strongly agree/25% agree** - The task enabled me to demonstrate my ability to speak and understand the language used for RT communications (SA, V)
					75% strongly agree/25% agree - The language elicited by the task was appropriate to the communicative demands of pilots in international RT (V, SA)

					50% strongly agree/50% agree - The task was appropriate to measure 'Interactions' according to ICAO rating scale (V) **25% strongly agree/50% agree/25% somewhat agree** - The task was of an appropriate length and the room arrangement and equipment used was appropriate (P)
Engagement	"Affirmative".(3)	"Yes".(3)	"Talking about the language, and the speeches that we had, yes".(3)	"Yes, I did".(3)	**50% strongly agree/50% agree** - I felt as if I was actually engaged in real RT communications as I carried out the task (IA) **50% strongly agree/50% agree** - I was involved in the task in terms of field specific knowledge and language knowledge (IA) **75% strongly agree/25% agree** - The interlocutor played the role of the ATCO adequately (IA, R)
Add/ remove sth	"No, it is okay".(3)	"It's better if the interlocutor say again "Now imagine that the situation has changed and now that's another situation".(2)	"I just think that the script and the prompt has to be worked on and has to be a little clearer for the test-taker".(2)	"Maybe just add some more information. The way it was made was good, in my opinion".(2)	
Suggestions/ comments	"Very good, very appropriate".(3)	"It should be very good if he said "Now imagine that you are taxiing", it's better for the pilot to understand the situation. That's my only complaint, not complaint, my advice".(2)	"Maybe some kind of cues during the task might be necessary… maybe there could be like cue number 1, cue number 2, and if the test-taker doesn't take action, then the interlocutor could say "We are on cue number 5", so he could find himself on the script and go ahead and do what he needs to do".(2)	"Just liked this way".(3)	

	50% strongly agree/50% agree - The task is likely to cause a positive impact on pilots' language training (IMP)

Note. Magnitude Coding: (0) Very negative, (1) Moderately negative; (2) Moderately positive, (3) Very positive (R) Reliability, (V) Validity, (SA) Situation authenticity, (IA) Interactional authenticity, (IMP) Impact, (P) Practicality

In addition, I added to each Likert scale question a quality of good testing practice (Douglas, 2000): Reliability (R), Validity (V), Situational Authenticity (SA), Interactional Authenticity (IA), Impact (IMP), and Practicality (P), to inform my discussion.

A lot of convergence was noted in the test-takers' answers and comments, which indicated an overall positive response to the tasks. They all agreed that the role-play tasks elicited language appropriate to the communicative demands of pilots in international radiotelephony (V), reflected the TLU domain of aviation communications over the radio (SA), and enabled test-takers to feel as if they were actually engaged in real RT communications (IA). However, even though test-takers had a generally positive feeling during task performance (IMP), their perceptions on 'feeling confident' varied a bit. This might be related to the need for more straightforward instructions and guidance to move from one phase of flight to the subsequent one (R), reported by participants as a more negative side of the tasks. In fact, the Likert scale question related to the information received to guide the test-taker through the interaction displayed the lowest mean value (M=4.5, SD=1.73). This confirms the test-takers' comments during the interviews, who said the transitions were a bit confusing and at times they were not sure when to start their turn. In addition, although not addressed in the interviews, there was consensus in relation to the positive impact the tasks are likely to cause on pilots' language training

(IMP), but some divergence on their opinions regarding the length of the tasks, equipment and room arrangement (P).

Overall, from test-takers' perspectives, the role-play tasks reflect pilots' communicative needs in RT communication to a great extent, but still require some adjustments in terms of clearer instructions to guide them throughout the sequential phases of the flight.

9.5.2 Research question 3.4

RQ 3.4) What are the interlocutors' insights on the draft pilot tasks, and, from their perspective, to what extent do they feel confident in following the role-play instructions and perceive the appropriateness of the task?

In order to answer RQ 3.4, the same process of comparing and contrasting concurrent sources of evidence was followed. A similar joint display was built including responses from face-to-face semi-structured interviews with interlocutors, added to their respective Magnitude Coding, and percentage analysis of responses to the Likert scale questions (see Table 9.6). An integrated data interpretation was possible, as themes discussed with interlocutors were entered along the vertical axis, and reduced findings from the two sources of data were included along the horizontal axis. On top of that, a quality of good testing practice - Reliability (R), Validity (V), Situational Authenticity (SA), Interactional Authenticity (IA), Impact (IMP), and Practicality (P) (Douglas, 2000) – was incorporated to each Likert scale question.

Table 9.6. Interlocutors' perceptions from individual interviews and Likert scale questions

Themes	Face-to-face semi-structured interviews				Likert scale questions
	Task 1 - Morning (ATCO)	Task 2 - Morning (Pilot)	Task 1 - Afternoon (ATCO)	Task 2 - Afternoon (Pilot)	

General feeling	"I felt very well, I think it is something I have been doing, I did for many years, and I didn't..." (3) "But I practiced before, it was okay".(2)	"I felt nervous because we had, that's a role that I was not used to perform, to interact with the test-taker, so I was kind of nervous to follow the script. Even though I took notes of what I was supposed to do, I made marks on my, on the lines I was supposed to read and everything".(0)	" I felt very confident, it was quite easy, good to work with..." (3)	"It was okay".(2)	**25% strongly agree/50% agree/25% somewhat disagree -** I was confident during task administration (IMP)
Difficulties	"No. only those transitions that were not very natural, for example: He was at the gate, and then all of a sudden he was ready for take off. So, that's not so natural".(1)	"I think the difficult I'd say, how harsh the test is ...rating level 7 or 8, I think the test-taker should be very proficient to run through it".(1)	"Yeah, there was one moment when the pilot, he used different kind of vocabulary that I was not expecting. Then, I thought, well, what do I have to say now? But it was only in one situation".(2)	"But there is a lack of instructions, there was supposed to have at least one pilot interaction that was not written. I think there should have some kind of instruction like how long should I wait to play the ATIS, ah..." (1)	**25% agree/ 25% somewhat agree/ 25% disagree/25% strongly disagree/** - The task was difficult to administer (R)
Task appropriateness	"Yeah, I think it is good".(3)	"I think it was very, very good".(3)	"Perfect, yes, very good practice".(3)	"It was appropriate, I just consider that this task might be kind of difficult for the average test-taker that he normally, maybe he won't get the flow of the thing so fast. We need to have more instructions".(2)	**75% strongly agree/25% agree** - The task enabled the test-taker to demonstrate his/her ability to speak and understand the language used for RT (SA, V) **50% strongly agree/50% agree** - I felt the test-taker was engaged in real RT communications and also involved in the task in terms of field specific knowledge and language knowledge (IA) **100% agree -** The language elicited by the task was appropriate to the communicative demands of pilots in international RT (V, SA) **50% strongly agree/50% agree - The** task was appropriate

					to measure 'Interactions' according to ICAO rating scale (V) and was of an appropriate length (P) **25% strongly agree/50% agree/25% somewhat agree** - The room arrangement and equipment used were appropriate (P)
Script information	"Yes, it was. I think as an interlocutor, it was".(2)	" I think it was sufficient to guide me through, but I think it was a little confusing sometimes, because you know, you cannot predict 100% what the candidate is gonna say, so...there is not an evasive action...rsrsrs".(2)	"Yes, it was good enough".(3)	"Yes".(3)	**25% strongly agree/50% agree/25% somewhat agree** - The information received was enough to guide me through the interaction (R)
Add/ remove sth	"Well, I already did, because I changed the information on the taxi, because the verb "to clear" is not used anymore with taxi information. So, I used "Taxi approved to holding point".(3)	"I think it was good to assess the proficiency, I think there is no need to add anything else". (3)	"It should be trained more just to get some gaps, but I think that in general terms it was perfect, it is very good, because it is like the reality, like it really happens"(3).	"No, no, it was okay". (3)	
Suggestions/ comments	"Ah, I think it is ok. It is a very good task, so provides information that makes the pilots use the vocabulary, aviation vocabulary according to real life situations. I think it is good".(3)	"I suggest to ...try to somehow predict different path the interlocutor should go through in case the candidate gives different answers and make it possible to guide the interlocutor in a different way".(2)	"Yeah, I think so. Well, within the.....it is necessary to brief the pilot in advance. You have to tell the pilot what is going to happen at each step, before showing the picture, okay, because I am not seeing the pilot, especially if he is wearing the earphones or not, okay".(2)	"Just the same....should have more information, more instructions to the test-taker". (2)	**75% strongly agree/25% agree** - The task is likely to cause a positive impact on pilots' language training (IMP) **25% strongly agree/ 75% somewhat agree/** - I felt the task instructions were clear to the test-taker (R)
Interlocutors' training	"Yes, I think pilots they are going to feel some difficulty because the role of the ATCO is a little bit different. But I think they are going to adapt because they have English proficiency, they already know aviation English, I think only the	"I think .....uh....I don't know, I think the interlocutor were able to handle that without much training. I'd say not a training, but a rehearsal would be needed, just to get used, but we have all the tools to perform that kind of role".(2)	"Just training, the training is going to be enough. It is not necessary...Yeah, rehearsing, that's right. Practicing, after a couple of days you are going to be ready for this, okay, without any problems"(3).	"They need to have specific training on each of the test versions, besides that I think it is okay".(2)	(R)

	procedures on how to conduct the interview".(2)

Note. Magnitude Coding: (0) Very negative, (1) Moderately negative; (2) Moderately positive, (3) Very positive (R) Reliability, (V) Validity, (SA) Situation authenticity, (IA) Interactional authenticity, (IMP) Impact, (P) Practicality

Some differences in opinion could be noticed within the group of interlocutors, but there was not much divergence of findings from the two sources of data. First, interlocutors with a background as ATCOs felt very confident during task administration (IMP), whereas those who were pilots had a different experience, especially one who reported not being used to interacting with test-takers. Therefore, a variety of opinions emerged concerning their perceived difficulty in doing the task (R), which was confirmed by the lowest mean value of the Likert scale question related to this issue (M=3.00, SD=1.83) in contrast to the others. In addition, most interlocutors found that task instructions were not very clear to test-takers (R), requiring more information or a type of briefing beforehand. Nevertheless, interlocutors were very positive in relation to task appropriateness and all agreed that the language elicited by the task was appropriate to the communicative demands of pilots in international RT communications (V, SA), enabling test-takers to engage in real RT communications in terms of field specific knowledge and language knowledge (IA). Concerning practical issues, although not questioned during the interviews, interlocutors considered that the task was of an

appropriate length (P) but their opinions were a bit different in relation to room arrangement and equipment used (P). A final comment was made about the need to train interlocutors (R), mainly the SMEs who are pilots, in order to play the role of ATCOs effectively.

In sum, from the perspective of interlocutors, the task is, to a great extent, appropriate to enable the test-taker to demonstrate his/her ability to speak and understand the language used for RT communications. Yet, some of them reported not feeling confident enough to follow the role-play instructions and to deal with the co-constructed nature of the interaction.

9.5.3 Research question 3.5

RQ 3.5) What are the raters' insights on the draft pilot tasks?

Evidence to answer this question was obtained from the qualitative comments made by AETEs who acted as observers during the pilot testing of the tasks, i.e., raters of the Brazilian language proficiency test for pilots, comprising ELEs, SMEs pilots, and SMEs air traffic controllers. My discussion draws on commentaries from the three focus groups, which were coded according to the mind map presented in Section 9.3.4.3. In addition, the qualities of good testing practice - Reliability (R), Validity (V), Situational Authenticity (SA), Interactional Authenticity (IA), Impact (IMP), and Practicality (P) (Douglas, 2000) frame my comments, focusing on key issues to inform future decisions on task improvement. As Douglas (2000) acknowledges, reference to these qualities is "necessary to the development of what is essentially a measurement device which must display accepted standards for good measurement" (p. 251).

To begin with, a lot of conflicting opinions were given in relation to the appropriateness of the tasks to the target population (V), i.e., Brazilian airplane pilots. One participant mentioned that the role-plays would be suitable "only for experienced international pilots" (ELE 1 – Group 1); another stated that "a Level 3 with lots of experience would cope with that, on the other hand a pilot with very high level of English but no experience would not cope with that" (ELE 3 – Group 1). A pilot raised two sides of this issue, saying that "pilots who just graduated from a flight school, who have little knowledge, they are not gonna be able to do this. It is expected of them to do this, because it's a day-by-day thing that you need to do all over the world, not only here", and adding that "if they [pilots] want to get the ICAO Level 4 so he can fly abroad, he needs to know how to do this. So, it's complicated, it's complicated (SME 6 – Group 2). Another pilot contributed to this discussion by stating that "somehow ICAO requirements are for both experienced and non-experienced pilots, it doesn't matter if they are airline pilots or general aviation pilots" (SME 2 – Group 1). An ATCO also corroborated this view mentioning that no matter if "pilots [are] dusting crops or flying an airliner, and he wants to be Level 4 he should know this kind of things, he should know this vocabulary, it's important for a Level 4 to know, for an ICAO Level 4 to know" (SME 3 – Group 1).

These comments lead to the discussion of another crucial, and yet controversial, topic in Aviation English test design, namely, the construct to be measured (V). Some participants in the morning group reacted negatively to the tasks in terms of what should be assessed, whereas the other two groups had a more positive perspective on that. For example, a pilot stated that "the problem is that [with] the design of the task, we are also assessing the aviation knowledge of the candidate doing that way (SME 1 – Group 1),

and another pilot added that “My first impression is that we stopped evaluating English and somehow like ah....we went [ ] to a more technical version of the exam” (SME 2 – Group 1). Even more debatable was the opinion of another pilot while referring to the production of plain English in contrast to standard phraseology: “We have very experienced pilots here, and they had to ask "Say again" a couple of times, because they didn't even take the whole clearance. I mean, we don't want to check that, we don't want to check if the pilot is able to write down a clearance and read back the clearance properly. Because, actually we are not getting anything from the pilot, correct?”(SME 4 – Group 1). However, it is important to emphasize that the language required for effective RT communication does not rely solely on plain English, but includes a range of standard expressions that are specific to this context of use and communicative strategies to cope with breakdowns in communication. Yet, the assessment of test-takers’ production in radiotelephony-based tasks should not be centered on ‘procedural appropriateness or technical correctness’, according to Doc 9835 (ICAO, 2010).

Excerpt 3 below discloses more evidence of this controversial issue:

Excerpt 3

SME 1(Group 1): I think this task would be suitable for maybe a Level 6 task, the role-play.

ELE 1(Group 1): Why would that be suitable for Level 6? I would say it could be the opposite, because in Level 6 we are testing idiomatic, and fluency, you know, not just the voice....it's completely different for me. A Level 6 technical ability to fly an airplane, or to manage the control tower, yes. But a Level 6 to speak English, I am not so sure.

All these comments reveal that there is still a lot to be done in the aviation industry in relation to achieving a clearer and common understanding of the construct to be measured (V). To what extent are the descriptors in the Level 6 scale relevant to the specific domain of RT communications in a multicultural context? What should we include in a test for pilots so that we can make inferences about their ability to communicate effectively with ATCOs in international RT, in routine and non-routine situations? As discussed in previous chapters, in the ICAO guidelines (2010), background knowledge, or technical knowledge, is not considered part of the construct in the specific assessment of "the ability to speak and understand the language used for radiotelephony communications" (p. 4-4). However, findings from Phase 2 of this MM study suggested that background knowledge is a crucial component of the construct in the Aviation English domain. This is supported by Douglas (2000), who emphasizes that background knowledge is an integral part of the specific purpose language ability, as appointed by Knoch (2014), who argues that in Aviation English testing, "because of the prominence of the use of standard phraseology [ ] the testing of language and technical knowledge cannot and should not be separated" (p. 85). In the third focus group discussion, a comment was made in that direction, as a suggestion to include the elicitation of more plain language while keeping the format of a radiotelephony-based task:

> I think the best thing would be to include more plain English in this task, because that's what we are testing, plain English in ....in radiotelephony interactions. ...mixed with phraseology. I don't how, that's challenging, but I think that would be perfect. If we could develop tasks that elicit a lot of plain English in the...live interaction. (ELE 3 – Group 3)

Still on the topic of test construct, in relation to the language functions and behaviors elicited by the tasks (V), there was a general agreement within and across groups, although much less discussed by participants. An ELE stated that

> concerning this language functions list that you gave us [ ] the task did elicit the ones that we have marked as number 4, so for example, read back, give repetition, request clarification, give clarification, announce readiness and availability, announce a problem, state the expected moment/duration of an action/event. (ELE 4 – Group 1)

In relation to the behaviors indicative of effective communication, included in the *Indicator Checklist*, many comments confirmed that "most of them" were noticed in the task, such as "they were polite, the way it should be, they were concise in their messages" (ELE 2 – Group 1), and "cooperating to each other" (SME 2 – Group 1g). However, one ATCO pointed out that "even the most impolite guy will have good behavior for a short time. He didn't have enough time to lose his patience" (SME 3 – Group 1), to which a pilot replied that "we shouldn't give them [test-takers] a hard time like just to see if they are going to be polite or not. [ ] I don't think it is fair to evaluate something like that" (SME 4 – Group 1). These last comments direct us to the issue of politeness in radiotelephony, which also raised distinct viewpoints, especially in the afternoon group, as shown in Excerpt 4 below:

> Excerpt 4
>
> SME 5 (Group 2): But it is not relevant, politeness. Not at all.
>
> SME 4 (Group 2): Yes, but not for the test.
>
> SME 5 (Group 2): Not for the real life, in radiotelephony. It's not relevant.

ELE 1 (Group 2): No, I don't agree, because yesterday in the workshop, there was a situation in that they were evaluating that the pilot was too bossy, so this is a matter of politeness somehow, and this was a matter that we spoke about, that the pilot was too bossy in the communication. Maybe this could be…

SME 6 (Group 2): If we take into consideration strict rules of communication in RT, politeness is not used in a day-by-day basis, should not be used in a day-by-day basis. So, when you enter a control center or a control tower and say "Airport A, good morning, this is ANAC 123 bla bla bla ", the good morning, please, could you, and this and that, that does not exist. [ ] We are polite, because we are polite people, we are educated, but this, from the strict point of view, this should not be used.

Indeed, according to radiotelephony manuals (e.g. ICAO, 2007), the use of markers of politeness, such as greetings, signoffs, thanking, etc, is not considered standard ICAO phraseology. However, although not standard, these expressions are common in pilot-ATCO interactions, and as Moder (2013) explains, other researchers (see Sänne, 1999 and Wyss – Bühlmann, 2005) reported that "controllers have positive attitudes toward such friendly, personal contact with flight crews" (p. 238) and also that the use of politeness markers and mitigated questions "are used strategically to clarify meanings, solve problems of understanding, and contribute to cooperative exchanges" (p.238). As discussed in the analysis of the six RT communication scenarios in Chapter 7 and mentioned by ELE 1 in Excerpt 4, sometimes it is not a matter of using a politeness marker or expression, but displaying an attitude and tone of voice that convey the message of being professional and polite. In the example given, the fact that the pilot

'was too bossy' may be perceived as impoliteness by his interlocutor, leading to less effective communications. On the other hand, Clark (2017) highlights that politeness markers may be used in local languages, reducing shared interpretation by all participants in communication. She adds that "participants in interaction can feel insulted if something goes wrong, for example if a politeness marker is misinterpreted. Feelings of insult and upset can be distracting, and reduce situational awareness, creating a threat to aviation safety" (p. 73). That said, it is apparent that controversy in relation to politeness in RT communications is not limited to the participants in this study, but instead is a debate in the whole industry. Moder (2013) concludes that "the extent to which safety may be aided or impeded by features coding relational aspects of language, in particular markers of politeness and cooperation, remains an open question" (p. 238).

In terms of situational authenticity (SA) of the tasks, many comments stated that they sounded really authentic. One pilot mentioned that "They are performing as pilots. This is real close to the real thing" (SME 1 – Group 1) and another pilot added that "But the way it was, it's like a real situation" (SME 7 – Group 2). In general, the ELEs also agreed with that, and one reported that "It [the task] reflects real RT communication" (ELE 3 – Group 3). However, one of the issues raised was the fact that in the same task the test-taker had to deal with two or three unexpected situations or problems, which in real life would not happen in the same flight. In this respect, divergent comments were made. From one side, and ELE mentioned that "The test covered two problems, one situation that was a heart attack and it also had a runway incursion, right? I think it's too much to have two unexpected events in only one situation" (ELE 5 – Group 1), which was corroborated by a pilot who stated that this would be "Far away from reality. A bad

luck pilot" (SME 1 – Group 1). On the other side, an opposing comment was made by an ELE, who said "I disagree with that, because this is a test, because it's not real life [ ] They are not flying but all situations could happen, we are just joining them for the purpose of testing!" (ELE 3 – Group 3).

Further, it is important to cite participants' observations that substantiate the involvement of test-takers with the task, in terms of field-specific knowledge and language knowledge, and how its characteristics may promote the engagement of test-takers in real RT communications (IA). One ELE noticed that "intonation patterns are completely different in this activity than they are in any other part of the test or in everyday speech. That's perhaps because we were assessing professionals [ ] the sound was so good, because it sounded like a real thing" (ELE 2 – Group 1). In this respect, a pilot responded that "there is some intonation while you speak over the radio [ ] I don't know how we get that, we get used to it, nobody teaches us 'You shall speak that way"' (SME 1 – Group 1), and an ATCO replied that this way of speaking is required "for security and conciseness....you have to be precise and concise" (SME 5 – Group 1). In addition, previous comments selected to emphasize the need of background knowledge to accomplish the task also reinforce the quality of interactional authenticity (IA). As O'Sullivan (2012) points out "The extent to which LSP assessment developers include the test taker's background knowledge of the target domain in their construct definition is a key element of the resulting assessment's interactional authenticity, since it is this aspect of a test task that makes it specific in the first place" (p. 74).

Considering the assessment criteria required to rate the tasks (V), participants expressed their opinions but also raised some concerns. First, in the focus group

conducted in the morning, while referring to the ICAO rating scales one pilot stated that "Actually, I think it's just interactions, maybe comprehension" (SME 4 – Group 1), and another mentioned that "only a little vocabulary and structure, because it's like a broken discourse, broken speech" (SME 2 – Group 1). When asked whether the behaviors in the checklist would be useful as assessment criteria, most of the responses were positive, and I selected some that illustrate that: "Yeah, not only useful, but we need that. The thing is ICAO should have done that in their rating scale" (ELE 3 – Group 3). Another ELE reacted positively by saying:

> I think it will be very useful, because the way the rating scale is ...it's impossible to assess what you have written in this list, this checklist of attitude, awareness, whatever, everything, so the scale does not cover it now, and if we need to check their ability to communicate under these new rules, then it will be very important to have that for us to mark. (ELE 2 – Group 2)

In a more comprehensive way, another comment emphasized the value of using the behaviors as part of the assessment criteria, which would potentially standardize the rating of task performances and, as a consequence, increase the reliability of test scores (R):

> I think the raters should have another script like ok, this is number ...wait… unexpected situation number 1, and number 2, all of them, the unintelligible piece of information, asking to repeat, everything should be pointed out and okay, how many of these situations did the candidate manage? And maybe differentiate them in terms of behavior, this behavior number 2 is this one here: accommodating to

> different accents. And then it could be easier for them to rate, they have this very clear ....not only holistically he interacted well. (ELE 3 – Group 3)

Regarding participants' worries about marking test-takers' performance, one issue raised refers to the need to differentiate between one performance to the other:

> As they said probably a Level 3 pilot would accomplish the task as well as a Level 4 or Level 5. So, it's difficult to differentiate one performance from the other and so maybe from that ....we should take into consideration like how much the candidate required from the interlocutor, and from that perspective you do need a very skillful interlocutor with content knowledge and background knowledge to be able to put him back on track. So, there would be something that would differentiate one behavior from ... one performance from the other, also maybe the number of repetitions, and I mean since we would have to raise or maybe to brainstorm some things that would differentiate one performance from the other. (ELE 4 – Group 1)

From the side of SMEs, one pilot raised questions of reliability, stating the following: "Like, this I think, is gonna be more subjective, you know, we would have more problems in order to actually rate, and accommodate [ ] I mean, maybe it could be possible, but it's gonna be more difficult for us, because now we have to be calibrated first of all, to rate the candidate" (SME 4 – Group 1). These comments indicate potential challenges raters may face to identify the behaviors and differentiate performances for rating purposes, which would require further training and standardization sessions before the main trial. Similarly, a pilot study of a purpose-built ELF assessment task (Harding & McNamara, 2017) also reported on lack of clarity about "what judges were attending to

in the performances" and the difficulties raters encountered to "capture these [ELF-like competences] for scoring/rating purposes" (p. 578).

The effect of interlocutors' behavior on test-takers' performance (R) and possible difficulties the SMEs may face during the role-play tasks were mentioned in all focus groups. However, the need for appropriate training to perform the role of ATCOs was a controversial topic, even among SMEs. An ATCO said that pilots should receive more training in "phraseology, because you know the pilot's part and you have to learn the controller's part.[ ] The pausing of the voice, the intonation, giving orders, right? Asking for clarification, etc" (SME 5 – Group 1). In the same lines, an ELE stated that "they [SMEs – pilots] will need to train as if ...like a two day course as an air traffic controller, I guess" (ELE 1 – Group 1), while a pilot had a more simplistic viewpoint, mentioning that "I don't think it demands training, but rehearsal", as they have a script to follow. However, another ELE raised concerns related to the nature of live interactions and her view on interlocutors' training:

> They [SMEs] need to be very well trained, they need to know ah ... we need to predict what the candidate will say, so lots of possibilities … Sometimes it is not possible but a lot of possibilities and teach them, so if the candidate doesn't ask for repetition, what should you do? We need to train all the possibilities we can imagine and tell them, if something else happens you are free to improvise. (ELE 3 .– Group 3)

As Douglas (2000) explained, "test takers have minds of their own, and may decide to respond in a way we simply did not anticipate; or perhaps they may not pay as close attention to the information we provided as we intend they should, or the information

may not have been as clear as we thought it was" (p. 62). Actually, this happened in the pilot testing of the tasks and demonstrates the need for skilled interlocutors, trained to deal with the unexpected in the most standardized way as possible. As was reported in the study on the OET extended speaking assessment criteria (O'Hagan, Pill & Zhang, 2016), interlocutors have a crucial role in allowing the candidate to "demonstrate some of the skills/behaviours associated with the new criteria … place[ing] a greater burden on interlocutor consistency which would need to be addressed in training to ensure candidates have fair and equal opportunities to demonstrate the associated skills/behaviours" (p. 210).

Relevant to mention at this point and worth of further considerations is the duration of the same role-play task in different administrations (R). For example, Task 1 in the morning (Group 1) took 15min 14s, whereas the same task in the afternoon (Group 2) took only 10min 29s. Interestingly, in both administrations the interlocutors were ATCOs, but test-takers, although very experienced, displayed different levels of understanding of the information provided in their role-play cards. As they reported in the interviews, guidelines throughout the task should be made clearer. Task 2, on the other hand, had less variation in length. In the morning it took 07min 41s, while in the afternoon its duration was of 08min 05s. In this case, both interlocutors were SME – pilots, playing the role of ATCOs.

In regards to task length, participants had different viewpoints. In general, many of them reported the tasks were too long, but two SMEs, a pilot and ATCO, had distinct perceptions. Excerpt 5 below expresses their opinion, compared to the perspective of an ELE:

Excerpt 5

ELE 2 (Group 2): No, too long. I am not a pilot, I had the impression it was too long. But, a real pilot is saying that it is ok.

SME 5 (Group 2): I think it was fair enough.

SME 7 (Group 2): For me it was okay, no problem.

However, task length leads to the discussion of another quality of good testing practice, practicality (P), which is related to the resources required to design, operationalize and to administer the task. One ELE raised a concern about the need for more time to listen to and rate performances, while other mentioned the challenges to generate parallel tasks with a variety of built-in complications: "Because actually you have to create many, many different situations so that we can have many versions generated by the computer, you know, by the system" (ELE 2 – Group 3). As in the current test for Brazilian pilots it is the ELE who interacts with the test-taker, the possibility of also having an SME interacting in the role-play task was debated:

> Another point that we should take into account is the cost-benefit of including this in the test. Because as the test is now, it is the interlocutor-candidate, and to add a second interlocutor… it's disruptive, so it has to be very, very carefully placed within the test for it to be actually worthwhile in generating more important information. (ELE 2 – Group 1)

Following along those lines, a lot of comments were made about equipment and room arrangement. Some participants were in favor of using the visual barrier, affirming that "when interacting as a pilot the barrier is good, makes [the task] more realistic" (SME 5 – Group 1), while others not so much, stating that "maybe you don't need to have like a

physical barrier over there, I think it creates like a gap" (SME 6 – Group 2). On the other hand, there was consensus on the use of headsets, for both test-taker and interlocutor, which may increase the complexity of the test setting.

Few comments were made about the impact this type of role-play task would have on test-takers, teaching practices and society (IMP). One ELE expressed her opinion referring to how Brazilian pilots would feel performing this task in a real test: "I think they would feel much better doing this [ ] I think their feedback would be very positive: 'Ah, I think now ANAC is really testing Aviation English'. That's my feeling" (ELE 3 – Group 3). Another, also mentioned a positive impact the task would cause, this time on teachers and their educational practices: "But then we have the washback effect, the trainers shape their teaching to.... and I think this is enough to increase the standards of our test and of the language in general" (ELE 5 – Group 1). Further, an SME made a comment about the impact that decisions based on test scores have on society: "but we have to show him [test-taker] that, if I am gonna give you a Level 4 for you to be Operational, it means you are operational to go to Paraguay and you are operational to go to JFK. So, that's an important issue" (SME 6 – Group 2).

Finally, some suggestions were presented to improve the overall quality of the task, such as to give the test-taker a good briefing before the task and some planning time. Others were proposed as a way to adapt the role-play tasks to the existing format of the Brazilian test, by having the ELE act together with the SME, presenting the situations and the visual input throughout the task:

> I think the ELE should be by his side, and with her script and ...the pilot is just worried about communicating with the SME and then there is a picture, she just

> says ‘this situation is happening to you right now’, so he doesn't get worried about ‘Oh my God, take turns, talking, and’ ...the ELE is there to support him. (ELE 3 – Group 3)

Lastly, an important recommendation for future trials: “testing different candidates with different levels, I think we have to be sure to get low level linguistic competence with lots of experience, and then a pilot, the other way round, on the opposite end, to make quite sure that there is, the language plays an important part and not just the experience” (ELE 2 – Group 1).

As a way to summarize findings from the focus group discussions, Table 9.7 was built with the number of coding references for each sub-category per focus group, including color coding for Magnitude codes, in order to visualize the balance between positive and negative comments across groups and the overall evaluative balance within the whole dataset.

The red color was attributed to cells in which the number of very negative (0) + moderately negative (1) comments outweighed the number of moderately positive (2) + very positive (3) comments (e.g., the sub-category ‘Target population’ received in Focus Group 1 three very negative (0) and four moderately negative (1) comments, totalizing seven on the negative side, as compared to zero moderately positive (2) and three very positive (3) comments, totalizing three on the positive side). When the opposite occurred, the green color was used (e.g., the same sub-category ‘Target population’ received in Focus Group 3 one very negative (0) and zero moderately negative (1) comments, totalizing one on the negative side, as compared to zero moderately positive (2) and three

very positive (3) comments, totalizing three on the positive side). When the negative side was equal to the positive side, the yellow color was applied to the cell.

Table 9.7. Coding references per sub-category and focus group, including results for Magnitude codes

Categories	Sub-categories	Focus group 1		Focus group 2		Focus group 3	
		# Coding references	Magnitude coding	# Coding references	Magnitude coding	# Coding references	Magnitude coding
Appropriateness	**Target population**	10	0=3; 1=4; 2=0; 3=3	6	0=0; 1=5; 2=1; 3=0	4	0=1; 1=0; 2=0; 3=3
	Task length	4	0=4; 1=0; 2=0; 3=0	7	0=2; 1=2; 2=2; 3=1	6	0=1; 1=3; 2=0; 3=2
	Room arrangement	10	0=4; 1=4; 2=1; 3=1	14	0=3; 1=4; 2=4; 3=3	9	0=2; 1=1; 2=4; 3=2
Construct	**Language functions**	1	0=0; 1=0; 2=0; 3=1	7	0=0; 1=0; 2=3; 3=4	3	0=0; 1=0; 2=1; 3=2
	Behaviors	5	0=0; 1=1; 2=2; 3=2	20	0=6; 1=3; 2=3; 3=8	4	0=0; 1=0; 2=0; 3=4
	Aviation English	26	0=14, 1=1; 2=5; 3=6	8	0=2; 1=0; 2=2; 3=4	22	0=0; 1=5; 2=6; 3=11
Challenges	**Test-takers**	14	0=10; 1=4; 2=0; 3=0	7	0=5; 1=2; 2=0; 3=0	7	0=3; 1=4; 2=0; 3=0
	Interlocutors	8	0=6; 1=2; 2=0; 3=0	4	0=0; 1=4; 2=0; 3=0	16	0=7; 1=9; 2=0; 3=0
Authenticity	**Close to the real thing**	15	0=0; 1=0; 2=5; 3=10	4	0=0; 1=0; 2=1; 3=3	9	0=0; 1=0; 2=3; 3=6
	Far from reality	5	0=3; 1=2; 2=0; 3=0	1	0=0; 1=1; 2=0; 3=0	5	0=1; 1=4; 2=0; 3=0
Rating	**Assessment criteria**	12	0=0; 1=2; 2=7; 3=3	14	0=0; 1=2; 2=4; 3=8	18	0=2; 1=2; 2=7; 3=7
	ELE rater x SME rater	7	0=4; 1=1; 2=0; 3=2	0		0	
	Concerns	10	0=8; 1=2; 2=0; 3=0	1	0=0; 1=1; 2=0; 3=0	1	0=0; 1=1; 2=0; 3=0
Suggestions	**Task improvement**	7	0=0; 1=1; 2=5; 3=1	17	0=0; 1=1; 2=12; 3=4	36	0=0; 1=5; 2=29; 3=2
	Current test	11	0=0; 1=2; 2=4; 3=5	4	0=0; 1=0; 2=3; 3=1	15	0=4; 1=3; 2=7; 3=1
	Next steps	3	0=0; 1=1; 2=2; 3=0	3	0=0; 1=0; 2=2; 3=1	2	0=0; 1=0; 2=2; 3=0
	Interlocutors' training	22	0=0; 1=4; 2=10; 3=8	6	0=0; 1=0; 2=3;3=3	11	0=0; 1=8; 2=3; 3=0

Note. Magnitude coding: 0= Very negative, 1= Moderately negative, 2= Moderately positive, 3= Very positive

Summary:	Positive	27
	Neutral	1
	Missing	2
	Negative	21
	Total # of cells	**51**

Despite the divergence of observers' perceptions noticed in some sub-categories, findings reveal the predominance of positive commentaries in relation to the draft tasks, as can be depicted from the color coding summary.

9.5.4 Research question 3.6

RQ 3.6) What language functions and behaviors indicative of effective communication arise during test-taker's performance?

In order to answer this question, a qualitative analysis of transcripts from actual task performances was conducted to identify the language functions produced by test-takers and interlocutors in the four task administrations, as well as the behaviors indicative of effective communication. As O'Sullivan (2012) advises, "in LSP assessment, the test developer should identify the range of functions expected in performance within a specific domain, ensure that a representative sample of these are elicited by the proposed tasks at the design and specification phase of test development, and then check from actual performances whether those predictions have been supported (p. 82). These results will later be compared to other sources of data.

First, considering the production of language functions, not all the functions from the list of communicative language functions associated with aviation (ICAO, 2010) were produced during the role-play tasks. This was expected, as the list is comprehensive and encompasses functions used in a variety of both common and uncommon situations. It is also the case that in a task or a test the construct is 'sampled' as much as possible in relation to the table of specifications (Cheng & Fox, 2017). As long as the 'sample' is adequate for making inferences from performances, the language tester is doing his/her job. In research, what makes a sample 'adequate' is that it is arguably representative of

the population of interest. This allows us to infer on the basis of the sample to the population as a whole. In the case of sampling a table of specifications, we look for evidence that enough of the construct has been operationalized by the test to infer from performances on the test to behaviours in the domain of interest. It is again a matter of evidence and argument. However, data from the transcripts indicate that many language functions used in RT communications were elicited throughout the tasks. In order to get an idea of how the four groups of language functions were represented in the task performances, the range of functions from each group produced by test-takers and interlocutors is reported in absolute numbers and percentages. From the first group, which covers functions directed towards triggering actions, 04 types of functions out of 28 were used (14.3%); from the second group, sharing information, 19 out of 53 (35.8%); from the third group, management of the pilot-controller relation, 03 out of 13 (23.1%); and from the last group, management of the dialogue, 09 out of 21 (42.8%). Evidence of a wider range of functions related to sharing information and managing the dialogue confirm their relevance to achieve the communicative goals in RT communication and also indicate that negotiation was necessary during the role-play tasks. Some examples are shown in Excerpt 6 and Excerpt 7 below, with language functions highlighted in red, which do not intend to distinguish between instances of standard phraseology and plain English:

> Excerpt 6 – Task 1 (afternoon)
>
> P: [Airport A] (name addressee), [ANAC 123, stand 21] (give information), [cleared to push … ah request push back and start up] (self-correct and request permission/approval).

I: [ANAC 123] (name addressee), [good morning] (greet) [due to traffic] (give reasons), [your take off time in 30 minutes] (state the expected moment of an action). [Push and start at your discretion] (suggest a course of action). [Say your intentions] (ask about intentions).

P: [So, just to confirm, the expected delay 30 minutes?] (request confirmation)

I: [That's right, confirm] (give confirmation).

P: [Okay] (acknowledge), [ANAC 123, request to maintain position] (request permission) [and waiting for push back and start up clearance in about 30 minutes. Standing by, ANAC 123] (state intentions).

Excerpt 7 – Task 2 (afternoon)

P: [Airport B approach] (name addressee), ANAC 123, [we have information Charlie] (give information).

I: [ANAC123] (name addressee), [roger] (acknowledge). [Descend to flight FL 060, fly direct ZULU VOR, altimeter setting… xrixri (intentionally unintelligible), expect YANKEE ILS approach to runway 20, report 20 miles out] (give permission/approval).

P: [ANAC 123, cleared to descend to FL 060, ah....VOR] (read back) [… can you say again the information, please?] (request repetition)

I: [ANAC123] (name addressee), [roger] (acknowledge). [Descend to flight FL 060, fly direct ZULU VOR, altimeter setting… xrixri (intentionally unintelligible), expect YANKEE ILS approach to runway 20, report 20 miles out] (give repetition).

P: [Okay] (acknowledge), [ANAC 123, cleared to descend to FL 060, direct to ZULU VOR, ILS Yankee, for runway 20, I will call you 20 miles out] (read back). [Can you confirm the altimeter setting, please?] (request confirmation)

I: [ANAC 123] (name addressee), [I say again, altimeter setting 1019] (give confirmation).

P: [1019, ANAC 123] (read back).

It is important to emphasize that a lot of repetition is expected in the communicative loop of RT communications, including read back and hear back, as a form of redundancy to confirm that crucial information has been correctly understood. However, the use of negotiation and clarification strategies to repair breakdowns in communication is vital to ensure clear and safe radio exchanges.

A summary of results from the analysis of language functions is presented in Table 9.8, totalizing the number of functions produced in each task administration, by test-takers and interlocutors separately, and also providing an overview of the language functions that were used most frequently across administrations. Only the functions detected in the transcripts were included in this table

Table 9.8. Analysis of language functions from actual task performances

Doc 9835 Appendix B	Language Functions	Task 1				Task 2				Total
		Morning		Afternoon		Morning		Afternoon		
		Pilot	ATCO	Pilot	ATCO	Pilot	ATCO	Pilot	ATCO	
1.1.1	Give an order		8		5		3		3	**19**
1.1.6	Announce compliance with an order	2		1		1		1		**5**
1.2.2	Agree to act	1								**1**
1.3.3	Suggest a course of action		1		1					**2**
1.4.1	Request permission/approval	3		5				2		**10**
1.4.2	Give permission/approval		9		6		6		6	**27**
2.1.1	Request information		1	1	1					**3**
2.1.2	Give information	5	2	2	3	2	1	2	1	**18**
2.1.4	Describe a state	1		1		2				**4**
2.1.5	Describe a changed state	1				1				**2**
2.1.7	Describe an action in progress	6		4		1		2		**13**
2.1.11	Describe the source of a problem	1		1						**2**
2.1.17	Ask about readiness/availability		3		2					**5**
2.1.18	Announce readiness/availability	6		3						**9**
2.1.20	Give reasons	3		2	1	1		2		**9**
2.1.24	Announce a problem	2		2						**4**
2.2.1	Announce an expected action/event	3		1						**4**
2.2.3	State the expected moment of an action		1	1						**2**
2.2.6	Ask about intentions		1		1					**2**
2.2.7	State intentions	5		2						**7**
2.3.1	Announce a completed action having an effect on the present			1						**1**
2.3.2	Announce a change			1		1		1		**3**
2.4.4	Give a report	1		1						**2**
2.4.6	Describe a sequence of past actions	1								**1**
2.5.2	State necessity	2				2				**4**
2.6.1	Ask about feasibility/capacity	1	1		1					**3**
2.6.2	Announce feasibility/capacity		1	1	1					**3**
2.6.3	Announce unfeasibility/incapacity	1								**1**
3.1	Greet/take leave	2	1	2	1					**6**
3.2	Respond to greeting/leave-taking		2							**2**
3.3	Thank	3				2				**5**
4.1	Name adressee (s)	11	20	9	16		12	7	12	**87**
4.2	Self-correct			1	1			2		**4**
4.7	Check understanding		2				1			**3**
4.6	Read back	10		8		10		10		**38**
4.8	Acknowledge	14	6	13	3	7	4	3	4	**54**
	Invitation to proceed with transmission*				4					**4**
	Declare understanding*			1						**1**
4.11	Correct a misunderstanding		1							**1**
4.12	Request repetition	4		2		1		2		**9**
4.13	Give repetition		3		2		3		2	**10**
4.14	Request confirmation	1		1		2		1	1	**6**
4.15	Give confirmation	3	3	1	1	3	1	1	1	**14**
	Totals:	**93**	**66**	**68**	**50**	**36**	**31**	**36**	**30**	**410**

Note. * Language functions not included in Doc 9835, Appendix B list

Considering the variety of functions used by pilots (test-takers) and ATCOs (interlocutors), pilots in Task 1, morning and afternoon, produced 26 different language functions, a wider range than what pilots produced in Task 2, i.e., 14 different functions in the morning and 13 in the afternoon. Regarding ATCOs, these numbers decreased, ranging from 18-17 in both administrations of Task 1 to an even lower number in Task 2, i.e., eight different functions.

In addition, from the last column in Table 9.8 we can identify the language functions that were produced most frequently during the four task administrations. Starting with the one most used, we can cite *name addressee, acknowledge, read back, give permission/approval, give an order, and give information.* Further, a comparison of the number of language functions produced by test-takers in Task 1 reveals a great difference between the morning group (i.e., a total of 93 functions) as opposed to the afternoon group (i.e., a total of 68 functions), whereas in Task 2 the number was the same, i.e., 36 for both groups. This may be explained by the differences in Task 1 length, reported for the two administrations, i.e., 15min 14s (morning) and 10min 29s (afternoon).

Secondly, regarding the list of behaviors indicative of effective communication (Appendix Y), findings from the transcripts of actual task performances suggest that many of them were elicited throughout the tasks. In Task 1 and Task 2, different behaviors were detected in the same utterance, and some of them appeared repeatedly throughout the role-plays, such as: *complying with the rules of use that characterize the domain (e.g., use of phraseology, read back/hear back, etc)* (1)[30], *demonstrating a*

[30] These numbers in parentheses refer to the number of each behavior in the *Indicator Checklist* (Appendix Y).

professional attitude and tone (2), *communicating effectively in routine and in unpredictable situations* (3), *producing and recognizing the language functions used in RT* (5), *complying with the safety-critical requirements of intelligibility* (8), *avoiding the use of jargon, idioms, slang, and colloquialism* (9), *eliminating idioms, cultural references and syntactic complexity from speech* (16), *demonstrating a shared responsibility for successful communication* (17), and *using an appropriate participation framework* (20). Excerpt 8 below illustrates an exchange in which there is evidence of these behaviors but also indication of the use of plain language (highlighted in yellow), defined as "the spontaneous, creative and non-coded use of a given natural language" (ICAO, 2010, p. 3-5). However, the first instance of plain language, i.e., "Okay", could be classified as the incorrect use of the standard expression "Roger"; the second suggests a less precise way to share information in order to meet the communicative requirements of an unexpected situation; and the third, i.e., "Thank you for the assistance" is associated with the use of politeness markers.

Excerpt 8 – Task 1 (morning)

P: [Okay] (indicators 4, 5). [Cleared to return to the gate via .....ah....BRAVO, to gate 21 (*self-repair), via BRAVO and HOTEL] (indicators 1, 2, 3, 5, 8, 9, 16, 17, 20). [The medical emergency looks like a passenger is having a heart attack or something like that. Can you give us some assistance upon arriving at the gate?] (indicators 2, 3, 4, 5, 7, 8, 9, 17, 20)

I: [ANAC 123, proceed to the gate, medical assistance will be provided] (indi cators 1, 2, 3, 5, 8, 9, 16, 17, 20, 21)

P: [Thank you for the assistance] (indicators 4, 5, 12).

Another example of a non-routine situation presented in Excerpt 9 provides evidence of the need to share and negotiate critical information, as well as to demonstrate tolerance and collaborative efforts:

Excerpt 9 – Task 1 (morning)

P: [Tower, this is ANAC 123. We are rejecting take off] (indicators 1, 2, 3, 5, 8, 9, 16, 17, 20) [due to another aircraft crossing the runway at the same time] (indicators 2, 4, 5, 8, 9, 16, 20)

I: [Roger, ANAC 123] (indicators 1, 2, 3, 5, 8, 9, 16, 17, 20) [Do you have enough space to backtrack on the runway and return to the threshold?] (indicators 2, 3, 4, 5, 7, 8, 9, 16, 17, 20, 21)

P: [No, negative] (indicators 1, 2, 3, 5, 8, 9, 16, 17, 20). [We need some more radius to do the turn than is available] (indicators 2, 3, 4, 5, 7, 8, 9, 16, 17, 20). [Request clearance to line up again] (indicators 1, 2, 3, 5, 8, 9, 16, 17, 20).

In addition to instances of plain language, evidence of skills to deal adequately with apparent misunderstandings and to accommodate the constraints of the context and the perceived ability of the hearer are shown in Excerpt 10 below:

Excerpt 10 – Task 2 (morning)

I: [ANAC123, roger. Descend to FL 060, fly direct ZULU VOR, altimeter setting… hurm (intentionally unintelligible), expect YANKEE ILS approach to runway 20, report 20 miles out] (indicators 1, 5, 9, 16, 20)

P: [Roger, cleared to level 050, confirm?] (indicators 1, 2, 3, 5, 8, 9, 16, 17, 19, 20).

I: [Yeah] (indicator 4, 5), [descend to FL 060] (indicators 1, 5, 16, 20)

P: [Roger, descending to FL 060 and expecting Zulu VOR, ILS Yankee, and altimeter setting…Say again [please] (indicators 4, 12), the altimeter, for ANAC 123?] (indicators 1, 2, 3, 5, 8, 9, 10, 16, 17, 19, 20, 21).

I: [ANAC 123, I say again, altimeter setting 1019] (indicators 1, 2, 3, 5, 8, 9, 16, 17, 18, 19, 20, 21)

P: [Roger, 1019] (indicators 1, 2, 3, 5, 8, 9, 16, 17, 20), [thank you] (indicators 4, 5, 12).

An overview of findings from the analysis of behaviors indicative of effective communication is presented in Table 9. 9, organized by task, administration and role-play participant, i.e., test-taker (pilot) and interlocutor (ATCO) separately. Figures in the table refer to the number of times a specific behavior was elicited by the task. However, repeated behaviors captured in the same utterance were counted only once.

Figures in Table 9.9 indicate that in the four task administrations there was evidence of more behaviors of effective communication in the performance of pilots (test-takers) than in their partners ATCOs (interlocutors). This finding was expected, as the task was built to include complications requiring the test-taker to negotiate meaning and deal with many unexpected situations. However, low values for the indicators of ELF competence and also for the Intercultural awareness/competence are related to the language background of the participants. Both test-takers, who played the role of pilots, and interlocutors, who played the role of ATCOs, were speakers of the same L1, Portuguese. In addition, they all knew each other and worked as AETEs in the same context of language proficiency assessment, thus, members of the same cultural group.

Table 9.9. Analysis of behaviors indicative of effective communication from transcripts of actual task performance

Skills/behaviors indicative of effective communication	Task 1				Task 2				Total
	Morning		Afternoon		Morning		Afternoon		
	Pilot	ATCO	Pilot	ATCO	Pilot	ATCO	Pilot	ATCO	
Indicators of Professional (AE) competence									
1.complying with the rules of use that characterize the domain (e.g., use of phraseology, read back/hear back, etc.)	33	23	30	19	15	13	20	12	**165**
2.demonstrating a professional attitude and tone	36	24	31	18	16	11	21	11	**168**
3.communicating effectively in routine and in unpredictable situations	35	21	31	17	15	11	21	10	**161**
4.using plain English (when appropriate) for aeronautical RT communication	22	4	9	3	7	0	3	1	**49**
5.producing and recognizing the language functions used in RT	39	28	31	20	16	13	21	13	**181**
Indicators of ELF competence									
6.accomodating to different accents and dialects	0	0	0	0	0	0	0	0	**0**
7.adapting linguistic forms to the communicative needs at hand	6	1	3	1	0	0	1	0	**12**
8.complying with the safety-critical requirements of intelligibility	36	22	31	17	16	11	21	11	**165**
9.avoiding the use of jargon, idioms, slang, and colloquialism	35	24	31	29	15	12	20	12	**178**
10. adjusting and aligning to different communicative systems (e.g., new patterns of phonology, syntax, discourse styles)	2	0	2	0	1	0	2	0	**7**
Self-repair[b]	3	0	1	1	0	0	3	0	**8**
Indicators of Intercultural awareness/competence									
11.showing openness and flexibility towards different cultural frames of reference (e.g., communication style, conflict management, face-work strategies, etc.)	0	1	0	1	0	0	0	0	**2**
12.engaging with politeness conventions	12	4	4	2	3	0	2	0	**27**
13.engaging with and negotiating sociocultural differences	4	0	0	0	0	0	0	0	**4**
14.showing willingness to cooperate and to relativize one's own values, beliefs and behaviors	2	5	0	2	0	1	1	0	**11**
15.accomodating to difference and to multilingual aspects of intercultural communications	0	0	0	0	0	0	0	0	**0**
Indicators of Interactional competence									
16.eliminating idioms, cultural references and syntactic complexity from speech	36	24	31	19	16	13	21	13	**173**
17.demonstrating a shared responsibility for successful communication	34	21	31	17	15	11	10	11	**150**
18.accomodating to the constraints of the context and perceived ability of the hearer	2	6	4	2	1	2	1	0	**18**
19.dealing adequately with apparent misunderstanding, by checking, confirming and clarifying	7	6	3	2	3	3	2	1	**27**
20. using an appropriate participation framework	35	22	31	20	17	12	21	13	**171**

21. demonstrating tolerance and collaborative efforts	6	13	4	6	2	2	2	1	**36**
Total	**385**	**249**	**308**	**196**	**158**	**115**	**193**	**109**	**1713**

Note. [a]Numbers here do not include instances of plain English used as incorrect phraseology (e.g., "Okay" and "Copied that" for "Roger"; "Yeah" and "That's right" for "Affirm". However, they do include politeness markers, such as "Good morning", "Good day", "Thank you", etc.

[b]Behavior not included in the *Indicator Checklist*, but included in the final construct framework

Although I acknowledge this as a limitation of the study, the tasks were pilot tested in the Brazilian context due to the ease of gathering these experts together, obtaining proper authorization to run the pilot testing and also to serve other purposes, such as to contribute to the improvement of the current Brazilian test for pilots by proposing a new interactive task, and also to increase the language assessment literacy of the participants. Therefore, evidence of specific behaviors that would emerge only in intercultural communications has not been detected in this pilot testing. Nonetheless, the role-play tasks indicate potential to elicit a full range of skills and behaviors indicative of ELF competence and Intercultural awareness/competence in contexts where the role-players do not share the same L1.

9.5.5 Research question 3.7

RQ 3.7) On the basis of expert judgment (raters), to what extent are:

- the draft pilot tasks likely to elicit the desired language functions and the behaviors indicative of effective communication?

- the language functions and behaviors perceived as important?

Evidence to answer these questions was gathered from the quantitative data collected through the use of the *Observation Checklist of language functions* and the *Indicator Checklist of skills/behaviors indicative of effective communication,* to which raters responded to. However, response rates were variable, for both instruments used,

and the reason why participants did not answer all the questions may be attributable to a number of factors: (a) the time allowed to complete the checklists was not enough; (b) the checklists were too long; (c) participants did not fully understand the criteria; or (d) participants were not interested or motivated to complete the task.

Nevertheless, participants' responses were unpacked so as to check what the trend is across the respondents for each behavior and language function, aiming to determine: (a) if the tasks have the potential to test the construct; and (b) the perception of importance of the construct components to the group of AETEs sampled, i.e., Brazilian ELEs, SME –pilots, and SME – ATCOs.

Response rate for the *Observation Checklist of language functions* associated with aviation was not only variable, but also very low. Participants were not able to fully apply the checklist in real time. Complete responses were received from five participants, all from the afternoon group. Therefore, findings will be presented for Task 1 and Task 2, administered only in the afternoon, up to the point that other responses from additional 13 participants, although incomplete, were available.

To begin with, Table 9.10 organizes the language functions from the first group, i.e., directed towards triggering actions. Figures in this table refer to the number of participants who identified (or not) a language function in the performance of test-takers. Ratings for the level of importance of each language function is also provided, considering the total number of participants (N=35).

Results displayed in the last line of Table 9.10 provide a general overview of what participants were able to capture in relation to the use of language functions. Considering

only valid responses, participants identified that 39.1% of language functions were demonstrated (rated as Yes) in Task 1 – Afternoon and 40.5% in Task 2 – Afternoon.

Table 9.10. Evidence of Group 1 language functions and their perceived importance according to raters

Communicative language functions 1.Directed towards triggering action	Task1 Afternoon (N=15)				Task 2 Afternoon (N=15)				Level of importance (N=35)				
	Yes	No	N/A	M	Yes	No	N/A	M	1	2	3	4	M
1.1 Orders													
Give an order (C)	11	1	1	2	13	0	0	2	0	0	1	28	6
Give an amended order (C)	9	2	1	3	9	2	1	3	0	5	3	20	7
Give a negative order (C)	3	2	8	2	2	3	8	2	0	0	3	25	7
Give alternative orders (C)	4	2	7	2	4	2	7	2	2	3	7	16	7
Cancel an order (C)	3	3	7	2	6	2	5	2	2	0	2	24	7
Announce compliance with an order (P)	13	0	0	2	13	0	0	2	3	5	1	19	7
Announce non-compliance with an order (P)	3	1	9	2	3	1	9	2	0	1	3	24	7
1.2 Requests and offers to act													
Request action by another (C/P)	7	4	1	3	7	4	1	3	1	2	9	11	12
Agree to act (C/P)	13	0	0	2	13	0	0	2	0	2	4	24	11
State reluctance/unwillingness to act (C/P)	0	9	4	2	1	9	3	2	6	4	8	7	10
Refuse to act (C/P)	1	10	2	2	1	10	2	2	3	0	1	19	12
Offer to act (C/P)	2	4	7	2	2	3	8	2	2	8	11	4	10
Accept an offer to act (C/P)	4	2	7	2	4	2	7	2	0	9	10	5	11
Refuse an offer to act (C/P)	2	6	5	2	2	6	5	2	1	8	8	7	11
1.3 Advice (markers for politeness)													
Request advice (P)	1	9	3	2	1	9	3	2	3	4	16	1	11
Give advice (P)	2	7	4	2	2	7	4	2	4	4	14	2	11
Suggest a course of action (C/P)	4	6	3	2	3	7	3	2	1	3	17	3	11
Suggest a solution to a problem (C/P)	2	9	2	2	2	9	2	2	1	2	18	3	11
Suggest alternative courses of action (C/P)	2	5	6	2	2	5	6	2	2	3	16	3	11
1.4 Permission/approval (markers for politeness, directness)													
Request permission/approval (P)	13	0	0	2	13	0	0	2	0	0	6	17	12
Give permission/approval (C)	13	0	0	2	13	0	0	2	0	1	3	19	12
Deny permission/approval (C)	2	6	5	2	2	7	4	2	0	1	2	20	12
Forbid (C)	0	7	6	2	0	7	6	2	2	0	4	17	12
1.5 Undertakings													
Undertake to give a service (C/P)	3	3	6	3	3	3	5	4	1	3	10	8	13
Agree to undertaking/decision (C/P)	3	3	6	3	3	3	5	4	0	2	12	8	13
Undertake to assist (C/P)	5	1	6	3	5	1	6	3	1	3	13	3	15
Undertake to contact/relay/report (C/P)	8	1	3	3	7	1	4	3	0	1	15	4	15
Announce a spontaneous decision to act (C/P)	7	5	1	2	8	3	2	2	0	3	9	8	15
Total	**140**	**108**	**110**	**62**	**144**	**106**	**106**	**64**	**35**	**77**	**226**	**349**	**299**

Note. (Yes) demonstrated; (No) not demonstrated; (N/A) not applicable; (M) missing values
Level of importance ranges from 1 to 4; (C) Controller; (P) Pilot; (C/P) Controller or pilot

These numbers are fairly low and may be due to the fact that the list includes a wide range of functions that may not be fully used by pilots and ATCOs in a regular basis. Second, task observers may have had difficulties to understand each of the functions in order to identify them properly in participants' performance. Actually, when questioned about the evidence of language functions in the tasks, focus group comments were very short and vague, generally a "Yes" without further details. Only one ELE listed some of the functions she captured in the task performances: "So, concerning this list, the task did elicit the ones that we have marked as number 4, so for example, read back, give repetition, request clarification, give clarification, announce readiness and availability, announce a problem, state the expected moment/duration of an action/event" (ELE 4 – Group 1). It is worth highlighting that the functions associated with stating reluctance/unwillingness to act (C/P), refusing to act (C/P), requesting/giving advice (P), and suggesting a solution to a problem (C/P) received a great number of "No's" in the checklists. On the other hand, there was a positive consensus in relation to: give an order (C), announce compliance with an order (P), agree to act (C/P), request permission/approval (P), and give permission/approval (C).

In terms of the perceived importance of language functions belonging to this first group, the overall response rate was 69.7%, considering the whole group of participants (N=35). Out of that, ratings for Level 4 (very important) and Level 3 (important) together account for 83.7% of all valid responses, suggesting that a great number of language functions in the list are considered relevant for effective communication in radiotelephony. However, it is worth highlighting an exception to this trend, the function related to stating reluctance/unwillingness to act (C/P).

Moreover, responses received still allow for the discussion, from the second group of language functions – Sharing information – the sub-group of 'Information concerning present facts'. Table 9.11 displays results from Task 1 and Task 2 pilot tested in the afternoon group. Observers responded if the task elicited (or not) the language functions in the list.

Table 9.11. Evidence of sub-group 2.1 language functions and their perceived importance according to raters

Communicative language functions 2.Sharing information	Task1 Afternoon (N=15)				Task 2 Afternoon (N=15)				Level of importance (N=35)				
	Yes	No	N/A	M	Yes	No	N/A	M	1	2	3	4	M
2.1 Information concerning present facts													
Request information (C/P)	12	0	1	2	12	0	1	2	0	0	4	19	12
Give information(C/P)	13	0	0	2	13	0	0	2	0	3	2	18	12
Request a detailed description (C/P)	4	6	3	2	0	9	4	2	3	6	8	6	12
Describe a state (C/P)	10	1	2	2	8	3	2	2	1	7	8	7	12
Describe a changed state (C/P)	6	4	3	2	9	2	2	2	1	3	9	10	12
Describe an unchanged state (C/P)	3	4	6	2	4	4	5	2	4	6	8	5	12
Describe an action in progress (C/P)	11	1	1	2	10	1	1	3	1	5	9	8	12
Describe a process (C)	4	4	5	2	4	4	5	2	4	6	11	3	11
Describe a procedure (C)	2	7	4	2	4	5	4	2	4	2	6	12	11
Describe aims/precautions (C/P)	3	5	5	2	3	5	5	2	1	2	14	6	12
Describe the source of a problem (C/P)	12	0	1	2	7	4	1	2	1	4	12	7	11
Describe a visual impression (C/P)	5	3	5	2	6	2	5	2	3	6	6	8	12
Quote rules (C)	1	7	5	2	1	7	5	2	8	6	4	5	12
Ask about needs/wishes (C/P)	4	6	3	2	1	7	5	2	4	6	7	7	11
Ask about preferences (C)	0	10	3	2	1	10	2	2	7	5	6	6	11
State preferences (P)	3	7	3	2	7	4	2	2	3	5	6	9	12
Ask about readiness/availability (C/P)	8	3	1	2	6	4	2	3	0	2	5	14	14
Announce readiness/availability (C/P)	8	3	2	2	7	4	2	2	0	1	5	15	14
Request reasons (C/P)	1	10	2	2	1	10	2	2	4	8	7	3	13
Give reasons (C/P)	3	9	1	2	3	9	1	2	3	8	7	5	12
Request instructions on how to do (P)	4	7	2	2	3	7	3	2	2	7	4	9	13
Give instructions on how to do (C)	4	8	1	2	4	8	1	2	1	5	4	13	12
Identify (C/P)	11	0	2	2	11	2	0	2	0	1	5	16	13
Announce a problem (C/P)	13	0	0	2	12	1	0	2	0	0	1	21	13
Total	**145**	**105**	**61**	**48**	**137**	**112**	**60**	**50**	**55**	**104**	**158**	**232**	**291**

Note. (Yes) demonstrated; (No) not demonstrated; (N/A) not applicable; (M) missing values
Level of importance ranges from 1 to 4; (C) Controller; (P) Pilot; (C/P) Controller or pilot

A general overview of what participants were able to detect in relation to the use of language functions from this sub-group can be found in the last line of Table 9.11. Considering valid responses, participants identified that 46.6% of language functions were demonstrated (rated as Yes) in Task 1 – Afternoon and 44.33% in Task 2 – Afternoon. These numbers are a bit higher than the ones reported in Table 9.12, but still low to serve as a good confirmation of their elicitation by the tasks. In addition, there was not much consensus about evidence of these language functions. In Task 1, the functions related to give information and announce a problem had a 100% agreement, whereas in Task 2, only the one related to giving information had agreement. Worth noting, the functions associated with asking about preferences (C) and requesting reasons (C/P) were considered not demonstrated by most participants.

Further, the overall response rate for the perceived importance of language functions belonging to the sub-group of sharing information concerning present facts was 65.4%, considering the whole group of participants (N=35). Out of that, ratings for Level 4 (very important) and Level 3 (important) together account for 71% of all valid responses, suggesting that a moderate number of language functions in this list is considered important for successful radiotelephony communications, in the observers' opinion.

Following the sequence of language functions in the checklist, the ones related to the remainder of group 2, group 3 and group 4 received very few responses. However, as group 4 includes the functions used in the management of the dialogue, and thus crucial for effective radiotelephony communications, findings will be presented and discussed according to what the trend suggests in relation to their evidence in task performance.

From Table 9.12 we can depict, despite a low response rate of 35.2% for both Task 1 and Task 2, that valid responses indicate nine instances of total positive agreement for the following language functions: *read back (C/P), check understanding (C/P), acknowledge*

Table 9.12. Evidence of Group 4 language functions and their perceived importance according to raters

Communicative language functions 4. Management of the dialogue	**Task1 Afternoon (N=15)**				**Task 2 Afternoon (N=15)**				**Level of importance (N=35)**				
	Yes	**No**	**N/A**	**M**	**Yes**	**No**	**N/A**	**M**	**1**	**2**	**3**	**4**	**M**
Name addressee(s) (C/P)	3	1	2	9	3	1	2	9	1	9	2	14	9
Self-correct (C/P)	3	1	2	9	3	1	2	9	0	0	3	27	5
Paraphrase (C/P)	1	2	3	9	1	2	3	9	0	3	5	22	5
Close an exchange (C/P)	2	1	3	9	2	1	3	9	5	3	13	10	4
Request response (C/P)	5	0	1	9	5	0	1	9	0	2	6	23	4
Read back (C/P)	6	0	0	9	6	0	0	9	0	0	2	29	4
Check understanding (C/P)	5	0	0	10	5	0	0	10	0	0	4	27	4
Acknowledge (C/P)	5	0	0	10	5	0	0	10	0	0	9	22	4
Check certainty (C/P)	4	1	0	10	4	1	0	10	3	2	6	19	5
Declare non-understanding (C/P)	5	0	0	10	5	0	0	10	0	0	0	31	4
Correct a misunderstanding (C/P)	4	1	0	10	4	1	0	10	0	0	0	31	4
Request repetition (C/P)	5	0	0	10	5	0	0	10	0	0	10	21	4
Give repetition (C/P)	5	0	0	10	5	0	0	10	0	4	7	20	4
Request confirmation (C/P)	5	0	0	10	5	0	0	10	0	1	11	19	4
Give confirmation (C/P)	5	0	0	10	5	0	0	10	0	3	6	22	4
Request clarification (C/P)	5	0	0	10	5	0	0	10	0	0	5	26	4
Give dis-confirmation (C/P)	0	4	1	10	0	4	1	10	0	2	3	26	4
Give clarification (C/P)	3	2	0	10	3	2	0	10	0	2	6	23	4
Relay an order (C)	1	3	1	10	1	3	1	10	0	0	11	18	6
Relay a request to act (C)	2	2	1	10	2	2	1	10	0	0	7	22	6
Relay a request for permission (P)	2	2	1	10	2	2	1	10	0	0	9	20	6
Total	**76**	**20**	**15**	**204**	**76**	**20**	**15**	**204**	**9**	**31**	**125**	**472**	**98**

Note. (Yes) demonstrated; (No) not demonstrated; (N/A) not applicable; (M) missing values
Level of importance ranges from 1 to 4; (C) Controller; (P) Pilot; (C/P) Controller or pilot

(C/P), check certainty (C/P), declare non-understanding (C/P), request repetition (C/P), give repetition (C/P), request confirmation (C/P), and give confirmation (C/P). Added to that, three other language functions in this group displayed a good agreement in terms of being elicited by the tasks: *request response (C/P), check certainty (C/P), and correct a*

misunderstanding (C/P). On the other hand, *give dis-confirmation (C/P)* was the function that received more "No" from the participants that responded to that specific part of the checklist.

Yet, a more complete understanding of the perceived importance of Group 4 language functions is possible, as the overall response rate for all observers (N=35) was 86.7%. Ratings for Level 4 (very important) and Level 3 (important) together account for 93.7% of all valid responses, suggesting that a high number of language functions in the list are considered relevant for effective communication in radiotelephony. However, it is worth highlighting an exception to this trend, the function related to *closing an exchange (C/P)*.

After presenting results on the extent to which the draft tasks are likely to elicit the desired language functions and their perceived importance according to raters' perspectives, my discussion now will focus on the behaviors of effective communication in the RT context.

Evidence of behaviors 1 to 21, as captured by raters, in relation to task (Task 1 and Task 2) and administration (morning and afternoon), is presented in Table 9.13. Figures in this table refer to the number of participants that identified (or not) a certain behavior in the performance of test-takers.

Table 9.13. Evidence of behaviors indicative of effective communication and their perceived importance according to raters

Skills/behaviors indicative of effective communication	Task 1								Task 2								Level of importance N=35				
	Morning (N=17)				Afternoon (N=15)				Morning (N=17)				Afternoon (N=15)								
	Yes	No	N/A	M	Yes	No	N/A	M	Yes	No	N/A	M	Yes	No	N/A	M	1	2	3	4	M
Indicators of Professional (AE) competence																					
1.complying with the rules of use that characterize the domain (e.g., use of phraseology, read back/hear back, etc.)	8	0	0	9	14	0	0	1	7	0	1	9	14	0	0	1	0	0	2	32	1
2.demonstrating a professional attitude and tone	8	0	0	9	14	0	0	1	8	0	0	9	14	0	0	1	0	4	12	18	1
3.communicating effectively in routine and in unpredictable situations	7	0	0	10	14	0	0	1	7	0	1	9	14	0	0	1	0	0	4	30	1
4.using plain English (when appropriate) for aeronautical RT communication	3	5	0	9	9	1	3	2	6	2	0	9	6	1	6	2	0	0	10	23	2
5.producing and recognizing the language functions used in RT	8	0	0	9	13	0	0	2	8	0	0	9	13	0	0	2	1	0	5	26	3
Indicators of ELF competence																					
6.accomodating to different accents and dialects	0	1	7	8	5	1	6	3	1	1	6	9	5	1	7	2	0	3	7	24	1
7.adapting linguistic forms to the communicative needs at hand	4	1	3	9	4	1	7	3	4	1	3	9	4	1	8	2	0	3	8	23	1
8.complying with the safety-critical requirements of intelligibility	8	0	0	9	11	0	1	3	8	0	0	9	11	0	1	3	0	0	4	29	2
9.avoiding the use of jargon, idioms, slang, and colloquialism	7	1	0	9	13	0	0	2	7	1	0	9	13	0	0	2	0	1	13	20	1
10. adjusting and aligning to different communicative systems (e.g., new patterns of phonology, syntax, discourse styles)	0	1	7	9	3	0	8	4	0	1	7	9	3	0	9	3	2	3	11	18	1
Indicators of Intercultural awareness/competence																					
11.showing openness and flexibility to different cultural frames of reference (e.g., communication style, conflict management, face-work strategies, etc.)	0	4	4	9	6	1	7	1	0	4	4	9	4	1	8	2	0	2	16	16	1
12.engaging with politeness conventions	8	0	0	9	6	1	6	2	8	0	0	9	5	1	7	2	3	14	11	6	1
13.engaging with and negotiating sociocultural differences	1	1	5	10	2	1	10	2	1	1	6	9	1	1	11	2	2	7	13	11	2
14.showing willingness to cooperate and to relativize one's own values, beliefs and behaviors	6	1	1	9	4	2	7	2	6	1	1	9	3	2	8	2	2	7	10	15	1
15.accomodating to difference and to multilingual aspects of intercultural communications	1	1	6	9	0	2	10	3	1	1	6	9	0	2	11	2	0	4	14	16	1

Indicators of Interactional competence																					
16.eliminating idioms, cultural references and syntactic complexity from speech	6	1	1	9	12	0	1	2	6	1	1	9	12	0	1	2	1	1	12	19	2
17.demonstrating a shared responsibility for successful communication	8	0	0	9	13	0	0	2	8	0	0	9	13	0	0	2	0	1	6	26	2
18.accomodating to the constraints of the context and perceived ability of the hearer	6	0	0	11	12	0	1	2	7	0	0	10	12	0	1	2	0	0	9	23	3
19.dealing adequately with apparent misunderstanding, by checking, confirming and clarifying	8	0	0	9	13	0	0	2	8	0	0	9	13	0	0	2	0	0	0	32	3
20. using an appropriate participation framework	7	1	0	9	11	1	1	2	8	0	0	9	11	1	1	2	0	0	10	22	3
21. demonstrating tolerance and collaborative efforts	8	0	0	9	12	1	0	2	8	0	0	9	12	1	0	2	0	2	7	23	3
Total	**112**	**18**	**34**	**192**	**191**	**12**	**68**	**44**	**117**	**14**	**36**	**190**	**183**	**12**	**79**	**41**	**11**	**52**	**184**	**452**	**36**

Note. (Yes) demonstrated; (No) not demonstrated; (N/A) not applicable; (M) missing values.
Level of importance ranges from 1 to 4; (M) missing values

Ratings for the level of importance of each behavior are also provided, so as to contribute to the interpretation of findings.

As can be inferred by the numbers in the last line of Table 9.13, the overall response rate for both administrations of Task 1 (morning + afternoon) was 64.8%. Considering only valid responses, participants identified that 68.3% of behaviors were demonstrated (Yes) in Task 1 – Morning and 70.5% in Task 2 – Afternoon. Results were very close in regards to response rate for both administrations of Task 2, which was 65.6%. In the same lines, participants considered that 70.1% of behaviors were demonstrated (Yes) in Task 2 – Morning and 66.8% in Task 2 – Afternoon. These findings suggest, in general, a moderate to fairly high elicitation of the behaviors in the proposed tasks, which agree with findings from the analysis of focus group discussions. For the three focus groups, Magnitude coding indicated a positive trend for the category "Construct", sub-category "Behaviors", as displayed in Table 9.7.

Looking at valid responses for individual behaviors, it is possible to say that in Task 1 – Morning, all participants responded "Yes" (i.e., 100% agreement) to 10 out of 21 behaviors, distributed across the domains in the following way: AE= 4 behaviors, ELF=1, ICA= 1, and IC= 4 behaviors. A slight difference was noted in Task 1 – Afternoon, where 100% agreement on the confirmation of behaviors was found only in 7 behaviors: AE= 4, ELF=1, ICA=0, and IC=2.

The same understanding of these numbers in regards to Task 2 reveals that, in the morning, all participants responded "Yes" (i.e., 100% agreement) to 9 out of 21 behaviors, distributed across the domains in the following way: AE= 2 behaviors, ELF=1, ICA= 1, and IC= 5 behaviors. However, in Task 2 – Afternoon, 100% agreement on the

confirmation of behaviors was found only in 7 behaviors: AE= 4, ELF=1, ICA=0, and IC=2. Despite that, throughout Table 9.13 it is also possible to find seven behaviors in which a good agreement (i.e., 99-80% of positive responses) was reached for Task 1 (morning + afternoon), and a similar situation for eight behaviors in Task 2 (morning + afternoon).

In spite of having role-players from the same cultural group and same L1, none of the 21 behaviors reached 100% agreement of not being demonstrated. But still, a few of them were considered "Not applicable" by a great number of participants, such as behavior 10= *adjusting and aligning to different communicative systems* (ELF), behavior 13= *engaging with and negotiating sociocultural differences,* and behavior 15= *accommodating to difference and to multilingual aspects of intercultural communications* (ICA). A comment from the focus group discussion confirms this finding: "But I think the ones, considering intercultural phrases, when you have …[name of a native speaker] there, for instance, you can elicit better, because he is a foreigner, but with them, the two of them are Brazilian" (ELE 1 – Group 2).

Yet, the trend was not very clear in relation to other behaviors. For example, little agreement was reached for behavior 4= *using plain English (when appropriate) for aeronautical RT communication* (AE), behavior 7= *adapting linguistic forms to the communicative needs at hand* (ELF) and behavior 11=*showing openness and flexibility to different cultural frames of reference (e.g., communication style, conflict management, face-work strategies, etc)* (ICA). In relation to the use of plain English, the confirmation of this divergence of opinions was revealed by the color coding in Table 9.7, where we find more negative comments from Focus group 1, compared to more positive ones for

Focus Groups 2 and 3. One behavior that displayed different perceptions in the morning and afternoon groups was behavior 12, *engaging with politeness conventions*. While there was a positive consensus in the morning group, for both Task 1 and Task 2, opinions of afternoon participants were mixed and divergent. Again, this finding was corroborated by the arguments about the use of politeness markers that emerged in Focus group 2, which were presented in Excerpt 4.

In Table 9.13 we also find information regarding the perceived importance of the 21 behaviors from raters' perspectives. Ratings for Level 4 (very important) and Level 3 (important) together account for 90.9% of all responses, suggesting that most behaviors in the list are deemed important for effective communication in radiotelephony. However, it is worth highlighting some exceptions to this trend in the sample analysed, such as behavior 10= *adjusting and aligning to different communicative systems (e.g. new patterns of phonology, syntax, discourse styles)*, from the ELF domain, and from the ICA domain, behavior 12= *engaging with and negotiating sociocultural differences*, behavior 13= *engaging with and negotiating sociocultural differences*, and behavior 14= *showing willingness to cooperate and to relativize one's own values, beliefs and behaviors.* Going deeper into these figures, it was possible to see that in the case of behavior 10, one SME – pilot and one ELE rated it as not important (Level 1). On the other hand, behavior 12 was rated as 1 (not important) by two SME – ATCOs and one SME – pilot; the same predominance of operational personnel not perceiving the importance of specific behaviors was found for behavior 13 (one SME – pilot and one SME – ATCO), and also for behavior 14 (two SME – pilots).

A comparison of the perceived importance of behaviors indicative of effective communication across different groups of raters, i.e., ELEs, SME – pilots and SME – ATCOs, is provided in Table 9.14. Mean values (*M*) and standard deviations (*SD*) for the raters' responses on Level of importance (1 to 4) are reported for each behavior and for the three groups of participants separately. Cells that contain the highest mean values, ranging from *M*=4.00 to *M*=3.80, were highlighted in **green** so as to provide a more straightforward identification of the behaviors considered more important by each group, but also across the three distinct groups. Consensus among the three groups was noted for behaviors 1 and 3 (AE), behavior 8 (ELF) and behavior 19 (IC). Similarly, but in the opposite direction, cells that contain the lowest mean values, ranging from *M*=2.00 to *M*=2.80, were highlighted in **red** to indicate the less important behaviors in the perception of ELEs, SME – pilots and SME – ATCOs. Agreement among the three groups was found for behavior 12 (ICA), which is related to the *engagement with politeness conventions*.

In conclusion, the pilot testing of Task 1 and Task 2 in an L1 context indicated that most behaviors that provide evidence of AE competence and Interactional competence were captured by the raters observing task performances. The same cannot be stated in regards to the indicators of ELF competence and also of the Intercultural awareness/competence, first, because of the constraints imposed by the similarity in the linguacultural background of test-takers and interlocutors, and second, because of the challenges to capture these behaviors in practice. As Harding and McNamara (2017) stated in relation to an ELF task, "what is observable through an analysis of discourse may be very difficult for raters to detect in practice" (p. 579).

Table 9.14. Comparison of perceived importance of behaviors indicative of effective communication across different groups of raters

Behaviors indicative of effective communication	Group 1 - ELEs		Group 2 - SME pilots		Group 3 - SME ATCOs	
	N	Importance (1 to 4)	N	Importance (1 to 4)	N	Importance (1 to 4)
Indicators of Professional (AE) competence						
1.complying with the rules of use that characterize the domain (e.g. use of phraseology, read back/hear back, etc.)	19	*M*=3.95, *SD*=0.23	11	*M*=3.91, *SD*=0.30	4	*M*=4.00, *SD*=0.00
2.demonstrating a professional attitude and tone	19	*M*=3.47, *SD*=0.70	11	*M*=3.18, *SD*=0.75	4	*M*=3.75, *SD*=0.50
3.communicating effectively in routine and in unpredictable situations	19	*M*=3.89, *SD*=0.31	11	*M*=3.82, *SD*=0.40	4	*M*=4.00, *SD*=0.00
4.using plain English (when appropriate) for aeronautical RT communication	18	*M*=3.67, *SD*=0.48	11	*M*=3.73, *SD*=0.47	4	*M*=3.75, *SD*=0.50
5.producing and recognizing the language functions used in RT	18	*M*=3.67, *SD*=0.77	10	*M*=3.90, *SD*=0.32	4	*M*=3.75, *SD*=0.50
Indicators of ELF competence						
6.accomodating to different accents and dialects	19	*M*=3.47, *SD*=0.77	11	*M*=3.73, *SD*=0.47	4	*M*=4.00, *SD*=0.00
7.adapting linguistic forms to the communicative needs at hand	19	*M*=3.42, *SD*=0.77	11	*M*=3.73, *SD*=0.47	4	*M*=4.00, *SD*=0.00
8.complying with the safety-critical requirements of intelligibility	18	*M*=3.83, *SD*=0.38	11	*M*=3.91, *SD*=0.31	4	*M*=4.00, *SD*=0.00
9.avoiding the use of jargon, idioms, slang and colloquialism	19	*M*=3.53, *SD*=0.51	11	*M*=3.45, *SD*=0.69	4	*M*=4.00, *SD*=0.00
10. adjusting and aligning to different communicative systems (e.g., new patterns of phonology, syntax, discourse styles)	19	*M*=3.26, *SD*=0.87	11	*M*=3.27, *SD*=1.01	4	*M*=3.75, *SD*=0.50
Indicators of Intercultural awareness/competence						
11.showing openness and flexibility to different cultural frames of reference (e.g., communication style, conflict management, face-work strategies, etc.)	19	*M*= 3.47, *SD*=0.61	11	*M*=3.18, *SD*=0.60	4	*M*=3.75, *SD*=0.50
12.engaging with politeness conventions	19	*M*=2.74, *SD*=0.81	11	*M*=2.55, *SD*=0.93	4	*M*=2.00, *SD*=1.15
13.engaging with and negotiating sociocultural differences	18	*M*=3.33, *SD*=0.77	11	*M*=2.64, *SD*=0.92	4	*M*=2.50, *SD*=1.00
14.showing willingness to cooperate and to relativize one's own values, beliefs and behaviors	19	*M*=3.10, *SD*=0.87	11	*M*=3.00, *SD*=1.09	4	*M*=3.50, *SD*=1.00
15.accomodating to difference and to multilingual aspects of intercultural communications	19	*M*=3.26, *SD*=0.73	11	*M*=3.36, *SD*=0.67	4	*M*=3.75, *SD*=0.50

Indicators of Interactional competence						
16.eliminating idioms, cultural references and syntactic complexity from speech	18	*M*=3.82, *SD*=0.39	11	*M*=3.64, *SD*=0.50	4	*M*=3.75, *SD*=0.50
17.demonstrating a shared responsibility for successful communication	18	*M*=3.72, *SD*=0.57	11	*M*=3.73, *SD*=0.47	4	*M*=4.00, *SD*=0.00
18.accomodating to the constraints of the context and perceived ability of the hearer	17	*M*=3.82, *SD*=0.39	11	*M*=3.55, *SD*=0.52	4	*M*=3.75, *SD*=0.50
19.dealing adequately with apparent misunderstanding, by checking, confirming and clarifying	17	*M*=4.00, *SD*=0.00	11	*M*=4.00, *SD*=0.00	4	*M*=4.00, *SD*=0.00
20. using an appropriate participation framework	17	*M*=3.70, *SD*=0.47	11	*M*=3.55, *SD*=0.52	4	*M*=4.00, *SD*=0.00
21. demonstrating tolerance and collaborative efforts	17	*M*=3.59, *SD*=0.71	11	*M*=3.73, *SD*=0.47	4	*M*=3.75, *SD*=0.50

Note. In green, behaviors considered more important by each group of raters. In red, behaviors considered less important.

However, findings discussed so far suggest that both tasks have the potential to test components of the construct considered relevant by aviation stakeholders to communicate effectively over the radio. This is also confirmed by test-takers' and interlocutors' opinions related to the usefulness of the task to elicit language that is appropriate to the communicative demands of pilots in international RT communications.

Up to this point, quantitative results from the pilot testing of Task 1 and Task 2 were reported separately from qualitative findings from the analysis of transcripts of task performances. Therefore, it is now crucial to integrate these two sources of data in order to demonstrate the "distinguishing characteristic of a mixed method versus multimethod study" (Guetterman & Salamoura, 2016, p. 163). Along these lines, two joint displays were built, presenting qualitative and quantitative results in adjacent columns as a way to visually highlight convergence and divergence.

First, a table was built including evidence of language functions in the transcripts (T) of task performances (QUAL column) and also evidence from the quantitative data (quan column) (see Table 9.15). However, due to the limited number of responses to the *Observation Checklist of language functions*, only data from Task 1 and Task 2 administered in the afternoon session were included. Quantitative results are reported in terms of the degree of agreement among raters using the *Observation Checklist*: Very good (VG – 100%), Good (G – 99-80%, positive or negative), Some (79-65%, positive or negative), and Little agreement (L – 64% or less).

Mutually confirming outcomes, i.e., convergence of QUAL and quan data, are highlighted in **green**. This happened in two situations: first, when referring to the presence of the language function in the transcript that was also captured by raters (i.e.,

T/VG; T/G+; and T/S+). For Task 1 – Afternoon this happened in 23 occasions, whereas for Task 2 – Afternoon, only in 13 occasions. And second, when referring to the convergence of data in relation to the absence of the language function in both strands (i.e., blank/S- and blank/G-). This happened in 20 occasions for Task 1 – Afternoon and 22 for Task 2 – Afternoon.

Table 9.15. Language functions – Evidence from transcripts and from *Observation Checklists*

Communicative language functions associated with aviation (ICAO, 2010)	**Task 1 After-noon**		**Task 2 After-noon**	
	QUAL	**quan**	**QUAL**	**quan**
1.Directed towards triggering action				
1.1 Orders				
Give an order (C)	T	G+	T	VG
Give an amended order (C)		S+		S+
Give a negative order (C)		L		L
Give alternative orders (C)		L		L
Cancel an order (C)		L		L
Announce compliance with an order (P)	T	VG	T	VG
Announce non-compliance with an order (P)		S-		S-
1.2 Requests and offers to act				
Request action by another (C/P)		L		L
Agree to act (C/P)		VG		VG
State reluctance/unwillingness to act (C/P)		S-		S-
Refuse to act (C/P)		S-		S-
Offer to act (C/P)		L		L
Accept an offer to act (C/P)		L		L
Refuse an offer to act (C/P)		L		L
1.3 Advice (markers for politeness)				
Request advice (P)		S-		S-
Give advice (P)		L		L
Suggest a course of action (C/P)	T	L		L
Suggest a solution to a problem (C/P)		S-		S-
Suggest alternative courses of action (C/P)		L		L
1.4 Permission/approval				
Request permission/approval (P)	T	VG	T	VG
Give permission/approval (C)	T	VG	T	VG
Deny permission/approval (C)		L		L
Forbid (C)		L		L
1.5 Undertakings				
Undertake to give a service (C/P)		L		L
Agree to undertaking/decision (C/P)		L		L
Undertake to assist (C/P)		L		L
Undertake to contact/relay/report (C/P)		S+		L
Announce a spontaneous decision to act (C/P)		L		L
2. Sharing information				

2.1 Information concerning present facts				
Request information (C/P)	T	G+		G+
Give information(C/P)	T	VG	T	VG
Request a detailed description (C/P)		L		S-
Describe a state (C/P)	T	S+		L
Describe a changed state (C/P)		L		S+
Describe an unchanged state (C/P)		L		L
Describe an action in progress (C/P)	T	G+	T	G+
Describe a process (C)		L		L
Describe a procedure (C)		L		L
Describe aims/precautions (C/P)		L		L
Describe the source of a problem (C/P)	T	G+		L
Describe a visual impression (C/P)		L		L
Quote rules (C)		L		L
Ask about needs/wishes (C/P)		L		L
Ask about preferences (C)		G-		S-
State preferences (P)		L		L
Ask about readiness/availability (C/P)	T	S+		L
Announce readiness/availability (C/P)	T	L		L
Request reasons (C/P)		S-		S-
Give reasons (C/P)	T	S-	T	S-
Request instructions on how to do (P)		L		L
Give instructions on how to do (C)		L		L
Identify (C/P)		G+		G+
Announce a problem (C/P)	T	VG		G+
2.2 Information concerning the future				
Announce an expected action/event	T	S+		S+
Ask about the expected moment/duration ...		L		L
State the expected moment/duration of an …	T	L		L
Ask about possible consequences of an action		L		L
State possible consequences of an action		L		L
Ask about intentions (C/P)	T	S+		G+
State intentions (C/P)	T	VG		G+
Request prediction (C/P)		L		L
Predict a future action/event (C/P)		L		L
Warn (C/P)		L		L
2.3 Information concerning immediate/recent past events				
Announce a completed action/event having an effect on the present (C/P)	T	L		L
Announce a change (C/P)	T	G+	T	S+
Announce a nearly completed action (C/P)		L		L
2.4 Information concerning the past				
Ask about past events (C/P)		L		L
Announce a past action/event (C/P)		L		L
Announce an avoided problem/incident (P)		L		L
Give a report (C/P)	T	S+		S+
Describe a previous communication (C/P)		L		L
Describe a sequence of past actions/events (C/P)		L		L
Request an explanation of a past action/event (C/P)		L		L
Give an explanation of a past action (C/P)		L		L
Indicate deductive reasoning (C/P)		L		L
2.5 Necessity				
Ask about necessity(C/P)		G+		G+
State necessity (C/P)		VG		VG
Announce a compulsory action (C/P)		L		L

Announce an inevitable action/event (C/P)		L		L
2.6 Feasibility/capacity				
Ask about the feasibility/capacity (C/P)	T	S+		L
Announce feasibility/capacity (C/P)	T	L		L
Announce unfeasibility/incapacity (C/P)		L		L
3. Management of the pilot-controller relation				
Greet/take leave (C/P)	T	S-		S-
Respond to greeting/leave-taking (C/P)		G-		G-
Thank (C/P)		G-		G-
Respond to thanks (C/P)		G-		G-
Complain (P)		S-		S-
Apologize (C/P)		S-		S-
Express dissatisfaction (C/P)		S-		S-
Reprimand (C)		G-		G-
Reject complaint/reprimand (C/P)		G-		G-
Express satisfaction (C/P)		S-		S-
Express concern/apprehension (P)		S-		S-
Reassure (C)		S-		S-
Encourage (C)		S-		S-
4. Management of the dialogue				
Name addressee(s) (C/P)	T	L	T	L
Self-correct (C/P)	T	L	T	L
Paraphrase (C/P)		L		L
Close an exchange (C/P)		L		L
Request response (C/P)		G+		G+
Read back (C/P)	T	VG	T	VG
Check understanding (C/P)		VG		VG
Acknowledge (C/P)	T	VG	T	VG
Check certainty (C/P)		G+		G+
Declare non-understanding (C/P)		VG		VG
Correct a misunderstanding (C/P)		G+		G+
Request repetition (C/P)	T	VG	T	VG
Give repetition (C/P)	T	VG	T	VG
Request confirmation (C/P)	T	VG	T	VG
Give confirmation (C/P)	T	VG	T	VG
Request clarification (C/P)		VG		VG
Give dis-confirmation (C/P)		G-		G-
Give clarification (C/P)		L		L
Relay an order (C)		L		L
Relay a request to act (C)		L		L
Relay a request for permission (P)		L		L

Note. T indicates that the language function was identified in the transcript of the role-play task.
VG, G+, S+, L, S-, and G- indicate the degree of agreement among raters using the *Observation Checklist*.
VG: Very good agreement; G+: Good agreement (positive); S+: Some agreement (positive)
L: Little agreement; S-: Some agreement (negative); G-: Good agreement (negative)
Green cells – convergence of QUAL and quan data
Red cells – divergence of QUAL and quan data

However, in a number of other instances there was very good, good, or some positive agreement among raters, but the language functions were not identified in the transcripts of task performances (i.e., blank/VG; blank/G+; and blank/S+). In these cases,

divergence is highlighted in red, representing 12 instances for Task 1 and 18 for Task 2. Similarly, divergence was also noted when the opposite occurred: evidence of the function was found in the transcripts but not detected by the raters (i.e., T/S-). This happened twice for Task 1 and only once for Task 2.

Generally, when there was little agreement among raters, the language function was not identified in the transcript. However, a few exceptions were noted, nine out of 119 occasions, when the functions were detected at discourse level.

The number of divergences reported in Table 9.15 may be related to the complexity of the checklist not only in terms of the number of language functions but also of the similarity in the description of some of them. Hence, this first attempt to apply the *Observation Checklist of language functions* disclosed the need to refine it based on the functions that were produced more frequently and also based on their perceived importance according to raters. In addition, the relevance of its use in the process of task design was made clear as a way to generate content-related evidence. As O'Sullivan, Weir and Savile (2002) noted, "by taking into account the expected response of a task (and by describing that response in terms of these functions) it will be possible to explore predicted and actual test task outcome" (p. 46). In addition, the investigation of how disparities in interlocutor behavior and adherence to scripts may affect test-taker performance, a concern raised by many focus group participants, might also be facilitated by the identification of the language functions produced.

Second, a joint display (see Table 9.16) was also designed to integrate results related to evidence of the behaviors of effective communication in radiotelephony found in the transcripts (T) of task performances (QUAL data) and also evidence from the

responses to *Indicator Checklists* (quan column). Findings are reported for Task 1 and Task 2 in both task administrations, morning and afternoon. Quantitative results are reported in terms of the degree of agreement among raters using the *Indicator Checklist*: Very good (VG – 100%), Good (G – 99-80%, positive or negative), Some (79-65%, positive or negative), and Little agreement (L – 64% or less).

Table 9.16. Behaviors of effective communication – Evidence from transcripts and from *Indicator Checklists*

Skills/behaviors indicative of effective communication	Task 1				Task 2			
	Morning		Afternoon		Morning		Afternoon	
	QUAL	quan	QUAL	quan	QUAL	quan	QUAL	quan
Indicators of Professional (AE) competence								
1.complying with the rules of use that characterize the domain (e.g. use of phraseology, read back/hear back, etc)	T	VG	T	VG	T	G+	T	VG
2.demonstrating a professional attitude and tone	T	VG	T	VG	T	VG	T	VG
3.communicating effectively in routine and in unpredictable situations	T	VG	T	VG	T	G+	T	VG
4.using plain English (when appropriate) for aeronautical RT communication	T	L	T	S+	T	S+	T	L
5.producing and recognizing the language functions used in RT	T	VG	T	VG	T	VG	T	VG
Indicators of ELF competence								
6.accomodating to different accents and dialects		G-		L		S-		L
7.adapting linguistic forms to the communicative needs at hand	T	L	T	L		L	T	L
8.complying with the safety-critical requirements of intelligibility	T	VG	T	G+	T	VG	T	G+
9.avoiding the use of jargon, idioms, slang and colloquialism	T	G+	T	VG	T	G+	T	VG
10. adjusting and aligning to different communicative systems (e.g. new patterns of phonology, syntax, discourse styles)	T	G-	T	S-	T	G-	T	S-
Indicators of Intercultural awareness/competence								
11.showing openness and flexibility to different cultural frames of reference (e.g., communication style, conflict management, face-work strategies, etc)		L		L		L		L
12.engaging with politeness conventions	T	VG	T	L	T	VG	T	L
13.engaging with and negotiating sociocultural differences	T	L		S-		S-		G-
14.showing willingness to cooperate and to relativize one's own values, beliefs and behaviors	T	S+		L		S+	T	L
15.accomodating to difference and to multilingual aspects of intercultural communications		S-		G-		S-		G-
Indicators of Interactional competence								
16.eliminating idioms, cultural references and syntactic complexity from speech	T	S+	T	G+	T	S+	T	G+
17.demonstrating a shared responsibility for successful communication	T	VG	T	VG	T	VG	T	VG
18.accomodating to the constraints of the context and perceived ability of the hearer	T	VG	T	G+	T	VG	T	G+
19.dealing adequately with apparent misunderstanding, by checking, confirming and clarifying	T	VG	T	VG	T	VG	T	VG
20. using an appropriate participation framework	T	G+	T	G+	T	VG	T	G+
21. demonstrating tolerance and collaborative efforts	T	VG	T	G+	T	VG	T	G+

Note. T indicates that the behavior was identified in the transcript of the role-play task.
VG, G+, S+, L, S-, and G- indicate the degree of agreement among raters using the *Indicator Checklist*.
VG: Very good agreement; G+: Good agreement (positive); S+: Some agreement (positive)
L: Little agreement; S-: Some agreement (negative); G-: Good agreement (negative)
Green cells – convergence of QUAL and quan data
Red cells – divergence of QUAL and quan data

Similarly to what was done in the joint display of language functions, mutually confirming outcomes, i.e., convergence of QUAL and quan data, are highlighted in **green**. They include instances in which the behavior was identified in the transcript and at the same time captured by raters (i.e., T/VG; T/G+; and T/S+). In addition, convergence of data in relation to the absence of the behavior in both strands (i.e., blank/S- and blank/G-) was also color coded in **green**. In total, convergent cases account for 67 out of 84 (79.7%) instances in the complete table. On the other hand, divergence was found in one occasion when there was some positive agreement among raters, but the behavior was not identified in the transcripts of task performances (i.e., blank/S+). Further, divergent findings were also noted when the opposite occurred: evidence of the behaviors were found in the transcripts but not detected by the raters (i.e., T/G-; and T/S-). This happened only in five occasions, highlighted in **red**. Considering the 17 cases when there was little agreement (L) among raters, eight of them included behaviors that were also not identified in the transcripts. Yet, in nine instances related to behaviors 4 (AE), 7 (ELF), and 12/13/14 (ICA), the behaviors were observed at discourse level but not detected by raters.

To conclude, the joint displays suggest that, in relation to the language functions, quantitative results from the *Observation Checklists* agree, to a moderate extent, with transcripts of task performances. Moreover, data integration also indicates that findings from the *Indicator Checklists of behaviors* agree with performance transcripts to a great extent and signpost the potential of the tasks to test the construct. Applying the *Observation Checklist of language functions* and the *Indicator Checklist of behaviors indicative of effective communication* in the phase of pilot testing proved very useful as

potential sources of evidence supporting the construct validity of the role-play tasks. Additionally, they assisted in the raters' familiarization and understanding of their components to be used later in the definition and trialing of appropriate assessment criteria. On top of that, applying the checklists added great value in highlighting what changes and adaptations needed to be made to the tasks in order to elicit the full range of functions and behaviors required for effective intercultural radiotelephony communications.

Having discussed results from the MM convergent parallel study covered by Phase 3, in the next chapter I will highlight the integration of the major findings from Phases 1, 2 and 3, aiming to address the research problem described in the introduction to this thesis. In addition, I will also discuss study limitations, directions for future research, implications for different stakeholders, and finalize with some concluding remarks.

Chapter 10 – Conclusions

Having presented in the previous chapters the method used in each Phase of this multiphase MM study, and having discussed results from Phase 1, Phase 2, and Phase 3, in this chapter I will refer back to the study's overarching research question in order to respond to it based on the integration of findings across all phases, as a way to summarize what has been done in this multiphase MM study. In addition, I will discuss limitations and propose next steps to further investigate the operationalization of the international RT construct into potential test tasks for use in language proficiency testing within the aviation workplace. Subsequently, I will consider some possible uses of the outcomes of this research by aviation stakeholders and highlight some implications of the findings for different fields of inquiry. Finally, I will conclude by reflecting on the learning process that I have experienced while conducting this complex study.

10.1 Integration of Findings

The integration of findings from the overall study is crucial in generating meta-inferences across all phases in order to address the research problem and answer the primary research question introduced at the beginning of this dissertation.

As discussed in Chapter 1, the main objective of this study was to contribute to the safety of pilot-ATCO intercultural communications through the exploration of the dimensions of awareness, knowledge, skills, and attitudes required for effective communication, relying on the perceptions of a range of international stakeholders. Additionally, this project aimed to increase the validity of inferences drawn from the results of specific purpose aviation English tests through the design and pilot testing of new tasks that operationalize the identified RT construct. As a result, to address the

research problem described above, the overarching research question that guided this study was:

> What are the communicative demands of pilots and ATCOs involved in intercultural RT communications that go beyond language proficiency; how can they be specified within a construct framework and operationalized as test tasks?

In order to answer this complex research question, a multiphase MM research design was chosen. As Ziegler and Kang (2016) explain, the value of MM in multiphase designs "lies in the use of meta-inferences generated within (or between) stages to inform the direction of the following stages, as well as generating meta-inferences from all strands across stages [phases]" (p. 77). Thus, Chapters 7, 8, and 9 reported mainly on the generation of meta-inferences within each phase of the study, Phases 1, 2, and 3, respectively. However, the existing interconnection between the three phases discloses the movement back and forth as an iterative process, which ultimately contributes to the generation of meta-inferences, as shown in Figure 10.1. This is in accordance with the ongoing nature of test development and validation processes, which are based on iterative feedback for revisions and continuous collection of evidence (e.g., Cheng & Fox, 2017; Davidson & Lynch, 2002; Saville, 2016; see also Figure 3.1)

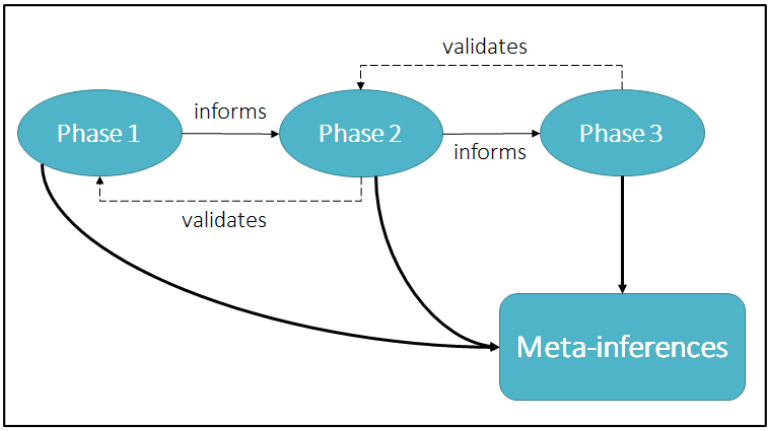


Figure 10.1. Interconnection between Phases 1, 2, and 3

Another way to display how connected all phases are, is to highlight the places of integration throughout the whole study. Table 10.1 provides this information, i.e., where the evidence of merging areas are situated, both within, between and across phases.

Table 10.1. Evidence for data integration within, between and across phases

Phase of the study	Meta-inferences
Phase 1	Chapter 7, Table 7.20. Phase 1 – MM exploratory sequential joint display
Between Phases 1 and 2	Chapter 8, Section 8.1 explains how Phase 1 informs Phase 2, and how Phase 2 validates Phase 1 (also detailed in Tables 8.10 and 8.11)
Phase 2	Chapter 8, Table 8.7. Final matrix of construct specification
Between Phases 2 and 3	Chapter 9, Sections 9.1 and 9.2 explain how Phase 2 informs Phase 3
	Chapter 9, Tables 9.15 and 9.16 detail how Phase 3 validates Phase 2
Phase 3	Chapter 9, Table 9.5 (test-takers' perspectives)
	Chapter 9, Table 9.6 (interlocutors' perspectives)
	Chapter 9, Table 9.7 (raters'/observers' perspectives)
	Chapter 9, Table 9.15 Language functions – Evidence from transcripts and from *observation checklists* (joint display)
	Chapter 9, Table 9.16 Behaviors of effective communication – Evidence from transcripts and from *indicator checklists* (joint display)
Integration across all Phases	Chapter 10, the overarching meta-inferences

As noted in Chapter 1 of this thesis, the rationale for the study grew out of concerns, voiced both empirically, and anecdotally, by aviation stakeholders and by the Aviation English testing community in Brazil (e.g. ANAC, 2016), about the nature of

communication in the aeronautical RT workplace and the challenges posed by the current ICAO testing policy. I was prompted by my sense as a test developer and by the literature review in the field of pilot-ATCO international communications and AE assessment, to consider that something was missing, i.e., that the construct of international RT communication was underrepresented in LPRs tests (Messick, 1989). The objective of such tests is to make decisions about pilots' and ATCOs' readiness to communicate in this intercultural workplace based on inferences about their ability to function efficiently as international RT communicators. Thus, underrepresenting the aviation RT construct may result in pilots and ATCOs mistakenly being granted a language proficiency license endorsement that allows them to fly internationally or to control international air traffic without the necessary communicative competencies. Equally threatening to the safety of aviation interactions over the radio, which was addressed in this study, is exempting native speakers of English from being formally tested and identifying them as being automatically competent to communicate in the specific context of RT communications.

As a result, I began my investigation in Phase 1 with an exploration of the intercultural RT communication context in order to identify intercultural factors that may affect the way pilots and ATCOs interact in the English language and to verify the extent to which those factors impact on safety, based on pilots' and ATCOs' perceptions. The qualitative analysis of six scenarios of authentic international RT communications disclosed a list of culturally influenced factors that may impact the safety of aviation communications over the radio. These factors were organized into a taxonomy, including six categories and 14 sub-categories, reported in parentheses: (a) power distance (power relations and deferential role); (b) face-work strategies (self-face concern and mutual-face

concern); (c) conflict management (conflictual direction, neutral direction and expectancy violations); (d) communication styles (directness and indirectness); (e) non-collaborative behavior (unprofessional tone, unprofessional attitude and non-compliance with rules); and collaborative behavior (professional attitude and supportiveness). By transforming the codes that originated within each sub-category into questions, as a building strategy, in the quantitative strand key aviation stakeholders answered an online survey. Quantitative findings indicate the situations that were perceived as the least and most frequent in participants' opinion, but confirmed that all of them do happen in international RT communications. Therefore, the scenarios selected appear to be representative of recurring instances of intercultural tension in the aviation workplace as evidenced by findings in Phase 1. Likewise, results suggest that pilots and ATCOs in the sample analyzed perceived, to a great extent, the potential threats of intercultural factors to safety. On top of that, it was also possible to find evidence from the survey open-ended responses to support the validation of the sub-categories from the provisional taxonomy (see Table 7.20 for Phase 1 joint display). Similarities but also some differences in perception across groups of participants (i.e., pilots vs. ATCOs, NSs vs. NNSs of English, male vs. females) were observed, as well as complex connections and relationships that exist among the recognized sub-categories. Participants' comments illustrated how some of the factors can pose challenges and even threats to safe RT communications, such as power relations, unprofessional attitudes, non-compliance with rules, differences in communication styles and conflict management, to name a few. These factors, not linguistically related, indicate that successful and effective

communications in this particular workplace context requires other competencies not addressed in prevailing models of communicative competence.

The interfaces of Aviation English and intercultural communications highlighted in Phase 1 suggested points of contact with other related disciplines. Thus, Phase 1 informed the selection of studies to be included in Phase 2 as part of a systematic review of theoretical and empirical research, from which three models of language use that account for the communicative demands of the international RT communication context were proposed. These models underscore the inclusion of the intercultural dimension, the use of AE as a lingua franca governed by the rules of the air traffic control system, i.e., background knowledge, in a dialogical and co-constructed manner. Four key areas of interest were identified, i.e., Aviation English, English as a lingua franca, Intercultural awareness/competence, and Interactional competence, and by drawing on previous research at the interface of AE and ELF, AE and ICA, and AE and IC, it was highlighted that in order to communicate effectively in a lingua franca setting, pilots and ATCOs, both NSs and NNSs of English, need to develop skills and competencies that go beyond language proficiency. Relevant features of the RT context that pilots and ATCOs should be aware of, know, use appropriately, and display as attitude for successful intercultural encounters over the radio were structured thematically according to the four major domains, generating the draft matrix of construct specification. Validation of the construct framework was conducted with 128 aviation stakeholders who participated in 26 focus group discussions. Findings revealed the components of the construct that were mentioned by the highest number of focus groups, i.e., *background knowledge*, followed by *professional tone and attitude* and *compliance with prescribed rules* and procedures

(*e.g., use of phraseology, readback/hearback*), which confirm results from previous studies (e.g., Douglas, 2014; Estival, 2018; Kim, 2012; Kim, 2018; Knock, 2014). The number of coding references for each component in the final matrix suggest that it is a very good representation of the international aviation RT construct, based on the perceptions of key stakeholders in the sample analyzed (see Table 8.8). These results also display the predominance of coding references for attitudes, awareness, skills and knowledge, in this order, which indicates that a lot more than just language proficiency is required for effective RT communications in multicultural contexts. These steps followed what Fox (2001) stated in relation to construct definition: "constructs are defined by networks of supporting theory and empirical research" (p. 270), which leads to meaningful construct-driven tests. In addition, the taxonomy of intercultural factors proposed in Phase 1 was also validated in Phase 2, based on aviation stakeholders' perceptions from focus group discussions (see Tables 8.10 and 8.11). The analysis revealed eight emerging sub-categories organized in two new categories, comprising *layers of culture* and *individual traits*. Not surprisingly, the sub-categories of *non-compliance with rules*, which includes the lack of adherence to standard phraseology and other prescribed procedures, *unprofessional* attitude and *unprofessional* tone were substantially mentioned during the discussions.

Having empirical evidence in my study from Phases 1 and 2 to suggest that important components of the construct are not included in the ICAO testing policy and, therefore, not addressed in current LPR tests, my response as a test developer was to further elaborate the model of test development that Fulcher and Davidson (2007, 2009) have described, aiming to provide an example of how the *Evidence models* and *Task*

models can be accomplished in a high-stakes specific purpose assessment context. As a result, Phase 3 addressed the operationalization of the specified construct as test tasks that can elicit the behaviors indicative of effective RT communication. The design of two draft pilot tasks and analysis of data from their pilot testing suggest that the aviation radiotelephony-specific communicative construct, as opposed to a language proficiency construct alone, can be operationalized as test tasks. The discussion of findings based on the qualities of good testing practice indicate so far that situational authenticity, interactional authenticity, practicality and impact of the tasks were generally perceived as positive by the research participants, i.e., those who acted as test-takers, interlocutors and observers. However, issues of reliability, in relation to the need for clearer instructions to guide test-takers throughout the tasks, and to the effect of interlocutors' behavior on test-takers' performance, were pointed out, as well as issues of validity, in terms of the controversial topic of what aspects of the construct should be measured in the specific purposes testing of pilots and ATCOs (Tables 9.5, 9.6 and 9.7). Yet, data integration in Phase 3 indicated a moderate agreement between the language functions detected in the transcripts of task performances and the responses from the *Observation Checklists* (Table 9.15), whereas a greater agreement was noticed between responses from the *Indicator Checklists of behaviors* and performance transcripts (Table 9.16), signposting the potential of the tasks to test the construct.

Finally, by exploring the multicultural context of international RT communications and by following the steps of the test development process, represented as layers and sub-layers of architectural documentation (Fulcher & Davidson, 2007, 2009), this study demonstrated how the communicative demands of pilots and ATCOs

can be specified within a construct framework and operationalized as test tasks. This was accomplished with the participation of key aviation stakeholders from different 'linguaculture' backgrounds in all phases of the study, taking into account what these domain experts considered as relevant for communicative effectiveness in pilot-ATCO interactions. Therefore, results corroborate the idea that, in LSP testing, it is crucial to give voice to domain experts in relation to what they consider relevant for successful communications (Douglas, 2000; Douglas 2004; Jacoby & McNamara, 1999; Knoch & Mcqueen, 2016; Pill, 2016).

In conclusion, not only did this study demonstrate how to move from a domain analysis to a more detailed specification of the construct, all the way through the development and initial pilot testing of tasks, but it also explained how to analyze what a test developer does along this process. Hence, I have gone so far in answering my research question, but not gone through all phases of the test development process. What follows would be the refinement of the tasks through further consultation with stakeholder groups and their actual trialing with live test-takers and raters.

10.2 Limitations and Directions for Future Research

Although I tried my best to conduct this research study based on rigorous methods of data collection and analysis, some limitations need to be addressed, which indicate the need to further investigate the issues addressed in this thesis. These are discussed below:

- in Phase 1, only a small number of interactions were analyzed in the qualitative strand, collected using purposive sampling. A future study employing random sampling and a wider variety of scenarios from different regions in the world

could reveal additional problems related to cultural differences and/or an indication of their actual frequencies;

- due to the small number of participants in the quantitative strand (Phase 1) who responded to the online survey, generalization of the findings to other contexts should be undertaken with caution;
- in Phase 2, the logistics and time constraints in conducting 26 focus groups in workshops held at two international conferences focusing on the use, training and testing of Aviation English posed some challenges to data collection. As in the same workshop room four to six groups were simultaneously discussing different scenarios of authentic international RT communications in a tight timeframe, only written transcripts were provided to participants in order to elicit their comments. Despite the fact that all scenarios were publicly available on aviation-related websites, such as www.youtube.com and www.planecrashinfo.com, further studies would benefit from the analysis of audio-recordings from actual RT communications, with the affordances of the aural input, conducted separately in small groups, and each moderated by the researcher;
- an attempt was made to conduct individual interviews via Skype with pilots and ATCOs using the audio-recordings of the six scenarios as triggers to the discussions, but technological limitations also restricted this option;
- in Phase 3, given the restrictions in participants and contexts, i.e., participants playing the role of test-takers (pilots) and interlocutors (ATCOs) were speakers of the same L1 and members of the same professional group, it was not possible to

verify if all behaviors related to ELF competence and Intercultural awareness/competence would be elicited by the draft tasks;

- considering that the AETEs who volunteered to be test-takers were all experienced pilots with high level of English proficiency, a comparison of how performances would differ across a range of language ability and professional experience remains to be investigated;
- the inclusion of the complete list of language functions associated with aviation in the *Observation Checklist* limited participants' ability to fully respond to them, thus impacting the response rate. Therefore, findings related to the language functions' perceived importance should be interpreted with caution. In future pilot testing of tasks, the use of a revised checklist that privileges the most relevant language functions and places upfront the ones related to the fourth group, i.e., management of the dialogue, may yield better results.

Therefore, following Cheng and Fox's (2017) understanding of continued collection of evidence of test function and use as a validity requirement, and Saville's (2016) explanation of "*validation as a process* of accumulating evidence to support claims made about the underlying constructs and the appropriate inferences that can be drawn from the results" (p. 18), future research would most probably include the following:

- pilot testing the tasks again, and possibly another task, which might be built from the matrix of construct specification, with a multicultural group of test-takers and interlocutors, where there is not the same sharing of cultural understandings and L1;

- comparing the outcomes from the mono-cultural context with the multicultural setting, and conducting the necessary revisions and adjustments emerging from this step;
- moving from the "intermediate stage" (O'Hagan, Pill & Zhang, 2016, p.199) of using the *Indicator Checklist of behaviors indicative of effective communication* to the generation of new assessment criteria as a basis for rater training and standardization. This could include having the raters apply this "intermediate" tool and then synthesize findings on the basis of their commentary and reaction to the live performances, repeated as required to the point where no new information and no new adjustments appear to be warranted in order to get to the final assessment criteria;
- addressing issues of interlocutor's training and consistency, including the proposed improvements to the guiding information provided in their role-play cards;
- officially trialing the task or tasks in the context of an actual assessment with real test-takers, introducing the assessment criteria that should apply; and
- continuing to validate the tasks whilst building the interpretive argument of the test. Although not the framework privileged in this study, there are other exemplars of how arguments for test use are built, such as Kane's (2006) model of the interpretive versus validity argument, and Chapelle, Enright, and Jamieson's (2008) use of a validity argument to guide test design.

Moreover, in relation to the Aviation English testing context of Brazilian civil aviation pilots, findings from this study suggest the potential of the proposed role-play tasks to be

incorporated into the existing test, *Santos Dumont English Assessment*. In addition to the future steps suggested above, the Brazilian pilot testing context would also require the need for further investigations, as detailed below:

- before the main trial of the tasks with real Brazilian test-takers, conducting further pilot testing in order to verify how the findings of this study coupled with the suggestions made by Brazilian AETEs can best be combined to maximize the benefits of including the task in the Brazilian testing context, aiming to increase the validity of inferences drawn from test results;
- continuing to collecting feedback from test-takers, interlocutors and raters based on the qualities of good testing practice (Douglas, 2000), aiming to improve the overall quality of the tasks. Although feedback received from the Brazilian testing community helped to enhance my understanding of the task and its usefulness, feedback from actual test-takers, i.e., the key stakeholder group, and raters actually rating their performances, will be crucial to inform decision-making.; and
- carefully considering the existing sections and task types in the Brazilian test for pilots, so as to proceed with the test development process with the final goal of maximizing content coverage and minimizing construct underrepresentation.

10.3 Meta-inferences: Implications for Aviation Stakeholders

Despite the limitations discussed in the previous section, the attempt described in this study to move from a language-only approach to a broader view of professional communicative competence for intercultural communications in aviation suggests an encouraging perspective to construct definition and task design in a highly specialized TLU domain, which has implications for different stakeholders in aviation.

In Chapter 6, I drew on Tashakkori and Teddlie`s (2008) conceptualization of meta-inference as one of the ultimate goals of MM research. They define it as "an overall conclusion, explanation, or understanding developed through an integration of the inferences obtained from the qualitative and quantitative strands of a mixed method study" (p. 101), which is, in the authors' opinion, the real source of the value supplemented by mixed methods. I agree with them. Below, I discuss the meta-inferences or the implications of this study.

First and foremost, findings from this study indicate how a better understanding of the communicative demands of pilots and ATCOs involved in intercultural RT communications can be used to enhance their alignment with the specific purpose language testing in this context, and consequently, promote a positive washback. The connection of teaching, learning and assessment is based on the concept of washback, considered within the scope of the impact of tests on stakeholders (Cheng, 2005; Wall, 2005). This means that if the construct of a test fails to address any important feature of the TLU domain, most probably it will not foster a positive impact on teaching and learning. Therefore, as policy-makers are the ones who define testing mandates, which ultimately shape how tests will be designed (Cheng & Fox, 2017), their role in the process is a central one. They should be guided by current theories of language use, best practices in language testing and assessment, and findings from empirical research, like the ones presented in this thesis, rather than by "politics and expediency" (Moder and Halleck, 2009). However, policy-makers have a different understanding and viewpoint from test developers, given that they belong to a different cultural group, such that previous recommendations for a revision of the ICAO testing policy that includes a re-

definition of the testing construct for both NSs and NNSs of English (ANAC, 2016; Douglas, 2014; Garcia, 2015; Kim, 2012, 2013; McNamara, 2012a) have not been accepted to date. As Harding and McNamara (2017) argue in relation to native speakers, ELF assessment and aviation, "the wide scale adoption of … practices – which often effectively exempt native speakers from needing to demonstrate language skills crucial to ELF communication – provides evidence of an institutionalized conservatism … around the place of the native speaker in the language assessment policy" (p. 579). Yet, the more evidence we can provide the aviation industry that in the dynamic and intercultural context of aviation RT high levels of language proficiency are not enough to communicate effectively and to cope with the challenges of interacting with members of different cultures, the more likely it is that changes will gradually begin to happen.

Second, as a different approach to the assessment of pilots and ATCOs is called upon to address the views of those really engaged in RT communications and the realities of using AE as a lingua franca, i.e., an approach that shifts its main focus from linguistic form to functional effectiveness, findings can be used by test developers in the specific context of aviation communications to enhance a better representation of the construct in task design and in complete test forms. Task design will inevitably require a close collaboration with SME – pilots and SME – ATCOs in order to include unexpected situations in radiotelephony that would require more negotiation of meaning and more opportunities of plain English usage to achieve the goals of communication. All these efforts would provide a means to increase the validity of inferences drawn from tests scores and consequently better decisions. As pointed out by Fox (2001), "what matters in judging the value of inferences drawn from a test is whether or not the responses to it are

indicative of those that commonly occur in the context to which we are attempting to generalize" (p.268). Wrong decisions in high-stakes contexts may bring potential negative consequences to stakeholders in the field and also to society at large, as explained by Elder et al. (2017):

> By not acknowledging all that is relevant for successful communication in real-world situations, decisions are made to exclude individuals from participation in professional settings who may in fact be competent to practice and to allow others access to professional practice whose actual competence may cause problems of communication, with potentially serious, even fatal consequences. (p. 19)

Third, as Aviation English raters and interlocutors have an important role in ensuring the reliability of test scores, findings from this study have implications for training and standardizing these stakeholders involved in the assessment of pilots and ATCOs. Firstly, the inclusion of professional criteria that may not be so easily captured by all raters indicates that a great deal of training would need to be conducted, especially taking into consideration that ELEs, SMEs – pilots and SME – ATCOs, who normally form the body of assessors in this context, might have a different interpretation of these criteria. In a similar token, comments from participants in this study suggested that interlocutors' behavior may have a great impact on test-takers' performance, implying the need for training for consistency, which ultimately would secure fairness and equal chances to test-takers.

Further, pilots and ATCOs, who are both the actual communicators in international RT interactions and, in the case of non-native speakers, the actual test-takers, would benefit from a greater awareness of the multiple factors that may impact

their exchanges over the radio and, also, of the possible criteria with which they would be judged in test task performances. However, in the case of native speakers of English, this would represent a significant shift in understanding (and accepting!) that they should also acquire AE as an additional language system (Bieswanger, 2016; Estival, 2016; Intemann, 2008), for AE is not their native language. As this study has shown and is corroborated by Estival (2018), "the non-compliance of Native English Speakers with the phraseology is arguably as much a problem for international aviation communication as the difficulties that non-native speakers of English may have" (p. 38). Not only that, but an awareness of their shared responsibility for effective communication, which inevitably encompasses the components of the construct detailed in the domains of English as a lingua franca, Intercultural awareness/competence and Interactional competence, emerges as crucial for their performance in this intercultural context. In this sense, both the taxonomy of intercultural factors (Table 8.10) and the *Indicator Checklist of skills/behaviors indicative of effective communication* in aviation radiotelephony (Appendix Y) have a lot of potential to be used as learning tools.

Likewise, Aviation English teachers, curriculum developers, and teacher trainers should be aware of those learning needs, so that critical features of real-world communications can be properly addressed in the training of pilots and ATCOs, and so minimize possible threats to the safety of international operations. In adopting a new direction to teaching practices and activities, i.e., one that includes AE, ELF, ICA and IC, teachers would most likely benefit from the outcomes of this research study, such as drawing on the categories of intercultural factors (Phase 1) and the components of the matrix of construct specification (Phase 2) to design appropriate training objectives,

activities and materials. Actually, I have started doing that in two practical workshops that I delivered in international conferences (Monteiro, 2018a; 2019). In the first, by letting participants base their discussions on my research findings from Phase 1, they brainstormed possible training activities to address the categories of intercultural factors (see a description of the activity and summary of results in Monteiro, 2018b). In a similar way, findings from Phase 2, i.e., the matrix of construct specifications, were used to trigger discussions among workshop participants on how they could develop training activities to raise awareness, impart knowledge, develop skills and improve attitudes in relation to construct components in each of the four domains.

Additionally, Human Factors specialists and CRM instructors would also have such inventories (Tables 8.7 and 8.10) as guidance to be incorporated into pilots' and ATCOs' training sessions, with the aim of improving their awareness of the complex relation between language, culture and communication. It is also important to highlight that, given the globalized nature of the aviation industry with multicultural and multilingual actors, findings from this study have implications to other activities that require communication and information sharing, either over the radio or face-to-face. These would include interactions between captain-first officer, captain-cabin crew, pilot-maintenance personnel, pilot-dispatcher, pilot-company, pilot-safety inspectors, and a long list of many others instances in which effective intercultural communication is key for safe outcomes.

Finally, applied linguists working in this field and AE researchers could further work towards the collection of more data, possibly replicating one or more phases of this study in other contexts, so that a broader understanding of what really matters for

effective communication, as valued by domain experts, could be achieved. Alternatively, moving forward to the investigation of sequential phases of the test development process, researchers in the field could gather more evidence related to the usefulness of similar tasks, the relation between predicted and actual test-takers' performances, application of professional assessment criteria, and a number of other validity studies.

10.4 Meta-inferences: Implications for Different Fields of Inquiry

This study is not only useful in the context it was undertaken, but it also serves as a model for test development endeavors in general, especially when dealing with a complex and abstract construct. It provides an example of how both theory and empirical data can feed all the steps a test developer needs to go through in terms of validating what is being measured. In relation to other contexts of specific purpose language assessment, this study underscores the value of involving domain experts and stakeholders in the process of exploring the communicative context of interest, specifying and validating the construct to be measured, and designing and pilot testing tasks that generate the evidence needed to make valid inferences about the test-taker's specific purpose language ability. In addition to other language assessment settings where ELF competence, intercultural awareness/competence and interactional competence would likely be part of the construct, this study serves as an exemplar of how these competencies could be operationalized in task design.

Furthermore, a similar context in which communications shall be clear, concise and unambiguous, involving speakers of different cultural backgrounds and with obvious safety implications in the event of misunderstandings, is maritime communications (Noble, Vangehuchten & Parys, 2011). Therefore, findings from this study could be

applied in this and similar workplace settings to help overcome breakdowns in communication related to cultural differences, aiming to increase cooperation and effective exchanges involving ship-to-ship, ship-to-shore, and communication among onboard crew members.

Finally, the methodological approach chosen to undertake this study confirms that MM research is well suited to address complex research problems and questions, by providing the MM researcher with a range of alternatives and a selection of methodological mixes to choose in order to design a study which will enable him/her to make stronger inferences and draw conclusions that will contribute to the field under investigation. Not only that, but the multiphase MM design described in this thesis illustrates how points of data interface can occur at the design level, methods level and interpretation level of a MM study, and how the language testing and assessment field can benefit from a more rigorous framework within which to conduct research.

Afterthoughts

As a non-native speaker of English and member of different cultural and professional groups, the experience of conducting this research study and writing this thesis helped me to understand the challenges that I myself face while communicating in intercultural settings. This greater intercultural awareness not only made me a better researcher but also a better communicator.

Throughout this endeavor, my attempt was to produce scientific knowledge in the field of specific purpose language assessment in aviation RT communications. Yet, it is important to highlight that findings presented in this study represent my own interpretation of data, which may have been shaped by the values of the communities of practice I belong to and also by my individual perceptions and expectations. Nevertheless, my conclusions were drawn from a reflection of data collected from multiple sources, representing a range of viewpoints and cultural frames of reference. On the other hand, I also gained a range of interdisciplinary knowledge that will certainly shape my future work as a test developer, item writer, rater, and interlocutor trainer.

As a personal reflection, I acknowledge that this thesis could have been smaller, in terms of having a reduced scope. The culmination of the thesis could have been the validation of the matrix of construct specification in Phase 2. However, my real goal was to do something that would not only be important to our field but also directly applicable. Therefore, I have included in Phase 3 the movement from the sub-layer of 'construct framework', to 'evidence models' leading up to 'task models'. As a result, two role-play tasks were designed for the assessment of aviation professionals and their detailed task

characteristics and accompanying task specifications developed, as a further contribution to the field.

I also learned the benefits of a pragmatic approach to research and methodology, through applying the best that different paradigms can offer in order to address my research questions. Especially considering MM research, my experience was a very positive one, despite the complex MM design that I chose to apply. It enabled me to understand how combining quantitative information to the more detailed and richer qualitative data, from participants' personal accounts and transcripts of naturally occurring discourse, added to enrich the understanding of the phenomenon under investigation. Without this methodological mix I would have missed relevant information to draw my conclusions. For instance, in Phase 1 it was possible to confirm the taxonomy of intercultural factors' sub-categories by integrating discourse data (QUAL), survey responses (quan) and open-ended comments (qual) from participants. In Phase 3, participants' perceptions on the tasks, added to the integration of data from task performances and from the degree of agreement between raters in relation to behaviors and language functions, revealed the potential of the tasks to operationalize the international RT-specific construct.

In closing, although this study cannot address all the challenges involved in developing a specific purpose language assessment in the context of international aviation RT communication, the ongoing investigation proposed here of how best to operationalize the construct valued by aviation experts as test tasks may contribute to increase the validity of inferences drawn from tests, thus enhancing the safety of international pilot-ATCO communications.

References

Alderson, J. C. (2007). The challenge of (diagnostic) testing: Do we know what we are measuring? In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 21−39). Ottawa, ON: University of Ottawa Press.

Alderson, J. C. (2011). The politics of aviation English testing. *Language Assessment Quarterly, 8*(4), 386−403.

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation.* Cambridge: Cambridge University Press.

Alptekin, C. (2002). Towards intercultural communicative competence in ELT. *ELT Journal, 56*(1), 57−64.

ANAC (2016). *The need for revision of Annex 1 Language Proficiency SARPS for pilots and air traffic controllers.* Working Paper A39-WP/249. Rio de Janeiro, Brazil: Agência Nacional de Aviação Civil-ANAC http://www.icao.int/Meetings/a39/Documents/WP/wp_249_en.pdf. Accessed 13 March 2017

ANAC (2018). *Instrução Suplementar Nº 183-000. Processo de credenciamento de pessoa jurídica para aplicação do exame de proficiência linguística da ANAC* (Revisão D). Rio de Janeiro, Brazil: Agência Nacional de Aviação Civil-ANAC.

ANAC (2018, May 30). *Santos-Dumont English Assessment-Instructions for candidates.* Retrieved from https://www.anac.gov.br/en/licenses-and-language-proficiency/santos-dumont-english-assessment-instructions-for-candidates

Aragão, B. F. (2018). 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 to air traffic controllers]. In M. V. R. Scaramucci, P. Tosqui-Lucks, & S. M. Damião (Eds.), *Pesquisas sobre Inglês Aeronáutico no Brasil* (pp. 243−269). Campinas, SP, Brazil: Pontes Editores.

Austin, J. (1962). *How to do things with words.* Oxford: Oxford University Press.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In J. Fox et al. (Eds.), *Language testing reconsidered* (pp. 41−71). Ottawa, ON: University of Ottawa Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.

Baker, W. (2009). The cultures of English as a lingua franca. *TESOL Quarterly, 43*(4), 567−592.

Baker, W. (2011). Intercultural awareness: Modeling an understanding of cultures in intercultural communication through English as a lingua franca. *Language and Intercultural Communication*, *11*(3), 197–214.

Baker, W. (2012). From cultural awareness to intercultural awareness: Culture in ELT. *ELT Journal, 66*(1), 62–70.

Baker, W. (2015). Culture and complexity through English as a lingua franca: Rethinking competences and pedagogy in ELT. *Journal of English as a Lingua Franca, 4*(1), 9-30.

Baker, W. (2016). Culture and language in intercultural communication, English as a lingua franca and English language teaching: Points of convergence and conflict. In P. Holmes & F. Derwin (Eds.), *The cultural and intercultural dimensions of English as a lingua franca (1)* (pp.70−89). Retrieved from http://site.ebrary.com/lib/oculcarleton/reader.action?docID=11152694&ppg=15

Baker, W. (2017). English as a lingua franca and intercultural communication. In J. Jenkins, W. Baker, & M. Dewer (Eds.), *The Routledge Handbook of English as a Lingua Franca* (Chapter 2). Retrieved from https://ebookcentral-proquest-com.proxy.library.carleton.ca

Bakhtin, M. M. (1986). The problem of speech genre. In C. Emerson, & M. Holquist (Eds.), *The problem of speech genres and other essays*. (V. W. McGee, Trans.) (pp. 60−102). Austin: University of Texas Press.

Berger, P.L. & Luckmann, T. (1966). *The social construction of reality*. London, UK: Penguin.

Bieswanger, M. (2016). Aviation English: Two distinct specialized registers? In C. Schubert, & C. Sanchez-Stockhammer (Eds.), *Variational text linguistics: Revisiting register in English* (pp. 67−85). http://site.ebrary.com/lib/oculcarleton/reader.action?docID=11214985&ppg=25. Accessed 18 July 2016.

Bieswanger, M. (2018, May). *Language awareness and intercultural competence as human factors in pilot-controller communication.* Workshop delivered at the International Civil Aviation English Association International Conference, Daytona Beach/FL, USA.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge, UK: Cambridge University Press.

Burgoon, J. K., & Hubbard, M. S. E. (2005). Cross-cultural and intercultural applications of expectancy violations theory and interaction adaptation theory. In W. B. Gudykunst (Ed.), *Theorizing about intercultural communication* (pp. 149−171). Thousand Oaks, CA: Sage.

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Clevedon (UK): Multilingual Matters.

Byram, M., Gribkova, B., & Starkey, H. (2002). *Developing the intercultural dimension in language teaching: A practical introduction to teachers.* Strasbourg: Council of Europe, Language Policy Division. Available at: http://www.Irc.cornell.edu/director/intercultural.pdf.

Canagarajah, A. S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly, 3*, 229–42.

Canale, M., & Swain, M. (1980). Theoretical basis of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1−47.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In C. Richards, & R. W. Schmidt (Eds.), *Language and communication* (pp. 2−27). London: Longman.

Carbaugh, D. (1988). Comments on ‘‘culture’’ in communication inquiry. *Communication Reports, 1*, 38–41.

Carbaugh, D. (2007). Cultural discourse analysis: Communication practices and intercultural encounters. *Journal of Intercultural Communication Research, 36*(3), 167−182.

Celce-Murcia, M. (1995). The elaboration of sociolinguistic competence: Implications for teacher education. In: J. E. Alatis, C. A. Straehle, & M. Ronkin (Eds.), *Linguistics and the education of language teachers: Ethnolinguistic, psycholinguistic, and sociolinguistic aspects.* Proceedings of the Georgetown University, Round Table on Languages and Linguistics, 2005. Georgetown University Press, Washington DC, pp 699−710.

Chapelle, C. A., Enright, M. K., & Jamielson, J. M. (2008). *Building a validity argument for the test of English as a foreign language.* New York: Routledge.

Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge: Cambridge University Press.

Cheng, L. & Fox, J. (2013). Review of doctoral research in language assessment in Canada (2006–2011). *Language Teaching, 46*, 518−544.

Cheng, L., & Fox, J. (2017). *Assessment for the language classroom*. London, UK: Palgrave Macmillan.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The M. I. T. Press.

Clark, B. (2017). *Aviation English research project: Data analysis findings and best practice recommendations* (CAP 1375). Civil Aviation Authority, West Sussex, UK.

Cogo, A. & Dewey, M. (2012). *Analyzing English as a lingua franca.* London: Continuum.

Cogo, A., & House, J. (2017). The pragmatics of ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge Handbook of English as a Lingua Franca* (pp. 210−223). London & New York: Routledge.

Cogo, A. (2016). "They all take the risk and make the effort": Intercultural accommodation and multilinguism in a BELF community of practice. In L. Lapriore, & E. Grazzi (Eds.), *Intercultural communication: New perspectives from ELF* (pp. 365−383). Roma: Roma TRE-Press.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: SAGE Publications.

Creswell, J. W., & Miller, D. (2000). Determining validity in qualitative inquiry. *Theory and Practice, 39*(3), 124−130.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. (2nd ed.). Thousand Oaks, CA: Sage.

Culpeper, J. (1996). Towards an anatomy of impoliteness. *Journal of Pragmatics, 25*, 349−367.

Cushing, S. (1995). Pilot-air traffic control communications: It's not (only) what you say, it's how you say it. *Flight Safety Foundation, V. 14*, No. 7, July.

Davidson, F. & Fulcher, G. (2012). Developing test specifications for language assessment. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to second language assessment* (pp. 36−75). Cambridge, UK: Cambridge University Press.

Davidson, F., & Lynch, B. K. (2002). *Testcraft: A teacher's guide to writing and using language test specifications.* New Haven and London: Yale University Press.

Davies, A. (2009). Assessing world Englishes. *Annual Review of Applied Linguistics, 29*, 80−89.

Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in Intercultural Education, 10*, 241−266.

Deardorff, D. K. (2009). Implementing intercultural assessment. In D. K. Deardorff (Ed.), *The Sage Handbook of Intercultural Competence* (pp. 477−491). Thousand Oaks: Sage.

Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.

Douglas, D. (2001). Three problems in testing language for specific purposes: Authenticity, specificity, and inseparability. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, … K. O-Loughlin (Eds.), *Experimenting with Uncertainty - Essays in Honour of Alan Davies* (pp. 45-52). Cambridge: Cambridge University Press.

Douglas, D. (2004). Assessing the language of international civil Aviation: Issues of validity and impact. *Proceedings from the International Professional Communication Conference*, IEEE Professional Communication Society (pp. 248−252). Minneapolis: IEEE.

Douglas, D. (2014). Nobody seems to speak English today: Enhancing assessment and training in aviation English. *Iranian Journal of Language Teaching Research, 2*(2), 1−12.

Douglas, D., & Myers, R. K. (2000). Assessing the communication skills of veterinary students: Whose criteria? In A. Kunnan (Ed.). *Fairness in validation in language assessment. Selected papers from the 19th language testing research colloquium*. *Studies in Language Testing 9*. Cambridge: Cambridge University Press, 60–81.

Elder, C. (2016). Exploring the limits of authenticity in LSP testing: The case of a specific-purpose language test for health professionals. *Language Testing, 33*(2), 147−152.

Elder, C., & Davies, A. (2006). Assessing English as a lingua franca. *Annual Review of Applied Linguistics, 26*, 282–301.

Elder, C., & Harding, L. (2008). Language testing and English as an international language: Constraints and contributions. *Australian Review of Applied Linguistics, 31*(3), 34.1–34.11. doi: 10.2104/aral0834.

Elder, C., & McNamara, T. (2016). The hunt for 'indigenous criteria' in assessing communication in the physiotherapy work place. *Language Testing, 33*(2), 153–174.

Elder, C., McNamara, T., Kim, H., Pill, J., & Sato, T. (2017). Interrogating the construct of communicative competence in language assessment contexts: What the non-language specialist can tell us. *Language & Communication*, *57*, 14−21.

Elliot, M., & Lim, G. S. (2016). The development of a new reading task: A mixed methods approach. In A. J. Moeller, J. W. Creswell & N. Saville (Eds.), *Second language assessment and mixed methods research* (pp. 233−268). Cambridge: Cambridge University Press.

Emery, H. J. (2014). Developments in LSP testing 30 years on? The case of Aviation English. *Language Assessment Quarterly, 11*(2), 198−215.

Estival, D. (2016). Aviation English. In D. Estival, C. Farris, & B. Molesworth (Eds.), *Aviation English: A lingua franca for pilots and air traffic controllers* (chapter 2). Retrieved from http://site.ebrary.com/lib/oculcarleton/reader.action?docID=11139932&ppg=16

Estival, D. & Farris, C. (2016). Aviation English as a lingua franca. In D. Estival, C. Farris & B. Molesworth (Eds.), *Aviation English : A lingua franca for pilots and*

air traffic controllers (chapter 1). Retrieved from http://site.ebrary.com/lib/oculcarleton/reader.action?docID=11139932&ppg=16

Estival, D., Farris, C. & Molesworth, B. (Eds.). (2016). *Aviation English: A lingua franca for pilots and air traffic controllers.* Routledge. http://site.ebrary.com/lib/oculcarleton/reader.action?docID=11139932&ppg=16. Accessed 27 May 2016.

Estival, D. (2018). What should we teach native speakers? In J. Roberts & D. Estival (Eds.), *The Proceedings of the International Civil Aviation English Association Conference* (pp. 37−46).

EUROCONTROL (2006). *Air-Ground Communication Safety Study: Causes and Recommendations.* Retrieved from https://skybrary.aero/bookshelf/books/162pdf

Fantini, A. E. (2000). A central concern: developing intercultural communicative competence. *School for International Training Occasional Papers Series. Inaugural Issue.* Retrieved from http://www.adam-europe.eu/prj/2935/prd/8/1/develop-I-com.pdf (accessed 4/10/13).

Farris, C. (2016). ICAO language proficiency requirements. In D. Estival, C. Farris & B. Molesworth (Eds.), *Aviation English : A lingua franca for pilots and air traffic controllers* (chapter 3). Retrieved from http://site.ebrary.com/lib/oculcarleton/reader.action?docID=11139932&ppg=16

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed method designs – Principles and practices. *Health Services Research 48* (6 Pt 2), 2134−2156.

Fox, J. D. (2001). *It's all about meaning: L2 test validation in and through the landscape of an evolving construct* (Unpublished doctoral dissertation). McGill University, Montreal, Canada.

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.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book.* London and New York: Routledge.

Fulcher, G., & Davidson, F. (2009). Test architecture, test retrofit. *Language Testing 26*(1), 123−144.

Galaczi, E., & Khabbazbashi, N. (2016). Rating scale development: A multistage exploratory sequential design. In A. J. Moeller, J. W. Creswell, & N. Saville (Eds.), *Second language assessment and mixed methods research* (pp. 208−232). Cambridge: Cambridge University Press.

Gallois, C., Ogay, T., & Giles, H. (2005). Communication accommodation theory. In W. B. Gudykunst (Ed.), *Theorizing about intercultural communication* (pp. 121−148). Thousand Oaks, CA: Sage.

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* (Unpublished master`s thesis). Lancaster University, Lancaster, UK.

Garcia, A. C. M. (2017, April). *The end-users' perceptions of test content: Is it working?* Workshop delivered at the International Civil Aviation English Association International Workshop, Dubrovnik, Croatia.

Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology? *Journal of Mixed Methods Research, 2,* 7−22.

Gudykunst, W.B., Lee, C.M., Nishida, T., & Ogawa, N. (2005). Theorizing about intercultural communication: An introduction. In W. B. Gudykunst (Ed.), *Theorizing about intercultural communication* (pp. 3−26). Thousand Oaks, CA: Sage.

Guetterman, T. C., & Salamoura, A. (2016). Enhancing test validation through rigorous mixed methods components. In A. J. Moeller, J. W. Creswell, & N. Saville (Eds.), *Second language assessment and mixed methods research* (pp. 153−176). Cambridge: Cambridge University Press.

Hall, J. K. (1999). A prosaics of interaction: The development of interactional competence in another language. In E. Hinkel (Ed.), *Culture in second language teaching and learning* (pp. 137−151). New York: Cambridge University Press.

Hall, B., & Howard, K. (2008). A synergistic approach: Conducting mixed methods research with typological and systemic design considerations. *Journal of Mixed Methods Research, 2*, 248−269.

Harding, L. (2008). *The use of speakers with L2 accents in academic English listening assessment: A validation study.* (Unpublished doctoral dissertation). University of Melbourne, Melbourne, Australia.

Harding, L. (2014). Communicative language testing: Current issues and future research. *Language Assessment Quarterly, 11*, 86–197.

Harding, L. (2015). Adaptability and ELF communication: The next steps for communicative language testing? In J. Mader, & Z. Urkun (Eds.), *Language testing: Current trends and future needs.* IATEFL TEASIG.

Harding, L., & McNamara, T. (2017). Language assessment: The challenge of ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge Handbook of English as a Lingua Franca* (Chapter 45). Retrieved from https://ebookcentral-proquest-com.proxy.library.carleton.ca

Hazrati, A. (2015). Intercultural communication and discourse analysis: The case of Aviation English. *Procedia - Social and Behavioral Sciences, 192,* 244−251.

Helmreich, R. L. (1994). Anatomy of a system accident: The crash of Avianca Flight 052. *International Journal of Aviation Psychology, 4*, 265−284.

Helmreich, R. L., & Merritt, A. C. (1998). *Culture at work: National, organizational, and professional influences*. London: Ashgate.

Hofstede, G. (1991). *Cultures and organizations: Software of the mind*. London: McGraw- Hill.

Hofstede, G., & Hofstede, G. J. (2005). *Cultures and organizations: Software of the mind* (2nd ed.) New York, NY: McGraw-Hill.

Holmes, P., & Dervin, F. (2016). Introduction – English as a lingua franca and interculturality: Beyond orthodoxies. In P. Holmes & F. Derwin (Eds.),

Languages for intercultural communication and education: The cultural and intercultural dimensions of English as a lingua franca (1) (pp.1−29). Retrieved from http://site.ebrary.com/lib/oculcarleton/reader.action?docID=11152694&ppg=15

Hutchins, E. (1995a). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive Science, 19*, 265-288.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics.* (pp. 53−73). Harmondsworth, UK: Penguin.

Hynnienem, N. (2014). The common European framework of reference from the perspective of English as a lingua franca: What we can learn from a focus on language regulation. *Journal of English as a Lingua Franca, 3*(2), 293 – 316.

ICAEA (n.d.). Guidelines: ICAO LPR Test Design Guidelines. Retrieved from https://www.icaea.aero/icao-lpr-tdg/guidelines/

ICAO (1998). *ICAO Doc 9683. Human Factors Training Manual*. Montreal, Canada: International Civil Aviation Organization.

ICAO (2004a). *ICAO Doc 9835. Manual on the implementation of ICAO language proficiency requirements* (1st ed.). Montreal, Canada: International Civil Aviation Organization.

ICAO (2004b). Cross Cultural Factors in Aviation Safety. *Human Factors Digest, Nº 16*. Montreal, Canada: International Civil Aviation Organization.

ICAO (2007). *ICAO Doc 9432*. *Manual of radiotelephony* (4th ed.). Montreal, Canada: International Civil Aviation Organization.

ICAO (2010). *ICAO Doc 9835. Manual on the Implementation of ICAO Language Proficiency Requirements* (2nd ed.). Montreal, Canada: International Civil Aviation Organization.

ICAO (2011). *Annex 1 to the Convention on International Civil Aviation – Personnel Licensing* (11th ed.). Montreal, Canada: International Civil Aviation Organization.

ICAO (2014). *Annex 10 to the Convention on International Civil Aviation – Aeronautical Telecommunications – Volume II Communication Procedures including those with PANS status* (6th ed.). Montreal, Canada: International Civil Aviation Organization.

ICAO (2016). *Report of the Technical Commission on Agenda item 35*. Montreal, Canada: International Civil Aviation Organization.

Intemann, F. (2008). "Taipei ground, confirm your last transmission was in English…?" – An analysis of Aviation English as a world language. In C. Gnutzmann & F. Intemann (Eds.), *The globalization of English and the English language classroom* (2nd ed., pp. 71−88). Tübingen: Narr Francke Attempto Verlag.

Jacoby, S., & Ochs, E. (1995). Co-construction: An introduction. *Research on Social Interaction, 28* (3), 171−183.

Jacoby, S., & McNamara, T. (1999). Locating competence. *English for Specific Purposes, 18*(3), 213−241.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.

Jenkins, J. (2006a). Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly, 40*(1), 157-181.

Jenkins, J. (2006b). The spread of English as an international language: A testing time for testers. *ELT Journal, 60*(1), 42–50.

Jenkins, J., Cogo, A. & Dewey, M. (2011). Review of developments in research into English as lingua franca. *Language Teaching, 44*(3), 281−315.

Jenkins, J. & C. Leung (2013). English as a lingua franca. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 4, pp. 1605–1616). Doi: 10.1002/9781118411360.wbcla047.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed method research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14−26.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp.17−64). Westport, CT: American Council of Education and Praeger Series on Higher Education.

Kay, M. (2015, April). *Cross-cultural communications: Implications for training and safety*. Paper presented at the International Civil Aviation English Association International Workshop, Istanbul, Turkey.

Kay, M. (2018, May). *Developing cross-cultural communicative competence*. Workshop delivered at the International Civil Aviation English Association International Conference, Daytona Beach/FL, USA.

Kecskes, I. (2014). *Intercultural pragmatics*. New York: Oxford University Press.

Khalifa, H., & Docherty, C. (2016). Investigating the impact of international assessment: A convergent parallel mixed methods approach. In A. J. Moeller, J. W. Creswell, & N. Saville (Eds.), *Second language assessment and mixed methods research* (pp. 269−295). Cambridge: Cambridge University Press.

Kim, M. (2005). Culture-based conversational constraints theory. In W. B. Gudykunst (Ed.), *Theorizing about intercultural communication* (pp. 93−117). Thousand Oaks, CA: Sage.

Kim, H. (2012). *Exploring the construct of aviation communication: A critique of the ICAO language proficiency policy* (Unpublished doctoral thesis). University of Melbourne, Melbourne, Australia.

Kim, H. (2013). 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), 103–10.

Kim, H. (2018). What constitutes professional communication in aviation: Is language proficiency enough for testing purposes? *Language Testing, 35*(3), 403–426. https://doi.org/10.1177/0265532218758127

Kim, H. & Billington, R. (2016). Pronunciation and comprehension in English as a lingua franca communication: Effect of L1 influence in international aviation communication. *Applied Linguistics*. Advanced online publication. doi: 10.1093/applin/amv075

Kim, H. & Elder, C. (2009). Understanding aviation English as a lingua franca: Perceptions of Korean aviation personnel. *Australian Review of Applied Linguistics, 32*(3), 23.1–23.17.

Kim, H., & Elder, C. (2015). Interrogating the construct of aviation English: Feedback from test takers in Korea. *Language Testing, 32*(2), 129−149.

Knoch, U. (2014). 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, 77−86.

Knoch, U. & Mcqueen, S. (2016). Language assessment for the workplace. In D. Tsagari, & J. Banerjee (Eds.), *Handbook of Second Language Assessment* (chapter 18). Retrieved from http://site.ebrary.com/lib/oculcarleton/reader.action?docID=11174254&ppg=315

Kramsch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal, 70*(4), 366−372.

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). New York, NY: Routledge.

Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Li, S., & Wang, H. (2018). Traditional literature review and research synthesis. In A. Phakiti, P. De Costa, L. Plonsky, & S. Starfield (Eds.), *Palgrave Handbook of Applied Linguistics Research Methodology* (pp. 123−144). Palgrave.

Lloyd Evans, G. (2013, April). *Cross-cultural flight crew and communication.* Paper presented at the International Civil Aviation English Association Forum, Paris, France.

Mathews, E. & Albritton, A. C. (2014). Language analysis: There's more to aviation English than radio conversations between pilots and air traffic controllers. *AeroSafety World, 9*(6), 22−25.

Matusov, E. (2007). In search of 'the appropriate' unit of analysis for sociocultural research. *Culture & Psychology, 13*(3), 307−333.

Maxwell, J. A., & Miller, B. A. (2008). Categorizing and connecting strategies in qualitative data analysis. In S. N. Hesse-Biber, & P. Leavy (Eds.), *Handbook of Eemerging Methods* (pp. 461−477). New York, NY: Guilford Press.

McMillan, D. (1998). *"...Say again?..." Miscommunications in air traffic control* (Unpublished master's thesis). Queensland University of Technology, Brisbane, Australia.

McNamara, T. (1996). *Measuring second language performance*. New York: Longman.

McNamara, T. (2011). Managing learning: Authority and language assessment. *Language Teaching, 44*(4), 500–515.

McNamara, T. (2012a). *At last: Assessment and English as a lingua franca*. Plenary talk at 5th International Conference of English as a Lingua Franca, 24 May, Istanbul.

McNamara, T. (2012b). English as lingua franca: the challenge for language testing. *Journal of English as a Lingua Franca, 1*(1), 199−202.

McNamara, T. F., & Roever, C. (2006). Validity and the social dimension of language testing. In *Language testing: The social dimension* (pp. 9−42) Malden, MA: Blackwell.

McNamara, T. (2007). Language assessment in foreign language education: The struggle over constructs. *The Modern Language Journal, 91*, 280−282.

Mell, J. & Godmet, C. (1997). *Aeronautical Radiotelephony Communicative Functions. Direction de la Navigation Aerienne*. DNA8 (F). Reprinted in Appendix B to the ICAO Doc 9835, with permission.

Merritt, A. C., & Helmreich, R. L. (1996). Human factors on the flight deck: The influence of national culture. *Journal of Cross-Cultural Psychology, 27*, 5−24.

Merritt, A. C. (2000). Culture in the cockpit: Do Hofstede's dimensions replicate? *Journal of Cross-Cultural Psychology, 31*(3), 283−301.

Messick, S. (1989). Validity. In R. L. Linn (Ed.). *Educational Measurement.* (3nd ed., pp.13−103). New York, NY: American Council on Education and Macmillan.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13−23.

Messick, S. (1996). Validity and washback in language testing. *ETS Research Report Series*, i–18. doi:10.1002/j.2333-8504.1996.tb01695.x

Mislevy, R. J., Almond, R. G. & Lukas, J. F. (2003). *A brief introduction to evidence-centered design*. Research Report RR-03-16. Princeton, NJ: Educational Testing Service.

Mislevy, R. J., Steinberg L. S., & Almond R. C. (2003). On the structure of assessment arguments. *Measurement: Interdisciplinary Research and Perspectives 1*(1), 3–62.

Moder, C. L. (2013). Aviation English. In B. Paltridge, & S. Starfield (Eds.), *The Handbook of English for Specific Purposes* (pp. 227−242). Malden, MA, USA: Wiley Blackwell.

Moder, C. L., & Halleck, G. B. (2009). Planes, politics and oral proficiency: Testing international air traffic controllers. *Australian Review of Applied Linguistics, 32*(3), 25.1-25.16. doi:10.2104/aral0925.

Moeller, A. (2016). The confluence of language assessment and mixed methods. In A. J. Moeller, J. W. Creswell, & N. Saville (Eds.), *Second language assessment and mixed methods research* (pp. 3−16). Cambridge: Cambridge University Press.

Monteiro, A. L. T. (2009). *Comunicações entre pilotos e controladores de vôo: fatores lingüísticos, discursivo-interacionais e interculturais [Pilot-controller communications: linguistic, discursive-interactional and intercultural factors]* (Unpublished master's thesis). Universidade Federal do Rio de Janeiro, Rio de Janeiro-RJ, Brasil.

Monteiro, A. L. T. (2012). Radiotelephony communications: Threats in a multicultural context. *Aviation in Focus, 3*(2), 44−66.

Monteiro, A. L. T. (2013, April). *Threats to oral comprehension in a multicultural context*. Paper presented at the International Civil Aviation English Association Forum, Paris, France.

Monteiro, A. L. T. (2014, November). *LPR for pilots and ATCOs: Overview of skills and competencies*. Paper presented at the International Civil Aviation English Association International Workshop, Buenos Aires, Argentina.

Monteiro, A. L. T. (2016a, March). *Pilot-air traffic controller interactions and intercultural communicative competence: An issue for non-native speakers alone?* Poster session presented at the 11th Annual Graduate Symposium of the Society of Applied Linguistics and Discourse Studies, Carleton University, Ottawa, Canada.

Monteiro, A. L. T. (2016b). *Exploring intercultural factors in international pilot-air traffic controller communications: Validating a taxonomy using mixed methods research*. Unpublished manuscript in ALDS 6102, Carleton University, Ottawa, Canada.

Monteiro, A. L. T. (2016c). *Exploring the role of intercultural awareness and interactional competence: The case of aviation English as a lingua franca.* Unpublished manuscript in ALDS 6109, Carleton University, Ottawa, Canada.

Monteiro, A. L. T. (2017, April). *Language, culture and effective radiotelephony communications: Analyzing case studies from aviation stakeholders' perspectives*. Paper presented at the International Civil Aviation English Association International Workshop, Dubrovnik, Croatia.

Monteiro, A. L. T. (2018a, May). *Exploring intercultural factors in international pilot-air traffic controller communications.* Workshop delivered at the International Civil Aviation English Association International Conference, Daytona Beach/FL, USA.

Monteiro, A. L. T. (2018b). Exploring intercultural factors in international pilot-air traffic controller communications: Validating a taxonomy using mixed methods research. In J. Roberts, & D. Estival (Eds.), *The Proceedings of the International Civil Aviation English Association (2018) Conference* (pp. 4−36). Daytona Beach/FL, USA.

Monteiro, A. L. T. (2019, May). *From a language-only approach to a broader view of communicative competence for intercultural communications in aviation.* Workshop delivered at the International Civil Aviation English Association International Conference. Tokyo, Japan.

Morrow, D. G., & Rodvold, M. (1998). Communication issues in air traffic control. In Molensky, M., & Stein, E. (Eds.), *Human factors in air traffic control* (pp. 421−456). San Diego: Academic Press.

Morse, J. M. (2015). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research*, *25*(9), 1212−1222.

Noble, A., Vangehuchten, L., & Van Parys, W. (2011). Communication for maritime purpose: Some exploratory results of a survey-based study on intercultural and linguistic features. *International Journal of Applied Linguistics*, *162*(1), 111−133.

O'Hagan, S., Pill., J. & Zhang, Y. (2016). Extending the scope of speaking assessment criteria in a specific-purpose language test: Operationalizing a health professional perspective. *Language Testing, 33*(2) 195−216.

Onwuegbuzie, A.J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools, 13*, 48−63.

Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of Mixed Methods in Social and Behavioral Research* (pp. 351−383).Thousand Oaks, CA: Sage.

O'Sullivan, B. (2012). Assessment issues in language for specific purposes. *The Modern Language Journal, 96* (Focus Issue), 71−88.

O'Sullivan, B., Weir, C. J., & Saville, N. (2002). Using observation checklists to validate speaking-test tasks. *Language Testing, 19*(1), 33–56. doi 10.1191/0265532202lt219oa

Pacheco, A. (2018, May). *Inter-cultural issues in air-ground communication: A case study – triggers for miscommunication*. Workshop delivered at the International Civil Aviation English Association International Conference, Daytona Beach/FL, USA.

Pallant, J. (2007). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows* (3rd ed.). Maidenhead (UK): Open University Press, McGraw Hill.

Pill, J. (2016). Drawing on indigenous criteria for more authentic assessment in a specific purpose language test: Health professionals interacting with patients. *Language Testing, 33*(2), 175−193.

Plough, I., Banerjee, J., & Iwashita, N. (2018). Interactional competence: genie out of the bottle. *Language Testing, 35,* 427−445.

Ragan, P. (2004). Cross cultural communication in Aviation. *Proceedings of the 14th European Symposium on Language for Special Purposes.* Surrey, BC, Canada.

Read, J. & Knoch, U. (2009). Clearing the air: Applied linguistic perspectives on aviation communication. *Australian Review of Applied Linguistics, 32*(3), 21.1−21.11.

Riazi, A. M., & Candlin, C. N. (2014). Mixed-methods research in language teaching and learning: Opportunities, issues and challenges. *Language Teaching, 47*(2), 135–173.

Roever, C., & Kasper, G. (2018). Speaking in turns and sequences: Interactional competence as a target construct in testing speaking. *Language Testing, 35*(3), 331–355. https://doi.org/10.1177/0265532218758128

Saldaña, J. (2009). *The coding manual for qualitative researchers*. Los Angeles: CA: Sage.

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). Thousand Oaks, CA: Sage.

Sänne, J. M. (1999). *Creating safety in air traffic control.* Lund: Arkiv Förlag.

Saville, N. (2016). Managing language assessment systems and mixed methods. In A. J. Moeller, J. W. Creswell & N. Saville (Eds.), *Second language assessment and mixed method research* (pp. 17-31). Cambridge: Cambridge University Press.

Scollon, R., & Scollon, S. B. K. (2001). *Intercultural communication: A discourse approach*. Malden, MA: Blackwell.

Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a lingua franca. *International Journal of Applied Linguistics*, *11*(2), 133–158.

Seidlhofer, B. (2009). Accommodation and the idiom principle in English as a lingua franca. *Intercultural Pragmatics, 6*(2), 195-215.

Shohamy, E. (2017). ELF and critical language testing. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge Handbook of English as a Lingua Franca* (Chapter 46). Retrieved from https://ebookcentral-proquest-com.proxy.library.carleton.ca

Snow, D. (2018). Intercultural communication in English courses in Asia: What should we teach about? In A. Curtis, & R. Sussex (Eds.), *Intercultural communication in Asia: Education, language and values* (pp. 55−70) Cham, Switzerland: Springer.

Sussex, R., & Curtis, A. (2018). Introduction. In A. Curtis, & R. Sussex (Eds.), *Intercultural communication in Asia: Education, language and values* (pp. 1−18) Cham, Switzerland: Springer.

Sweeney, E. & Zhu, H. (2010). Accommodating toward your audience: Do native speakers of English know how to accommodate their communication strategies

toward nonnative speakers of English? *Journal of Business Communication, 47*(4), 477-504. doi: 10.1177/0021943610377308.

Tajima, A. (2004). Fatal miscommunication: English in aviation safety. *World Englishes, 23*(3), 451−470.

Tashakkori, A., & Teddlie, C. (2008). Quality of inference in mixed methods research: Calling for an integrative framework. In M. M. Bergman (Ed.), *Advances in mixed methods research: Theories and applications* (pp. 101−119). London: Sage.

Tashakkori, A., & Teddlie, C. (2010). Putting the human back in ‘‘Human Research Methodology’’: The researcher in mixed methods research. *Journal of Mixed Methods Research, 4*(4), 271−177.

Taylor, L. (2005). Washback and impact. *ELT Journal, 59*(2), 154−155.

Taylor, L. (2006). The changing landscape of English: Implications for language assessment. *ELT Journal, 60*, 51–60.

Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods in Social and Behavioral Research* (pp. 3-50).Thousand Oaks, CA: Sage.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed method research*. Thousand Oaks, CA: Sage.

Ting-Toomey, S. (2005). The matrix of face: An updated face-negotiation theory. In W. B. Gudykunst (Ed.), *Theorizing about intercultural communication* (pp. 71−92). Thousand Oaks, CA: Sage.

van Batenburg, E. S. L., Oostdam, R. J., van Gelderen, A. J. S., & de Jong, N. H., (2018). Measuring L2 speakers' interactional ability using interactive speech tasks. *Language Testing, 35*(1), 75−100.

Vogt, W. P. (2007). *Quantitative research methods for professionals.* Boston, MA: Pearson Education, Inc.

Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. Cambridge, UK: Cambridge University Press.

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach.* Basingstoke, UK: Palgrave Macmillan.

Wu, D. D., & Li, C. (2018). Emotional branding on social media: A cross-cultural discourse analysis in global brands on twitter and weibo. In A. Curtis, & R. Sussex (Eds.), *Intercultural communication in Asia: Education, language and values* (pp. 225−240) Cham, Switzerland: Springer.

Wyss-Bühlmann, E. (2005). *Variation and co-operative communication strategies in air traffic control English.* Bern: Peter Lang.

Youn, S. J. (2013). *Validating task-based assessment of L2 pragmatics in interaction using mixed methods* (Doctoral thesis). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3577270)

Young, R. (2011). Interactional competence in language learning, teaching and testing. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 426-443). New York: Routledge.

Zhu, H. (2015). Identifying research paradigms. In H. Zhu (Ed.), *Research methods in intercultural communication: A practical guide*. Retrieved from http://web.b.ebscohost.com/ehost/ebookviewer/ebook/bmxlYmtfXzEwOTIzNjFfX0FO0?sid=a0496496-674d-4929-a765-68771daea080@sessionmgr102&vid=0&format=EB&lpid=lp_3&rid=0

Ziegler, N., & Kang, L. (2016). Mixed methods designs. In A. J. Moeller, J. W. Creswell, & N. Saville (Eds.), *Second language assessment and mixed methods research* (pp. 51−83). Cambridge: Cambridge University Press.

Appendix A - ICAO Rating Scale

LEVEL	PRONUNCIATION ASSUMES A DIALECT AND/OR ACCENT INTELLIGIBLE TO THE AERONAUTICAL COMMUNITY.	STRUCTURE RELEVANT GRAMMATICAL STRUCTURES AND SENTENCE PATTERNS ARE DETERMINED BY LANGUAGE FUNCTIONS APPROPRIATE TO THE TASK.	VOCABULARY	FLUENCY	COMPREHENSION	INTERACTIONS
EXPERT 6	Pronunciation, stress, rhythm, and intonation, though possibly influenced by the first language or regional variation, almost never interfere with ease of understanding.	Both basic and complex grammatical structures and sentence patterns are consistently well controlled.	Vocabulary range and accuracy are sufficient to communicate effectively on a wide variety of familiar and unfamiliar topics. Vocabulary is idiomatic, nuanced, and sensitive to register.	Able to speak at length with a natural, effortless flow. Varies speech flow for stylistic effect, e.g. to emphasize a point. Uses appropriate discourse markers and connectors spontaneously.	Comprehension is consistently accurate in nearly all contexts and includes comprehension of linguistic and cultural subtleties.	Interacts with ease in nearly all situations. Is sensitive to verbal and non-verbal cues, and responds to them appropriately.
EXTENDED 5	Pronunciation, stress, rhythm, and intonation, though influenced by the first language or regional variation, rarely interfere with ease of understanding.	Basic grammatical structures and sentence patterns are consistently well controlled. Complex structures are attempted but with errors which sometimes interfere with meaning.	Vocabulary range and accuracy are sufficient to communicate effectively on common, concrete, and work related topics. Paraphrases consistently and successfully. Vocabulary is sometimes idiomatic.	Able to speak at length with relative ease on familiar topics, but may not vary speech flow as a stylistic device. Can make use of appropriate discourse markers or connectors.	Comprehension is accurate on common, concrete, and work related topics and mostly accurate when the speaker is confronted with a linguistic or situational complication or an unexpected turn of events. Is able to comprehend a range of speech varieties (dialect and/or accent) or registers.	Responses are immediate, appropriate, and informative. Manages the speaker/listener relationship effectively.
OPERATIONAL LEVEL 4	**Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding.**	**Basic grammatical structures and sentence patterns are used creatively and are usually well controlled. Errors may occur, particularly in unusual or unexpected circumstances, but rarely interfere with meaning.**	**Vocabulary range and accuracy are usually sufficient to communicate effectively on common, concrete, and work related topics. Can often paraphrase successfully when lacking vocabulary in unusual or unexpected circumstances.**	**Produces stretches of language at an appropriate tempo. There may be occasional loss of fluency on transition from rehearsed or formulaic speech to spontaneous interaction, but this does not prevent effective communication. Can make limited use of discourse markers or connectors. Fillers are not distracting.**	**Comprehension is mostly accurate on common, concrete, and work related topics when the accent or variety used is sufficiently intelligible for an international community of users. When the speaker is confronted with a linguistic or situational complication or an unexpected turn of events, comprehension may be slower or require clarification strategies.**	**Responses are usually immediate, appropriate, and informative. Initiates and maintains exchanges even when dealing with an unexpected turn of events. Deals adequately with apparent misunderstandings by checking, confirming, or clarifying.**
PRE-OPERATIONAL 3	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation and frequently interfere with ease of understanding.	Basic grammatical structures and sentence patterns associated with predictable situations are not always well controlled. Errors frequently interfere with meaning.	Vocabulary range and accuracy are often sufficient to communicate on common, concrete, or work related topics but range is limited and the word choice often inappropriate. Is often unable to paraphrase successfully when lacking vocabulary.	Produces stretches of language, but phrasing and pausing are often inappropriate. Hesitations or slowness in language processing may prevent effective communication. Fillers are sometimes distracting.	Comprehension is often accurate on common, concrete, and work related topics when the accent or variety used is sufficiently intelligible for an international community of users. May fail to understand a linguistic or situational complication or an unexpected turn of events.	Responses are sometimes immediate, appropriate, and informative. Can initiate and maintain exchanges with reasonable ease on familiar topics and in predictable situations. Generally inadequate when dealing with an unexpected turn of events.
ELEMENTARY 2	Pronunciation, stress, rhythm, and intonation are heavily influenced by the first language or regional variation and usually interfere with ease of understanding.	Shows only limited control of a few simple memorized grammatical structures and sentence patterns.	Limited vocabulary range consisting only of isolated words and memorized phrases.	Can produce very short, isolated, memorized utterances with frequent pausing and a distracting use of fillers to search for expressions and to articulate less familiar words.	Comprehension is limited to isolated, memorized phrases when they are carefully and slowly articulated.	Response time is slow, and often inappropriate. Interaction is limited to simple routine exchanges.
PRE-ELEMENTARY 1	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.

Appendix B - Ethics Board: Initial clearance (Project number: 103859)

Carleton University
Research Ethics Office
Research Ethics Board
511 Tory, 1125 Colonel By Drive
Ottawa, ON K1S 5B6 Canada
Tel: 613-520-2517, ethics@carleton.ca

Ethics Clearance Form – New Clearance

This is to certify that the Carleton University Research Ethics Board has examined the application for ethical clearance. The REB found the research project to meet appropriate ethical standards as outlined in the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Human, 2nd edition*, and the *Carleton University Policies and Procedures for the Ethical Conduct of Research*.

Date of Clearance: December 14, 2015
Researcher: Janna Dorothy Fox (Primary Investigator)
Ana Lucia Tavares Monteiro (Student Research: Ph.D. Student)
Department: Faculty of Arts and Social Sciences\Linguistics and Language Studies (School of)
University: Carleton University
Project Number: 103859
Project Title: Exploring Intercultural Factors in International Pilot-Air Traffic Controller Communications: Validating a Taxonomy Using Mixed Methods Research [Ana Tavares Monteiro]
Funder (if applicable): N/A

Clearance Expires: May 31, 2016

All researchers are governed by the following conditions:

Annual Status Report: You are required to submit an Annual Status Report to either renew clearance or close the file. Failure to submit the Annual Status Report will result in the immediate suspension of the project. Funded projects will have accounts suspended until the report is submitted and approved.

Changes to the project: Any changes to the project must be submitted to the Carleton University Research Ethics Board for approval. All changes must be approved prior to the continuance of the research.

Adverse events: Should a participant suffer adversely from their participation in the project you are required to report the matter to the Carleton University Research Ethics Board. You must submit a written record of the event and indicate what steps you have taken to resolve the situation.

Suspension or termination of clearance: Failure to conduct the research in accordance with the principles of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, 2nd edition and the Carleton University Policies and Procedures for the Ethical Conduct of Research may result in the suspension or termination of the research project.

Louise Heslop
Chair, Carleton University Research Ethics Board

Andy Adler
Vice-Chair, Carleton University Research Ethics Board

Appendix C - Ethics Board: Change to Protocol clearance (Project number: 103859)

Research Compliance Office
511 Tory | 1125 Colonel By Drive
Ottawa, Ontario K1S 5B6
613-520-2600 Ext: 2517
ethics@carleton.ca

CERTIFICATION OF INSTITUTIONAL ETHICS CLEARANCE

The Carleton University Research Ethics Board-A (CUREB-A) has granted ethics clearance for changes to protocol to the research project described below and research may now proceed.
CUREB-A is constituted and operates in compliance with the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS2).
Ethics Clearance ID: Project # 103859
Project Team Members: **Mr. Janna Fox (Primary Investigator)**
Ana Lucia Tavares Monteiro (Student Research: Ph.D. Student)
Project Title: **Exploring Intercultural Factors in International Pilot-Air Traffic Controller Communications: Validating a Taxonomy Using Mixed Methods Research [Ana Tavares Monteiro]**
Funding Source (if applicable):
Effective: **April 20, 2017** Expires: **May 31, 2017**

Please email the Research Compliance Coordinators at ethics@carleton.ca if you have any questions or if you require a clearance certificate with a signature.
CLEARED BY: **Date: April 20, 2017**
Andy Adler, PhD, Chair, CUREB-A
Shelley Brown, PhD, Vice Chair, CUREB-A

Appendix D - Ethics Board: Renewal (Project number: 103859)

Research Compliance Office
511 Tory |1125 Colonel By Drive
Ottawa, Ontario K1S 5B6
613-520-2600 Ext: 4085
ethics@carleton.ca

CERTIFICATION OF INSTITUTIONAL ETHICS CLEARANCE

The Carleton University Research Ethics Board-A (CUREB-A) at Carleton University has renewed ethics approval for the research project detailed below. CUREB-A is constituted and operates in compliance with the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS2).

Title: Exploring Intercultural Factors in International Pilot-Air Traffic Controller Communications: Validating a Taxonomy Using Mixed Methods Research [Ana Tavares Monteiro]

Protocol #: 103859

Project Team Members: Mr. Janna Fox (Primary Investigator)
Ana Lucia Tavares Monteiro (Student Research: Ph.D. Student)

Department and Institution: Faculty of Arts and Social Sciences\Linguistics and Language Studies (School of), Carleton University

Funding Source (If applicable):

Effective: April 20, 2017 Expires: April 30, 2018.

Restrictions:

This certification is subject to the following conditions:

1. Clearance is granted only for the research and purposes described in the application.

2. Any modification to the approved research must be submitted to CUREB-A. All changes must be approved prior to the continuance of the research.

3. An Annual Application for the renewal of ethics clearance must be submitted and cleared by the above date. Failure to submit the Annual Status Report will result in the closure of the file. If funding is associated, funds will be frozen.

4. A closure request must be sent to CUREB-A when the research is complete or terminated.

5. Should any participant suffer adversely from their participation in the project you are required to report the matter to CUREB-A.

6. It is the responsibility of the student to notify their supervisor of any adverse events, changes to their application, or requests to renew/close the protocol.

Appendix E - Ethics Board: Initial Clearance (Project number: 107816)

Carleton UNIVERSITY
Canada's Capital University
Office of Research Ethics and Compliance
5110 Human Computer Interaction Bldg | 1125 Colonel By Drive
| Ottawa, Ontario K1S 5B6
613-520-2600 Ext: 2517
ethics@carleton.ca

CERTIFICATION OF INSTITUTIONAL ETHICS CLEARANCE

The Carleton University Research Ethics Board-A (CUREB-A) has granted ethics clearance for the research project described below and research may now proceed. CUREB-A is constituted and operates in compliance with the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS2).

Ethics Protocol Clearance ID: Project # 107816

Project Team Members: Ana Lucia Tavares Monteiro (Primary Investigator)
Janna Fox (Research Supervisor)

Project Title: Exploring Intercultural Factors in International Pilot-Air Traffic Controller Communications: Validating a Taxonomy Using Mixed Methods Research (Phase 4) [Ana Lucia Tavares Monteiro]

Funding Source (If applicable):

Effective: **September 21, 2017** Expires: **September 30, 2018**.

Restrictions:

This certification is subject to the following conditions:

1. Clearance is granted only for the research and purposes described in the application.
2. Any modification to the approved research must be submitted to CUREB-A via a Change to Protocol Form. All changes must be cleared prior to the continuance of the research.
3. An Annual Status Report for the renewal of ethics clearance must be submitted and cleared by the renewal date listed above. Failure to submit the Annual Status Report will result in the closure of the file.If funding is associated, funds will be frozen.
4. A closure request must be sent to CUREB-A when the research is complete or terminated.
5. Should any participant suffer adversely from their participation in the project you are required to report the matter to CUREB-A.

Failure to conduct the research in accordance with the principles of the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans 2nd edition* and the *Carleton University Policies and Procedures for the Ethical Conduct of Research* may result in the suspension or termination of the research project.

Please contact the Research Compliance Coordinators, at ethics@carleton.ca, if you have any questions or require a clearance certificate with a signature.

CLEARED BY: **Date: September 21, 2017**

Andy Adler, PhD, Chair, CUREB-A

Bernadette Campbell, PhD, Vice-Chair, CUREB-A

Appendix F - Ethics Board: Renewal (Project number: 107816)

Carleton UNIVERSITY
Canada's Capital University

Office of Research Ethics
5110 Human Computer Interaction Bldg | 1125 Colonel By Drive
Ottawa, Ontario K1S 5B6
613-520-2600 Ext. 2517
ethics@carleton.ca

CERTIFICATION OF INSTITUTIONAL ETHICS CLEARANCE

The Carleton University Research Ethics Board-A (CUREB-A) at Carleton University has renewed ethics approval for the research project detailed below. CUREB-A is constituted and operates in compliance with the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS2)

Title: Exploring Intercultural Factors in International Pilot-Air Traffic Controller Communications: Validating a Taxonomy Using Mixed Methods Research (Phase 4) [Ana Lucia Tavares Monteiro]

Protocol #: 107816

Project Team Members: **Ana Lucia Tavares Monteiro (Primary Investigator)**
Prof. Janna Fox (Research Supervisor)

Department and Institution: Faculty of Arts and Social Sciences\Linguistics and Language Studies (School of), Carleton University

Funding Source (if applicable):

Effective: September 24, 2018 Expires: September 30, 2019

Please ensure the study clearance number is prominently placed in all recruitment and consent materials: CUREB-A Clearance # 107816.

Restrictions:

This certification is subject to the following conditions:

1. Clearance is granted only for the research and purposes described in the application.
2. Any modification to the approved research must be submitted to CUREB-A. All changes must be approved prior to the continuance of the research.
3. An Annual Application for the renewal of ethics clearance must be submitted and cleared by the above date. Failure to submit the Annual Status Report will result in the closure of the file. If funding is associated, funds will be frozen.
4. A closure request must be sent to CUREB-A when the research is complete or terminated.
5. Should any participant suffer adversely from their participation in the project you are required to report the matter to CUREB-A.
6. It is the responsibility of the student to notify their supervisor of any adverse events, changes to their application, or requests to renew/close the protocol.
7. Failure to conduct the research in accordance with the principles of the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans 2nd edition* and the *Carleton University Policies and Procedures for the Ethical Conduct of Research* may result in the suspension or termination of the research project.

Upon reasonable request, it is the policy of CUREB, for cleared protocols, to release the name of the PI, the title of the project, and the date of clearance and any renewal(s).

Please email the Research Compliance Coordinators at ethics@carleton.ca if you have any questions.

CLEARED BY: Date: September 24, 2018

Bernadette Campbell, PhD, Chair, CUREB-A

Natasha Artemeva, PhD, Vice-Chair, CUREB-A

Appendix G - Letter of authorization to trial test tasks at ANAC

FEDERATIVE REPUBLIC OF BRAZIL - NATIONAL CIVIL AVIATION AGENCY
Av. Pres. Vargas 850, Centro - Rio de Janeiro/RJ – Brazil
Tel. +55-21-35015409

August 17, 2017

Ms. Ana Lúcia Tavares Monteiro
Carleton University
1125 Colonel By Dr, Ottawa, ON K1S 5B6

Subject: Test tasks trialing during ANAC examiners' recurrent course

Dear Ana Lúcia Tavares Monteiro,

As the Coordinator of the Language Proficiency Group at the Brazilian Civil Aviation National Agency (ANAC), and as the Coordinator of the "ANAC Aviation English Language Proficiency Examiners –Recurrent Course", I am writing you to formally invite you to join us on the last day of the course, September 29th 2017, in São José dos Campos, S.P., Brazil.

Your participation as a researcher will be very welcome, especially considering that you have designed Aviation English test tasks that need to be trialed by language proficiency examiners, both language experts and aviation experts (pilots and air traffic controllers), who are the target audience of this course. The participants are ANAC employees and interlocutors/raters from ANAC accredited institutions, who are experienced in applying/rating tests of pilots' language proficiency. Therefore, their comments and perceptions will be valuable to your current research, as your experience in academia will bring useful insights to our group of examiners.

You have my full consent and authorization to trial the new test tasks, by collecting audio, video and written data. Participants will be advised that they will need to sign consent forms individually prior to the commencing of the trialing.

We look forward to welcoming you at ANAC.

Best regards,

Atila de Rezende Jordão

Language Proficiency Assessment Group Coordinator

Appendix H - Consent Form (Phase 1 – Online survey)

Carleton UNIVERSITY
Canada's Capital University

CUREB clearance #: 103859

Title: Exploring Intercultural Factors in International Pilot-Air Traffic Controller Communications: Validating a Taxonomy Using Mixed Methods Research

Date of ethics clearance: December 14, 2015

Ethics Clearance for the Collection of Data Expires: May 31, 2016

This is a pilot study on cultural differences in the context of international radiotelephony communications. This study aims to identify culturally influenced factors that can affect the way pilots and air traffic controllers interact on the radio using the English language as well as to verify the extent to which experienced professionals perceive the potential threats of intercultural factors to the safety of radiotelephony communications. **The researcher for this study is Ana Lúcia Tavares Monteiro, a PhD student in the Applied Linguistics and Discourse Studies Program at Carleton University, located in Canada.** She is working under the supervision of Professor Janna Fox and Professor Natasha Artemeva in the School of Linguistics and Language Studies.

This study involves one 20 minute survey that will take place online. You will answer a questionnaire with a number of closed and open-ended questions. I will not ask any questions regarding personal experiences with cultural differences in radiotelephony communications, and urge you not to provide personal accounts of details about possible personal instances of misunderstandings on the radio.

While this research might involve mild psychological and social discomfort, your responses will not be used to cause you any harm or difficulty in your profession. No evaluation of performance will be carried out and your responses will not be shared with professional colleagues or supervisors, thus causing no impact on your social status or career. You should not feel obliged to answer any question that makes you feel uncomfortable or distressed.

You have the right to end your participation in the survey at any time, for any reason, up until you hit the "submit" button. You can withdraw by exiting the survey at any time before completing it. It will not be possible to withdraw from the study after the survey is submitted, because your responses will not be identifiable.

The company hosting the online survey is SimpleSurvey, which is based in Canada. SimpleSurvey is a fully designed, developed and hosted-in-Canada survey software tool. Data physically resides on Canadian soil and is protected by Canadian strict laws governing the privacy and confidentiality of data and information.

All data will be encrypted and will be deleted from the survey tool once the survey is complete. Research data will be accessible by the researcher, the research co-supervisors and the survey company. No names will be linked to any of the data provided. Although IP addresses are recorded by the survey tool as a default, to prevent IP tracking and to make sure survey results are anonymous, the researcher will edit collector options before sending the survey, by turning on "Anonymous Responses".

Once the project is completed, all research data will be kept until the completion of my PhD studies and potentially used for other research projects on this same topic. At the end of five years, all research data will

be deleted.

All participants may benefit from the study by exploring their own perceptions and understandings of how cultural differences may affect their language behavior and by raising awareness of the potential threats of intercultural factors to the safety of radiotelephony communications in English.

The results of this research may be presented in conferences or published in professional journals; however, the findings will be reported in an aggregated manner, which may sometimes be exemplified by cautiously selected anonymous quotes.

If you would like a copy of the finished research project, you are invited to contact the researcher by email to request an electronic copy.

The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. Should you have questions or concerns related to your involvement in this research, please contact:

CUREB contact information:
Professor Louise Heslop, Chair
Professor Andy Adler, Vice-Chair
Carleton University Research Ethics Board
Carleton University
511 Tory
1125 Colonel By Drive
Ottawa, ON K1S 5B6
Tel: 613-520-2517
ethics@carleton.ca

Researcher contact information:
Ana Lúcia Tavares Monteiro
School of Linguistics and Language Studies
Carleton University
Email: anatavaresmonteiro@cmail.carleton.ca

Co-Supervisors contact information:
Prof. Janna Fox
School of Linguistics and Language Studies
Carleton University
Email: janna.fox@carleton.ca

Prof. Natasha Artemeva
School of Linguistics and Language Studies
Carleton University
Email: natasha.artemeva@carleton.ca

By clicking “SUBMIT”, you consent to participate in the research study as described above.

Appendix I - Consent Form (Phase 2 – Focus groups discussions)

Carleton UNIVERSITY
Canada's Capital University

CUREB clearance #: 103859

Title: Exploring Intercultural Factors in International Pilot-Air Traffic Controller Communications: Validating a Taxonomy Using Mixed Methods Research

Date of ethics clearance: April 20, 2017

Ethics Clearance for the Collection of Data Expires: April 30, 2018

This consent form indicates that my participation in this study is entirely voluntary. As a result, if at any time during the course of this focus group discussion I find the questioning objectionable, I am able to choose not to answer or make any comments. I am also aware that the following group discussion will be audio-recorded and the subsequent inter-groups discussion will be moderated and recorded by the researcher, Ana Lúcia Tavares Monteiro. I understand that once the recordings have been transcribed, they will be destroyed at a date no later than a month after the focus group takes place. In addition, given the precautions that will be taken to protect my identity, by keeping all responses and comments confidential and not attributing any response or comment to me in future dissertation, other publications or presentations, I agree that the potential personal risk involved in my participation in this study is minimal.

I additionally understand that for the purposes of this study I will be asked to review a transcript of pilot-ATC communication and respond to six questions, as a group. I understand that the collection of data for this study involves no deception. By signing below, I realize that I am indicating acceptance of all of the terms and methods of data collection listed above.

I am aware that one signed copy will be retained by Ana Lúcia T. Monteiro while a second copy will be provided for my own records. The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600 ext. 2517 or via email at ethics@carleton.ca).

Participant's name: _____

_____ _____
Signature of participant Date

_____
Signature of researcher

Appendix J - Consent Form (Phase 3 – Task pilot testing/role of interlocutor)

Carleton UNIVERSITY
Canada's Capital University

CUREB clearance #: 107816

Title: Exploring Intercultural Factors in International Pilot-Air Traffic Controller Communications: Validating a Taxonomy Using Mixed Methods Research

Date of ethics clearance: September 21, 2017

Ethics Clearance for the Collection of Data Expires: September 30, 2018

I _____, choose to participate in a study on cultural differences and construct definition in the context of international radiotelephony communications. This phase of the study aims to verify the operationalization of what was considered relevant by stakeholders to be included in an aviation English test in terms of language and communication, by trialing draft tasks to be used in the assessment of pilots' English proficiency for this occupational context. The researcher for this study is Ana Lúcia Tavares Monteiro, a PhD candidate in the Applied Linguistics and Discourse Studies Program at Carleton University, who is working under the supervision of Professor Janna Fox in the School of Linguistics and Language Studies.

I understand that my participation will consist of playing the role of an interlocutor in the administration of a role-play task (15 min), filling in a feedback form (10 min) and participating in an individual semi-structured interview conducted by the researcher (15 min). **With my consent, I agree that the task administration will be observed by Aviation English Testing Experts (AETEs) and also video-recorded.** I understand that no evaluation of my performance will be carried out, as it is the task that is being evaluated.

I understand that the video-recording will be kept secure by the researcher, but the recordings and transcriptions will be coded to protect my anonymity. I agree that my video-recording may be used in publications, presentations and for interlocutor/rater/researcher training purposes. In that case, my identity will be protected by obscuring any identifiable features from the video (e.g. by facial blurring and voice alteration). I further understand that the video-recordings will be destroyed after five years of the completion of the researcher's PhD studies, regardless of whether they will be used in publications, presentations or for training.

Following the task administration, I agree that the individual semi-structured interview will be audio-recorded and that the recordings and transcriptions will be coded to protect my anonymity. This anonymity would extend to any data used in dissertations or other publications, and presentations. I understand that once the audio-recording has been transcribed, it will be destroyed at a date no later than a month after the interview takes place.

I acknowledge that the potential risk involved in my participation in this study is minimal, as precautions will be taken to protect my identity, by keeping all responses and comments anonymous, and allowing me to request that certain responses not be included in the final project.

I understand that I have the right to end my participation in the study at any point during the task administration and interview. However, all data that I provide up to the point of withdrawal will be retained for analysis.

I further understand that:

- All research data, including video/audio-recordings and any notes will be encrypted. Any hard copies of data (including any handwritten notes or USB keys) will be kept in a locked cabinet at the

researcher's home office. Research data will only be accessible by the researcher and the research supervisor.

- Once the project is completed, all research data will be kept until the completion of the researcher's PhD studies and potentially used for other research projects on this same topic. At the end of five years, all research data will be securely destroyed (electronic data will be erased and hard copies will be shredded).

- If I would like a copy of the finished research project, I am invited to contact the researcher to request an electronic copy which will be provided to me.

- The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. If I have any ethical concerns with the study, I should contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600 ext. 2517 or via email at ethics@carleton.ca).

Researcher contact information:
Ana Lúcia Tavares Monteiro
School of Linguistics and Language Studies
Carleton University
Email: anatavaresmonteiro@cmail.carleton.ca

Supervisor contact information:
Prof. Janna Fox
School of Linguistics and Language Studies
Carleton University
Email: janna.fox@carleton.ca

_____ _____
Signature of participant Date

_____ _____
Signature of researcher Date

Appendix K - Consent Form (Phase 3 – Task pilot testing/role of test-taker)

Carleton UNIVERSITY
Canada's Capital University

CUREB clearance #: 107816

Title: Exploring Intercultural Factors in International Pilot-Air Traffic Controller Communications: Validating a Taxonomy Using Mixed Methods Research

Date of ethics clearance: September 21, 2017

Ethics Clearance for the Collection of Data Expires: September 30, 2018

I _____, choose to participate in a study on cultural differences and construct definition in the context of international radiotelephony communications. This phase of the study aims to verify the operationalization of what was considered relevant by stakeholders to be included in an aviation English test in terms of language and communication, by trialing draft tasks to be used in the assessment of pilots' English proficiency for this occupational context. The researcher for this study is Ana Lúcia Tavares Monteiro, a PhD candidate in the Applied Linguistics and Discourse Studies Program at Carleton University, who is working under the supervision of Professor Janna Fox in the School of Linguistics and Language Studies.

I understand that my participation will consist of playing the role of a test-taker in the administration of a role-play task (15 min), filling in a feedback form (10 min) and participating in an individual semi-structured interview conducted by the researcher (15 min). **With my consent, I agree that the task administration will be observed by Aviation English Testing Experts (AETEs) and also video-recorded.** I understand that no evaluation of my performance will be carried out, as it is the task that is being evaluated.

I understand that the video-recording will be kept secure by the researcher, but the recordings and transcriptions will be coded to protect my anonymity. I agree that my video-recording may be used in publications, presentations and for interlocutor/rater/researcher training purposes. In that case, my identity will be protected by obscuring any identifiable features from the video (e.g. by facial blurring and voice alteration). I further understand that the video-recordings will be destroyed after five years of the completion of the researcher's PhD studies, regardless of whether they will be used in publications, presentations or for training.

Following the task administration, I agree that the individual semi-structured interview will be audio-recorded and that the recordings and transcriptions will be coded to protect my anonymity. This anonymity would extend to any data used in dissertations or other publications, and presentations. I understand that once the audio-recording has been transcribed, it will be destroyed at a date no later than a month after the interview takes place.

I acknowledge that the potential risk involved in my participation in this study is minimal, as precautions will be taken to protect my identity, by keeping all responses and comments anonymous, and allowing me to request that certain responses not be included in the final project.

I understand that I have the right to end my participation in the study at any point during the task administration and interview. However, all data that I provide up to the point of withdrawal will be retained for analysis.

I further understand that:

- All research data, including video/audio-recordings and any notes will be encrypted. Any hard copies

of data (including any handwritten notes or USB keys) will be kept in a locked cabinet at the researcher's home office. Research data will only be accessible by the researcher and the research supervisor.

- Once the project is completed, all research data will be kept until the completion of the researcher's PhD studies and potentially used for other research projects on this same topic. At the end of five years, all research data will be securely destroyed (electronic data will be erased and hard copies will be shredded).

- If I would like a copy of the finished research project, I am invited to contact the researcher to request an electronic copy which will be provided to me.

- The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. If I have any ethical concerns with the study, I should contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600 ext. 2517 or via email at ethics@carleton.ca).

Researcher contact information:
Ana Lúcia Tavares Monteiro
School of Linguistics and Language Studies
Carleton University
Email: anatavaresmonteiro@cmail.carleton.ca

Supervisor contact information:
Prof. Janna Fox
School of Linguistics and Language Studies
Carleton University
Email: janna.fox@carleton.ca

_____ _____
Signature of participant Date

_____ _____
Signature of researcher Date

Appendix L - Consent Form (Phase 3 – Task pilot testing/role of rater)

Carleton UNIVERSITY
Canada's Capital University

CUREB clearance #: 107816

Title: Exploring Intercultural Factors in International Pilot-Air Traffic Controller Communications: Validating a Taxonomy Using Mixed Methods Research

Date of ethics clearance: September 21, 2017

Ethics Clearance for the Collection of Data Expires: September 30, 2018

I _____, choose to participate in a study on cultural differences and construct definition in the context of international radiotelephony communications. This phase of the study aims to verify the operationalization of what was considered relevant by stakeholders to be included in an aviation English test in terms of language and communication, by trialing draft tasks to be used in the assessment of pilots' English proficiency for this occupational context. The researcher for this study is Ana Lúcia Tavares Monteiro, a PhD candidate in the Applied Linguistics and Discourse Studies Program at Carleton University, who is working under the supervision of Professor Janna Fox in the School of Linguistics and Language Studies.

I understand that my participation will consist of observing the administration of a role-play task and filling in an *observation checklist of language functions* in real time (15 min), filling in an *Indicator Checklist of skills/behaviors* (15 min) and participating in a focus group discussion moderated by the researcher (40 min). With my consent, I agree that focus group discussions will be audio-recorded. I understand that once the recording has been transcribed, the audio-recording will be destroyed at a date no later than a month after the focus group takes place.

I acknowledge that the potential risk involved in my participation in this study is minimal, as precautions will be taken to protect my identity, by keeping all responses and comments confidential and not attributing any response or comment to me in future dissertation, other publications or presentations.

I understand that I have the right to end my participation in the study at any point during the task administration and focus group. However, all data that I provide up to the point of withdrawal will be retained for analysis.

I further understand that:

- All research data, including audio-recordings and any notes will be encrypted. Any hard copies of data (including any handwritten notes or USB keys) will be kept in a locked cabinet at the researcher's home office. Research data will only be accessible by the researcher and the research supervisor.

- Once the project is completed, all research data will be kept until the completion of the researcher's PhD studies and potentially used for other research projects on this same topic. At the end of five years, all research data will be securely destroyed (electronic data will be erased and hard copies will be shredded).

- If I would like a copy of the finished research project, I am invited to contact the researcher to request an electronic copy which will be provided to me.

- The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board,

which provided clearance to carry out the research. If I have any ethical concerns with the study, I should contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600 ext. 2517 or via email at ethics@carleton.ca).

Researcher contact information:
Ana Lúcia Tavares Monteiro
School of Linguistics and Language Studies
Carleton University
Email: anatavaresmonteiro@cmail.carleton.ca

Supervisor contact information:
Prof. Janna Fox
School of Linguistics and Language Studies
Carleton University
Email: janna.fox@carleton.ca

_____ _____
Signature of participant Date

_____ _____
Signature of researcher Date

Appendix M - Transcript of Scenario 1

Background information: With eight planes in sequence for runway 28 at Dublin, a pilot flying an A319 interacts with a female tower controller at Dublin at its peak, mid-day time period.

		Scenario 1 – Transcript (Available at https://www.youtube.com/watch?v=uWg7IpphPc8)
1	ATCO	[ ] 845, there's traffic behind you waiting to depart, if you're not ready…I'm going to bring you onto the runway to vacate on 34.
2	PILOT	Ok, um, just to the point of order we're ready just waiting the cabin to be ready…it takes about two minutes from the time we were talking to you just before.
3	ATCO	Ok, you need to enter runway 28 please, turn right onto runway 34 to vacate.
4	PILOT	Standby, [ ] 845.
5	ATCO	[ ] 845 proceed onto runway 28 now and vacate right onto runway 34, there is traffic behind you waiting to depart.
6	PILOT	Yep, we've just got a phone call standby…
7	PILOT	And [ ] 845, we are actually fully ready.
8	ATCO	[ ] 845, line up runway 28 and wait.
9	PILOT	Line up and wait 28, [ ] 845.
10	ATCO	And for future reference [ ] 845, err, I suggest you advise the, err, ground controller…that you are unable to take departure yet and you shouldn't really be taxiing out to E1 when you're not ready for departure…because there is traffic behind, waiting, that is ready.
11	PILOT	Standby.
12	ATCO	[ ] 845, are you fully ready for departure?
13	PILOT	Affirm [ ] 845, we're just doing the checks whilst err, whilst you keep talking over us, standby.
14	ATCO	Ok, negative! Turn right please onto runway 34. That's the third time I've asked you to vacate onto runway 34 if you are not ready, turn right onto runway 34.
15	PILOT	Madam, we are fully ready, we're just trying to complete the checklists, but err, you just keep interrupting our checks, standby.
16	ATCO	Negative! Turn right onto runway 34, I've asked you three times now! Turn right, to vacate onto runway 34.
17		(Dublin control speaking to Ryanair aircraft)
18	ATCO	Except when I instruct you three times, to vacate because there's traffic behind you.
19	PILOT	[ ] 845, roger, what would you like us to do now Madam?
20	ATCO	As I said, I'll let you depart, winds 250 degrees, 17 knots, runway 28 cleared takeoff.
21	PILOT	Cleared takeoff, [ ] 845.
22		(After departure, Dublin control gave the pilot a phone number to call the ground at Heathrow. Two aircraft had to go around.)

Appendix N - Transcript of Scenario 2

Background information: Kennedy Ground control and tower interacts with [ ] 7997 (PILOT 1) and [ ] 020 (PILOT 2), at JFK International Airport.

		Scenario 2 (Available at https://www.youtube.com/watch?v=2t_NT7aUrE0)
1	PILOT 1	Uh, Ground, uh… morning, [ ] 7997, requests taxi, we are on Sierra Bravo.
2	ATCO	[ ] 7997 Heavy, uh runway 22R, taxi uhm…actually, I tell you what: follow an American aircraft coming from the left side, follow him coming from the left and then on (…) left on Alpha short of Whiskey behind him.
3	PILOT 1	Roger, that's to follow the American coming from our left to right and then left on Alpha short of Whiskey.
4	ATCO	That's correct, [ ], thank you.
5	PILOT 1	Thanks.
6	PILOT 1	Uh, [ ] 7997, confirm that's a right turn here on the Alpha?
7	ATCO	(Silence)
8	PILOT 1	Ground, [ ] 7997, could you give me further taxi clearances?
9	ATCO	Sir, Alpha short of taxiway Whiskey. I need you to continue on Alpha.
10	PILOT 1	Alpha, the first right (…) me now?
11	ATCO	Just continue the aircraft towards you know (…) British Airways 74 off your right side.
12	PILOT 1	Okay, thanks uh …okay, Alpha short of Whiskey, thanks.
13	ATCO	Actually, [ ], when you get to Whiskey, sir, you can turn left and cross 13L, then turn right onto taxiway Charlie.
14	PILOT 1	Ok, when I get to Whiskey turn left, cross the runway 13L and then turn right onto Charlie, I think, [ ] 7997 Heavy.
15	ATCO	That's correct, sir, you know, I just wanna, if you don't mind, I need to confirm the first two or three fixes on your route, uh…On the, the, everything is the same, make sure (…) go MERIT intersection then Hartford, and then Partham, and then after that it's as filed.
16	PILOT 1	Okay, uh, [ ] 7997 (…) our flightplan we go direct to …Partham, but…on the SITA, as cleared earlier, we go, follow (…) to Hartford and Partham, and then, we join the route.
17	ATCO	That is correct, yeah, (…) vectors to MERIT intersection, then direct Hartford, direct Partham, and then, it's, you know, as previous cleared and that's filed.
18	PILOT 1	Roger, thanks.
19	ATCO	…[ ] 7997 Heavy, change (…) frequency 119.1.
20	PILOT 1	19.1, good day.
21		(...)
22	PILOT 1	Uh, tower, morning, [ ] 7997 Heavy…
23	ATCO	(Silence)
24	PILOT 1	Tower, there's [ ] 7997.
25	ATCO	[ ], what's up?
26	PILOT 1	Uh, just check that uh…we're clear to cross yeah? 13L?
27	ATCO	Affirmative, Sir, cross 13L, then turn right on taxiway Charlie.
28	PILOT 1	13L, right on Charlie, thanks.
29	ATCO	Kennedy tower, good morning, [ ] 020, on…approaching 22L.
30	PILOT 2	[ ] 020, wind calm, runway 22L, cleared to land.

31	ATCO	Cleared to land 22L, [ ] 020 Heavy.
32		(...)
33	PILOT 1	And…[ ] 7997, uh…just want to confirm you the point before Hartford, could you give me the name again, please?
34	ATCO	[ ] you gonna kill me, what do you want now?
35	PILOT 1	Okay, ground, (…) checking (…) our routing just the point before Hartford and Partham, could you give me the point again?
36	ATCO	Now sir, you've been given a change of frequency, you'd be talking to the same guy all night long, see? You're going back for a million questions, but let's go over it: MERIT intersection, that's spelled: Mike Echo Romeo India Tango; direct Hartford, that is Hotel Foxtrot Delta; direct Partham, that is Papa Uniform Tango, and then as filed. Do you have any further questions about your route, your taxi route, the route you gotta fly, anything else?
37	PILOT 1	Not for now, sir, thanks.
38	ATCO	…now. I'm sure in 30 seconds you'll have another one, but continue to the runway.
39	PILOT 1	Okay.
40		(...)
41	PILOT 2	Getting to Juliette, [ ] 020 heavy.
42	ATCO	[ ] 020, once you get to Juliette Hotel, 22R, remain this frequency.
43	PILOT 2	Confirm for [ ] 020 Heavy?
44	ATCO	Oh, boy…[ ] 020, hold short of 22R, this frequency.
45	PILOT 2	Holding short 22R, sir, [ ] 020 Heavy.
46	PILOT 2	It's been a very long day, [ ] 020 Heavy, holding short of 22R.
47	ATCO	Oh, it's been a very long night.
48	PILOT 2	Ok, no problems.

Appendix O - Transcript of Scenario 3

Background information: A passenger aircraft was put in a series of extended holding patterns as it approached New York. The crew informed APPR they were running out of fuel but did not declare an emergency and were cleared to land. After a missed approach and during a go-around, the plane ran out of fuel and crashed in a wooded area (73 out of 158 aboard were killed).

TWR = JFK Tower APPR = New York Approach Controller FE = Flight Engineer in the cockpit
PILOT = First Officer communicating with ATC
FO = First Officer communicating with captain in the cockpit

Note: In this Scenario, only the most relevant excerpts have been selected, not the complete transcription. A few utterances from the cockpit interaction involving the Captain, the First Officer (FO) and the Flight Engineer (FE) were included to add understanding of the situation.

		Scenario 3 (Available at http://www.planecrashinfo.com/cvr900125.htm)
1	PILOT	Executing a missed approach, [ ] zero five two heavy.
2	TWR	[ ] zero five two heavy, roger, ah, climb and maintain two thousand, turn left, heading one eight zero.
3	TWR	[ ] zero five two, you are making a left turn, correct sir?
4	PILOT	That's right to one eight zero on the heading, and, ah, we'll try once again. We're running out of fuel.
5	TWR	Okay.
6	Captain	*(Advise him we are emergency!)*
7	Captain	*(Did you tell him?)*
8	FO	*(Yes sir.)*
9	FO	*(I already advised him.)*
10	TWR	[ ] zero five two heavy, continue the left turn, heading one five zero, maintain two thousand.
11	PILOT	One five zero, maintaining two thousand, [ ] zero five two heavy.
12	TWR	[ ] zero five two heavy, contact approach on one one eight point four.
13	PILOT	One one eight point four.
14	PILOT	Approach, [ ] zero five, ah, two heavy, we just missed a missed approach, and ah, we're maintaining two thousand and five on the...
15	APPR	[ ] zero five two heavy, [ ], good evening, climb and maintain three thousand.
16	Captain	*(Advise him we don't have fuel.)*
17	PILOT	Climb and maintain three thousand, and ah, we're running out of fuel, sir.
18	APPR	Okay, fly heading zero eight zero.
19	PILOT	Flying heading zero eight zero, climb to three thousand.
20	Captain	*(Did you already advise that we don't have fuel?)*

21	FO	*(Yes sir, I already advise him, hundred and eighty on the heading. We are going to maintain three thousand feet, and he's going to get us back.)*
22		*(Some time later…)*
23	APPR	[ ] zero five two heavy, turn left, heading zero seven zero.
24	PILOT	Heading zero seven zero, [ ] zero five two heavy.
25	APPR	And [ ] zero five two heavy, ah, I'm going to bring you about fifteen miles northeast, and then turn you back onto the approach, is that fine with you and your fuel?
26	PILOT:	I guess so, thank you very much.
27	Captain	*(What did he say?)*
28	FE	*(The guy is angry.)*
29		(Some time later…)
30	PILOT	Ah, can you give us a final now? [ ] zero five two heavy.
31	APPR	[ ] zero five two, affirmative sir, turn left, heading zero four zero.
32	PILOT	[ ] zero five two heavy, left turn two five zero, and ah, we're cleared for ILS.
33	APPR	[ ] fifty two, climb and maintain three thousand.
34	PILOT	Ah, negative sir. We just running out of fuel. We okay three thousand. Now okay.
35	APPR	Okay, turn left, heading three one zero sir.
36	PILOT	Three one zero, [ ] zero five two.
37		(Some time later…)
38	APPR	[ ] fifty two, fly heading of three six zero please.
39	PILOT	Okay, we'll maintain three six zero now.
40	APPR	Okay, and you're number two for the approach. I just have to give you enough room so you can make it without, ah, having to come out again.
41	PILOT	Okay, we're number two and flying three six zero now.
42	APPR	Thank you sir.
43	APPR	[ ] zero five two heavy, turn left, heading three three zero.
44	PILOT	Three three zero on the heading, [ ] zero five two.
45	APPR	[ ] zero five two, turn left, heading two five zero, intercept the localizer.
46	PILOT	Roger.
47		(Some time later…)
48	PILOT	[ ] zero five two, we just, ah, lost two engines and, ah, we need priority, please.
49	ATCO	[ ] zero five two, turn left, heading two five zero, intercept the localizer.
50	PILOT	Roger.
51	APPR	[ ] zero five two heavy, you're one five miles from the outer marker, maintain two thousand until established on the localizer. Cleared for ILS two two left.
52	PILOT	Roger, [ ].
53	Captain	*(Did you select the ILS?)*
54	FO	*(It is ready on two.)*
55		[End of recording]

Appendix P – Transcript of Scenario 4

Background information: JFK Ground controller, in New York, interacts with two pilots: one from [ ]503 (United Arab Emirates) and the other from [ ] 222 (South Korea).

		Scenario 4 - Transcript (Available at https://www.youtube.com/watch?v=ZWOOKQlEe5s)
1	ATCO	[ ] 503, where you park?
2	PILOT	Bravo 28, Sir.
3	ATCO	Not taxiway, the LETTER!
4	PILOT	Oh negative sir, we are on 22R holding short of Foxtrot.
5	ATCO	What taxiway do you enter the ramp?
6	PILOT	Okay, so we just exit the runway and we're holding short of Foxtrot on 22R.
7	ATCO	You are not listening to what I'm asking you. What taxiway do you enter the ramp?
8	PILOT	I'm not on the ramp yet, sir.
9	ATCO	What taxiway do you enter the ramp. Tell me. What letter?
10	PILOT	Okay we can enter at KILO for [ ] 503.
11	ATCO	That's what I need get out of you. We talked like 6 times. Straight ahead and hold short of HOTEL, sir.
12	PILOT	Straight ahead, hold short of HOTEL, roger.
13		(...)
14	ATCO	[ ] 222. Turn right here. Turn left on to 22R, and hold short of Juliet.
15	PILOT 2	(Silence)
16	ATCO	[ ] 222?
17	PILOT 2	Go ahead?
18	ATCO	Turn right, left on 22R. Hold short of Juliette.
19	PILOT 2	Okay, right turn, then 22R, holding short Juliette, [ ] 222.
20		(Some time later...)
21	ATCO	[ ] 503 follow [ ]222, hold short Juliette on the runway.
22	PILOT	Yes, we'll follow the [ ], and next time I would like you to be polite with me. Thank you.
23	ATCO	Okay, but if I got to talk to you 6 times, and I got all other people I got to talk to, and you don't know what I'm saying.
24	PILOT	(…)…nice day, polite with me. All right?
25	ATCO	Are you impolite with me?
26	PILOT	I'll make a report.
27	ATCO	Go ahead!

Appendix Q - Transcript of Scenario 5

Background information: A pilot interacts with the ATCO at Ezeiza International Airport, Buenos Aires, in a situation that generated an operational limitation for the aircraft with regard to fuel endurance.

		Scenario 5 - Transcript (Available at https://www.youtube.com/watch?v=Lpxc0B-t5tM)
1	PILOT	Ezeiza, [ ] 417.
2	ATCO	(Controller mumbling) Ascend to 045 level, uh transition level 045 and proceed to chart # 5.
3	PILOT	Say all that again cause I didn't get a word of it…
4		(...)
5	PILOT	You confirm the ILS 35 is serviceable?
6	ATCO	Runway 35, it's available, affirmative.
7	PILOT	The ILS is available?
8	ATCO	Affirmative, Sir.
9		(...)
10	PILOT	Could you please confirm we are cleared to ILS 35?
11	ATCO	Authorized ILS 35.
12	PILOT	[ ] 417, could you confirm the ILS is serviceable, we are not receiving it.
13	ATCO	ILS to 35….out of service, 417.
14	PILOT	417,Roger. We are unable to continue this approach. I inquired about the ILS on runway 35 before. I want you to know that I asked you many times if the ILS 35 was operative and you said it was. Tell me, how can it not function anymore?
15	ATCO	(Ezeiza does not respond)
16		(Some time later…)
17	PILOT	I am declaring a low fuel.
18	ATCO	(Ezeiza does not respond)
19	PILOT	Ezeiza, [ ] 417, did you copy?
20	ATCO	Affirmative, Sir. Report stablished on localizer…
21	PILOT	OK, we are taking heading 310 and I would like to see the authorities on the ground.
22	ATCO	(Ezeiza does not respond)
23	PILOT	Did you copy my request about seeing the people on the ground?
24	ATCO	(Ezeiza does not respond)
25	PILOT	Still low fuel. So I want priority for landing.
26	ATCO	(Ezeiza does not respond)
27	PILOT	Did you copy, [ ] 417?
28	ATCO	Affirmative, 417.
29	PILOT	I will be filling a report about this charade because it's quite amazing what happened.
30	ATCO	Can you repeat, please?
31	PILOT	I'll be filling a report against you, what happened is truly, truly amazing.

Appendix R - Transcript of Scenario 6

Background information: A pilot requests assistance to an air traffic controller at John Wayne Tower, California.

		Scenario 6 - Transcript (Available at https://www.youtube.com/watch?v=JYROSTV_KVg)
1	PILOT	And [ ], we are VFR under the clouds right now. And if you could give me a (inaudible).
2	ATCO	You're not familiar with this airspace?
3	PILOT	Yes sir, I'm very familiar with this airspace. But just coming through the clouds now it would be easier if you just give me my heading for a moment.
4	ATCO	What kind of NAV equipment do you have on board?
5	PILOT	Slant Uniform, VOR sir.
6	ATCO	[ ] fly heading 150. Vectors Mile Square Park.
7	PILOT	Okay, we are currently 150 sir. Thank you sir, just wanted a little help. Thank you.
8	ATCO	Well, let me give you some advice. We are really busy. We've got one controller working all the airspace and a lot of inbounds coming in, the last airliners coming into John Wayne. I probably don't always have time to hold your hand. Sorry to say that, but that's the truth.
9	PILOT	25 years I have been flying this airspace sir. I've never had a controller talk to me like that.
10	ATCO	Well, you are welcome to call me on the phone.
11	PILOT	Love to!

Appendix S – Second Cycle Coding: organization of codes into sub-categories and categories

MAJOR THEME	CATEGORIES	SUB-CATEGORIES	CODES – Coder 1
Intercultural factors in international pilot-ATCO communications	Power Distance	Power relations	Resorting to higher-level authority (3); Using own authority (2); Requesting equal treatment (1); Reprimanding/Warning (1)
		Deferential role	Addressing the interlocutor as "Sir/Madam"(11); Agreeing to act despite limitations (3); Avoiding disagreement/demanding (3); Using excessive politeness in emergencies (2)
	Face-work strategies	Self-face concern	Trying to restore face loss (1); Defensive move (shame) (5); Reluctant to ask for clarification (2); Concern for own interest (1)
		Mutual-face concern	Need to correct previous information (1)
	Conflict management	Neutral direction	Accommodating conflict style (10); Avoiding conflict style (5)
		Conflictual direction	Dominating/competing conflict style (4); Aggressive conflict style (13)
		Expectancy violations	Information change (1); Non work-related expression (1); Personal insult (1); Interlocutor's unpredictable behavior (1)
	Communication styles	Directness	Directness (19); Objectivity/assertiveness (1); Concern for clarity (2); Concern for efficiency (15)
		Indirectness	Ambiguous question (2); Conflicting information (1); Confusing/unclear statement (3); Failure to declare emergency (4); Lack of confidence(1)
	Non-collaborative behavior	Unprofessional tone	Showing annoyance/anger (2); Showing impatience (4); Showing sarcasm (2); Showing arrogance/rudeness (7)
		Unprofessional attitude	Unwillingness to help (1); Buying time (2); No time to support x time to reprimand (1); Non-accommodating to interlocutor's needs (3); Failure to question severity of problem (1); Impoliteness: seeking disagreement (1); Impoliteness: Making the other feel uncomfortable (1); Impoliteness: accusing/blaming the other (5); Threatening (3)
		Non-compliance with rules	Stating readiness when not ready (1); Non-compliance with orders (1); Use of non-standard phraseology (6)
	Collaborative behavior	Professional attitude	Concern for potential complications (3); Double-checking/being cautious (6); Compliance with order (5); Politeness (6); Requesting ATCO's support
		Supportiveness	Advice on appropriate course of action/behavior (1); Support to ATCO's request (1); Support to pilot's request (1)

Appendix T - Quantitative strand – Questionnaire structure

Example of questions for pilots – Section II							Research question
	(1= Never; 6= Very frequently)						
11.1 How often do you encounter air traffic controllers who avoid getting involved in conflicts or arguments?	1	2	3	4	5	6	RQ1
	(1= Not important; 6= Very important)						
11.2 If you selected 2 or higher: In your view, how important is this?	1	2	3	4	5	6	RQ2
11.3 Please comment.							
Example of questions for air traffic controllers – Section III							**Research question**
	(1= Never; 6= Very frequently)						
28.1 How often do you encounter pilots who show impatience and/or sarcasm in their speech?	1	2	3	4	5	6	RQ1
	(1= Not important; 6= Very important)						
28.2 If you selected 2 or higher: How important, in your view, were these events as potential threats to safety?	1	2	3	4	5	6	RQ2
28.3 Please comment.							

- **Section I: Demographic information:**

Gender/age range/years of experience as pilot or ATCO/nationality/first language/second or other languages/language mostly spoken in RT/years of experience in international operations/level of education

- **Section II:**

Q10.1 How often do you encounter pilots/ATCOs who, in a conflict situation, attempt to accommodate and restore neutral communication?

Q11.1 How often do you encounter pilots/ATCOs who avoid getting involved in conflicts or arguments?

Q12.1 How often do you encounter pilots/ATCOs who are concerned with both parties' images and group interests?

Q13.1 How often do you encounter pilots/ATCOs who are concerned with clarity and efficiency?

Q14.1 How often do you encounter pilots/ATCOs who are concerned with safety and potential complications?

Q15.1 How often do you encounter pilots who comply with ATCOs' orders/ATCOs who comply with rules?

Q16.1 How often do you encounter pilots/ATCOs who speak in a professional tone?

Q17.1 How often do you encounter pilots who are aware of ATCO's needs/ATCOs who are aware of pilots' needs and willing to help?

- **Section III:**

Q18.1 How often do you encounter pilots/ATCOs who demonstrate excessive authority or superiority in their speech?

Q19.1 How often do you encounter pilots/ATCOs who resort to higher-level authority to solve a conflict?

Q20.1 How often do you encounter pilots/ATCOs who respond in a deferential/submissive style and use excessive politeness?

Q21.1 How often do you encounter pilots/ATCOs who avoid any kind of disagreement or demand in their speech?

Q22.1 How often do you encounter pilots/ATCOs who are concerned with preserving their own images and interests?

Q23.1 How often do you encounter pilots/ATCOs who dominate or compete during an argument?

Q24.1 How often do you encounter pilots/ATCOs who engage in upfront and aggressive conflicts?

Q25.1 How often do you encounter pilots/ATCOs who violate your expectations of a standard flow of communication?

Q26.1 How often do you encounter pilots/ATCOs who speak in a confusing and unclear way?

Q27.1 How often do you encounter pilots/ATCOs who are reluctant to share critical information about a fact/state?

Q28.1 How often do you encounter pilots/ATCOs who show impatience and/or sarcasm in their speech?

Q29.1 How often do you encounter pilots/ATCOs who show annoyance and/or arrogance in their speech?

Q30.1 How often do you encounter pilots/ATCOs who do not accommodate to less proficient speakers'

Q31.1 How often do you encounter pilots/ATCOs who seek disagreement and/or make the other feel

Q32.1 How often do you encounter pilots/ATCOs who are unprofessional and/or unwilling to help?

Q33.1 How often do you encounter pilots/ATCOs who use non-standard phraseology?

Q34.1 How often do you encounter pilots who do not comply with ATCO's orders/ATCOs who do not comply with rules?

Q 35. Please, share any additional comments regarding other communication issues that you feel pose a threat to safety.

Appendix U - Oral Script for Recruitment

CUREB clearance #: 103859

Hello, my name is Ana Lúcia Tavares Monteiro and I am a PhD student in the Applied Linguistics and Discourse Studies Program at Carleton University, located in Canada. I am currently working on a research project under the supervision of Professor Janna Fox.

I am inviting you to participate in a study entitled *Exploring Intercultural Factors in International Pilot-Air Traffic Controller Communications: Validating a Taxonomy Using Mixed Methods Research*. This study aims to investigate how the cultural background of interlocutors may affect the way they interact on the radio using the English language as well as to explore the construct of international radiotelephony communication, from aviation stakeholders' perspectives.

This study will involve approximately 60 minutes of your time, during a small group discussion (30 minutes) and an inter-group discussion (30 minutes). You will be asked to review a transcript of pilot-ATC communication and respond to six questions, as a group. Other groups will be doing the same thing. Subsequently, your group will be asked to discuss the responses to the six questions with other focus group participants in a plenary discussion. With your consent, discussions will be audio-recorded. Once the recording has been transcribed, the audio-recording will be destroyed.

The risk to participants is deemed to be minimal. I will take precautions to protect your identity, and this will be done by keeping all responses and comments confidential in any publications or presentations.

The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research (Clearance number: 103859). If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600 ext. 2517 or via email at ethics@carleton.ca).

You can also reach me at anatavaresmonteiro@cmail.carleton.ca. My supervisor can be reached at janna.fox@carleton.ca. Do you have any questions or need clarification?

If you would like to volunteer, please fill in the consent form available in room xx.

Thank you for your attention.

Ana Lúcia Tavares Monteiro

Appendix V - Recruitment Poster

Carleton UNIVERSITY

Volunteers needed for focus group discussions and interviews!

Participate in a research study on the communicative needs of pilots and air traffic controllers in the multicultural context of aviation radiotelephony communication

(Ethics clearance number: 103859)

To participate in this study, you must be:

- ✓ A pilot or air traffic controller; or
- ✓ An aviation English teacher/material developer; or
- ✓ An aviation English examiner/test developer; or
- ✓ A researcher in the field of aviation communications/aviation English teaching or testing.

Native and non-native speakers of English are welcome!!

This is a 60-minute study. You will be asked to analyze pilot-air traffic controller communications and answer questions based on your perceptions and experience. The risk to participants is deemed to be minimal, as the audio-recordings will be destroyed after their transcription and all responses will be kept confidential in any future publications or presentations.

The ethics protocol for this project has been reviewed and cleared by the Carleton University Research Ethics Board. If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600 ext. 2517 or via email at ethics@carleton.ca).

If you would like to volunteer, contact the researcher **Ana Lúcia Tavares Monteiro** by email at anatavaresmonteiro@cmail.carleton.ca.

Appendix W - Letter of Introduction

Carleton UNIVERSITY
Canada's Capital University

CUREB clearance #: 103859

Title: Exploring Intercultural Factors in International Pilot-Air Traffic Controller Communications: Validating a Taxonomy Using Mixed Methods Research

Date of ethics clearance: 20 April, 2017

Ethics Clearance for the Collection of Data Expires: 30 April, 2018

Dear Sir or Madam,

My name is **Ana Lúcia Tavares Monteiro** and I am a PhD student in the Applied Linguistics and Discourse Studies Program at Carleton University, located in Canada. I am currently working on a research project under the supervision of Professor Janna Fox.

I am writing to you today to invite you to participate in a pilot study on cultural differences in the context of international radiotelephony communications. The purpose of this project is to generate a categorization of intercultural factors that can affect the way pilots and air traffic controllers interact on the radio using the English language. First, the study will explore some culturally influenced factors that can cause misunderstandings and threaten the safety of air-ground communications, such as differences in communication styles, individualism/collectivism, power distance, reluctance to declare emergency, face saving, attitudes, and politeness, among others. Secondly, the study will investigate experienced pilots' and air traffic controllers' perceptions of the potential threats to radiotelephony communications that may result from those factors, including as participants both native and non-native speakers of English. Further, in order to triangulate initial findings from phases 1 and 2, the perceptions of other aviation stakeholder groups (e.g. aviation English teachers, examiners/test developers, or researchers in the field of aviation communications) will be elicited, regarding the communicative needs of pilots and air traffic controllers in the multicultural context of aviation radiotelephony.

You have been invited to take part in this study because you represent one of the stakeholders listed above. If you decide to take part, your responses will contribute to our understanding of how cultural differences may impact air-ground communications and to the construct definition of international radiotelephony communication.

This study involves your participation in one of the following data collection methods:

a) a 30 minute group discussion and a 30 minute inter-group discussion, moderated by the researcher;
b) a 60 minute focus group discussion moderated by the researcher; or
c) a 60 minute individual semi-structured interview conducted by the researcher.

You will be asked to analyze pilot-air traffic controller communications and answer questions based on your perceptions and experience. With your consent, discussions will be audio-recorded. Once the recording has been transcribed, the audio-recording will be destroyed at a date no later than a month after the focus group/interview takes place.

The risk to participants is deemed to be minimal. I will take precautions to protect your identity, and this will be done by keeping all responses and comments confidential and not attributing responses or comments to any participant in future dissertation, other publications or presentations.

However, in the case of focus groups, although I will safeguard the confidentiality of the discussion to the best of my ability, the nature of focus groups prevents me from guaranteeing that other members of the group will do so. Please respect the confidentiality of the other members of the group by not repeating what is said in the focus group to others, and be aware that other members of the group may not respect your confidentiality.

You have the right to end your participation in the study at any time, for any reason, up until a month after the focus group or interview takes place. You can withdraw by emailing the researcher or the research supervisor. If you withdraw from the study, all information you have provided will be immediately destroyed.

All research data, including audio-recordings and any notes will be encrypted. Any hard copies of data (including any handwritten notes or USB keys) will be kept in a locked cabinet at the researcher's home office. Research data will only be accessible by the researcher and the research supervisor.

Once the project is completed, all research data will be kept until the completion of my PhD studies and potentially used for other research projects on this same topic. At the end of five years, all research data will be securely destroyed (electronic data will be erased and hard copies will be shredded).

All participants may benefit from the study by exploring their own perceptions and understandings of how cultural differences may affect language behavior and by raising awareness of the potential threats of intercultural factors to the safety of radiotelephony communications in English. Moreover, a clearer definition of the construct of international radiotelephony communication may benefit aviation English teachers, examiners, test developers, and researchers.

The results of this research may be presented in conferences or published in professional journals; however, the findings will be reported in an aggregated manner, which may sometimes be exemplified by cautiously selected anonymous quotes.

If you would like a copy of the finished research project, you are invited to contact the researcher to request an electronic copy which will be provided to you.

The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600 ext. 2517 or via email at ethics@carleton.ca).

Researcher contact information:
Ana Lúcia Tavares Monteiro
School of Linguistics and Language Studies
Carleton University
Email: anatavaresmonteiro@cmail.carleton.ca

Supervisor contact information:
Prof. Janna Fox
School of Linguistics and Language Studies
Carleton University
Email: janna.fox@carleton.ca

_____ _____
Signature of participant Date

_____ _____
Signature of researcher Date

Appendix X - Workshop Handout

Workshop Title: Language, culture and effective radiotelephony communications: Analyzing case studies from aviation stakeholders' perspectives

Presenter: Ana Lúcia Tavares Monteiro
Organization: Carleton University (Canada) and ANAC (Brazil)

Group 1: Please write the number of each type of participant in your group.
If anyone has overlapping roles, include him/her in the option that best represents his/her main activity:
( ) pilots ( ) ATCOs ( ) aviation English teachers ( ) examiners/test developers ( ) researchers ( ) regulators ( ) other: _____

How is your group composed of in terms of language background?
( ) native speakers of English ( ) non-native speakers of English

Background information: With 8 in sequence for runway 28 at Dublin, a British Airways pilot flying an A319 interacts with a female tower controller at Dublin at its peak, mid-day.

		Scenario 6 - Transcript (Available at https://www.youtube.com/watch?v=JYROSTV_KVg)	Comments
1	PILOT	And [ ], we are VFR under the clouds right now. And if you could give me a (inaudible).	
2	ATCO	You're not familiar with this airspace?	
3	PILOT	Yes sir, I'm very familiar with this airspace. But just coming through the clouds now it would be easier if you just give me my heading for a moment.	
4	ATCO	What kind of NAV equipment do you have on board?	
5	PILOT	Slant Uniform, VOR sir.	
6	ATCO	[ ] fly heading 150. Vectors Mile Square Park.	
7	PILOT	Okay, we are currently 150 sir. Thank you sir, just wanted a little help. Thank you.	
8	ATCO	Well, let me give you some advice. We are really busy. We've got one controller working all the airspace and a lot of inbounds coming in, the last airliners coming into John Wayne. I probably don't always have time to hold your hand. Sorry to say that, but that's the truth.	
9	PILOT	25 years I have been flying this airspace sir. I've never had a controller talk to me like that.	
10	ATCO	Well, you are welcome to call me on the phone.	
11	PILOT	Love to!	

Questions to guide your discussion:	**Comments:**			
1) Is communication effective? How do pilots and/or ATCs assess this interaction?				
2) How does the cultural background of interlocutors affect the outcome of this communication?				
	Awareness	Knowledge	Skills	Attitudes
3) How might the interaction have been improved, in terms of awareness, knowledge, skills, and attitudes?				
4) How could the identified awareness, knowledge, skills, and attitudes be addressed in terms of teaching and learning activities aiming at more effective and safer communications?				
5) Are all identified awareness, knowledge, skills, and attitudes aligned with the current ICAO testing policy?				
6) How could they be operationalized into Aviation English test design?				

Appendix Y - Indicator Checklist of skills/behaviors indicative of effective communication

Instructions

Complete the checklist for Task 1 and Task 2. Circle the best option for each item: "Yes", "No", or "N/A". Then, rate the level of importance of each indicator, from 1 (not important) to 4 (very important).

In the role-play task there is evidence of the test-taker…

Indicators of **Professional (AE) competence**	**Task 1**	**Task 2**	**Level of importance**
1.complying with the rules of use that characterize the domain (e.g. use of phraseology, readback/hearback, etc)	Yes No N/A	Yes No N/A	1 2 3 4
2.demonstrating a professional attitude and tone	Yes No N/A	Yes No N/A	1 2 3 4
3.communicating effectively in routine and in unpredictable situations	Yes No N/A	Yes No N/A	1 2 3 4
4.using plain English (when appropriate) for aeronautical RT communication	Yes No N/A	Yes No N/A	1 2 3 4
5.producing and recognizing the language functions used in RT	Yes No N/A	Yes No N/A	1 2 3 4
Indicators of **ELF competence**	**Task 1**	**Task 2**	**Level of importance**
6.accomodating to different accents and dialects	Yes No N/A	Yes No N/A	1 2 3 4
7.adapting linguistic forms to the communicative needs at hand	Yes No N/A	Yes No N/A	1 2 3 4
8.complying with the safety-critical requirements of intelligibility	Yes No N/A	Yes No N/A	1 2 3 4
9.avoiding the use of jargon, idioms, slang and colloquialism	Yes No N/A	Yes No N/A	1 2 3 4
10. adjusting and aligning to different communicative systems (e.g. new patterns of phonology, syntax, discourse styles)	Yes No N/A	Yes No N/A	1 2 3 4
Indicators of **Intercultural awareness/competence**	**Task 1**	**Task 2**	**Level of importance**
11.showing openness and flexibility to different cultural frames of reference (e.g., communication style, conflict management, face-work strategies, etc)	Yes No N/A	Yes No N/A	1 2 3 4
12.engaging with politeness conventions	Yes No N/A	Yes No N/A	1 2 3 4
13.engaging with and negotiating sociocultural differences	Yes No N/A	Yes No N/A	1 2 3 4
14.showing willingness to cooperate and to relativize one's own values, beliefs and behaviors	Yes No N/A	Yes No N/A	1 2 3 4
15.accomodating to difference and to multilingual aspects of intercultural communications	Yes No N/A	Yes No N/A	1 2 3 4
Indicators of **Interactional competence**	**Task 1**	**Task 2**	**Level of importance**
16.eliminating idioms, cultural references and syntactic complexity from speech	Yes No N/A	Yes No N/A	1 2 3 4
17.demonstrating a shared responsibility for successful communication	Yes No N/A	Yes No N/A	1 2 3 4
18.accomodating to the constraints of the context and perceived ability of the hearer	Yes No N/A	Yes No N/A	1 2 3 4
19.dealing adequately with apparent misunderstanding, by checking, confirming and clarifying	Yes No N/A	Yes No N/A	1 2 3 4
20. using an appropriate participation framework	Yes No N/A	Yes No N/A	1 2 3 4
21. demonstrating tolerance and collaborative efforts	Yes No N/A	Yes No N/A	1 2 3 4

Appendix Z - Observation Checklist of language functions

Communicative language functions associated with aviation (ICAO, 2010)		Task 1	Task 2	Level of importance
1. Directed towards triggering action	**1.1 Orders**			
	Give an order (C)	Yes No NA	Yes No NA	1 2 3 4
	Give an amended order (C)	Yes No NA	Yes No NA	1 2 3 4
	Give a negative order (C)	Yes No NA	Yes No NA	1 2 3 4
	Give alternative orders (C)	Yes No NA	Yes No NA	1 2 3 4
	Cancel an order (C)	Yes No NA	Yes No NA	1 2 3 4
	Announce compliance with an order (P)	Yes No NA	Yes No NA	1 2 3 4
	Announce non-compliance with an order (P)	Yes No NA	Yes No NA	1 2 3 4
	1.2 Requests and offers to act			
	Request action by another (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Agree to act (C/P)	Yes No NA	Yes No NA	1 2 3 4
	State reluctance/unwillingness to act (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Refuse to act (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Offer to act (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Accept an offer to act (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Refuse an offer to act (C/P)	Yes No NA	Yes No NA	1 2 3 4
	1.3 Advice (markers for politeness)			
	Request advice (P)	Yes No NA	Yes No NA	1 2 3 4
	Give advice (P)	Yes No NA	Yes No NA	1 2 3 4
	Suggest a course of action (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Suggest a solution to a problem (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Suggest alternative courses of action (C/P)	Yes No NA	Yes No NA	1 2 3 4
	1.4 Permission/approval (markers for politeness, directness)			
	Request permission/approval (P)	Yes No NA	Yes No NA	1 2 3 4
	Give permission/approval (C)	Yes No NA	Yes No NA	1 2 3 4
	Deny permission/approval (C)	Yes No NA	Yes No NA	1 2 3 4
	Forbid (C)	Yes No NA	Yes No NA	1 2 3 4
	1.5 Undertakings			
	Undertake to give a service (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Agree to undertaking/decision (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Undertake to assist (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Undertake to contact/relay/report (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Announce a spontaneous decision to act (C/P)	Yes No NA	Yes No NA	1 2 3 4
2. Sharing information	**2.1 Information concerning present facts**			
	Request information (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Give information(C/P)	Yes No NA	Yes No NA	1 2 3 4
	Request a detailed description (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Describe a state (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Describe a changed state (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Describe an unchanged state (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Describe an action in progress (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Describe a process (C)	Yes No NA	Yes No NA	1 2 3 4
	Describe a procedure (C)	Yes No NA	Yes No NA	1 2 3 4
	Describe aims/precautions (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Describe the source of a problem (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Describe a visual impression (C/P)	Yes No NA	Yes No NA	1 2 3 4

	Quote rules (C)	Yes No NA	Yes No NA	1 2 3 4
	Ask about needs/wishes (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Ask about preferences (C)	Yes No NA	Yes No NA	1 2 3 4
	State preferences (P)	Yes No NA	Yes No NA	1 2 3 4
	Ask about readiness/availability (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Announce readiness/availability (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Request reasons (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Give reasons (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Request instructions on how to do (P)	Yes No NA	Yes No NA	1 2 3 4
	Give instructions on how to do (C)	Yes No NA	Yes No NA	1 2 3 4
	Identify (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Announce a problem (C/P)	Yes No NA	Yes No NA	1 2 3 4
	2.2 Information concerning the future	Yes No NA	Yes No NA	
	Announce an expected action/event	Yes No NA	Yes No NA	1 2 3 4
	Ask about the expected moment/duration of an event (C/P)	Yes No NA	Yes No NA	1 2 3 4
	State the expected moment/duration of an action/event (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Ask about possible consequences of an action/event (C/P)	Yes No NA	Yes No NA	1 2 3 4
	State possible consequences of an action/event (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Ask about intentions (C/P)	Yes No NA	Yes No NA	1 2 3 4
	State intentions (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Request prediction (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Predict a future action/event (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Warn (C/P)	Yes No NA	Yes No NA	1 2 3 4
	2.3 Information concerning immediate/recent past events			
	Announce a completed action/event having an effect on the present (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Announce a change (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Announce a nearly completed action (C/P)	Yes No NA	Yes No NA	1 2 3 4
	2.4 Information concerning the past	Yes No NA	Yes No NA	
	Ask about past events (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Announce a past action/event (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Announce an avoided problem/incident (P)	Yes No NA	Yes No NA	1 2 3 4
	Give a report (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Describe a previous communication (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Describe a sequence of past actions/events (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Request an explanation of a past action/event (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Give an explanation of a past action (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Indicate deductive reasoning (C/P)	Yes No NA	Yes No NA	1 2 3 4
	2.5 Necessity			
	Ask about necessity(C/P)	Yes No NA	Yes No NA	1 2 3 4
	State necessity (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Announce a compulsory action (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Announce an inevitable action/event (C/P)	Yes No NA	Yes No NA	1 2 3 4
	2.6 Feasibility/capacity			
	Ask about the feasibility/capacity (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Announce feasibility/capacity (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Announce unfeasibility/incapacity (C/P)	Yes No NA	Yes No NA	1 2 3 4

3. Management of the pilot-controller relation	Greet/take leave (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Respond to greeting/leave-taking (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Thank (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Respond to thanks (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Complain (P)	Yes No NA	Yes No NA	1 2 3 4
	Apologize (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Express dissatisfaction (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Reprimand (C)	Yes No NA	Yes No NA	1 2 3 4
	Reject complaint/reprimand (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Express satisfaction (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Express concern/apprehension (P)	Yes No NA	Yes No NA	1 2 3 4
	Reassure (C)	Yes No NA	Yes No NA	1 2 3 4
	Encourage (C)	Yes No NA	Yes No NA	1 2 3 4
4. Management of the dialogue	Name addressee(s) (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Self-correct (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Paraphrase (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Close an exchange (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Request response (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Read back (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Check understanding (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Acknowledge (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Check certainty (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Declare non-understanding (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Correct a misunderstanding (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Request repetition (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Give repetition (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Request confirmation (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Give confirmation (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Request clarification (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Give dis-confirmation (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Give clarification (C/P)	Yes No NA	Yes No NA	1 2 3 4
	Relay an order (C)	Yes No NA	Yes No NA	1 2 3 4
	Relay a request to act (C)	Yes No NA	Yes No NA	1 2 3 4
	Relay a request for permission (P)	Yes No NA	Yes No NA	1 2 3 4

Appendix AA - Characteristics of the proposed role-play task for pilots (Task2)

(Based on the framework of specific purpose test task characteristics – Douglas, 2000)

Rubric		
Objective	To demonstrate the ability to speak and understand the language used for radiotelephony communications	
Procedures for responding	Speak to interlocutor, interaction will be video-recorded	
Structure		
Number of tasks	5 small role-plays	
Relative importance	Role-plays of equal importance	
Distinction between tasks	Quite distinct – new phases are announced	
Time allotment	1-minute preparation; 1 minute to listen to ATIS; 2 minutes for each role-play	
Evaluation		
Criteria	ICAO rating scale: pronunciation, structure, vocabulary, fluency, comprehension, interactions; Test-takers are aware of the *Indicator Checklist of skills/behaviors of effective communication.*	
Procedures	Two raters (1 ELE and 1 SME) use a standard form to score performance independently after the test; third rater in case of disagreement between final levels 3 and 4; all categories scored on a scale of 1 – 6.	
Input		
Prompt		
LSP context	Test-taker's card	Interlocutor's card
Setting	Aeronautical RT communication	Aeronautical RT communication
Participants	Pilot	ATCO
Purposes	Communicate with ATCO from en-route (outbound from Airport A) to taxi (at Airport B) Explicit: Listen to ATIS information. Contact ATCO and reply to his messages in an appropriate way, taking into account the phase of the flight and the information given (in both routine and non-routine situations) Implicit: manage the dialogue and the pilot-ATCO relation	Explicit: Interact with the pilot according to the phase of the flight and the scenario given. Explicit: Produce an intentionally ambiguous utterance/Indicate failure to comprehend/Request clarification/check understanding/request repetition
Form/Content	Oral interaction with no visual contact; Implicit: routine information exchange in each phase of the flight Explicit: Non-routine situations to handle	Oral interaction with no visual contact; Implicit: routine information exchange in each phase of the flight Explicit: Non-routine situations to handle
Tone	Professional manner	Professional manner
Language	ICAO standardized phraseology and plain English for aeronautical communication; Communicative language functions for aviation	Same as for candidate
Norms	Implicit: Pilot/ATCO interaction over the radio; elements of radio transmitting techniques	Same as for candidate
Genres	International aeronautical radiotelephony communications	Same as for candidate

Identification of problems	Deal with: medical emergency during taxi/runway incursion at take-off/depressurization en-route Negotiate meaning	Provide information, directions and assistance to pilot Simulate non-understanding/ produce intentionally ambiguous utterance/request repetition
Input data		
Format	Aural and visual	
Vehicle	Aural - Taped: Genuine recording of (or simulated) ATIS information Aural - Live: oral input from interlocutor Visual: two pictures portraying unexpected situations	
Length	Prompt: 20 lines Recording: 0.5 minutes Oral input: 4-5 minutes	
Level of authenticity		
Situational	Fairly high: common type of situations to pilots; pictures portray real scenarios	
Interactional	High to moderate: some interlocutors (SME-pilots) are not highly trained in their role as ATCO	
Expected response		
Format	Oral	
Type	Extended	
Response content		
Language	ICAO standardized phraseology and plain English for aeronautical communication; language functions associated with aviation; strategies related to Aviation English competence, ELF competence, Intercultural awareness/competence and Interactional competence.	
Background knowledge	Aviation knowledge, RT communication rules and procedures	
Level of authenticity		
Situational	High situational authenticity: The task shares many features of TLU situation	
Interactional	High to moderate: The task engages an appropriate discourse domain in test-takers (aeronautical RT communications)	
Interaction between input and response		
Reactivity	Highly reciprocal: adaptation on both sides as necessary for mutual comprehension	
Scope	Moderately broad: must process information on card	
Directness	Fairly indirect: must use background knowledge	
Assessment		
Construct definition	Refer to the matrix of construct specification, which details the dimensions of interest (awareness, knowledge, skills and attitudes) across the domains of Aviation English, English as a lingua franca, Intercultural awareness/competence, and Interactional competence	
Criteria for correctness[a]	ICAO Rating Scale descriptors: pronunciation, structure, vocabulary, fluency, comprehension, interactions. Expanded assessment criteria which values strategies related to Aviation English competence, ELF competence, Intercultural awareness/competence and Interactional competence.	
Rating procedures	Two raters (1 ELE + 1 SME) rate the sample independently after the test terminates, by listening to the audio/video recording. A third rater is used only in case of disagreement between final levels 3 and 4.	

Note. [a] Assessment criteria needs to be expanded based on trialing data accumulated over time as part of the task validation process.

Appendix BB - Draft task specifications

Spec #: ST-RP 01

Title: Speaking task to be included in the pilots' language assessment: Role-play

Purpose
The purpose of this test task is to assess pilots' ability to communicate with air traffic controllers over the radio within the high-stakes context of international aviation radiotelephony, using the English language. The objective is to make decisions about pilots' readiness to communicate in this intercultural workplace setting based on inferences about their ability to function efficiently as international communicators, leading to a licence endorsement. That is, a certification of pilots by determining whether they are competent enough to perform their communicative duties in international aviation operations.

Characteristics of the TLU situation
Pilots communicate with ATCOs over the radio. The complexity of tasks performed by pilots and ATCOs requires a joint cooperative effort, including distribution of knowledge and high levels of coordination with artifacts and technological tools. Communications happen during routine and non-routine situations; speakers are separated in space and have no visual contact; transmitting techniques allow only one speaker at a time, who engage in receptive, productive, interactive and mediating activity, from different perspectives and standpoints. One pilot interacts with several ATCOs throughout distinct phases of the flight, whereas one ATCO talks to several pilots in the same radio frequency.

Characteristics of the TLU tasks
Pilots perform a range of communicative tasks while interacting with ATCOs in distinct positions and locations, which vary according to the phase of the flight. Generally, they communicate with clearance delivery (pre-flight), ground controller, (push-back and taxi), tower controller (take-off and departure), departure controller (climb), area controller (en-route), approach controller (approach), tower controller (final approach and landing), and ground controller (taxi). Among the communicative language functions associated with aviation (ICAO, 2010) performed by pilots, it is possible to cite some as examples: request permission/approval; request and give information; describe an action in progress; announce readiness/availability; announce a problem; state intentions; read back; acknowledge; request repetition; give confirmation, etc. They are organized into four categories: (a) directed towards triggering actions; (b) sharing information; (c) management of the pilot-controller relation; and (d) management of the dialogue.

Characteristics of the test-takers
Commercial airline pilots, who are qualified to fly under Instrument Flight Rules (IFR), both men and women. They come from a variety of national, linguistic and cultural backgrounds, including those who speak English as L1, L2 or as an additional language. Varying levels of communicative English ability, background knowledge, flying experience and exposure to intercultural communications. Test-takers are expected to interact with an interlocutor who is proficient in the English language, experienced in radiotelephony communications and who, ideally, is not familiar with his/her L1[31].

Construct Definition
The construct to be measured refers to the specific purpose language ability required to communicate effectively with air traffic controllers from different linguacultural backgrounds over the radio. It was specified based on the analyses conducted in Phases 1 and 2 of this multiphase MM study, drawing on what aviation stakeholders value as important for safe and effective radiotelephony communications. That is, an aviation radiotelephony-specific communicative construct, as opposed to a language proficiency construct, is what the task intends to measure. It comprises test-takers' awareness, knowledge, skills and attitudes in the domains of Aviation English, English as a Lingua Franca, Intercultural awareness/competence and

[31] Although the task was pilot tested with a group of Brazilian test-takers and interlocutors, the task was designed for an international community of users.

Interactional competence. Although not directly assessed in the test-taker performance, it is important to highlight the role of background knowledge to accomplish the task. For more details, refer to the matrix of construct specification within the aviation radiotelephony domain (see Table 8.8).

General Description of the Task

The task is a role-play activity in which the test-taker interacts as a pilot in the context of international aeronautical radiotelephony communication, with a visual barrier. He is required to respond orally to an interlocutor, who plays the role of several air traffic controllers in sequential phases of a flight, during 12-14 minutes. The task begins with a listening component of an Automatic Terminal Information Service (ATIS), defined as "the automatic provision of current, routine information, to arriving and departing aircraft" (ICAO, 2007, p. 1-6), to which pilots normally listen to before initial contact with ATCOs. As in real life, it gives the test-taker a lot of contextual information about the airport conditions in the flight planning phase. The test-taker listens to it using headsets and takes notes. Then, he starts interacting according to the contextual information in his role-play card in five sequential mini role-plays. He contacts clearance delivery (pre-flight), ground controller, (push-back and taxi), tower controller (take-off) and departure controller (climb). Throughout the phases of the flight, the pilot is expected to elicit and respond to information provided by the ATCO, as well as to give information and state intentions using appropriate language functions and standard expressions. Unexpected situations and linguistic complications are intentionally included in the task, which requires the use of plain language and communicative strategies to manage the interaction, negotiate meaning, and accommodate to his/her interlocutor.

Rubric

Test-takers are told that they will take the role of a pilot in a role-play task with an interlocutor who will perform the role of several air traffic controllers in different phases of the flight. They are also told that taking notes is permitted during the task and that they will have: (a) one minute to have a look at the prompt, which contains contextual information about their flight and important directions to be followed in order to complete the task, requiring a certain degree of specific background knowledge to be processed.; (b) one minute to listen to ATIS information and take notes; and (c) two minutes for each of the five mini role-plays, which are of equal importance. Test-takers use headsets while listening to ATIS information and, ideally, continue using headsets with a microphone throughout the entire role-play, in addition to the visual barrier placed between the interlocutors. Pilots are aware of the assessment criteria in the ICAO rating scale, and also about the criteria which values strategies related to Aviation English competence, ELF competence, Intercultural awareness/competence and Interactional competence.

Input

- **Prompt:** The prompt includes information to identify the specific context of aeronautical RT communication set by the task, including aircraft call sign, departure and landing airports, type of flight (IFR), and current location of the aircraft at the airport. It also contains the phases of the flight included in the task and a set of directions to be followed by the test-taker in each of them, for both routine and non-routine situations. Some of the information is implicit, e.g., routine information exchange in each phase of the flight while others are explicit, e.g., non-routine situations to handle. These situations should vary in topic in the design of distinct tasks, but should still be appropriate for the phase of the flight and aim for an equivalent level of difficulty.

- **Input data:** Input data that test-takers need to process is in two different formats: aural and visual. Aural input includes the following vehicles: (a) taped - one genuine recording of ATIS information or simulated recording with an international accent and appropriate background noise (0.5 minutes); and (b) live - oral input from interlocutor (4-5 minutes). As for the visual input, the task presents the test-taker with three pictures that portray real scenarios of unexpected and/or emergency situations that pilots may face during a flight. They need to elicit different topics and be appropriate to the specific phase of the flight. The prompt also includes written information that should not exceed 20 lines in total.

Expected Response
Test-takers responses are video recorded. They are expected to respond orally to their interlocutors using ICAO standard phraseology and plain English for aeronautical communication whenever necessary, and also using the required language functions associated with the communicative demands in each phase of the flight. Moreover, not only aviation background knowledge and radiotelephony communication rules and procedures are required in order to communicate effectively and efficiently with ATCOs, but also it is expected that test-takers demonstrate a set of behaviors indicative of ELF competence, intercultural awareness/competence and interactional competence. Although test-takers are not actually flying an airplane, as the task shares many features of the TLU situation, i.e., includes a sequence of communicative activities they are used to perform in a real flight, it can be said that it has high situational authenticity. As for interactional authenticity, the prompt and input data provided help to engage test-takers in real RT communications and involve them in the task in terms of field specific knowledge and language knowledge.

Interaction between input and response
In terms of reactivity, the task can be said to be highly reciprocal, as the test-taker receives and gives immediate feedback to his interlocutor about the effectiveness of his communication, requiring adaptation and negotiation on both sides for mutual comprehension. The task is also moderately broad in scope, requiring the test-taker to process different but related types of input with contextualized information and to produce an extended oral response. In addition, as the task engages a specific purpose discourse domain in the test-takers and requires a great deal of field specific background knowledge in order to complete it, we can say that it is fairly indirect.

Assessment
In order to assess test-takers' performance in such a highly specialized professional domain, the assessment criteria included in the ICAO rating scale descriptors – pronunciation, structure, vocabulary, fluency, comprehension and interactions – seem not to address all of the features that aviation stakeholders consider as relevant for successful interactions in real-life communications over the radio. Therefore, the assessment criteria needs to be expanded based on the *Indicator Checklist of skills/behaviors indicative of effective communication* in pilot-ATCO interactions, which values strategies related to Aviation English competence, ELF competence, Intercultural awareness/competence and Interactional competence.

Two raters (1 ELE + 1 SME) rate the sample independently after the test terminates, by listening to the video recording. A third rater is used only in case of disagreement between final levels 3 and 4 of the ICAO Rating scale. Raters receive initial training in advance of rating and recurrent training once a year.

Sample Task
See Attachment A for an excerpt of the role-play task **SLT-RP 01** (one phase of the flight only), including a sample of the test-taker's card and the corresponding interlocutor's card.

The task should reflect the qualities of good testing practice (Douglas, 2000)[32]:
Reliability (R), Validity (V), Situational Authenticity (SA), Interactional Authenticity (IA), Impact (IMP), Practicality (P)[33]
The Task must:

1. Ensure that scores represent consistent measures of the construct (R)
2. Be aligned with test purpose, construct definition, and assessment criteria (V)
3. Include prompts and directions that target the construct to be measured (V)
4. Be aligned with the directions and with all forms of input data (V)
5. Provide information about students' strengths and weaknesses in regards to the construct (V)

[32] Douglas (2000) has based his list of the qualities of good language testing practice on the qualities of test usefulness in Bachman and Palmer (1996), but makes a distinction between them because he has "focused the qualities a bit differently so as to make them more relevant to LSP testing" (p. 114).
[33] Adapted from Williams' (2017) task specification for "Parallel Integrated Writing Task for Diagnostic Assessment: Hyperloop".

6. Ensure the relevance and appropriate coverage of test content (V)
7. Be centered on topics that are aviation-specific and likely to appear in the context of radiotelephony communication between pilots and ATCOs (V, SA)
8. Reflect the target language use domain (SA)
9. Elicit radiotelephony-specific terminology (e.g. standard phraseology and plain English for aeronautical communications) (V, SA, IA)
10. Elicit an appropriate range of language functions typical to the genre (e.g. request information, state intentions, request permission, announce a problem, etc) (V, SA, IA)
11. Elicit behaviors indicative of ELF competence, intercultural awareness/competence and interactional competence (V, SA, IA)
12. Engage an appropriate discourse domain (aeronautical RT communication) in the test- takers (IA)
13. Involve test-takers in the task in terms of field specific knowledge and language knowledge (IA)
14. Have a low perceivable negative impact on test-takers' learning or well-being (IMP)
15. Include topics that are not likely to elicit any negative emotional response (IMP)
16. Have a positive impact on safe and effective communications between speakers of different language backgrounds, including those who speak English as L1 (IMP)
17. Be administrated practically (required human and technological resources are available, can be scored within a reasonable period of time) (P)

Specification Supplement

Consult the following resources for more information:

- ICAO Rating Scales and holistic descriptors (ICAO, 2011)
- ICAO list of communicative language functions associated with aviation (ICAO, 2010)
- Matrix of construct specification within the aviation radiotelephony domain (Table 8.8)
- *Indicator Checklist of skills/behaviors indicative of effective communication* (Appendix Y)
- Characteristics of the proposed role-play task for pilots (Task 1: Table 9.2 and Task 2 – Appendix AA)

Appendix BB – Attachment A - A sample of the role-play task SLT-RP 01

(Take-off phase) including an excerpt of the interlocutor's card and the corresponding test-taker's card

Tower controller	Take-off		
			ATCO: ANAC 123, 'AIRPORT A' TOWER. ARE YOU READY FOR IMMEDIATE DEPARTURE?
			PILOT: AFFIRM, ANAC 123.
			ATCO: ANAC 123, CLEARED TO LINE UP AND IMMEDIATE TAKEOFF, RUNWAY 10, WIND 070 DEGREES AT 10KT.
			PILOT: CLEARED TO LINE UP AND IMMEDIATE TAKEOFF, RUNWAY 10, ANAC 123.
			(Show picture)
		Runway incursion	
			PILOT: ANAC 123, ABORTING TAKE OFF. THERE IS A VEHICLE ON THE RUNWAY.
		Attention: The interlocutor produces an intentionally ambiguous utterance which must be negotiated by the test taker.	*ATCO: ANAC 123, ROGER. DO YOU HAVE ENOUGH SPACE TO BACKTRACK ON THE RUNWAY AND RETURN TO THE THRESHOLD?*
			Pilot reads this information in his role-play card: "Respond positively".
			PILOT: ROGER, BACKTRACKING TO THE THRESHOLD.
			Pilot reads this information in his role-play card: "You have just taken off".
		Attention: The interlocutor produces an intentionally unintelligible piece of information.	*ATCO: ANAC 123, AIRBORNE AT 22. CONTACT 'AIRPORT A' DEPARTURE ON FREQUENCY 119.20 (unintelligible).*
			PILOT: SAY AGAIN THE DEPARTURE FREQUENCY FOR ANAC 123? OR: SAY AGAIN, PLEASE?
			ATCO: ANAC 123, I SAY AGAIN, CONTACT DEPARTURE ON FREQUENCY 119,20.
			PILOT: ANAC 123, CONTACT 'AIRPORT A' DEPARTURE ON 119.20.

Interlocutor's card

Tower controller	Take-off	PILOT	ATCO
		Interact appropriately.	
		Interact appropriately.	
		Based on the picture you receive, interact with ATCO.	
		Interact appropriately.	
		Respond positively to the controller.	
		Interact appropriately.	
		You have just taken off.	
		Interact appropriately.	
		Interact appropriately.	

Test-taker's card

Note. Interlocutors are guided by a script with possible responses but may need to adapt according to the demands of the communicative situation.

Appendix CC - Semi-structured interview questions

For AETEs who played the role of test-takers:

1) How did you feel during task administration?
2) Did you have any difficulty to follow the guidelines provided in the role-play card? If yes, please explain.
3) Overall, do you think the task was appropriate to elicit the language used for radiotelephony communications?
4) Did you feel as if you were actually engaged in real radiotelephony communications as you carried out the task?
5) In your opinion, would you add or remove anything from the task to make it more authentic for the assessment of pilots' aviation English proficiency?
6) Do you have any additional comments and/or suggestions regarding the task?

For AETEs who played the role of interlocutors:

1) How did you feel during task administration?
2) Did you have any difficulty to administer the task? If yes, please explain.
3) Overall, do you think the task was appropriate to elicit the language used for radiotelephony communications?
4) Was the information you received in the role-play card enough to guide you through the interaction?
5) Were the prompts to "Check understanding", "Produce an intentionally ambiguous utterance which must be negotiated by the test-taker", "Indicate failure to comprehend a test-taker's utterance, requesting clarification" and "Request repetition" challenging for you, as an interlocutor?
6) In your opinion, would you add or remove anything from the task to make it more authentic for the assessment of pilots' aviation English proficiency?
7) Do you have any additional comments and/or suggestions regarding the task?
8) What type of training should be provided to interlocutors (pilots and air traffic controllers) in order to best administer this task? (consider standardization)

Appendix DD - Feedback form for AETEs: role of test-takers

Please circle the level of agreement that corresponds to your opinion on the questions below: 1= Strongly disagree to 6= Strongly agree.

To what extent:	**Level of agreement**
1) …were the task instructions clear to you?	1 2 3 4 5 6
2) …was the information you received in the role-play card enough to guide you through the interaction?	1 2 3 4 5 6
3) …did the task enable you to demonstrate your ability to speak and understand the language used for radiotelephony communications?	1 2 3 4 5 6
4) …were you confident during the task administration?	1 2 3 4 5 6
5) … was the language elicited by the task appropriate to the communicative demands of pilots in international radiotelephony?	1 2 3 4 5 6
6) … was the task appropriate to measure "Interactions" according to ICAO rating scale?	1 2 3 4 5 6
7) …was the room arrangement and equipment used appropriate to the administration of the task?	1 2 3 4 5 6
8) …was the task of an appropriate length?	1 2 3 4 5 6
9) …did the interlocutor play the role of the air traffic controller adequately?	1 2 3 4 5 6
10) …did you feel as if you were actually engaged in real radiotelephony communications as you carried out the task?	1 2 3 4 5 6
11) …were you involved in the task in terms of field specific knowledge and language knowledge?	1 2 3 4 5 6
12) … is the task likely to cause a positive impact on pilots' language training?	1 2 3 4 5 6

Appendix EE – Feedback form for AETEs: role of interlocutors

Please circle the level of agreement that corresponds to your opinion on the questions below: 1= Strongly disagree to 6= Strongly agree.

To what extent:	**Level of agreement**
1) …did you feel the task instructions clear to the test-taker?	1 2 3 4 5 6
2) …was the information you received in the role-play card enough to guide you through the interaction?	1 2 3 4 5 6
3) …was the task difficult to administer?	1 2 3 4 5 6
4) …were you confident during the task administration?	1 2 3 4 5 6
5) … was the language elicited by the task appropriate to the communicative demands of pilots in international radiotelephony?	1 2 3 4 5 6
6) …did the task enabled the test-taker to demonstrate his/her ability to speak and understand the language used for radiotelephony communications?	1 2 3 4 5 6
7) … was the task appropriate to measure "Interactions" according to ICAO rating scale?	1 2 3 4 5 6
8) …was the room arrangement and equipment used appropriate to the administration of the task?	1 2 3 4 5 6
9) ... was the task of an appropriate length?	1 2 3 4 5 6
10) …did you feel the test-taker was actually engaged in real radiotelephony communications as you administered the task?	1 2 3 4 5 6
11) …did you feel the test-taker was involved in the task in terms of field specific knowledge and language knowledge?	1 2 3 4 5 6
12) … is the task likely to cause a positive impact on pilots' language training?	1 2 3 4 5 6

Appendix FF - Demographic information

Aviation English Testing Expert (AETE):

1. Gender: ( ) Male ( ) Female
2. Age range: ( ) Under 25 years ( ) 25-35 ( ) 36-45 ( ) More than 45
3. Years of experience as an Aviation English examiner (rater/interlocutor): (Check one)

( ) Less than 2 years ( ) 2-4 years ( ) 4-6 years ( ) 6-8 years ( ) more than 8 years

4. Your field of expertise:
   ( ) English Language Expert ( ) Subject Matter Expert – Pilot
   ( ) Subject Matter Expert – Air traffic controller
5. Nationality: _____
6. First language (L1) or native language (s): _____
7. Second language (L2) or other language(s): _____

8. Highest level of education: (check one)

( ) Secondary School ( ) College ( ) University ( ) Master Degree

( ) Other _____

Appendix GG - Focus group questions: AETEs who played the role of raters

1) Overall, do you think the task was appropriate to elicit the language functions used for RT communications?

2) Overall, do you think the task was appropriate to elicit the behaviors indicative of effective RT communication?

3) Would you add or remove anything from the task to make it more authentic for the assessment of pilots' aviation English proficiency?

4) In terms of assessment criteria, would it be easier to rate the performance of test-takers with more detailed descriptors, relevant to the requirements of the task?

5) To what extent do you think interlocutors' behavior could affect the performance of test-takers and, as a result, the assessment of their performance?

6) What type of training should be provided to interlocutors (SMEs') in order to best administer this task?

7) Was the task of an appropriate length?

8) Was the room arrangement and equipment used appropriate to the administration of the task?

9) Do you have any additional comments and/or suggestions regarding the assessment of task performance?