

## University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g., Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.



**UNIVERSITY OF SOUTHAMPTON**

**FACULTY OF HUMANITIES**

School of Modern Languages

**Identifying Context-Sensitive English Language Use  
Needs for a University Exit Test:  
a Case Study in Durango, México**

by

**Cecilia Araceli Medrano Vela**

Thesis for the degree of Doctor of Philosophy

September 2023



# UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF HUMANITIES

School of Modern Languages

Doctor of Philosophy

## **Identifying context-sensitive English language use needs for a university exit test: a case study in Durango, México.**

By Cecilia Araceli Medrano Vela

This research examines to what extent the language use needs of the local work environment can be considered when designing a university English language test.

The study takes place at a Mexican university that has established the CEFR B1 level as a university exit requirement. This requirement has become a funnel, leading some students to modify their graduation plans until it is fulfilled.

A questionnaire and a short semi-structured interview were applied to explore the face validity of the in-house developed General English Language Test. International work domain language descriptors at the B1 level were used to identify the workplace language needs within the local context. Professionals provided specific examples of language use activities within the workplace.

The face validity of the test in use does not appear to meet the expectations of test takers, particularly about the domain in which the test is focused. Results reveal that although test takers are interested in using the language for social purposes, the professional and academic domains are more relevant and useful to their professional language use needs. The specific examples of language use activities allow a better understanding of the more commonly used competences in the local professional context.

A Local Context-Sensitive Language Test (LCSLT) proposal is presented under an argument-based approach to test validation. Design patterns for language assessment have been adapted to allow mediation between the target language use domain and the test tasks.

## **Table of contents**

<b>ABSTRACT</b>	i
<b>Table of contents</b>	ii
<b>Table of tables</b>	vii
<b>Table of Figures</b>	ix
<b>List of abbreviations</b>	xi
<b>Declaration of authorship</b>	xii
<b>Acknowledgements</b>	xiii
<b>Dedication</b>	xiv
<b>1 Introduction and research context.....</b>	<b>1</b>
1.1 Introduction	1
1.2 The rationale for the study	1
1.3 Purpose and significance of the study	2
1.4 Research questions	3
1.5 The English language is needed in a globalising world.	5
1.5.1 English at a macro level	5
1.5.2 English at a micro level	6
1.6 Language use and language testing	7
1.6.1 Language use in context, a jagged skills profile	9
1.6.2 English as a foreign language	11
1.6.3 Testing English as a foreign language	11
1.6.4 English in Latin America	12
1.6.5 English in México	13
1.7 The local context	16
1.7.1 Higher Education in Mexico	16
1.7.2 Local context – Durango, México	16

1.7.3	Employability	17
1.7.4	English proficiency	17
1.7.5	English at the University	18
1.7.6	The students	18
1.7.7	The language classes	19
1.7.8	The current test	19
1.8	Structure of the thesis	21
1.9	Summary	23
<b>2</b>	<b>English for specific purposes and needs analysis.....</b>	<b>24</b>
2.1	Introduction	24
2.2	The purpose of learning English	24
2.3	English for specific purposes	25
2.4	Specificity within a specific purpose	26
2.4.1	EAP or EPP at university?	29
2.4.2	Approaches to Needs Analysis in ESP	30
2.5	Needs analysis in the workplace	34
2.5.1	Tasks and needs analysis.	35
2.5.2	Informants	38
2.6	Motivation	40
2.7	Summary	44
<b>3</b>	<b>Language testing for specific purposes .....</b>	<b>45</b>
3.1	Introduction	45
3.2	Language testing	45
3.3	Tests for specific contexts of use	46
3.4	Testing English for the academic context	46
3.4.1	Testing for the Professional Workplace	48
3.4.2	Tests for more than one purpose	52
3.4.2.1	Face validity	54

3.5	Test use	55
3.6	Validating the use of a language assessment tool	59
3.6.1	The Interpretation/use argument	60
3.6.2	The Assessment Use Argument	62
3.6.3	The Language Assessment for Professional Purposes model	67
3.7	Tasks in language testing	69
3.8	Proposal for a Local Context-Sensitive Language Test (LCSLT)	71
3.9	Test task design patterns	73
3.10	Summary	77
<b>4</b>	<b>Research Methodology .....</b>	<b>79</b>
4.1	Introduction	79
4.2	Research design	79
4.3	Data collection methods	84
4.4	Research participants, sample selection and size	86
4.4.1	Sample selection process	87
4.4.1.1	Test takers	88
4.4.1.2	University students	90
4.4.1.3	Workplace professionals	91
4.5	Design of the data-gathering instruments	92
4.5.1	Questionnaire 1	93
4.5.2	Semi-structured interview	98
4.5.3	Questionnaire 2 (Q2)	99
4.5.3.1	Questionnaire 2a (Q2a)	101
4.5.3.2	Questionnaire 2b (Q2b)	102
4.5.3.3	Questionnaire 3	102
4.5.4	Summarizing research design	103
4.6	Ethics	105
4.7	The role of the researcher	105



4.8	Data analysis	107
4.8.1	Preparing data for analysis	107
4.8.2	Analysis of quantitative data	107
4.8.2.1	Questionnaire 1	108
4.8.2.2	Questionnaire 2a	110
4.8.2.3	Questionnaire 2b	111
4.8.2.4	Comparison of results -- Q2a and Q2b	111
4.8.3	Analysis of qualitative data	112
4.8.3.1	The open-ended question at the end of Q1	112
4.8.3.2	Semi-structured interviews for test takers	115
4.8.3.3	Supplementary interviews	116
4.9	Summary	117
<b>5</b>	<b>Understanding the test-taker population .....</b>	<b>118</b>
5.1	Introduction Phase I results	118
5.2	Student voices	118
5.3	Test content	124
5.3.1	Section 1. What test takers think the test assessed	125
5.3.1.1	Workplace or academic-related language	125
5.3.1.2	Receptive skills, social/general purpose domain	127
5.3.1.3	Productive skills, social/general purpose domain	128
5.3.2	Section 2: What test takers think should be assessed	130
5.3.2.1	The importance of the work domain	141
5.3.2.2	The value of the academic domain	142
5.3.2.3	The interest in the social domain	144
5.3.2.4	The relevance of language skills	146
5.3.2.4.1	Understanding written English is necessary.	150
5.3.2.4.2	Understanding others is important.	152

5.3.2.4.3	Writing in English in the workplace	154
5.3.2.4.4	Oral interaction	156
5.3.3	Section 3. Preparing for the test	160
5.3.4	Section 4: General opinions about the current test	161
5.4	Summary	167
<b>6</b>	<b>Local context language needs.....</b>	<b>171</b>
6.1	Introduction Phase II	171
6.2	Work-related language competences	171
6.2.1	Speaking and listening-related language competences	173
6.2.2	Listening and speaking-related language competences	178
6.2.3	Listening-related language competences	180
6.2.4	Reading-related language competences	181
6.2.5	Writing-related language competences	184
6.3	Most relevant competences within the workplace	190
6.3.1	Speaking and listening	190
6.3.2	Listening and speaking, and listening-only competences	192
6.3.3	Reading competences	193
6.3.4	Writing competences	195
6.4	Summary	196
<b>7</b>	<b>Considerations for a local context-sensitive test.....</b>	<b>197</b>
7.1	Introduction	197
7.2	The argument-based approach to validation	198
7.3	Test development process	201
7.3.1	Initial planning	202
7.3.2	Identifying the local language use needs	203
7.3.2.1	Domain analysis: identifying TLU activities	206
7.3.2.2	Speaking and listening competences	207

7.3.2.3 Listening & speaking and listening competences	208
7.3.2.4 Reading competences	209
7.3.2.5 Writing competences	210
7.3.3 Design statement	211
7.3.4 Blueprint	216
7.3.5 From the LCSLT Design patterns through Operational use	219
7.4 Justifying the use of an LCSLT	225
7.4.1 Introduction	225
7.4.2 The Local Context-Sensitive Certification Test (LCSCCT)	226
7.5 Main Research Objective (MRO)	229
<b>8 Conclusions</b> .....	<b>230</b>
8.1 Introduction	230
8.2 Summary of findings	231
8.3 Contribution to the research area	233
8.4 Implications	234
8.5 Constraints and limitations	235
8.6 Future directions	238
<b>Bibliography</b> .....	<b>239</b>
<b>Appendix I Tests accepted to fulfil the school-exit requirement in non-native English-speaking universities</b> .....	<b>280</b>
<b>Appendix II Test Takers Consent Form</b> .....	<b>281</b>
<b>Appendix III Questionnaire 1</b> .....	<b>282</b>
<b>Appendix IV Semi-structured interview guide</b> .....	<b>283</b>
<b>Appendix V CENNI</b> .....	<b>284</b>
<b>Appendix VI Questionnaire 2a</b> .....	<b>285</b>
<b>Appendix VII Questionnaire 2b</b> .....	<b>288</b>

<b>Appendix VIII Transcription conventions .....</b>	<b>290</b>
<b>Appendix IX Short semi-structured interview example in Spanish and English version .....</b>	<b>291</b>
<b>Appendix X Statistics Q1 .....</b>	<b>293</b>
<b>Appendix XI Q2 Statistics .....</b>	<b>303</b>
<b>Appendix XII Example of language use test task .....</b>	<b>307</b>

## Table of tables

Table 3.1. Design pattern attributes and definition, adapted from (Wei et al., 2008)....	76
Table 4.1 Overview of research design.....	83
Table 4.2 Number of participants grouped by disciplinary area.....	90
Table 4.3 Q1, Section 1. What test takers think the test assessed.....	94
Table 4.4 Q1 Section 2. What test takers think the test needs to include and the expected need to use the four skills within a discipline-related job.....	95
Table 4.5 Q1, Section 3, preparation for the test.....	96
Table 4.6 Q1, Section 4. General perceptions about the test.....	97
Table 4.7 Q1, Section 5 Factual information from test takers.....	98
Table 4.8 Q2b Likert scale options participants chose according to the frequency that competences are acquired in the workplace. ....	103
Table 4.9 Reliability statistics for Q1 .....	108
Table 4.10 Competences in Q2 according to the skill(s) required for their fulfilment. ....	110
Table 4.11 Examples of participant coding .....	115
Table 6.1 Most relevant speaking and listening competences per disciplinary group	191
Table 6.2 Most relevant listening and speaking and listening only competences per disciplinary group.....	193
Table 6.3 Most relevant reading competences per disciplinary group .....	194
Table 6.4 Most relevant writing competences per disciplinary group.....	196
Table 7.1 Most frequently used speaking and listening competences for disciplinary groups .....	208
Table 7.2 Most frequently used listening and speaking and listening competences for disciplinary groups.....	209

Table of tables

Table 7.3 Most frequently used reading competences for disciplinary groups .....	210
Table 7.4 Most frequently used writing competences for disciplinary groups .....	210
Table 7.5 Design Statement; Competences and task types .....	213
Table 7.6 LCSLT general blueprint .....	218
Table 7.7 LCSLT Blueprint for the health sciences .....	220
Table 7.8 <i>Common3</i> design pattern to assess reading; understanding instructions .	222
Table 7.9 <i>Common3</i> design pattern to asses spoken interaction; making questions of fact-finding nature .....	224

## Table of Figures

Figure 1.1 The three ‘circles’ of English (Crystal, 2003).....	8
Figure 2.1 General and specific domains of language use.....	25
Figure 2.2 Areas of ESP taken from Basturkmen 2010, pg. 6.....	27
Figure 2.3 Assessment Specificity Continuum .....	28
Figure 3.1 Specificity of assessment construct and assessment use (Taken from Knoch and Macqueen, 2016, p. 294).....	58
Figure 3.2 Assessment development and use, stages 1-5. Bachman and Palmer (2010).....	63
Figure 3.3 Inferential links from consequences to assessment performance (L. Bachman & Damböck, 2018) adapted from (L. Bachman & Palmer, 2010).....	65
Figure 3.4 LAPP test development cycle. Knoch and Macqueen (2020).....	68
Figure 3.5 Inferences in a validity argument. Knoch and Macqueen (2020).....	68
Figure 4.1 Summary of research design .....	104
Figure 4.2 Themes and codes for voluntary comments to the open-ended question at the end of Q1.....	114
Figure 5.1 Theme and code compulsory nature of the test taken.....	119
Figure 5.2 What test takers think the school exit test assessed .....	129
Figure 5.3 What test takers think the test should measure.....	132
Figure 5.4 Interview domain preference per disciplinary group .....	139
Figure 5.5 Interest in language skills per disciplinary group .....	
Figure 5.6 Number of participants per option selection for C34d (reading).....	150
Figure 5.7 Number of participants per option selection for C34c (listening).....	153
Figure 5.8 Number of participants per option selection for C34c (writing) .....	155

Table of tables

Figure 5.9 Number of participants per option selection for C34c (speaking)..... 157

Figure 5.10 Voluntary written comments coding at “More time”..... 165

Figure 7.1 Initial planning stage, adapted from Bachman and Palmer.....203

Figure 7.2 Development process of a Local Context-Sensitive Language Test.....205

Figure 7.3 Justification of a local context-sensitive school-leaving English language test  
.....227



**List of abbreviations**

ALTE	Association of Language Testers in Europe
BALEAP	British Association of Lecturers in English for Academic Purposes
CEFR	Common European Framework of Reference
CENEVAL	National Degree Certification
CENNI	Certificado Nacional de Nivel de Idioma
CLB	Canadian Language Benchmarks
EF	Education First
EFL	English as a Foreign Language
ELASH	English Language Assessment System for Hispanics
ELF	English as a Lingua Franca
FCE	First Certificate in English
IELTS	International English Language Testing System
NES	Native English Speakers
non-NES	Non-Native English Speakers
TOEFL iBT	Test of English as a Foreign Language internet-Based Test
TOEIC	Test of English for International Communication
TLU	Target Language Use
TLUD	Target Language Use Domain
OET	Occupational English Test
PTE	Pearson Test of English
SEP	Secretaría de Educación Pública

## Declaration of authorship

I, **Cecilia Araceli Medrano Vela**

declare that the thesis entitled

Identifying context-sensitive English language use needs for a university exit test: a case study in Durango, México,

and the work presented in it are my own and has been generated by me as the result of my own original research. Therefore, I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission;

Signed: \_\_\_\_\_ Date: \_\_\_\_\_

## **Acknowledgements**

First, I want to thank God because He made the opportunity to cross my path possible, and the conditions were given for me to participate in this educational adventure. Secondly, He was always there along this journey. I would not be writing these lines if He had not given me what I needed to reach this point.

I am very grateful to both of my supervisors. Prof. Chris Sinclair helped me through the first Milestones of this academic journey. Upon his change of activities, my former advisor, Prof. Vicky Wright, became my supervisor, who patiently, supportively and wisely guided me through the last stages of the study. Her thought-provoking questions during our many lively discussion sessions were instrumental in completing this paper.

I want to thank PRODEP (Programa de Desarrollo Profesional Docente) for supporting my studies through a scholarship from the Ministry of Education in México SEP. I also thank my employer, the Universidad Juárez del Estado de Durango (UJED), for giving me this opportunity for professional growth.

I want to thank all the participants in my study, those known to me and all those I may never see again but who were willing to dedicate a few minutes to answer a questionnaire or respond to an interview.

I have been incredibly fortunate to have an understanding and supporting family. Their unconditional love encouraged me to go on. My husband Martin would always encourage me. My daughter Alejandra and my son Diego Iván, who have grown during my years of study, have been my driving force to see the end of this venture. Finally, I want to thank my brother Sergio, who was always patient, understanding, and encouraging during these years.

## **Dedication**

To my beloved parents

To Alejandra and Diego Iván

To Sergio

# 1 Introduction and research context

## 1.1 Introduction

This research examines the extent to which local English language needs within the workplace can be considered when designing a high-stakes university graduation test.

The study takes place in a public university in northern Mexico. Through a mixed methods approach, this case study explores the language needs of the workplace to design an English language test. The University decided to establish a language graduation requirement due to the global relevance of English. Quantitative and qualitative data is collected from students (in their last year of studies), professionals in the workplace and test takers. This data helps identify the most relevant and commonly used English language use competences within the local context. The data obtained is used to support the design of a proposal for a Local Context-Sensitive Language Test (LCSLT).

This chapter will first look at the rationale, followed by the research questions and the purpose and significance of the study. Then, to guide the reader on the role of English in the local context, the chapter goes from general aspects, such as the role of English in a globalised world, to particular aspects, such as the local context. In the path towards the local context, language use and language testing will be discussed. The chapter ends with the summary, preceded by the structure of the thesis.

## 1.2 The rationale for the study

As a former test development coordinator in the University's language department, I received various opinions and attitudes towards the (locally designed) accreditation and certification tests. While delivering results, it was surprising to learn that not all test takers see English as an enrichment for their future work and private life. Students had been previously informed of the new university language requirement for the degree. However, they could not imagine to what extent learning the English language would be relevant for their later professional activity. While students did not universally reject general English classes, the school leaving test was met with some resistance, particularly when the test scores hampered their academic journey. Comments such as "I have no plans to go to the

United States”, “My future job does not require English”, or “We never speak to Americans” were common.

The high-stakes nature of the test had a negative impact on the academic life of students who could not graduate with their cohort because they failed the test. Comments from test-takers indicated they considered there was little or no relation between the language assessed and the activities and tasks they expect to encounter in their future professional life. Such comments were not entirely surprising given that the test influenced their plans for their future career. Their dissatisfaction sparked in the researcher the idea that it was not English *per se* that was seen as irrelevant to their career development but rather the type of English being assessed. It was assumed that if the test content was related to their interests, i.e., related to their field of study, English might have been considered useful for their future professional activities.

English tests are designed as a response to the needs of a specific context. They are designed with a specific audience and context of language use (Bachman and Palmer, 1996; Douglas, 2000). It is widely recognised that language varies according to the context of use (Bachman, 1990; Brunfaut, 2014; Douglas, 2000; Hutchinson and Waters, 1987; Jordan, 1997; O’Sullivan, 2011). By their very nature, standardised tests have not considered the needs of a local context. In addition, it is unlikely that one test will cover the language needs of several disciplines. The purpose of a school leaving test is to predict, to the extent possible, the ability of graduates to use the language in their future professional jobs. In that case, the University needs a test that corresponds to the language requirements of the local professional work context. The design of an internal test is essential, provided that the test results are intended to show test takers’ ability to use English in their professional environment.

### **1.3 Purpose and significance of the study**

This case study research examines the language needs within the local professional context of use to inform the design of an English language test. Students must demonstrate that they have a B1 level of the CEFR (Common European Framework of Reference for Languages: Learning, Teaching, Assessment; COE, 2001) to meet the requirements for a university degree. The current test, developed and administered locally, focuses on English for general purposes.

By considering the voices of senior students, test takers and professionals in the workplace, the study aims to help build a better understanding of the language needs of the local context within the professional work environment. Identifying the language use activities professionals need for job-related purposes helps define and develop a language test proposal.

The results and suggestions of this study may be relevant in contexts such as Latin America or other parts of the world. Contexts in which English is a valuable tool for specific purposes, although day-to-day communication takes place in the country's native language.

## 1.4 Research questions

A mixed-methods exploratory research (Nunan, 1992) addresses this study's main research objective (MRO), which is to propose the specifications of a high-stakes university exit test that is sensitive to the local context.

To this end, it is first necessary to examine what test takers think of the in-house developed test and whether they think changes are needed. For this purpose, the first research question (RQ1) is presented:

RQ1. From the point of view of test takers, what is the face validity of the current school exit test in terms of content, test preparation, timing, and difficulty?

A university exit test relevant to the local context requires the identification of the language use needs of the domain of interest. These needs are identified from two perspectives: participants in the workplace and future test takers. Two other research questions are presented. The second research question (RQ2) is:

RQ2. To what extent are work-related language needs expectations of the students aligned with current professional language needs within the workplace?

Answering RQ2 requires addressing sub-question 2a (RQ2a) and sub-question 2b (RQ2b). These sub-questions are:

RQ2a. What are the work-related language competences students think their future professional job will require?

RQ2b. What are professionals' current work-related language needs in the local working context?

The comparison of data provided by RQ2a and RQ2b identifies the relevant language competencies for both groups of participants. However, this is not specific enough to build a test design proposal that reflects the local language needs in the professional workplace. Therefore, a third research question (RQ3) is presented, which seeks to learn more about the specific kind of activities professionals carry out using English in the workplace. RQ3 is stated as follows:

RQ3: What are some of the typical domain-related tasks that professionals use?

Finally, identifying the specifications for the university exit test requires exploring a language test validation framework that allows the incorporation of local language needs. This inquiry is addressed through a fourth research question (RQ4), which is:

RQ4: To what extent does an argument-based approach to test validation offer a flexible framework to address specific testing needs?

The study has been organised into three phases. RQ1 is placed in Phase I, which examines the face validity of the current test, mainly concerning the domain addressed by the test. In Phase II, RQ2 and RQ3 are found. Phase II identifies the most commonly used language competences within the local workplace and looks at the examples of specific language use activities in the workplace. Analysis of data collected from test takers, students and workplace professionals provides insight into the role English plays in the local professional work environment. Finally, in Phase III, RQ4 is addressed. In Phase III, the results inform a framework proposal for a local context-sensitive language test.

This research highlights the relevance of correspondence between test content and the real-life target context of language use, not only in terms of test tasks but also in considering jagged skills and how language skills are used in real life. These aspects are particularly relevant when test results are used to meet university degree language requirements. As the number of institutions setting a language certificate requirement increases worldwide (Cronquist and Fiszbein, 2017; Mansfield and Poppi, 2012; UAA, 2012; Universidad de Granada, 2017), it is compelling that the language taught and tested is relevant to everyone involved. It has been found that tests are not always used for their designed purposes.



## 1.5 The English language is needed in a globalising world.

Whether the spread of English facilitated globalisation or globalisation the spread of English (Graddol, 2000), the presence of English in various scenarios is undeniable. The ever-increasing use of English for business, academic, entertainment, social and tourism purposes worldwide (*ibid.*) has become the current driver of change at both macro and micro levels (EF, 2013). At the macro level, English can influence a country's development, while at the micro level, the language skills of individuals can make a difference in their socioeconomic status (Cambridge English, 2016a; Education First, 2011).

As English continues to be the common language of today's globalised international environment (Graddol, 2000), its impact has become the focus of international business attention. The English Proficiency Index (EPI) ranks countries according to the language proficiency level of their population. Public and private stakeholders are increasingly investing large amounts of resources (both economic and human) in learning English, as the link between a country's development and the English proficiency of its population has been well documented (EF, 2013).

A few decades ago, English was the means of communication between native speakers and speakers of other languages. Currently, most English learners are native speakers of other languages. This group of language users outnumbers native speakers and speakers of English as a second language (ESL) (Jenkins & Leung, 2016). In today's globalised world, where the use of English among native speakers of other languages continues to increase, a snowball effect is perceived: "...the more people use it, the more useful it becomes" (Education First, 2020, p. 4). Perhaps its relevance to countries and individuals has never been greater.

### 1.5.1 English at a macro level

Recent technological advances have contributed to the role English plays in the international environment and vice versa (Mexicanos Primero, 2015). In the communication era, when the internet has reduced geographic distance between places (Niño-Puello, 2013), and when globalisation has favoured an international environment within transnational corporations, English is the most widely used language for day-to-day interaction (Graddol, 2006). A study conducted by Cambridge English in 2016 shows that

for international companies, English is just as important for native-speaking countries as it is for countries with another native language (Cambridge English, 2016b).

Many developing countries, such as international outsourcing companies, require foreign investment to trigger their emerging markets. For some transnational companies, the local population's English level is essential for establishing branches abroad (Education First, 2011). A high level of language proficiency within a developing country's population can provide the boost its economy needs (Education First, 2016). According to the 2011 edition of Education First (EF), there is a strong correlation between English proficiency and a country's gross national income (2011, p. 7). It, therefore, appears that the more a country and its citizens invest in English language learning, the better the socioeconomic conditions of the country and its citizens. English is increasingly becoming a gateway to poverty reduction in many countries (Crystal, 2003).

Among the 14 Latin American countries included in the EF-reported 2016 study, México is one of the countries with a low proficiency index (Education First, 2016). According to Education First, "(P)oor English remains one of the key competitive weaknesses of Latin America" (EF, 2013).

For developing countries, English means more than foreign economic investment. It is also the language used for research, international collaboration, technology and information sharing (Education First, 2015b). A country unable to contribute to international research or access the latest innovations due to poor English proficiency may not quickly join the global international exchange of information, goods and services (Education First, 2011; Education First, 2016b; Education First, 2020; Mexicanos Primero, 2015).

### 1.5.2 English at a micro level

It has been observed that the impact English can have on individual economies is a worldwide phenomenon. Studies conducted in different parts of the world show its importance in the work environment, as Sarudin et al. (2013) reported. They found that "(M)any employers are hesitant to hire graduates due to their poor proficiency in English" (p.74) despite being highly qualified in their field. According to Velázquez (2015), mastering English offers individuals economic and professional opportunities. From this, it is suggested that knowledge of English can be a life-changing door-opener in different situations.

English has evolved from being less of a privilege to an essential skill for various activities (Education First, 2011). It has also been linked to better job opportunities (Crystal, 2003; Heredia & Rubio, 2015) and better salaries (Education First, 2013; Mexicanos Primero, 2015). For example, people who speak English at work may have better-paid jobs. Consequently, household income has been associated with English proficiency (BCEI, 2015); better proficiency translates into higher income (Education First, 2011).

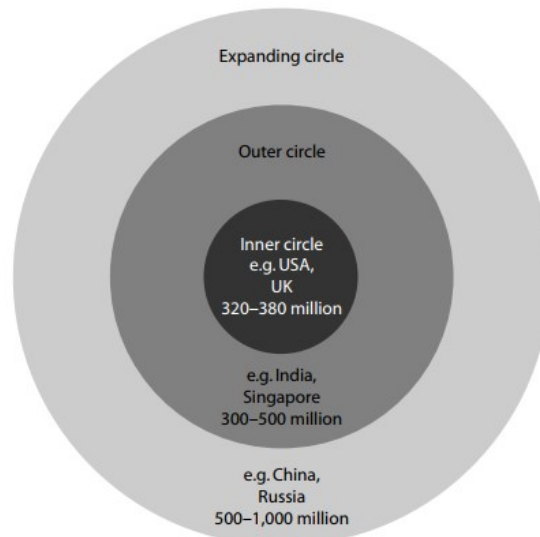
## 1.6 Language use and language testing

### **Diverse communication environments**

On the premise that languages are used for communication purposes, learners benefit when the focus of English learning, teaching and testing is on using the language for such purposes (Savignon, 2017). Individuals would be able to communicate in English if, in addition to language skills, they develop the skills that enable them to communicate successfully (COE, 2001). The skills to be developed relate to the communicative needs of the speaker. These needs could include the language required for everyday life in another country or interacting with speakers of other languages by exchanging ideas, thoughts and feelings (COE, 2001). That is, the required language depends on the context in which it is needed. For this study, context refers to the boundaries of the situation in which language is used. Context refers to “the social, physical, and temporal situation the language activity is taking place in” (Douglas, 2000, p. 42).

Although English can be considered a global language, it plays a different role in the countries where it is used or studied. In 1992, Kachru categorised the use of English worldwide into three concentric circles. In inner-circle countries, such as the United Kingdom, the United States of America and Australia, English is used as a native language (ENL). English is a Second Language (ESL) in Outer Circle countries such as Puerto Rico, Malaysia, the Philippines, Singapore and South Africa (Graddol, 2000). Countries within the outer circle have a first language, and English is a second or additional language. In a third circle, the expanding circle, English is a Foreign Language (EFL). This topic is discussed further below (see section 1.6.2 below). Russia and China are examples of countries within the expanding circle (Crystal, 2003). Figure 1.1 below shows that the number of speakers in the outer circle is more significant than those in the inner circle. México is part of the

expanding circle. This ample use of English means it is the *de facto* language for all communication purposes.



The three 'circles' of English

Figure 1.1 The three 'circles' of English (Crystal, 2003)

The way English is used varies significantly among a large number of speakers around the world. While English is the language of everyday use (ENL) for some speakers, it can be an alternative language for others. The decision to prefer English over another language can depend on the context and purpose of its use (Fang, 2017). For the most significant number of speakers, those within the expanding circle, English may not be needed daily, as in México. Individuals within the expanding circle use English primarily to communicate “with speakers from other countries” (Graddol, 2000, p. 11), and language proficiency can vary from native-like to relatively poor (Crystal, 2003).

The context in which English is used is closely related to the purpose of its use. Native English speakers use the language for purposes different from those of non-native English speakers (Mansfield and Poppi, 2012).

### 1.6.1 Language use in context, a jagged skills profile

The type of language, skills and proficiency level required in the international environment varies depending on the situation in which they are used. Using the language in the real world does not require equal use of each of the four skills, nor does it require the same level of proficiency. It has been observed that it is common for learners to exhibit different proficiency levels in different language descriptors (North *et al.*, 2016). A differentiated language competence profile is not only observed in foreign or second language learning, recognising that “all knowledge of a language is partial” (COE, 2001, p. 169). The level of proficiency of native speakers of a language also varies across language components.

In real life, skills are not only developed heterogeneously but are not equally used or demanded (Beacco and Byram, 2016). In some contexts, spoken interaction activities can be more linguistically challenging than written ones. In other contexts, the same level of proficiency may be required for all four skills (Stevens, 2005). The Council of Europe and the Association of Language Testers in Europe have pointed out that the frequency of language use and the proficiency required are determined by the language use requirements of the context in which it is used. They found that different contexts, such as the “...personal, public, occupational and educational” (COE & ALTE, 2016, p. 18) domains are likely to place different linguistic demands on speakers of other languages.

A study conducted between 2014 and 2015 revealed that although French employers considered the four skills required across different departments, foreign workers may need to use some skills more often than others (Langues et employabilité & Erasmus+, 2015). For some workers, “...communicating on the telephone (69%) and receiving and sending an email (61%)” (*ibid.*, p. 3) were frequent activities performed in a foreign language (not necessarily English). Employees from different departments would most frequently “...use the foreign language when operating software and web programs (32%), working overseas (28%) and working as part of a team (26%)” (*ibid.*, pg.3). The frequency and proficiency of each of the skills depend on the specific activities that the individual needs to carry out.

Cambridge English also reported a differentiated skills profile in the 2016 study conducted in 38 countries worldwide. Results showed that although employers valued mastery of the four skills, reading was considered the most important (in 11 industries), followed by speaking (in 9 industries). The former is considered “essential for maintaining professional knowledge, as it’s the language most often used in international journals, contracts and

instructions” (Cambridge English, 2016. pg. 16). It was pointed out that reading is an essential skill in countries where English is not an official language. Speaking is more relevant for the service industries such as tourism, where interaction with the customer is essential. Speaking is the most critical skill in countries where English is an official or de facto official language (Cambridge English, 2016). Different language level requirements were identified within the organisations participating in this study. This differentiated level was related to the type of industry rather than the company's size (*ibid.*).

Another study carried out in 2000 found that when recruiting foreign employees, European multinationals expect an advanced level of English and consider “oral skills (are) far more important than written ones” (Didiot-Cook et al., 2000, p. 3) if the employees have it as a second foreign language. An example is a survey conducted in Silicon Valley, California, in the United States of America. The results show that employers demand solid oral communication skills, as employees, whether native English speakers or not, in business and high-tech industries are expected to make oral presentations and participate in meetings and negotiations (Stevens, 2005). According to this study, universities and teachers are seen as influencing factors in graduates' employment and promotion opportunities. Teachers and institutions are expected to help students develop the language skills that employers value (*ibid.*). Higher education institutions need to know and understand the language skills required in the workplace so that students' language development work is relevant to their future careers.

Similar results have also been reported for medical students, as not all skills are equally required (Javid, 2014) or considered necessary in the Iranian context (Aliakbari & Boghayeri, 2014). For engineers in Malaysia, the ability to understand subject-specific written texts is considered very important (Shamsudin et al., 2013).

The jagged skills profile that characterises language needs in the workplace has encouraged the design of courses that meet the needs of a specific target workplace (Chairat, 2016; Kassim and Ali, 2010; Mohamed *et al.*, 2014; Moslehifar and Ibrahim, 2012). To address this important issue, a test that allows stakeholders to choose the skills in which their employees should be assessed according to the company's needs was developed (British Council & Aptis, 2016). Testing the skills relevant to their work environment allows EFL employees to focus on improving the genuinely relevant aspects.

Acknowledging that jagged language profiles reflect real-world language use across all abilities can significantly benefit learners. It would allow progress in the various skills to be recognised. This would also allow students to be acknowledged at a higher level for some skills than others “rather than delaying accreditation until learners (were) are equally proficient in all four skills” (Ashton, 2008, p. 177). Learners could focus their efforts and maximise the study time available to achieve the “language skills that employers often say they need (for example, a higher level in speaking and reading than in writing)” (Tinsley, 2012, p. 1). Correctly identifying the differentiated skills profile that graduates need upon graduation would help them concentrate on what is relevant to the job. Language tests for university graduation or hiring purposes could focus on the skills relevant to the workplace, so results are useful to predict learners’ ability to fulfil job-related tasks.

### 1.6.2 English as a foreign language

The most significant number of English learners since the 19<sup>th</sup> century (Graddol, 2006) is in places where English as a Foreign Language (EFL) is studied (Lowenberg, 2002). Although English is not the language used for day-to-day communication, and despite its restricted use and non-official status (Kachru, 1992; Lowenberg, 2002), it “can be made a priority in a country’s foreign language teaching” (Crystal, 2003, p. 4). English is the prioritised language in several Latin American countries, where efforts have been made to increase the English proficiency index of the population (Education First, 2016).

### 1.6.3 Testing English as a foreign language

The terms “language testing” and “language assessment” are synonyms in language measurement literature. Richards and Schmidt (2010) note that “(T)he term ‘testing’ is often associated with large-scale standardised tests” (p. 36). Similarly, Fulcher and Owen (2016, p. 110) state that external testing is mainly done for certification purposes. On the other hand, assessment is viewed as the process by which data is collected to make decisions about learning progress or language proficiency (L. Bachman & Palmer, 2010; H. D. Brown, 2004; Fulcher & Owen, 2016; J. C. Richards & Schmidt, 2010). Language assessment data can be collected in several ways, one of which is testing. Chappelle and Voss argue that while the terms test and assessment can have different meanings depending on the context in which they are used, they “both refer to systematic procedures for gathering data from test takers, from which interpretations are made to assign scores that are used for making decisions”

(2021, p. 2). Validation is required in both situations, so the terms test and assessment have the same meaning for this paper.

Standardised English language tests are the way through which speakers of EFL demonstrate language proficiency. International examination institutions such as Cambridge English (UCLES, n.d.) or the Educational Testing Service (ETS, n.d.) offer these. The Cambridge English: First (FCE), offered by the former institution, can be taken in any state within México (UCLES, n.d.) as it is the Test of English as a Foreign Language Internet-Based Test (TOEFL iBT), which is offered by the latter. The TOEFL iBT can be taken at any of the numerous locations in México; frequent test sessions are offered throughout the year (ETS, n.d.). These testing institutions have the most extensive presence in México, although there are other testing institutions, such as Pearson (Pearson Inc., 2020) and College Board (College Board, 2015a). The number of test centres and sessions offered in different locations indicates the significant existing demand for language certificates nationwide.

The introduction of the “*Certificado Nacional de Nivel de Idioma*” – CENNI (National Certificate of Language Level) (SEP, n.d.) shows the importance given to foreign languages by the Mexican education authorities. The CENNI (see Appendix V) results from the Federal Ministry of Public Education project. It provides a national reference framework for language assessment and certification. Two international language descriptors, the CEFR (COE, 2001) and the Canadian Language Benchmarks (CCLB, 2012), were the primary reference sources from which the 20-level language descriptor scale was developed. CENNI presents a list of international language tests and the test results accepted for certification at each level. Although CENNI is not exclusively intended for English certification, the number of options for certifying CENNI levels of English is greater than for other languages. CENNI aims to improve the quality of language teaching (SEP, n.d.), especially English.

#### 1.6.4 English in Latin America

The need for college graduates to be able to interact with today’s globalised world (Cambridge English, 2016a) has encouraged academics around the world to modify English learning programs to accommodate them (Cronquist & Fiszbein, 2017; EF, 2013; Education First, 2016b).

In Latin America, the universities “still function almost entirely in the national languages” (Hamel et al., 2016, p. 4), mainly Spanish and Portuguese in Brazil. Despite this, higher



education institutions are increasingly incorporating English (De Wit et al., 2005; Dussel, 2005; UCol, 2016; UNAM, 2018), either in their programs (Hamel et al., 2016) or as a degree requirement (BUAP, 2019; Reyes Fierro et al., 2008; UAA, 2012; UdG, 2011).

Latin American countries encourage the learning of English as they acknowledge its value as a tool to participate in the global economy (Hamel *et al.*, 2016). Various initiatives have been launched to support English teaching (Education First, 2016). Colombia (Peña Dix & de Mejía, 2012), Argentina, Chile, Uruguay (Dussel, 2005), Brazil, Bolivia, Perú, Cuba, Venezuela, México (Porto, 2014), Ecuador, Honduras, Costa Rica and Paraguay (Sánchez & Diez, 2014) have recently examined their national education policies related to English at different levels of education. Efforts, significant investments and socioeconomic links with the USA and Canada have not been enough as the adult population's language skills remain low, except for Argentina, which has remained above low proficiency (Education First, 2016) since the 2011 edition of Education First. According to the 2016 edition of Education First, for countries to see positive results from their investments and efforts, teaching practice in the Latin American classroom may need to change its current focus on accuracy to favour learning English for communication purposes. This comment suggests that the language taught may not reflect the way language is used in the real world.

High levels of economic inequality and the urgent need for reform in education systems (Education First, 2016) are seen as part of the challenges confronted by Latin America. This problematic situation was reported to have worsened in 2018, as it is “the only region in the world to have experienced a decline in average adult English skills since 2017” (Education First, 2015, p. 32). México was one of the countries with the most significant declines in the region.

#### 1.6.5 English in México

In 2015, the British Council (BC) and Education Intelligence (EI) reported the results of a study conducted in México. They found that although 80% of the companies participating in the study said that the primary language for internal communication is Spanish, the only other language with a percentage greater than 2% (Portuguese and French) is English. Of 110 employers in the study, 33% use English as the primary language for internal communication (BCEI, 2015). Although most Mexicans may not need English for internal

communication in the workplace, some may need to use it to communicate and share information with the international global community (Education First, 2016).

Like in other EFL contexts, the level of English proficiency within México varies greatly depending on the context in which the language is required. A university student does not have to give oral presentations or write essays in English for subject classes like every student in an English-speaking environment must do. However, he may be required to read books or articles in English (Dudley-Evans and St John, 1998). The need to use English depends not only on the University but also on the subject and its teacher. A survey conducted at the Meritorious University of Puebla (BUAP) in Puebla, a state south of Mexico City, found that most teachers consider English to be the most important foreign language in their subject. Although most teachers claim to be proficient in a foreign language, most do not dare to teach their subject in another language (English or any other they may know) (Facultad de Lenguas, n.d.). Therefore, English is not expected to be required for educational purposes beyond reading subject-related texts during undergraduate studies. Among the documents on language policies in Mexican universities available on the internet, the BUAP is the only one conducting a study before language policy implementation.

Although Mexico borders the United States of America to the north, for several decades since its introduction in 1926, English as a subject in Mexican public education has only been offered at the secondary level (Mexicanos Primero, 2015). English has been progressively integrated into public primary education since 1992 through a state-funded program (Ramírez Romero, 2015). The signing of the North American Free Trade Agreement (NAFTA) in 1994 (USTR, n.d.) was, perhaps, the spark that started a gradual awareness of the importance of English, particularly in the business sector (EF, 2013).

Gradually, state-funded English programs in other states followed until the federal government announced the implementation of the PNIEB (National English Program in Basic Education), now PRONI (Programa Nacional de Inglés) (SEP, 2017b). These changes reflect the federal government's awareness that "contemporary society... demands citizens with the necessary competences to incorporate into a constantly changing globalised world" (SEP, 2011, p. 12).

Mexican documents and initiatives on language policy in education do not explicitly state which language should be taught as an additional language. However, at different levels of

education, individuals need and desire to communicate with their international peers and integrate into a competitive global job market ( SEP, 2017c). English Teacher development programs have been funded (Navarro Leal et al., 2009) according to the policies of federal authorities, which are renewed every six years (Mexicanos Primero, 2015). Efforts to improve teachers' language skills have not been enough, as an overwhelming number of primary school English teachers have not yet reached the proficiency level required to teach at this level (Moore, 2013), as have those teaching at the secondary level (Mexicanos Primero, 2015). Survey results suggest that English speakers are a minority in México (AltoNivel, 2015), and the lowest proportion of English speakers are found in the country's southeast (Consulta Mitovsky, 2013). A low level of attainment in English at all levels of education results from the absence of a clearly articulated national language program across levels of education (IMCO, 2015).

Most students come to University with a low level of English, where English is rarely used outside of their English classes (if taken). Four to five years at University is not enough to learn English if it is not a priority in the study programmes. It is also not enough if the English lessons do not focus on what is needed or useful during school years or for future professional work. Students may not be intrinsically motivated to learn English. Meeting the language requirements does not always provide the necessary extrinsic motivation to make an effort. The relevance or usefulness of the language being taught can be the detonator needed to change the mindset of students who need to put in extra effort to reach the expected level of English.

The work environment is considered the area where English has the most significant impact, as a country's business success and economic development are directly related to its workforce's English proficiency (IMCO, 2015). Despite this, in most higher education institutions in México, the language taught or assessed is not related to the language of the working environment.

Although a large number of foreign investments come from the United States of America (USA), other countries such as Japan, Germany (Export Enterprises SA, 2019), Belgium and Holland (Reyes Solís, 2018) are also countries with strong economic presence in México. According to this, Mexicans need English in business and production to communicate with native English speakers and speakers of other languages.

## 1.7 The local context

### 1.7.1 Higher Education in Mexico

Mexico has a free education policy (Congreso Constituyente, 1993), so tuition fees are only paid in private institutions. Most universities charge an enrolment fee every semester. In many Mexican states like Durango, this fee is much less than the cost of an international language test that includes speaking.

### 1.7.2 Local context – Durango, México

The study takes place at the public University of the state of Durango. Durango is both the state and the capital in the northern part of México.

The city of Durango is approximately 1,000 km. away south of the North American border city of El Paso, Texas. *Paisanos* is the word used to refer to Mexicans who immigrated to the United States of America a few years ago. Some of them have learned English quite well and have become bilingual, allowing them to take jobs that require English. Paisanos take every opportunity to visit their loved ones in their hometown. When our *paisanos* visit Durango, sometimes they bring along relatives who speak little or no Spanish, making communication difficult even among some family members.

Famous for its deadly venomous scorpions and its film-making attractiveness, Durango is dubbed the *Land of Scorpions* (The Mazatlán Post, 2018) or *Land of Movies* (Notimex, 2017) and attracts both domestic and international tourists year-round.

Even though Durango is not a large, industrialised, multilingual, and multicultural city, it is not isolated from the rest of the world. Its professional workforce must be able to participate in the knowledge debate of the international community for the benefit of its own local communities. Joining the global discussion requires language knowledge and skills related to their professional practice. Universities may be able to anticipate the degree to which graduates will be prepared by taking their local language needs into account when designing a school exit language test.

### 1.7.3 Employability

The recently opened Durango-Mazatlán highway has reduced risk and travel time from around 6 - 7 hours to an average of 3.5 hours through complicated mountainous terrain. The expected impact has benefited Durango's economic and tourism sectors (Presidencia de la República, 2012). The new route to Mazatlán, which takes travellers across the Baluarte Bridge, the tallest cable-stayed bridge in the Americas, is a tourist attraction in its own right (Mazatlántoday.net, 2019).

Despite being one of the largest states in México (Saber es práctico, 2014), Durango is among the moderate to low competitiveness states (IMCO, 2018), with agriculture, metalworking, timber processing and business support services as main economic activities (*The 5 Main Economic Activities of Durango*, n.d.). It is rich in minerals and ranks second in gold and silver production (*ibid.*), although the largest mining companies are Canadian-owned. Few outsourcing companies have settled in the city of Durango. These include YAZAKI Corporation (Japanese), DELPHI (Irish), Prysmian (Italian) and DAWS (Korean-North American), all of which are involved in electronics, wiring and harness building, except for Carhartt (North American), which has outdoor activities textile industry.

People of Durango, university students included (see SSc59, HSc24 and HSc 89 below), recognise the increasing presence of tourists and individuals due to Durango's outsourcing businesses. The increasing presence of foreigners for tourism or business purposes requires that locals communicate in English with speakers of other languages.

### 1.7.4 English proficiency

Despite its location, outsourcing companies and visitors, the number of learners of English and proficient speakers are limited. An online study by the British Council in 2015 found that only 40% of participants from Durango said they had learned English. This position is the third lowest among the 32 Mexican states, tied with the neighbouring state of Zacatecas. Although the number of participants from each state is inconsistent, this result shows that overall, people in Durango have low engagement in learning English. The 2015 Mexico's English Proficiency Index (EPI) ranks Durango 11<sup>th</sup> out of 32 Mexican states (Education First, 2015).

The role of English in the development of individuals and their communities has influenced the decisions of local authorities, which have emphasised the need for people to learn English (UJED, 2015). State authorities are supporting the growth of international tourism and establishing outsourcing companies in the city's outskirts. These changes have gradually affected people's perceptions of the need to use English.

### 1.7.5 English at the University

The study takes place at a comprehensive public university covering social sciences, healthcare sciences, natural and hard sciences.

Following the need Mexican education authorities have to prepare individuals "capable of dealing with the communicative challenges of the globalised world" (SEP, 2011, p. 12), in early 2008, the University Board approved introducing an English language learning and testing programme for all majors. This program is henceforth called the University English Program (UEP). The UEP established a B1 level (according to the CEFR) certification as a graduation requirement (Reyes Fierro, 2008).

Before UEP, only the medicals bachelor's program included two semesters of English in its curriculum. These English courses are called *Technical English* and should address Medical English. However, students state that these courses are held infrequently and are mostly limited to lists of medical-related words they need to translate.

### 1.7.6 The students

Even though there are students from middle and high-economic backgrounds, most come from middle- and low-income families, often with two family members contributing to the family income. An additional burden for parents of students from nearby towns is the cost of living away from home. Several students are self-sufficient or work part-time to help with expenses.

Many undergraduates have not been able to meet the language requirement. As a result, the percentage of university graduates has fallen since the introduction of this regulation. The high-stakes nature of the test negatively impacts the academic life of students who cannot graduate with their cohort. This requirement has become a degree award gatekeeper for many students. Comments from former test takers (see 1.2 above) indicate

that English is not perceived as useful. Other comments show participants consider that the language assessed is not related to activities and tasks that are relevant or useful in their future professional lives (see 5.2 below). Such statements were unsurprising since the test affected their future career plans.

#### 1.7.7 The language classes

The UEP offers language courses, but the students can also prepare themselves in other ways. They can enrol at a language centre within or outside the University or prepare for the test independently. Enrolling in UEP courses is much cheaper than enrolling in a language centre, whether inside or outside the University. This, and the fact that UEP courses are offered at their school and often within their school schedule, can impact the number of students who enrol in UEP classes.

The general English textbooks used for these courses are said to be based on the CEFR (COE, 2001) levels and language descriptors. The four skills are expected to be developed in language classes.

The language courses focus on English for general purposes from A1 to B1 levels of the CEFR (COE, 2001).

#### 1.7.8 The current test

The socioeconomic profile of most university students was a determining factor for not considering the possibility of adopting an international test to assess the language degree requirement. A locally developed test was needed to determine whether undergraduates had the language proficiency established by the University (1.7.5 above) to graduate.

Time and resource constraints did not allow a study to be conducted, nor external guidance and support to develop the test. Therefore, the test, rather than being the result of a needs analysis (Bachman and Palmer, 1996), was developed based on the teachers' best judgment. When determining the purpose of the examination, neither the current language learning needs of the students nor their language needs when entering professional life were considered. The construct to be assessed was taken from the Self-assessment checklists from the Swiss version of the European Language Portfolio (Schneider & North, 2000).

The director of the School of Languages appointed some teachers to work on developing the in-house general English B1 level (of the CEFR) test. It measures vocabulary, grammar and the four skills: listening, speaking, reading, and writing. The written and oral components are separate but applied within the same exam session. The speaking component is carried out in pairs and assessed by two evaluators (one is also the interviewer, and the other just listens). The interviewer first asks questions to test-takers, and later, they are asked to interact with each other in a role-play task type followed by a reaching an agreement task type. Finally, they are individually asked to comment on a specific topic.

Students from different disciplinary areas can be found in a testing session, as they previously registered for the day and time of their choice. To be admitted into the testing facilities, the identity of the participants is verified. The test takers need to show their national identity card to the invigilator when their names are called. The exam sessions take place in a large room that may fit around 50 students.

Several teachers are the invigilators during the test session, watching tests are answered individually, and only the pencil and eraser provided to students are used during the test. Students are instructed to put their pencils down when they have finished answering the test and are ready to complete the listening component. When at least eight students are ready to take the listening component, the exam is interrupted for all students to take the listening component. Instructions for the listening component are read aloud in English, and questions are answered before the audio file is played twice. When the audio stops, all participants can continue answering the test. Early finishers can review their answers before submitting their papers to the invigilators.

When the students have completed the written component, they are taken to a waiting area and grouped with another participant to take the oral component together. They are called by an invigilator when a pair of evaluators is ready to receive them. Test takers are dismissed from the facilities after the speaking component.

Most students take the locally designed B1 test, although some international tests (at the B1 level) are also accepted to meet the language requirement (see Appendix I). Few students choose the latter option, mainly because of the cost. Most of those who do are likely to be in the process of applying for a higher education scholarship.



The current university test aims to determine if undergraduates have acquired the general English B1 level of the CEFR (COE, 2001). Students could have prepared for the test by taking the classes offered at the University, which are also based on the general English B1 level of the CEFR. Considering that the purpose of higher education institutions is to prepare individuals to join the workforce, it is argued that the University is not fully meeting this goal as the language courses and the test does not take into account “the communicative challenges of the globalised world” (SEP, 2011, p. 12) in the working environment. A mismatch is perceived between the purpose of higher education institutions (to prepare discipline-specific professionals able to interact with their international peers) and the content of the university exit test, as the characteristics of communication in the workplace are not considered.

## 1.8 Structure of the thesis

The present thesis is organised as follows:

Chapter One introduces the thesis and discusses the rationale and the research questions, preceded by the purpose and significance of the study. Then, going from the general to the particular, the chapter takes the reader from the macro level of the role, need and use of English to the local context and its characteristics. The chapter ends with a summary preceded by the current section, which guides the reader through the organisation of the thesis.

Chapter Two deals with English for Specific Purposes (ESP). The purpose for which English is learned, and the branches into which ESP is divided are discussed. Next, the needs analysis is considered, focusing on the workplace, which is the most relevant domain to this research. Tasks and informants are the other two discussed aspects related to needs analysis. Finally, the role of motivation within ESP is discussed.

Chapter Three addresses language testing for specific purposes. Tests designed for academic and professional purposes are discussed, followed by the use of tests for purposes other than those that motivated their design. One of the most widely discussed principles in language testing is validity. This testing principle is discussed, arguing that the use of a test may question the validity of the results if this is different from the use for which the test was designed. Argument-based approaches are then discussed, followed by the use of tasks in language assessment for specific purposes. These led to the articulation of

a proposal for a local-context-sensitive language test (LCSLT). Finally, and in line with the argument-based approach of the LCSLT and before the summary, a proposal for design patterns to assess language for specific purposes is presented.

Chapter Four describes the methodological aspects of the study, describing in detail the research design. The description of the participants follows, and then the approach to sample selection and sample size are addressed. Next, the development of the data collection instruments is detailed. Ethics and the role of the researcher are also addressed before discussing the way data was analysed.

Chapter Five looks at data results from applying Questionnaire 1 and carrying out a semi-structured interview. Graphs and excerpts from the opinions of test takers are embedded in the discussion of the results. The answer to Research Question (RQ1) 1 is addressed in this chapter.

Chapter Six discusses the data analysis results of applying Questionnaires 2 (Q2) and 3 (Q3) and carrying out a supplementary interview while answering Q3. Questionnaire 2 is applied to students (Q2a) and professionals (Q2b). Data results from Q2 (Q2a and Q2b) and Q3 are compared to identify the most highly ranked work domain language competences for both groups of participants. A better understanding of results is achieved through the voices of professionals and some test takers by providing examples of the language use activities they carry out in the workplace. Answers to Research Questions 2 (RQ2) and 3 (RQ3) are presented.

Chapter Seven looks at the aspects to consider for a context-sensitive test. This is followed by the design statement for a local context-sensitive English language certification test, under the assessment use argument approach to test validation. In line with the argumentative approach to test validation, an example of how design patterns for language assessment are adapted to be used for domain-specific test task design related to the language descriptors of the CEFR is presented. The Main Research Objective (MRO) is addressed.

Finally, Chapter Eight presents the conclusions of the study. A summary of the findings is presented. These are followed by the contribution to the research area and the constraints and limitations of the study. Finally, areas in which further work is required are discussed.

Appendices are found after the references.

## 1.9 Summary

The purpose of the present chapter was to help the reader understand the context in which the study took place. This was done starting from a macro perspective, looking at the role of English worldwide and its impact at different levels. The differences in how language is used in diverse contexts were addressed by looking at it from a global perspective. Within the Latin American context, going through México, Durango, and finally, the situation at the local University was addressed. The characteristics of Durango help understand the context where participants study, work or are likely to work.

The perceived relevance of a language influences attitudes towards learning and the test used as a graduation requirement. A language user-oriented approach establishes an essential relation between the language needs of test takers and test contents. This approach entails considering the type of language and tasks required in the real world during the test design process. The research questions seek to understand better the test-taker population and their language needs within the local working environment. Data results analysis is used to inform the design of a context-sensitive high-stakes language test.

The structure of the thesis was also presented in this chapter.

The following chapter discusses the use of English for specific purposes (ESP) and the process through which data is collected to inform (ESP) courses or tests. Motivation is also discussed as an essential factor in language learning.

## **2 English for specific purposes and needs analysis**

### **2.1 Introduction**

This chapter first addresses the purpose that individuals have in learning English. While some individuals know why they are learning English, others decide to learn it because they find it useful despite not having yet identified a specific purpose. The distinction between learning English for specific purposes and general purposes is first discussed below. We will then discuss English for specific purposes (ESP) and look at subdivisions within ESP and the levels of specificity. The following section discusses needs analysis, the process used in ESP to identify the characteristics of the ESP situation. It then examines needs analysis (NA) in the workplace, followed by tasks presented as a good alternative to language assessment. Finally, as part of the NA topic, we consider the informants, who they are and their role in the NA process.

The final topic discussed is motivation and its relation to learning English.

### **2.2 The purpose of learning English**

The ever-increasing use of English for various purposes and interests (Education First, 2011) created a new learner population who could identify the purpose of learning English (Hutchinson and Waters, 1987). According to the CEFR Companion Volume, learners' needs to communicate in the real world fall into one of "four domains of language use: public, personal, occupational and educational" (COE, 2018, p. 52). Most adult language learners can identify a purpose for learning English (Hutchinson and Waters, 1987).

Although it is impossible to predict or determine all contexts in which language may be used, identifying the most common domains of language use has been found useful, particularly for teaching and learning purposes. Just as the CEFR differentiated the domains of language use mentioned above, the ALTE (Association of Language Testers in Europe) presented a series of 'CanDo' statements organised within the "Social and Tourist, Work and Study" (COE, 2001, p. 245) domains. Even though language competences, or language descriptors, have been organised within domains of language use, these competences are not exclusive to the domain in which they are categorised. Competences

may overlap between domains, e.g. individuals engage in social interaction, albeit at different levels, in domains other than the personal (COE, 2001).

English, which refers to a language use domain, is called English for Specific Purposes (ESP). When English learned is not related to a domain, it is referred to as English for General Purposes (EGP).

### 2.3 English for specific purposes

As opposed to General English, which is regarded as detached from the learner's language needs (Hutchinson and Waters, 1987), English for Specific Purposes (ESP) is concerned with the reason for the language being learned. The language needs will determine the type of language required within the specific context in which it will be used (Dudley-Evans & St John, 1998). As ESP focuses on the needs of the learners, "tasks are prescribed by their study or work situation" (Dudley-Evans & St John, 1998, p. 1), yielding narrower contents than those which may characterise English for General Purposes (EGP).

ESP aims to help learners develop the abilities needed to understand or produce the language required in the target context of language use. While aims will be more explicitly defined and the content will be narrower in ESP, the division between EGP and ESP is unclear (Barnard and Zemach, 2003). A language continuum has been used to



Figure 2.1 General and specific domains of language use

represent EGP and ESP's position concerning identified needs (Brunfaut, 2014). The language considered general interest unrelated to a particular purpose would be found at the EGP extreme. At the opposite end, the language related or tied to a particular purpose of language use (ESP) would be placed, as Figure 2.1 above shows.

Most language course books would be placed at the EGP end, as they are designed to be used with a diverse target population worldwide. English for Specific Purpose courses are designed to fit the needs of a particular group of learners, i.e., at the university level (Vidal, 2005), such as medical English (Javid, 2014), English for mathematicians (Ferrari Fermín & Torrealba M, 2009) or for architects (Di Bella & Batista, 2007).

The specific purpose is not only addressed by language courses. Several tests have been designed to meet specific needs, mainly for the academic and working environments.

## 2.4 Specificity within a specific purpose

During the past few decades, scholars have seen the work of ESP practitioners respond to “a definable need to communicate in English” (Hutchinson & Waters, 1987b, p. 57) in academic and working environments (Dudley-Evans and St John, 1998). More work and research have been done in English for Academic Purposes (EAP) than in English for Occupational Purposes (EOP) or English for Professional Purposes (EPP); this is because, presumably, EAP takes place in academic institutions where work in this area is promoted and required (Flowerdew and Peacock, 2001).

Within a domain of language use, e.g., English for medicine, it is necessary to identify whether it will be used for academic or working purposes. Certain commonalities may be identified between the academic and the working environments, though other situations are likely specific to the academic or the working contexts. Consequently, English for academic purposes (EAP), English for professional purposes (EPP) and English for occupational purposes (EOP) have been sub-divided to attend to the needs of a general target group of language users, such as English for General Academic Purposes (EGAP), or to a more specific learner population, such as English for Specific Academic Purposes (ESAP). While EGAP incorporates “the skills and language that are common to all disciplines” (Dudley-Evans & St John, 1998, p. 41), ESAP focuses on the features that are specific to a discipline and distinguish it from others (*ibid.*). This division is exemplified in Table 2.1 below.

Just as Figure 2.1 above represents a continuum with EGP and ESP at each end, the areas within ESP shown in Table 2.1 below can be represented in a continuum, as shown in Figure 2.2 below. In this ESP continuum, EGAP, EGPP or EGOP are found at one end of it, while ESAP, ESPP or ESOP are placed at the opposite extreme, as presented by Knoch

and Macqueen (2016). Tests at the right end of the Assessment specificity continuum “are designed to closely simulate the real-world tasks of specific professions” (ibid., p. 293). By doing so, they can provide a more accurate estimation of performance when the test taker interacts in the real world.

<b>Branch</b>	<b>Subbranches</b>	<b>Example</b>
English for Academic Purposes (EAP)	English for General Academic Purposes (EGAP)	English for academic writing
	English for Specific Academic Purposes (ESAP)	English for law studies
English for Professional Purposes (EPP)	English for General Professional Purposes (EGPP)	English for the healthcare sector
	English for Specific Professional Purposes (ESPP)	English for Nursing
English for Occupational Purposes (EOP)	English for General Occupational Purposes (EGOP)	English for the hospitality industry
	English for Specific Occupational Purposes (ESOP)	English for hotel receptionists

Table 2.1 Areas of ESP taken from Basturkmen 2010, pg. 6

Practicality may be an element for selecting EGAP over ESAP. While the former encompasses language knowledge and skills common to all or several disciplines, its contents are relevant for a wider learner population. The shared content for a wider population makes it more cost-effective for institutions (de Chazal, 2012). In deciding whether to choose ESAP or EGAP, Dudley-Evans and St John (1998) consider some situations that will benefit from a combination of ESAP and EGAP.

Whether general or specific content within the discipline or area of interest, ESP holds the advantage that by considering those aspects that characterize the target situation of language use, learners can focus on what is relevant for them, maximizing language learning time (Tickoo, 1988).

Even though English for Social Purposes was distinguished as another area of interest for language learning (Jordan, 1997), these contents are frequently included within the ESP content because they may be considered part of the language required to function effectively in an English-speaking environment (Jordan, 1997). In 2002, the results of an extended European research project launched by the ALTE revealed that English for Social and Tourist Purposes (ESTP) was another language learning purpose (ALTE, 2002) for many individuals.



Figure 2.2 Assessment Specificity Continuum

Whichever purpose a learner has for learning English, it would be difficult to establish a clear-cut division of domains. O’Sullivan (2012) argues it is possible to “identify particular aspects of language use as being specific to a given context (such as vocabulary, syntax, rhetorical organization)” (pg. 73). However, he also contends that it is not possible to clearly state where the boundaries to the language that is used in that context are, as these do not exist. Domains often overlap or interact with each other. Not all language activities during student life are necessarily related to learning activities, as individuals also use the language for social interaction within the institution. Similarly, not all communication in the workplace is necessarily job-related; thus, individuals, as community members within an institution, may also engage in social interaction (COE, 2001). Besides the relevant ESP content, including some competences from English for Social and Tourist Purposes will be helpful for a worker or student travelling to an English-speaking country for training or academic purposes (Davies, 2001).



#### 2.4.1 EAP or EPP at university?

Universities within the expanding circle include English as part of their learning programs because they consider it essential for the future studies of their students (Barjesteh & Shakeri, 2013). They may also do it because they want their graduates to be able to express themselves in oral and written form as well as understand spoken and written texts on topics related to their field of expertise (Ferrari Fermín and Torrealba M, 2009; Tratnik, 2008). Some studies have found that students would benefit from having more subject-specific vocabulary in their English courses (Ferrari Fermín and Torrealba M, 2009), especially to enhance subject-specific reading comprehension (Shamsudin et al., 2013).

Students may need EAP to access discipline-related content; hence, universities may focus on helping students develop these skills (Di Bella and Batista, 2007; Ferrari Fermín and Torrealba M, 2009; Guerrero Cárdenas, 2008; Vidal, 2005). Helping university students “access the research literature” (J. Flowerdew & Peacock, 2001, p. 8) in English is one of the main reasons that higher education institutions in EFL environments such as Latin America include EAP within their undergraduate programs. However, it has been observed that EAP courses do not provide the skills that the working environment requires (Flowerdew, 2005; Sarudin et al., 2013). Consequently, some institutions address both needs: those of the academic and the work environments (Vidal, 2005).

The study conducted by García-Ponce (2020), who works at a state university located in central México, revealed a mismatch between the type of English taught at the university and the expectations and needs of employers and students. The result of the focus groups with nine employers of locally-based multinational companies revealed that the English proficiency level of university graduates was insufficient for them to perform the different activities the workplace required. On their behalf, university authorities, teachers and students agreed on the need to incorporate English related to the workplace.

Addressing the necessary language, whether EAP or EPP, EGAP or ESAP, brings many benefits to all stakeholders involved (Dudley-Evans & St John, 1998); therefore, it is necessary to define clearly their language needs. ESP does this through a needs analysis (Dudley-Evans, 2000). Conducting a needs analysis allows the ESP practitioner to identify the language required in the specific context of use.

### 2.4.2 Approaches to Needs Analysis in ESP

A needs analysis process identifies the specific language needs of a group of learners (J. C. Richards & Schmidt, 2010). Brown (2016) defined needs analysis as “the systematic collection and analysis of all information necessary for defining and validating a defensible curriculum” (p. 4). In this case, a defensible curriculum refers to the content that meets the requirements of the situation in which stakeholders are interested. For Brown, stakeholders are those “who have a stake or interest in the curriculum (for example, teachers, students, administrators, and parents)” (ibid, p.4). Even though needs analysis is a common practice in language curriculum development (Basturkmen, 2010; Brown, 2009; Serafini et al., 2015), it is also used in the identification of language needs for testing purposes (Bachman and Palmer, 1996; Davies, 2001; Weir, 2005). Departing from the above definition and considering its use for testing purposes, needs analysis could be defined as the systematic collection and analysis of all information necessary and available to define and validate the link between test performance and the target language use situation.

Woodrow (2017) and Flowerdew (2013) have argued that needs analysis is the starting point in course design. Flowerdew states that it is through NA that it is possible to identify the “what” and the “how” of course development, including assessment and evaluation (ibid.). Long considers all language courses should be the result of a comprehensive needs analysis and be regarded to be for specific purposes. These would be differentiated by the exactness with which learner needs can be identified, ranging from none or limited identification of needs to very detailed, as would be the situation of young learners and most adult learners, respectively (M. H. Long, 2005). According to Brown (2016), NA is an essential component of ESP; that is, it is from identifying the specific needs of a group of learners that ESP is based on.

Richard West (1994) describes how the term ‘analysis of needs’ first emerged in the 1920s in India when Michael West became interested in identifying what learners would need to use English for in the target situation, as well as how to make more efficient use of the time available for learning. According to Braine (2001), the concept of analysing the needs of the learner does not seem to have been retaken for some decades, as he claims that for several decades, “course design in English language teaching may have been based mainly on teachers’ intuition of students’ needs” (p. 195). During the early 70s, the Council

of Europe introduced a formal analysis of the type of language individuals may require in various real-life situations (Council of Europe, n.d.; Trim, 2007).

The identification of learner needs has been approached through several angles. The vast and narrow-angle approaches refer to the degree of specificity of the language to be addressed (Basturkmen, 2010). Proponents of a wide-angle approach, especially for English for Academic Purposes (EAP), consider a “common-core language and language-learning strategy” (D. D. Belcher, 2006, p. 138) would provide learners with the tools that allow them to transfer the language knowledge and skills to other domains and situations. A wide-angle approach states learners from different disciplines can take the same EAP course (Basturkmen, 2010; Belcher, 2009), as this approach does not look into vocabulary that pertains to a specific discipline or language domain. Instead, it may be considered that knowing 2,000 high-frequency words and the 570 words in the Academic World List gives learners around “90% coverage of the running words in most academic texts” (Coxhead & Nation, 2001, p. 260). This approach considers the content does not need to be related to a specific subject area, and any carrier content will “serve the purposes of most language learners” (D. D. Belcher, 2006, p. 138). It is expected that learners would be able to “transfer the knowledge and skills they gain from a wide-angled course” (Basturkmen, 2010b, p. 54) to their subject of interest, just as it is the focus of English for General Academic Purposes.

On the other hand, Belcher (2006) argues choosing a narrow-angle approach is based on considering that domain-related content and tasks would be much more interesting for the learner and the outcome of learning would be immediately available to be used. In favour of the narrow-angle approach, Hyland (2006) claims defining the ‘common core’ of the wide-angle approach may not be easily achieved, mainly when meaning and use of language are introduced. English for dentistry studies or veterinary studies could be examples of the narrow-angle approach, as their language and tasks would be more related to disciplines of dentistry or veterinary. Hyland argues that ESP pertains to “the literacy skills which are appropriate to the purposes and understandings of particular communities” (Hyland, 2002, p. 386). Basturkmen stated that “learning is more likely to occur” (2010, p. 11) when addressing the content related to the disciplinary area the students are interested in; that is, the topics related to their university studies or work. The narrow-angle approach considers learners will use the language in their specific area of development rather than in a broad domain (ibid. p. 54).

Hyland (2006) argues that different disciplines may not share discourse characteristics. In contrast, Brown (2016) highlights how areas among different disciplines may overlap. As a strong advocate of the narrow-angle approach, Hyland considers that universities offer more wide-angled than narrow-angled courses due to economic resources (Hyland, 2002). It could be thought that an EGAP course is more cost-effective than an ESAP one, as the former can be taken by students regardless of their disciplinary area, while the latter may have courses with very few students.

While identifying the language needs, the needs analyst must ponder whether a wide or narrow-angle approach better fits the specific situation of interest. According to Graves (2016), needs analysis for course design purposes consists of gathering information when a course or program starts, also known as the present situation (Dudley-Evans & St John, 1998), and “about possible or expected final outcomes” (Graves, 2016, p. 87). The latter is the language learning objectives. The course is designed to bridge the gap between the information obtained at the beginning, the present situation analysis and the language learners are expected to know at the end of the course. Chambers claimed that the analysis of the language required to communicate in the target situation, a target situation analysis, defines the focus of learning and teaching of an ESP course (1980).

An active discussion about the types of needs that need to be carried out or taken into account has been taking place for several years (Berwick, 1989; Braine, 2001; Brindley, 1989; Brown, 2016; Dudley-Evans and St John, 1998; Flowerdew, 2013; Huhta et al., 2013; Hutchinson and Waters, 1987). Brown contends that the term needs can be interpreted in various ways according to those involved and their context (2016). Needs can be thought of as “*wants, desires, necessities, lacks, gaps, expectations, motivations, deficiencies, requirements, requests, prerequisites, essentials, the next step* and  $x + 1$  (where  $x$  is what students already know, plus the next step, or 1)” (ibid., pg. 13).

Brown (2016) considers these different conceptualizations of needs to be comprised of four categories of points of view on needs. He calls them *needs viewpoints*. These are the democratic, discrepancy, analytic, and diagnostic views. The democratic view can be summarized as considering what most stakeholders want, desire, expect, request or are motivated about. Stakeholders include the students, teachers, administrators, or any other interested party. Data collected from all the relevant parties will likely provide a sounder representation of the actual language needs (West, 1994). The discrepancy view looks at the difference between what students can currently do with the language and what they are

expected to be able to do in the ESP situation. This point of view looks at deficiencies, lacks, gaps, and requirements of students in terms of language (J. D. Brown, 2016). The analytic view is grounded on the Second Language Acquisition theory, which focuses on the hierarchy of language learning. It is referred to as  $x + 1$ , where  $x$  is the current knowledge of a student's language, and  $1$  represents the following step in language learning. A significant drawback of this viewpoint is we do not have a comprehensive understanding of how languages are learned or the order in which language is acquired. Finally, the diagnostic view seeks to identify the elements of ESP that are the most important and will do the most harm if not addressed. This viewpoint looks at the students' necessities, essentials, or prerequisites: those encountered in the situation of language use or those essential for students to continue to the following stage. These four viewpoints are not mutually exclusive; thus, Brown suggests the needs analyst may ponder considering all four of them, pondering their usefulness at different stages of the needs analysis process.

These viewpoints aim at describing the meaning of the word *needs*, which Brown considers could be interpreted differently by people according to their context. Once this has been defined, it is pertinent to consider how needs will be analysed. Brown (2016) presents eleven options for language analysis, which are: target-situation use analysis, target-situation linguistic analysis, target-situation learning analysis, present-situation analysis, gap analysis, individual differences analysis, rights analysis, classroom-learning analysis, classroom teaching analysis, means analysis and language audits. Among these, the type of analysis relevant to this paper is the target-situation use analysis, which will be discussed below.

The target situation use analysis, as its name states, seeks to identify the language required to be used in the specific context of language use upon finishing a period of language instruction (Brown, 2016; Woodrow, 2017). The language used within the specific discourse community where the group of learners of interest will engage is analysed to identify its specific language events. Provided that the learner is expected "to use the target language in real communication" (Branden, 2006, p. 2), learners should be prepared to use the language to perform tasks in non-classroom situations, "for example, ordering from a menu, or completing a job application" (Graves, 2016, p. 82).

## 2.5 Needs analysis in the workplace

Greenberg (2012) claims that it is a common practice in the United States that companies offer language training for workers. This training needs to be tailored to the communication needs of the company. The needs analysis results frequently shape courses designed for the workplace, as companies in the same industry may consider the language needs of their workers to be different. Greenberg describes two companies within the shipyard industry that requested different learning goals for their workers. On the one hand, one required training in occupational safety and health administration, while the other company's learning goal was on writing and grammar. In identifying the language needs, "interviewing subject-matter experts is essential" (2012, p. 180). O'Sullivan (2012) argues, "(T)he test taker lies at the heart of any test" (p.80) as it is the test taker the person we aim to draw inferences from their test results. Having the test taker at the centre is especially relevant because, as argued by Swales (1990), the specific characteristics of tasks are defined by each institution; that is, they have their own ways of doing. Long (2005) argues that lessons are more likely to serve the needs of the workplace when expert insiders are involved in analysing needs.

Greenberg (2012) emphasizes the need to identify the learning outcomes and prioritize them according to their relevance. She exemplifies this by talking about when conducting a needs analysis for a construction company, and she was informed that falls from ladders were the most frequent cause of injuries. Ladder safety became one of the essential learning objectives for the course she developed. This identification of language needs is an example of the diagnostic viewpoint (J. D. Brown, 2016) discussed above, where the contents of a course are defined by the essential language needs, those which would do the most harm if not considered.

Hutha and collaborators (2013) challenged existing views towards needs analysis which have been language-centred, focusing on the four skills (reading, listening, speaking and writing) and the functions and notions of the target situation. They refer to this as a traditional, "first generation approach" (p. 14) to needs analysis. Their proposal, a "second generation needs analysis" (ibid.), adopts a task-based approach, conceiving "the task as the primary unit of needs analysis" (p. 15). This approach, they contend, is aligned with the action-oriented approach of the CEFR (COE, 2001). Hutha and collaborators (2013) report on a Europe-wide enterprise to develop "professional, field-specific language proficiency

profiles” (p. 26). The project's first stage involved an ethnographic-like approach to needs analysis that used qualitative and quantitative methods. The qualitative method, structured interviews, prevailed over the structured questionnaires. The former asked participants, among other data such as background information, about the communication situations they frequently encounter, the level of demand these place on the language user, the importance of the skills, the people they interact with and other details. Project partners from Bulgaria, Finland, Germany, Hungary, the Netherlands, and Poland used a set of interview questions to gather data for the development of the CEF (Common European Framework) Professional Profiles. This project sought to consider the relationship between language and culture not represented in the CEFR (Risager, 2007 in Huhta et al., 2013). Data collected was compared to the CEFR language descriptors from the domain, context, text, location, and participants' perspectives, finding the CEFR scales and descriptors too broad and unspecific to relate and represent the communication situations identified in the Project. The outcome of this project is a set of “field-specific language and communication descriptors for technology, business, health and social care and law” (p. 30), which have been used for course design purposes.

### 2.5.1 Tasks and needs analysis.

Tasks are defined as the real-life activities people perform daily that may or may not require the use of language (COE, 2001; Long, 2015). Nunan (2004) distinguishes what he calls *real-world or target tasks* and *pedagogical tasks*. The former happens in a non-learning environment, while the latter refers to activities that take place in a classroom, “only indirectly related to real-life tasks and learner needs” (COE, 2001, p. 157).

For learning and teaching purposes, Long defines tasks as the “real world communicative uses to which learners will put the L2 beyond the classroom – the things they will *do* in and through the L2” (M. H. Long, 2016, p. 6) (emphasis in the original). Bachman and Palmer (2010) define language use tasks as those activities in which language is used to accomplish communicative goals in situations other than language learning or teaching. From an ESP perspective, for Huhta and collaborators (2013), tasks are the communicative activities learners “are familiar with from their professional environments” (p. 9) but need to be taught to accomplish in the target language.

The Canadian Language Benchmarks (CLB) defines a language task as “a communicative “real world” instance of language use to accomplish a specific purpose in a particular context” (2012, p. ix). This definition is more specific than the first two in that the aim and environment in which language is used are established, linking the purpose for language use to the context in which communication occurs. The definition of task presented in the Common European Framework of Reference for Languages: Learning, Teaching, and Assessment (CEFR) is even more specific about the contexts in which language is used. The CEFR defines tasks as “a feature of everyday life in the personal, public, educational or occupational domains” (COE, 2001, p. 157).

For this research, the task is defined as the communicative use to which learners put the L2 in the personal, public, educational, or occupational domains within real-world everyday life activities. Tasks may require several competences to fulfil several steps or sub-tasks that lead to an expected outcome or goal achievement (COE, 2001). As a unit of analysis, among others, a task poses the advantages of being familiar to domain experts and compatible with theories about how adult learners learn languages. Another advantage is that the descriptors that were created by informed individuals are most likely to be task-based (M. Long, 2015).

A study conducted by Serafini et al. (2015) looked at the reports of NAs carried out between 1984 and 2014 for English for specific purpose learners in different contexts. They aimed to identify the tasks causing communication breakdowns within the institution that requested assistance to solve the situation. According to their study, even though the design, methods and procedures employed by researchers have improved over the years, there are still areas in which increased rigour would enhance the validity and reliability of the outcome of the NAs. They argue in favour of a task-based approach to needs analysis, using tasks as the units of analysis of the language needs of learners. They contend considering the outcomes of NA research in task-based language teaching (TBLT) would contribute to methodological improvements in NA practice. Using tasks and consistency in using relevant contextualized techniques could contribute to more valid, reliable findings. Among their methodological recommendations, they stress the importance “of ensuring that tasks identified by NA are really those required of learners to function successfully at work or other settings” (ibid. pg. 21). According to their research, this could be achieved by interviewing domain experts and trainees, both to gather initial data for designing the interview and once the instrument has been piloted, probing the generalizability of findings



so these could be used to design a questionnaire. After applying the piloted version of the questionnaire to the target population, findings were triangulated, and sources and methods were compared. It was also suggested that questionnaire findings be triangulated with findings from other sources, such as follow-up interviews.

The methodological recommendations of Serafini et al. (2015) emerged from their large-scale study carried out with international scientists at an American national research institution. The study was carried out by a faculty member (one of the authors) and several post-graduate students in 2010. Among the areas of improvement in their study, they stated a Likert scale could help differentiate the difficulty and frequency of target task criticality. They also mention open-ended responses in questionnaires, as relevant additional information can emerge through this means. Another aspect to consider is the use, whenever possible, of stratified random samples, as they consider the samples of convenience who volunteered to participate in their study may have been biased due to familiarity with the communication difficulties experienced by the international scientists.

The authors acknowledge that needs analysts may sometimes need to compromise due to practical constraints, and their recommendations may be applicable only in an ideal situation, which is often not the case. However, they argued their suggestions might be adaptable to different situations and invite researchers to opt for NA practices as rigorous as the situation allows.

Alsamadani (2017) reports on a study to identify the contents of an ESP course for Saudi engineering students. He argues that during the course design development process, it is pertinent to consider the language tasks students will require upon leaving university. He contends this will give them the knowledge and skills to function effectively in the workplace. The study involved 200 second-year industrial and civil engineering students and 25 ESP and subject-matter teachers. Students were asked about the importance of carrying out certain tasks, grouped by skills, such as, among others, listening to lectures, participating in discussions, reading journal articles, and answering class quizzes and exams. Teachers were asked about their perception of the importance of language skills for their students, their engineering studies and future working situations. Results revealed students and teachers had different perceptions of the importance of skills. While listening was the least important skill for students, for teachers, it was the second most important. Writing was the most relevant skill for students, while reading was the most important for teachers. Reading had second place in importance for students. Results from students and

teachers did not coincide. This mismatch was explained by claiming that responses given by students seemed not to consider their specific needs for academic or professional purposes. Even though the article mentions that, because of the study, recommendations were given for improving the ESP course for engineering students, these are not specified.

Another study carried out to identify the target language use situation for course design purposes was reported by Benavent and Sánchez-Reyes (2015) in the context of Spanish law enforcers. A target situation analysis was conducted to identify law enforcers' communicative real-world language requirements. Upon analysing the different specialist units within fifteen Spanish police stations, a questionnaire was designed, containing the competences their professionals required to fulfil the general tasks of an ordinary day. The questionnaire was sent to 130 other police stations in Spain to determine the extent to which the competences selected were relevant for other Spanish law enforcers. The needs analysis outcome was critical in defining a practical syllabus that would provide police officers and inspectors with the necessary tools to carry out their daily job tasks in another language. Authentic material was used to design classroom tasks that emulated real-life situations for the students.

The findings from a study conducted in Malaysia revealed the importance of considering content expert knowledge when deciding the contents of an EOP course, as Malaysian employers highly value English. The study aimed to explore the extent to which the EOP course offered at the University of Malaysia met the expectations of internal stakeholders in preparing students for the business working environment. The group of 177 internal stakeholders was composed of undergraduate students, instructors, and academic administrators of the EOP course offered at the university. Results revealed closer communication between instructors and domain insiders could make up for instructors' insufficient knowledge of the language needs of the business organizations. Carrying out a needs analysis was one of the recommendations given to course organizers, considering this activity would contribute to meeting the expectations of stakeholders (Sarudin *et al.*, 2013).

### 2.5.2 Informants

When carrying out a TSA, including all the possible data sources (Chambers, 1980; Porcher, 1983 in West, 1994) contributes to a more accurate identification of needs

(Chambers, 1980). Even though linguistically well prepared, language teachers may not always be familiar with the target context of language use. It is, therefore, necessary to identify the stakeholders who are familiar with the language needs of the context of interest (Bachman and Palmer, 1996) and are in the position of providing details about the specific language activities involved.

Informants may be the students, the teachers, school administrators, field specialists or other stakeholders from non-academic institutions (Hoadley-Maidment in West, 1994). Knoch and Macqueen (2020) use the term “domain insider” to refer to “professionals with training and experience in a profession” (p.3). This term will be used interchangeably with professionals, employees, or subject experts. West (1994) states that the natural expectation of the purpose of needs analysis is to be “for the benefit of the *user*, i.e. the student or trainee” (p. 6). This expectation implies that, whenever possible, the latter should be part of the informants.

Even though ESP practitioners are advised to consider as many sources of information as possible (Chambers, 1980) when interpreting results, they need to be cautious about a particular group of informants imposing constraints on the final (Jones, 1991) outcome. On the other hand, while learners are not considered to be well prepared to provide relevant information concerning the language required, they may have certain expectations (Chambers, 1980; West, 1994) that may be worth considering. Similarly, employers may not be informed in detail of employees' specific language needs (Chambers, 1980). However, employers and employees are considered essential in identifying the language requirements of the target context of language use (Acedo Dominguez & Edwards Rokowski, 2002). Results from the study conducted in Malaysia considered that including domain insiders within the company in the NA process could provide a closer insight into the needs of the working environment, contributing to improved satisfaction of stakeholders on the course outcomes (Sarudin et al., 2013).

The ESP practitioner needs to ponder data gathered by the relevant parties to set priorities (Chambers, 1980); thus, Jones (1991) considers it not unusual for stakeholders to play a decisive role.

Provided that the ultimate goal of carrying out a needs analysis is for the benefit of the learner (West, 1994) who has been placed at the centre of the process, it is of the utmost importance to reconcile views or needs (Brindley, 1989; O'Sullivan, 2012). Hutchinson and

Waters (1987) argue, "(T)here is little point in taking an ESP approach, which is based on the principle of learner involvement, and then ignoring the learners' wishes and views" (p. 58). Learners will feel more motivated and interested when they notice their opinions and preferences are considered. Chovancová (2014) reports conducting an NA study to bring fresh ideas to the traditional course outline for legal English courses. The study examined pre-service students' expectations regarding the language they would need to use after graduation. Chovancová concludes that although pre-service students may not have clear ideas of their work-related English language requirements, it was important that course designers considered their needs and wants. The study reveals these "are crucial for increasing the students' motivation" (p. 43); however, she contends that more information needs to be gathered from other informants, i.e., professionals, to better understand the target situations. Once again, the degree to which the views of the learners are included needs to be pondered.

## 2.6 Motivation

Identifying the purpose for learning a language may directly impact a learner's motivation to study it (Dudley-Evans and St John, 1998). It is well-known and accepted that motivation is the driving force behind success in learning, regardless of what is being learned. Learners must want to learn it to succeed (Harmer, 2001). Not only in learning another language but in any learning, motivation is at the centre of the influential factors in goal achievement (Dörnyei, 1994). Language learning motivation, situated in the attitude theory within the social psychology discipline, "provides the primary impetus to initiate L2 learning and later the driving force to sustain the long and often tedious learning process" (Dörnyei, 2005, p. 65). Presumably, learners will be motivated to learn the language that will allow them to fulfil their identified specific language needs (Hutchinson and Waters, 1987).

An identified purpose for learning the target language could be motivated by internal or external purposes. For some learners, studying English could be reduced to passing an exam. In such cases, learning the language is not motivated by its perceived usefulness but is seen as a requirement to be fulfilled. The extent to which learners are aware of the relevance the language being learned has for their current or future activities "would improve the learners' motivation and thereby make learning better and faster" (Hutchinson & Waters, 1987b, p. 8). The same could be said about language testing. The degree of

preparation and general attitude towards a language test may be directly associated with the degree to which the test language is considered useful or relevant for test takers.

The environment surrounding learners may generate extrinsic motivation, given that their needs are “imposed by the reality of the language situation” (Richterich, 1972, p. 4).

Unmotivated learners may think they do not need to learn English. However, they may also be unaware “of the (language) needs they may have later in life” (M. Ellis & Johnson, 1994, p. 80). Therefore, they remain unmotivated to learn the language.

Language motivation research in the twentieth century focused on the social psychological approach, primarily led by Gardner (2019) and his Canadian collaborators (Gardner & MacIntyre, 1993; Gardner & Tremblay, 1994; Gardner, 2001). This approach is grounded on the perspective that motivation to learn a second language (L2) distinguishes it from motivation to learn other school subjects in that, in the latter, the learner is not expected to acculturate into the L2 group in the learning process (Al-Hoorie, 2017). This type of motivation, called *integrative* orientation, proposed in 1972 by Gardner and Lambert (in Al-Hoorie, 2017), presupposes the learner has the desire “to identify with the target language community and adopt their distinctive speech behaviours and styles” (Ushioda & Dörnyei, 2012, p. 396). A different kind of motivation towards learning another language, the *instrumental* orientation, is related to the perceived “practical value and advantages of learning a new language” (Gardner and Lambert, 1972 in Ushioda and Dörnyei, 2012, p. 397).

However, in the twenty-first century, there has been a focus on the psychological theory of “possible selves” (Markus & Nurius, 1986) from which Dörnyei develops a version of the “L2 Motivational Self System” which centres on the language learners’ vision of themselves in the future (Dörnyei, 2009; Ushioda & Dörnyei, 2012). The L2 Motivational Self System comprises three dynamic concepts: a) the *Ideal L2 Self*, b) the *Ought-to Self*, and c) the *L2 Learning Experience*. The Ideal L2 Self represents wishes and desires that the learner envisages becoming because of L2 learning. The Ideal L2 Self can be a powerful motivator when an ideal vision of the self exists, when the learner can create an image of the self, i.e., successfully using the language. The Ought-to Self is concerned with the desire to meet the expectations of others or to avoid negative consequences and represents the traits an individual believes he ought to possess. The Ought-to Self may be in discrepancy with the individual’s wishes or desires. Finally, the L2 Learning Experience refers to the context and situation in which learning occurs. The L2 Learning Experience may be influenced by the

teacher, peer learners, the content of classes and the individual's feeling of success (Dörnyei, 2009). The L2 Motivational Self System is conceptualised as two self-guides (the ideal and the ought-to) and the learning experience.

The concept of the L2 Motivational Self-System considers that all actions performed by individuals emerge from purpose in professional or everyday contexts. This purpose is frequently expressed in terms of goals which "give meaning and direction" (Dörnyei, 2019a, p. 58) to the actions of individuals.

In the study conducted to identify the language needs of Spanish-speaking law enforcers to work in international settings (see 2.5.1 above), Benavent and Sánchez-Reyes (2015) state that "a syllabus that is based on target situations is motivating ESP apprentices who see its practical relevance" (p. 147). They also state they engage in learning and gain confidence to interact with professional peers in the real world. It could be hypothesized that this overall positive response to the syllabus is related to an ideal L2 self who successfully uses the language within their professional field.

Research carried out by Lamb (2004) reports high motivation to learn English among Indonesian children between 11 and 12 years old participants. In this study, it was almost impossible to distinguish between the two traditional constructs: instrumental and integrative motivation. As a result of findings, Lamb argues that in the current globalized world, the ever-growing use of English for international communication in varied multicultural environments for various purposes, the desire to acculturate with a particular Anglophone culture fades away. The term integrative motivation loses sense as individuals may see their future selves as part of an international community. He contends that for these children, the concept of the self may be reframed towards a bicultural self "which incorporates an English-speaking globally-involved version of themselves in addition to their local L1-speaking self" (p. 1). A similar finding was reported by Weger (2013) in a study aimed at exploring the learning motivation of international adult learners during a 20-hour-per-week course in a United States (US) based English learning program. Findings reveal that even though these international students chose to study English in the US, their motivation to study English was "more closely associated with developing an international identity and meeting a variety of everyday personal needs, namely entertainment, media and travel" (p. 99) than with integrating into the target language community.

Green (2014) argues that because assessments are generally associated with rewards of different types (job opportunities, attention from the teacher, grades or prizes, among others), they can serve as powerful motivators. However, he warns that assessments may have the opposite effect. Learners who cannot get the grades they aspire to or fail to pass the test repeatedly may become discouraged and end up dropping out all effort to continue learning. A research review by Harlen and Crick (2003) on the relationship between testing and motivation for learning revealed that assessment and testing practices may negatively affect students' motivation for learning.

Purpura argues that social psychological factors such as motivation "may have a significant impact on test scores, suggesting that language knowledge may be necessary, but not a sufficient condition for 'good' language test performance" (2004, p. 94). Concerning test task content, Alderson argues that the level of engagement of test takers with a text may be influenced by its type. He states that "texts that carry meaning for readers, that interest them that relate to their academic background, leisure interests, intellectual level and so on might motivate a deeper reading than the traditional anodyne or even contentless texts" (2000, p. 29). The way test tasks are addressed is influenced by the motivation or the interest of the test taker in the content of the task itself (Weir, 2013).

Test-taking motivation has been defined by Schunk and collaborators (2008 in Penk et al., 2014) as "an active process by which goal-oriented activity is initiated and maintained" (p. 4). Test-taking is an activity which is specific to students. The desire to take part in such activity was defined by Baumert and Demmrich as "the willingness to engage in working on test items and to invest effort and persistence in this undertaking" (2001, p. 441) and argue that this factor may present a threat to validity, especially when results are compared among different parts of the world. According to Sundre and Kitsantas (2004), student test-taking motivation is critical in sparking performance that accurately represents learning. Knekta (2017) argues that if students cannot find a reason to complete a test or consider they will not be able to complete it, they may not be willing to make an effort to respond. A lack of desire to actively demonstrate actual language ability will yield results that underrepresent student learning. A study conducted by Bai (2020) in China regarding the relationship between learning motivation and test performance revealed that when test takers had a negative motivation towards the test, this prevented learning and negatively affected test performance.

Fulcher (2009) argues that individual test-taking motivation is beyond the control of the test designer. Even though, according to Knehta (2017), high-stakes tests are assumed to motivate learners to complete them, some students may not be able to perform at their best “because of factors such as test anxiety or low expectancies” (p. 1). This situation could lead to test results that do not accurately represent the students’ knowledge. However, adverse reactions from a test-taker group towards the content or format of the test have been identified to be related to language learning and test-taking motivation. Henning (1987) referred to this as *response validity*, a term that has apparently, not been retaken by the testing research community. He stated that “(I)f examinees do not approach the test-taking situation in the expected manner, the results may prove to be invalid. This may occur if examinees are insincere, insecure, or hostile to the-taking situation” (p. 92). Haggerty and Fox (2015, p. 12) argue that the test-taking environment could negatively affect some students’ motivation to learn English. According to Fulcher, test designers should try, to the extent possible, to avoid or reduce hostile or adverse reactions to the test or the testing situation. On the contrary, test developers want test takers to have positive views towards the test or the test-taking situation and be motivated to do their best. Therefore, they should try to do what is possible to harness motivation towards the test and the process around it (Fulcher, 2009).

## 2.7 Summary

Most adult learners of English can identify the purpose for which they need or want to learn the language. English for Specific Purposes aims to help students learn the language they require or are interested in. Learning what’s important enhances learner motivation and uses time more efficiently by concentrating on the relevant language. Tests are also designed to target specific language use purposes. The language to be learned or assessed is identified through a needs analysis process, within which it is crucial to gather data from all stakeholders involved. Even though students or test takers are not frequently considered, their participation is critical if needs analysis is expected to benefit the language user.

Test designers may reduce or eliminate negative results in response validity by considering the opinion of test takers towards the test. If adverse reactions to test content or test tasks are observed, efforts made to modify these, to the extent possible, may positively impact the response validity of test takers.



## **3 Language testing for specific purposes**

### **3.1 Introduction**

This chapter looks into language testing for specific purposes, addressing testing for academic purposes, working purposes and multipurpose tests. These topics lead to a discussion on tests used for purposes other than those for which they were designed, followed by face validity and validation of the use of a test. Within the latter, the argument-based approach to test validation will be looked at through three frameworks for language test validation: the interpretation/use argument, the assessment use argument (AUA) and the language assessment for professional purposes (LAPP). Next, tasks in language testing will be discussed. A proposal for a Local Context-Sensitive Language Test (LCSLT) will then be presented. The LCSLT aims to consider the language needs of the local environment. A discussion of adapting design patterns in language assessment as a layout for domain-related test task design follows. The chapter ends with a summary.

### **3.2 Language testing**

Bachman (1990) states that language tests are generally designed to fulfil an identified purpose of a defined context of use and involve a specific group of language users. They are commonly used to aid the decision-making process about test performance of individuals (M. Kane, 2013) in many institutions and situations around the globe. These decisions have consequences on the lives of individuals and society (L. F. Bachman, 1990; L. F. Bachman & Palmer, 1996; Weir, 2005). Using test scores for decision-making processes is expected to bring beneficial consequences for stakeholders (L. Bachman & Palmer, 2010).

Test designers and administrators need to be prepared to justify using an assessment in a particular situation with a particular group of test takers. Stakeholders need to be convinced that the uses of assessment results are justified (L. F. Bachman, 2015).

The need to use English in a variety of contexts has brought about many learners who need evidence of language proficiency for a wide range of purposes.

### 3.3 Tests for specific contexts of use

For Douglas (2000), specific purpose language testing aims to “make inferences about individuals’ abilities to use language in specific academic, professional, or vocational fields” (p. 41). These inferences are achieved by looking at the performance of test tasks that incorporate the features of the target situation of language use (Bachman and Palmer, 1996; Douglas, 2000). Tests for the academic, professional or vocational fields are expected to differ from the other, as, according to McNamara (2000), a test designed for a field is expected to incorporate the characteristics of the field of interest. Seemingly, Basturkmen and Elder (2004) argue that specific purpose language testing is grounded “on the assumption that different domains of language use draw on different areas of knowledge and are associated with distinct varieties of language” (p. 681). In this sense, a test designed for a given domain is expected to comprise language related to that context.

Specific purpose language tests are mainly designed for the academic and the working environment (Basturkmen & Elder, 2004) and are “narrowly focused with tasks designed to simulate the demands of particular real-world situations” (p. 680). Therefore, a test for the academic environment is expected to engage the test taker with knowledge and skills that pertain to the academic environment rather than those found in a working context

### 3.4 Testing English for the academic context

Within the academic context, the aim of a test may be to diagnose the candidates’ language knowledge, skills or abilities or to provide feedback on areas of improvement, such as the DIALANG (Alderson, 2005). Language tests are also designed to aid higher education institutions within English-speaking countries in the international candidate screening process. The EFL background of candidates prompts academic institutions to ensure they are capable of “reading or listening to source texts, as they would in real academic tasks in English” (Fulcher & Davidson, 2007, p. 231). Two examples of tests used for this purpose described below are discussed according to their inception.

The Test of English as a Foreign Language (TOEFL) was developed in the early 1960s to assess the English proficiency of non-native speakers interested in enrolling in English-medium academic institutions (TOEFL, 2020) in the United States of America. Even though by 1997, the test had not gone without revisions (McNamara, 2001), it was still reflecting “a

structuralist view of language” (Wall & Horák, 2011) as a large proportion of the test was devoted to language knowledge and the two receptive skills included (reading and listening) were tested independently. Despite many international students being admitted with high TOEFL test scores, academic institutions were highly concerned about their “insufficient writing and oral communication skills to participate fully in academic programs” (Jamieson et al., 2000, p. 3). This led to a significant revision of the test during the late 1990s (McNamara, 2001) and early 2000s, aiming to design a test that would “measure examinees’ English language proficiency in situations and tasks reflective of university life in North America” (Jamieson et al., 2000, p. 14). Extensive work was carried out to identify the academic tasks that better reflected the communicative competence required in the academic context (ETS, 2011). The construct on which the revised TOEFL was built was defined as “communicative language proficiency for academic life” (Chapelle et al., 1997, p. 1). This construct acknowledges that many academic tasks require the use of both receptive (listening and reading) and productive (speaking and writing) skills in an integrated or supplementary form (TOEFL, 2020). Even though the new iBT TOEFL, designed for academic purposes, is not subject specific, and could not be placed close to the specific (right) end of the continuum (see Figure 2.2 above) (Knoch & Macqueen, 2016), its development is grounded on a construct that attempts to replicate the way language is used in academic contexts.

On the other hand, the ELTS (English Language Testing Service), the predecessor of the IELTS (International English Language Testing System), was designed in the 1970s by the British Council to measure the extent to which foreign students met the language required to study at a university in the United Kingdom “represented a real departure from the structurally focused approach of previous decades” (Davies, 2008, p. ix). The latter was possible because of the needs-based approach adopted to identify the communicative demands international students meet in a very practical way. The ELTS was an ESP test that allowed students to select the subject area of their interest according to the discipline they were applying for (Alderson & Clapham, 1992). Even though the ELTS was an innovative modular test that yielded a subject-specific diagnostic result with high face validity, it also presented measurement, theoretical and practical issues (Davies, 2008). The ELTS had to be shortened and simplified (from six to four specialized modules) upon revising its validity, practicality, and reliability in 1989. At that time, the involvement of the International Development Program Education Australia (IDPEA) with the British Council and the University of Cambridge Local Examinations Syndicate (UCLES) in managing the

test led to the addition of the word *international*, and the ELTS became the IELTS. This partnership gave the test an international perspective and precluded the content of the test from being country-specific (Charge & Taylor, 1997). Continuous research motivated further changes (Davies, 2008), moving the initial ELTS, once placed somewhere closer to the specific purpose end of the continuum (see Figure 2.2 above), towards its opposite end. The current IELTS offers two strands: academic and general training and migration (Davies, 2008). From being an English for specific academic purpose (ESAP), the ELTS evolved towards the IELTS. The IELTS is an English for general academic purpose test (EGAP). This evolution placed the IELTS towards the general end in Figure 2.2 above (Knoch & Macqueen, 2016), reflecting the complexities of such an endeavour.

International students wishing to study at an English-speaking university are the target test-taker population of the two EGAP tests discussed. They were designed to meet the needs identified by the target academic institutions. However, such needs differ from those of academic institutions where English is not the medium of instruction. Even though university students, like those attending the UJED, need English to access the most recent information published in their field of interest, their language needs are not the same as those of students applying to study at an English-speaking university.

### 3.4.1 Testing for the Professional Workplace

The purpose of tests for the working environment was to assess the language competence of professionals whose field qualifications were gained in a context where English is not used daily. Migration or international requirements to use English within the workplace stimulated the development of these tests (Knoch & Macqueen, 2016).

These “tests are generally designed for smaller, more homogeneous groups of professionals” (ibid. pg. 293) or an identified local target population, as is the case of the Test of English for International Communication (TOEIC) (ETS, 2012). Other tests are designed for a broader group of test takers, such as the APTIS (British Council and Aptis, 2016), which “is designed for a wide, mainly young adult and adult population” (O’Sullivan, 2015, p. 9).

Although the TOEIC was initially developed for a Japanese businessmen target population in the late 1970s (Chapman, 2004; Woodford, 1982), it is now claimed to be aimed at non-native English speakers who work within international settings (Powers & Powers, 2015).

The TOEIC, which was, in essence, a listening and reading (LR) test, went through a redesign process in 2006 to align it “with current theories of language proficiency” (Schedl, 2010, p. 2.2), a year in which the speaking and writing (SW) test was introduced (Im & Cheng, 2019, p. 316). The second revision, announced in 2015, incorporated different accent varieties of English and types of current written communication, such as electronic messaging (Im and Cheng, 2019). Even though the changes made to reflect better the real-life use of English, Im and Cheng argue “there is a growing need to expand TOEIC constructs to fit the real-world language demands of international workplace contexts” (2019, p. 322). They claim that although the test has been in use for more than 30 years measuring everyday language proficiency of individuals whose mother tongue is other than English, the extent to which results of the TOEIC Speaking reflect real-life language use in the target domain has not been clearly established. They also argue that it is unknown if the validity evidence gathered for extrapolating purposes, consisting of self-assessment reports from Korean and Japanese test takers, is sufficient to claim its validity in other contexts, such as Brazil or Taiwan. O’Sullivan (2012) is sceptical about the claims made on the purpose of assessment of TOEIC when he states that it has “never seriously attempted to reflect the business domain, and instead test(s) general language knowledge” (ibid. p. 82). O’Sullivan also argues that incorporating the SW papers assesses general language production rather than business-related language. As the TOEIC is not related to a specific working context but intends to be relevant for working environments in different countries, it would be found towards the general end of the continuum in Figure 2.2 above (Knoch & Macqueen, 2016).

The APTIS, a much younger test, also developed for the international working environment, is claimed to be useful for hiring or assessing English performance of employees whose mother tongue is other than English (British Council & Aptis, 2016). Test users are offered five subtests to choose from (one for each skill and a core one focusing on grammar and vocabulary). These subtests offer a variety of tasks and response formats through which it is possible to collect ample and varied evidence of performance that provides a precise and reliable estimate of the language ability of the candidate (O’Sullivan, 2015). The APTIS system is innovative because it allows test users to select the subtests that best meet their testing needs. This innovation makes the test more accessible, as even though all candidates must take the core subtest (grammar and vocabulary), clients decide on the combination of papers of their interest, paying only for what is needed. The flexibility of the APTIS system allows the test designer team to further meet particular testing requirements

by working with the client for a localised test to be designed. O’Sullivan claims it is possible to make the test “appropriate to the particular context and domain of language uses” (p. 10).

A study conducted in 2016 to compare the APTIS and the GEPT (General English Proficiency Test), a famous language test used in Taiwan, revealed more than 80% of test takers had favourable opinions about the core subtest, as well as about the listening, reading, and writing subtests. They considered the test results “could reflect their English ability”(Wu et al., 2016, p. 31). Test scores for the speaking subtest were lower than the other subtests; not surprisingly, test takers' perception of the extent to which test results reflected their language ability was also lower (74%). Around 85% of test takers consider the topics, vocabulary and sentence structure found in the tests to be relevant to them since they are likely to encounter them for everyday or workplace purpose communication (ibid.). In sum, the study reports overall positive face validity among test takers for the abovementioned aspects.

The current APTIS test does not focus on any particular area or discipline. Hence, it is placed towards the general ESP end of the continuum discussed in Figure 2.2 above.

The Occupational English Test (OET) was another assessment instrument designed to measure the level of English of non-native English speakers who wish to work in an English-speaking medium (Arkoudis et al., 2009). The OET has been measuring the readiness of internationally trained healthcare professionals to function effectively in the Australian workplace (Elder et al., 2013) since the late 1980s (OET, n.d.). Tests for 12 health professions were developed, such as nursing, medics, physiotherapy, dentistry, dietetics, and optometry, among others, for listening, speaking, reading and writing, emphasizing assessment of successful communication within the healthcare context (OET, n.d.). The OET is no longer exclusive to the Australian healthcare target population, owned by the Cambridge Boxhill Language Assessment Trust, and is now recognised by Australian, New Zealander and Singaporean healthcare-related authorities. Its place in the continuum of Figure 2.2 is towards the specific end of it, as test tasks are more closely related to the activities observed directly in the TLU healthcare domain (Knoch & Macqueen, 2016).

Also designed for the healthcare profession, the Canadian English Language Benchmark for Nurses (CELBAN) has been assessing “the threshold English language proficiency of

internationally educated nurses” (IENs) (Lewis & Kingdon, 2016) since mid-2005 (CCLB, 2006). The need for more adequate test results than those being obtained from “English proficiency tests (e.g. TOEFL, IELTS, MELAB) previously used to assess the language proficiency of IENs” (IENs Internationally Educated Nurses) (Lewis & Kingdon, 2016, p. 70) led to the design of this occupation-specific language test. A test designed based on the analysis of the target language use of nursing that dealt with its content and context linguistic demands and was validated with IENs was deemed necessary (ibid.). The CELBAN is also placed somewhere closer to the right (discipline-specific) end of the continuum in Figure 2.2 above.

A study conducted in 2003 on assessing the language skills of international medical graduates (IMGs) in their integration into the Canadian physician workforce recommended the development of the existing Objective Skills Clinical Examination (OSCE) into an “assessment tool that combines the assessment of both clinical skills and language proficiency for successful practice in Canada” (ibid. p. 5). Field specialists collaborated to identify the tasks that medical residents were expected to face to develop an assessment reflecting real-world interaction within the medical context. Even though an international assessment tool (such as IELTS or TOEFL) was used as part of the initial screening process, its role was to determine the readiness of IMGs to prepare to take an assessment that “is a better indicator of the language benchmarks required for medical practice” (Watt et al., 2003, p. 36) in Canada.

Successful communication between speakers of different mother tongues is desirable within the workplace, but this becomes more important in contexts such as healthcare or aviation. Successful communication within the international aviation context is vital to avoid incidents that could have fatal consequences (Alderson, 2009). The specific language of pilots and air traffic controllers is known as standard phraseology. Standard phraseology is or should be, well-known and followed by all interlocutors in this area of communication. However, there are occasions when its use is inadequate. During emergencies, presumably under stressful circumstances, proficiency in English, the language used for international aviation communication, is of the utmost importance. The ELPAC (English Language Proficiency for Aeronautical Communication) test was designed for the aviation industry by EUROCONTROL (a European Organization for the Safety of Air Navigation). Its purpose is to assess the English language proficiency of air traffic controllers that meets the

language proficiency requirements of the International Civil Aviation Organization (ICAO) (Enright, 2007).

The previous experience of EUROCONTROL when the PELA (Proficiency Test in English Language for Air Traffic Controllers) (Douglas, 2000) Test (for student controllers) was designed in the early 1990s provided a starting point for the development of ELPAC. Between the end of 2004 and the beginning of 2005, the Needs Analysis used for the design of PELA was analysed and found to continue to be valid for the communication of air traffic controllers. Two papers were designed for the ELPAC: an internet-accessible listening comprehension test and an oral interaction paper. The latter considers communication with no visual support and a face-to-face component. EUROCONTROL currently offers the ELPAC ATC (air traffic controllers), ELPAC pilots and ELPAC level 6. The ATC and pilot tests have been designed to assess the ability to communicate effectively at levels 4 and 5 of ICAO's 2011 language proficiency requirements of ICAO. Level 6 test aims to assess the ability to communicate effectively by dealing with ambiguity and clarifying misunderstandings with a pilot or controller who is either less proficient or comes from a different cultural background (EUROCONTROL, 2018).

The ELPAC test can be placed towards the specific end of the ESP Assessment Specificity Continuum (see Figure 2.2 above), as the language assessed serves for aeronautical communication context purposes.

The tests discussed above for academic and workplace purposes have been designed to meet an identified need. While some require certain specialist background knowledge, others were designed to relate to a broad range of areas of knowledge, disciplines or registers within the academic or workplace domain (Knoch & Macqueen, 2016). There are, however, test results from other tests which are claimed to be relevant for several purposes, i.e., academic and workplace purposes.

### 3.4.2 Tests for more than one purpose

Whether a test has been designed considering a broad international test taker population like the IELTS General or the B2 First (formerly known as Cambridge English: First -FCE), others, like the ELASH, have been designed based on a test taker population that share a common feature, such as the mother tongue.



While the IELTS Academic was designed to aid the selection of candidates applying to study at British universities, the IELTS General module claims its test results can be used to assess the language to “train or study at below degree level” (UCLES, 2017, p. 3). It is also claimed to be relevant for immigration to an English-speaking country. The target test-taker population would be adults and young adults whose first language is other than English.

The IELTS and the B2 First share the same test-taker target population and the use of their test results. The B2 First, as the IELTS, assesses the four skills. Both tests are composed of four papers, one per skill. The difference between these two in terms of components is that the B2 First includes a use of English component within the reading paper. The speaking paper is also different because candidates speak to the examiner in the IELTS, and the interview is recorded. The speaking test for B2 First has been designed under a different format, as the interview is done in pairs, two evaluators are present in the testing room, and the interview is not recorded. A holder of a B2 First certificate is claimed to be able to “use everyday written and spoken English for work and study purposes” (Cambridge English Language Assessment, 2016, p. 2). Both tests are claimed to be relevant for the academic and the working contexts.

ELASH (English Language Assessment System for Hispanics), designed to test proficiency in English of native speakers of Spanish, was first applied at the end of 1999 (College Board, 2008). The College Board, an American institution, designed two levels for ELASH. Both ELASH 1 and ELASH 2 contain a listening section, a language use section, an indirect writing section and a reading comprehension section. The difference between ELASH 1 and ELASH 2 is that within the reading comprehension section, ELASH 1 includes a vocabulary component, while ELASH 2 includes an idiomatic expressions component (College Board, 2015a; College Board, 2015b). The test was designed in a way that it is possible to assess “the positive and negative interference between the two languages in areas such as word order and cognates” (College Board, 2008, p. 22). A distinctive characteristic of ELASH is that, as it is designed for Spanish speakers, the instructions for carrying out the tasks are in Spanish (College Board, 2015b).

The College Board claims the results of the ELASH can be used for several purposes, such as to determine or certify command of the English language of Spanish speakers. ELASH is also claimed to be useful to place students in different courses according to their command of English or as additional criteria for special program entrance and/or to evaluate progress in language learning (College Board, 2008, 2015a). In 2010, ELASH was first benchmarked

to the Common European Framework of Reference (CEFR) (COE, 2001b) and subsequently revised in 2017 (College Board, 2018).

These tests have not been designed to meet the needs of a specific target group of test takers or the context of language use; therefore, they are not likely to be considered tests for specific purposes. Their place in Figure 2.2 above will not be found towards the specific end of the continuum.

#### 3.4.2.1 Face validity

One of the most important and widely discussed concepts in language testing is validity. The term validity provides information regarding whether “the test measure(s) what it is intended to measure” (Phillips et al., 2020, p. 187). Even though validity is usually addressed empirically, there is a type of validity that refers to how the test looks. If a layperson questions or doubts that a test is a test, if the test does not appear to measure what it is intended to measure, then the face validity of the test may be considered to be low (Phillips et al., 2020).

According to Brown (2004), test takers’ “feeling” of a test as “valid” has a strong influence on the examinees’ attitude towards the test. If a test is perceived to be “unfair” due to the difficulty of the tasks, the amount of time allocated for its completion, the type of language assessed, or the lack of clarity of task instructions, among other factors, test takers’ confidence and anxiety could be negatively affected. Test takers’ attitudes could be affected due to this “seemingly superficial indicator of validity” (Phillips et al., 2020, p. 187). Test takers may feel encouraged to become better language learners if they have a positive attitude towards a test (Dawadi, 2021). A test may build on test takers’ motivation to do their best when considered good, while a test perceived as bad may cause test takers to answer the test carelessly (Phillips et al., 2020).

Taking into account the voices or insights of test takers is not a common practice, as reported by Jin (2023) when searching for articles between 1984 and 2021 published in the *Language Testing* journal. Face validity is frequently undermined as a type of validity worth considering. Opportunities for test takers to participate in language assessment policies and practices are scarce, much less in large-scale assessments (Jin, 2023). However, face validity “has important implications for other aspects of the test’s validity” (Phillips et al., 2020, p. 187). Even though test takers are seen as the target of assessment instead of

valuable sources of insights (Jin, 2023), they are the ones who experience the consequences of test development and test use (Hamid et al., 2019).

### 3.5 Test use

Language tests are used as gate-keepers or door-openers (L. F. Bachman & Purpura, 2008) for a variety of activities and situations in many different contexts. The decisions based on test scores have consequences for those involved (McNamara, 2010). The consequences of test use were addressed by Messick in his influential discussion in favour of a unified concept of validity in 1989 (L. F. Bachman, 2005; McNamara & Roever, 2006; Messick, 1989). He was concerned about the interpretations and test uses, claiming these needed to be justified to all stakeholders (Messick, 1989). He argued in favour of pondering the potential consequences of test use whenever the test was considered for a new purpose. Not doing so could result in unintended consequences, an undesired outcome that opposes the testing aims (Messick, 1980).

Language assessment for academic purposes aims to determine if an individual possesses the language skills required upon entering university. The aim of assessing language for the workplace is to ascertain the extent to which an individual has the language skills to carry out the professional duties required in the workplace (O'Loughlin, 2008). Tests are developed to be used for an identified purpose with a specific target test-taker population. However, some tests are being used for purposes other than those which initially motivated their design.

Watt and collaborators (2003) claimed that international assessment tools (such as IELTS or TOEFL) developed and used as university entrance tests did not consider the language and skills required for medical practice. They claim they “are at best measures of general language ability, and not measures of language proficiency” (ibid. p. 16). Therefore, in assessing the language skills of international medical graduates in their integration into the Canadian physician workforce, they recommended using an assessment tool that better reflected the communication demands placed on physicians.

Tests for academic purposes such as IELTS (Davies, 2008) and TOEFL (TOEFL, 2020) are appropriate to predict performance within an academic environment, as they were designed to assess. Using test scores of a test designed for purpose X to make decisions about purpose Y, without the required validation, questions the validity of using test scores

of X for Y purpose (Knoch & Macqueen, 2016). Elder (2017) argued about the usefulness of tests designed for university admission purposes as exit tests because it "...remains uncertain ...the utility of the information they provide for employers" (pg. 279).

An example of the use of a test for a purpose other than the one for which it was designed is reported by Qian (2007). The University Grants Committee (UGC) of Hong Kong considered that adopting an exit English language test for all undergraduate students would motivate students to improve their English proficiency. In addition, the intention of the UGC in adopting a standardized English test as a school exit test was to aid in a job search or further studies of students. The IELTS Academic, designed for the university entrance screening process, was chosen mainly for its international status and used as the university exit test. However, it was concluded that there was little ground to believe that the "test can claim with confidence to serve both purposes at the same time" (Qian, 2007, p. 32).

Whether it is student motivation or a different reason, it is a fact that many non-native English-speaking higher education institutions around the world have decided to establish an English language test as a graduation requirement. On many occasions, this school exit requirement is fulfilled using international tests. Most of these tests were not designed for the use they are being put to by these academic institutions, as Appendix Table 1 below shows.

There are many tests used to fulfil a university's exit language requirement. However, these tests were likely designed to meet different purposes, such as university admission in an English-speaking environment, placement, or assessment of General English performance. Some of them were designed for multi-purpose assessment purposes. Few focus on the work domain.

Tests that aim to assess one domain, i.e., the academic or the business domain, are designed to elicit language about the domain they want test scores to relate to (Knoch & Macqueen, 2016). The language these test tasks elicit will be specific to the domain they want to make inferences about (L. F. Bachman & Palmer, 1996). Tests that claim to meet several needs will include topics, language and tasks that represent and produce language that is common across several domains (Knoch & Macqueen, 2016). Bachman and Palmer (1996) argue that the level of correspondence between the TLU domain and test tasks is low in tests that aim to meet the needs of different domains.

Knoch and Macqueen present two intersecting continua to represent their argument that language for a specific purpose “(LSP) tests must be conceptualised, not just in terms of specificity of their constructs aligned with their theoretical or intended purposes, but also the specificity of their *actual* uses” (Knoch & Macqueen, 2016, p. 294). The interaction between the test construct and the use the test is put to is represented in each quadrant in Figure 3.1 below. The horizontal continuum represents the test construct's specificity degree, while the vertical continuum illustrates the real interpretation and how test scores are used. These could be the same or different from the purpose that drove the test design. A test like the OET, designed for the healthcare profession, is used to obtain official physician registration. A specific domain language test results are used for a specific activity related to the domain the test aimed to assess. A test like the IELTS on the left of the vertical continuum can be used for university entrance or a more specific purpose: being officially registered as a physician. The main concern in these examples “is the justifiability of the fit between the test construct and the test interpretation/use” (Knoch & Macqueen, 2016, p. 295). As the IELTS exemplifies, a test may be given different uses in terms of specificity. It is more common to find a general test construct being used for a specific purpose, as in the top left quadrant, than to give a specific test construct a more general use, as in the bottom right quadrant (the last time the Cambridge English Legal was offered was in December 2016) (*International Legal English (ILEC)* | *British Council*, n.d.). Knoch and colleagues discuss examples of the former.

In 2016, a study explored the suitability of using the IELTS as a university exit or a professional entry test for accounting and engineering professionals (Knoch et al., 2016). The writing demands between the last year of study of accounting and engineering professionals and their first year in the workplace were compared. Participants also looked at IELTS writing samples. Stakeholders interviewed include students in the final year of study, lecturers, employers, graduates, and board members. Results reveal that “(m)ost stakeholders agreed that graduates entering the workforce are unprepared for the writing demands in their professions” (ibid., p.1). The writing demands at university, and the workplace differed in genre and the review and editing processes entailed. They also found that new graduates had difficulty in appropriately addressing different audiences. While having two writing tasks was positively perceived by some stakeholders, others argued two timed tasks were not representative of the professional writing demands, especially in engineering.

Qian (2007) explored the selection process of a university exit test in Hong Kong, where the IELTS was chosen. On analysing its suitability for establishing readiness for professional workplace communication, it was found that the IELTS Academic did not include components that were specifically oriented toward the workplace. Qian argued that even though an individual performed well during his time at the university, that did not ensure good performance in the workplace. He stated that communication for academic purposes and within the workplace differed in “sets of constructs, strategies and registers, among other things” (ibid. p. 32). The IELTS Academic was found not to be suitable as a university exit test.

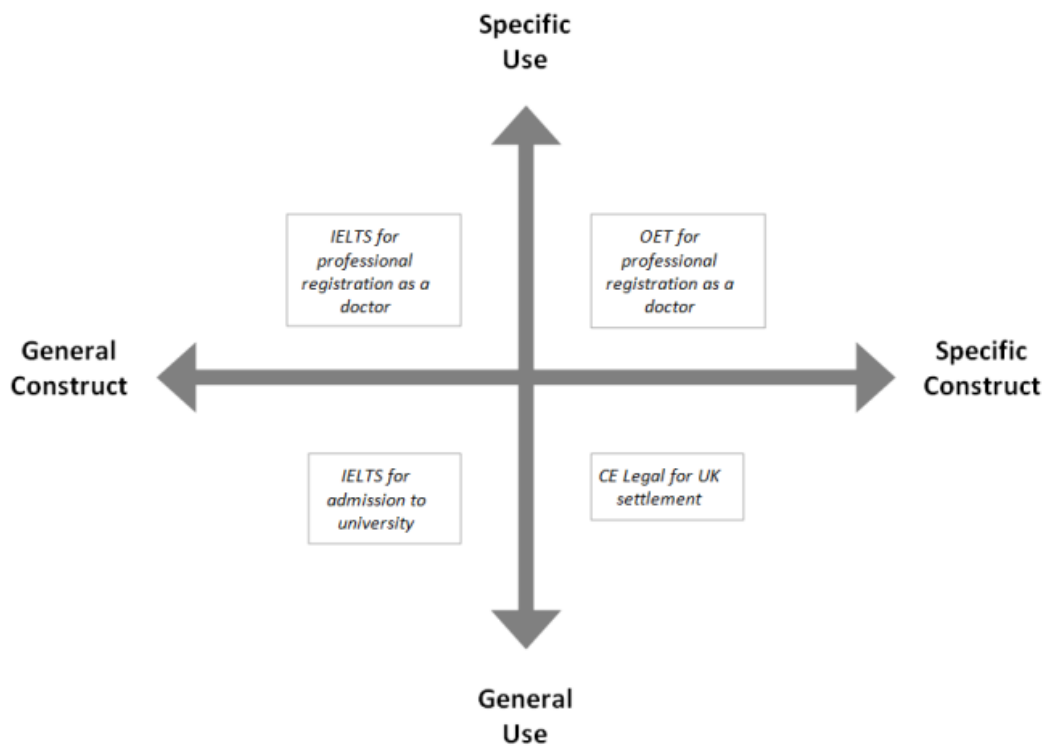


Figure 3.1 Specificity of assessment construct and assessment use (Taken from Knoch and Macqueen, 2016, p. 294)

In 2017, the Australian Government Department of Immigration and Border Protection (DIBP) received a report where the use of IELTS as a general English proficiency test as part of the citizenship application process was discussed (Knoch et al., 2017). Knoch and colleagues argue that the IELTS General was initially designed for people going to the UK

to receive training (i.e., police officers, firemen, etc.); therefore, “while called ‘general’, is not general English language skills but a test of academic English” (p. 3). They argued that “the highly developed literacy skills measured by the IELTS” (ibid. p. 3) are not necessary for everyday communication or to become part of Australian society. If a general English test was required to demonstrate readiness for citizenship, they suggested using an assessment instrument focused on language for communication, emphasising listening and speaking. According to the authors, such a test would better meet the language assessment needs for citizenship application.

Bachman and Palmer (2010) argue that the premise behind the use of language assessments is that they will result in outcomes or consequences that benefit stakeholders. They argue that the needs of the different stakeholders need to be met by language assessment fairly. Their outcome needs to be beneficial. Making high-stakes decisions based on test results (Bachman, 2005) that do not provide information regarding performance in the target language use situation (TLUS) (Bachman and Palmer, 2010) may result in unintended negative consequences for those involved. Test developers and users need to clearly understand what the instrument being used is assessing and its relevance to the context in which the language will be used (*ibid*). They need to be held accountable for using the results of a test in a decision-making process. The validation process focuses on gathering the necessary evidence and interpreting it to support the decision made (Cook *et al.*, 2015).

### 3.6 Validating the use of a language assessment tool

The relevance of language assessment tools lies in providing data in the form of scores used to make claims about the language performance of individuals (M. Kane, 2013). Such claims need to be evaluated through a validation process. According to Bachman (2015), in real-world language assessment, “all validation is local” (p. 4) as it is related to “a *particular* group of test takers, and in a *particular* setting” (ibid) (emphasis in the original).

Validation can be done through an argument-based approach (Bachman, 2005; 2010; Cook *et al.*, 2015; Kane, 1992; 2013) to validation. The argument-based approach to validation is rooted in Toulmin’s (2003) argumentation model (Chapelle & Lee, 2021), which consists of a sequence of arguments starting with evidence and facts that serve as grounds for a claim. The claim and grounds, which could be data, are linked by warrants (Toulmin,

2003). These three elements, claim, grounds and warrants, are the basic unit of argument concepts that have provided a framework used across various disciplines (Chapelle & Lee, 2021).

The most influential works on an argument-based approach to validation have been presented by Kane (1992; 2013), Kane, Crooks and Cohen (1999) for the educational assessment community in general and by Bachman (2005; 2015), Bachman and Palmer (2010), Chapelle (2008; 2011; 2021), Chapelle, Enright and Jamieson (2010), Chapelle and Voss (2013), and Knoch and Chapelle (2018) for the language testing circle. In general, an argument-based approach to validation consists of a logical argumentation chain specifying the assessment tool's characteristics from which test scores will be drawn based on performance. These arguments allow test developers to state how test results will be interpreted and the uses these test results will be given (Bachman, 2005, 2015; Bachman and Palmer, 2010; Kane, 1990, 1992, 2013; Kane et al., 1999).

Drawing on language assessment work presented by Chapelle (2008), Knoch and Chapelle (2018) and Knoch and Elder (2013), a model for the assessment for professional purposes is presented by Knoch and Macqueen (2020). Assessment for professional purposes is an area of interest for the present study. The models presented by Kane, Bachman, Knoch and Macqueen are discussed below as they inform the current study.

### 3.6.1 The Interpretation/use argument

Kane (1992) argues that validity is not about the test nor the test scores, but it is related to the way test scores are interpreted and states that interpreting test scores is to explain their meaning and implications. Drawing on earlier work on validity argumentation, Kane (1992; 2011) presented the interpretive argument (IA) approach for score interpretation. This approach consists of a series of arguments based on inferences and assumptions within the interpretation and use of test scores (M. T. Kane, 1992). The initial articulation of the IA emphasised interpretations, while uses seemed to be undermined. Therefore, the framework was revisited to become the "interpretation/use argument" ("IUA") (M. Kane, 2013). The purpose of the IUA is to "lay out a rationale for whatever claims are being made by the interpretation and use" (ibid., p. 9). The two-step process of this framework begins by stating the claims. These claims will build the IUA. The following step is to evaluate the claims. This is done through a validity argument, which assesses the coherence and



plausibility of the argumentative chain within the IUA. The IUA includes inferences for the evaluation, generalization, explanation, extrapolation, and utilization of test scores (ibid.).

The development of the IUA involves articulating the claims and the rationale supporting them. This argumentation is done by closely looking at the chain of inferences (of the claims being made) that link the observed test performance to the interpretation and use of scores. Claims, which are supported by warrants, are also analysed. Warrants frequently use evidence (known as backing) to legitimize the claim. The assumptions on which the claims rest are also assessed.

The validity argument, which is the evaluation of the claims made, that is, the evaluation of the IUA, will be as valid as the IUA is thorough and coherent. The network of arguments, claims, inferences, assumptions, warrants and backing must be plausible, clear and coherent (M. Kane, 2013).

Xi and Davies (2016) argue that the primary purpose of the work done by Kane “is to provide a generalized framework and mechanism for test validation rather than to explicate the exact chain of validity inferences” (p. 64). That is why Johnson (2011) argues that “constructing and evaluating an interpretive argument is still a complex and time-consuming task” (p. 36).

The testing community has discussed the argument-based approach to validation presented by Kane (1990; 1992; 2013) and Kane and associates (M. Kane et al., 1999) and provided grounds for undertaking empirical research. Sireci (2013) highlights the relevance of Kane’s contribution (2013), though he argues in favour of simplifying the approach. Lane (2014) used this approach to look into the outcomes of assessment programs, and Youn (2015) used it when exploring the evaluation of pragmatic competence in interaction in an EAP setting. Knoch and Chapelle (2018) explored the rating processes based on Kane’s model, proposing this aspect of assessment can also be part of the validity argument.

In 2008, Chapelle, Enright and Jamieson carried out research to develop a validity argument for the TOEFL iBT® using Kane’s argument-based approach to validation. While using the framework, they added a new inference. The domain description inference at the beginning of the chain of arguments was thought to add support to the argument that observed performance of the TOEFL iBT® tasks are representative of the domain and candidates possess the knowledge, abilities and skills required in an English medium

university. Several papers have been written reporting on different aspects of this research (Chapelle et al., 2010; Chapelle, 2011; Chapelle & Voss, 2013).

### 3.6.2 The Assessment Use Argument

The Assessment Use Argument (AUA) was proposed by Bachman in 2005, drawing on existing (M. T. Kane, 1992; M. Kane et al., 1999; Mislevy, Steinberg, et al., 2003) argument-based proposals for validity.

The initial proposal of the AUA differed from the IUA proposed by Kane in that Bachman furthers the interpretive argument by linking it to the decisions made based on interpretation (uses) (L. F. Bachman, 2005). This second argument, the utilization argument, incorporates the consequences of test use and other aspects not included in other models, such as test fairness and ethics (Johnson, 2011).

This initial AUA proposal comprising a validity argument and a utilization argument evolved “into a single overarching validity argument” (Johnson, 2011, p. 42), where four claims (discussed below) constitute the backbone of the validity argument (L. Bachman & Palmer, 2010; L. Bachman & Damböck, 2018). The AUA presents a unified framework that guides test development and test use, as opposed to other approaches that deal only with test validation, such as the argument-based approach (M. Kane, 2013) or with test development, such as the Evidence-Centred Design approach (ECD) (Mislevy & Yin, 2012; Xi & Davies, 2016). The test design process within the AUA is divided into five stages: initial planning, design, operationalization, trialling, and assessment use. These stages are shown in Figure 3.2 below.

Each stage is continually revised to ensure the decisions made at each stage align with those previously made. The two-way arrows between the AUA backing and the stages in test development reveal the iterative nature of the process until the interpretations stage is reached.

The test design approach of the AUA for test developers begins by considering the intended outcomes and the consequences of test use and works backwards, thinking about the steps to follow and decisions to be made during the test development process to arrive at the expected beneficial consequences (Xi & Davies, 2016).

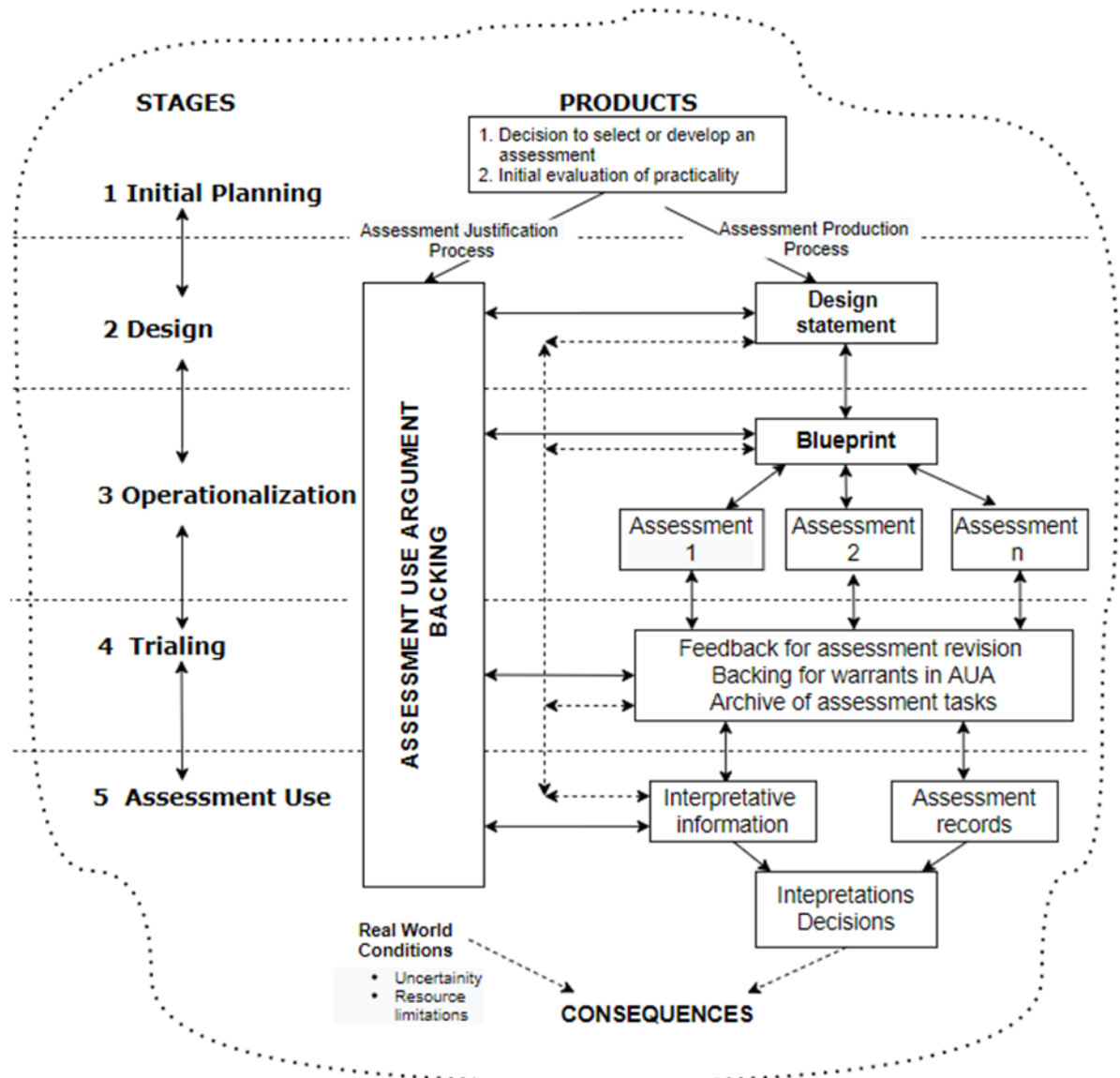


Figure 3.2 Assessment development and use, stages 1-5. Bachman and Palmer (2010)

The starting point in evaluating an AUA is looking at the claims made about the performance of test takers (Xi & Davies, 2016). Bachman and Damböck (2018) consider assessment justification is the process to follow “to *demonstrate* the extent to which the intended uses of an assessment are justified” (emphasis in the original) (p. 29). They compare such process to the one followed by a lawyer, presenting arguments to convince a jury or a judge in a legal trial. In building a case, the lawyer may argue in favour of the defendant's innocence. His argument will consist of a series of claims about or related to

the defendant. The lawyer presents arguments to convince a jury or a judge in a legal trial. The lawyer will present evidence to support the claims being made.

In the same way, in assessment justification, an argument is presented. This argument will consist of a series of claims supported by backing. Backing is the data or information supporting the claims about the assessment. The intended uses of the assessment of interest are justified through the claims made and backing presented (L. Bachman & Damböck, 2018).

The assessment development and assessment use stages are both guided by the four claims: consequences, decisions, interpretations, and assessment. These stages are shown in

Figure 3.3 below. When the decision to use a test has been made, the initial planning considers the beneficial consequences the test is expected to bring (consequences). The next step downward looks at the decisions that need to be considered to achieve the desired beneficial consequences (decisions). This step leads to identifying the language abilities that are relevant for the decisions that will be made and the kind of data that needs to be collected to be interpreted as indicators of the language ability of interest (interpretations). Finally, the type of performance that will be relevant to be interpreted as an indicator of the language ability of interest is determined. Assessment tasks are determined based on the type of performance we are interested in observing that represents the language abilities identified (assessment). Recording assessment performance is also determined based on language ability interpretation and assessment tasks.

The first claim refers to consequences that ought to benefit all interested parties. This claim specifies the consequences not only of using the test but also of the decisions made. Consequences are expected to be of no harm to stakeholder groups, mainly to test takers and teachers. The three aspects of this claim that can affect test takers consider their feelings, motivation, and decisions. Positive consequences can be promoted by asking test takers about “their perceptions of the assessment and assessment tasks” (L. Bachman & Palmer, 2010, p. 107). Teachers may be affected by the use of a test due to washback. Assessment use also has an impact on educational institutions and society.

The second claim focuses on decisions expected to be value-based and comply with legal requirements. Decisions are expected to be equitable. This claim addresses the social aspects that come into play when assessments are used to make decisions about individuals within society.

Claim 3 focuses on interpretations and seeks to ensure they are meaningful, impartial, generalizable, relevant, and sufficient. The components of the third claim relate test content and test tasks to the TLU domain and the neutrality of interpretations. Meaningfulness refers to the transparency of the information provided to stakeholders about the test aim and how it is disclosed. Within the aspect of impartiality, issues of fairness are addressed. Generalizability, relevance, and sufficiency pertain to the correspondence between test tasks and TLU tasks. Within generalizability, the target language use domain is described.

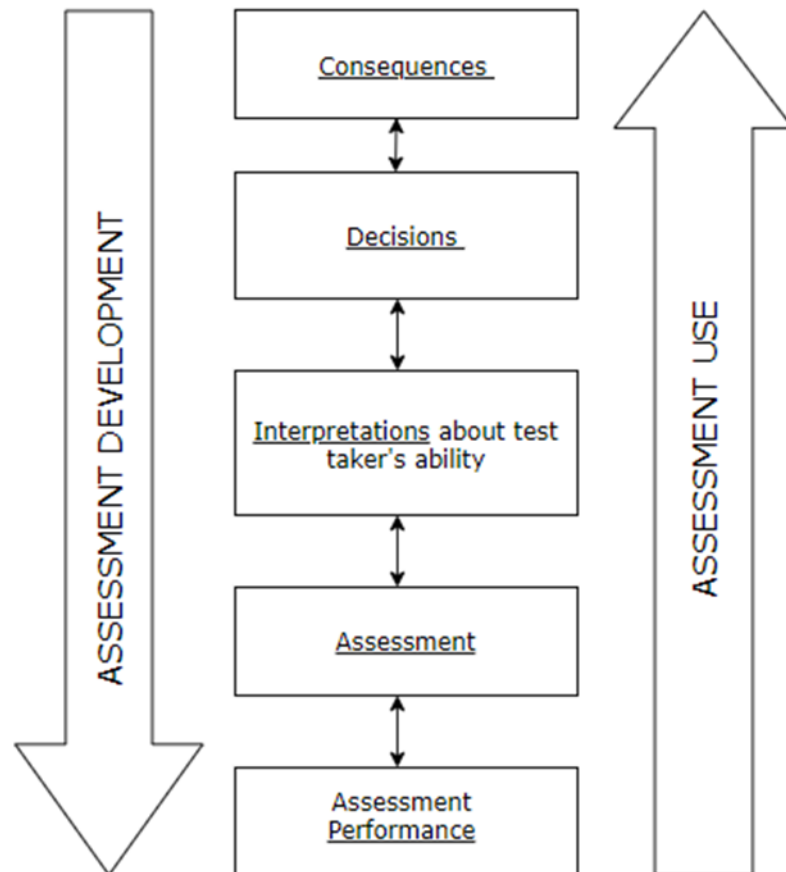


Figure 3.3 Inferential links from consequences to assessment performance (L. Bachman & Damböck, 2018) adapted from (L. Bachman & Palmer, 2010)

Claim 3 focuses on interpretations and seeks to ensure they are meaningful, impartial, generalizable, relevant, and sufficient. The components of the third claim relate test content and test tasks to the TLU domain and the neutrality of interpretations. Meaningfulness refers to the transparency of the information provided to stakeholders about the test aim and how it is disclosed. Within the aspect of impartiality, issues of fairness are addressed. Generalizability, relevance, and sufficiency pertain to the correspondence between test tasks and TLU tasks. Within generalizability, the target language use domain is described.

Claim 4 addresses the consistency of the results and ensures that these are not affected by aspects such as the version of the test taken, the time the test is taken, the way the test is scored, or the rater or raters involved.

The concern raised by Messick (1995) about the use of tests and the social consequences are included in the AUA. Warrants pertaining to test fairness, which refers to “equitable treatment” and the “absence of bias” (L. Bachman & Palmer, 2010, p. 128), are integrated into support of the four claims contained in the framework. A particular strength of the AUA is the “clear and logical grouping of validity inferences” (Xi & Davies, 2016, p. 66). This characteristic facilitates the gathering of evidence in support of Claim 3, which refers to the interpretations based on test results. Decisions and consequences are separated. These are also considered a strength of the AUA since not all test consequences are related to score-based decisions (*ibid*).

Xi and Davies (2016) argue that replacing specialized vocabulary with more intuitive terminology to refer to concepts and considerations related to language testing and validation can facilitate communication with a general audience. In this regard, they also argue that confusion can arise due to using new terminology to replace the words used to refer to existing concepts. Furthermore, they state that although collecting evidence to support claims is more transparent than with the IUA or the ECD, the amount of evidence that needs to be collected can be overwhelming for teachers and practitioners.

The AUA proposal has been used in several research studies, such as the review of the Pearson Test of English Academic (Wang et al., 2012), where the purpose of the study was to examine the degree to which the use being given to the test could be justified to stakeholders. In another study, Schmidgall, Getman and Zu (2017) used this model to

explore the design of a screener test for the reading component of the TOEFL. The AUA framework has also been adapted by Mann and Marshall (2010) for the context of bilingual sign education. The authors state that they were able to adapt the framework to meet their needs to improve the identification of some of the difficulties that deaf children face when starting their sign language education at an older age. It was also possible for them to address these difficulties.

### 3.6.3 The Language Assessment for Professional Purposes model

For Knoch and Macqueen (2020), validity is about interpreting results fairly and meaningfully for a particular language use situation and specific uses. Their concern with the interpretation and use of tests, drawing on the work of Kane (1992; 2013) and Bachman and Palmer (2010), is applied in the development of the Language Assessment for Professional Purposes (LAPP) model. While agreeing with the position of Bachman and Palmer (2010) and consistent with other proposals presented by Fulcher and Davidson (2007) and Mislevy and Yin (2012) to address testing consequences as the starting point for any test validation project, they draw their proposal from other comparable language assessment models: Chapelle, (2008), Knoch and Chapelle (2018) and Knoch and Elder (2013). The frameworks presented by Knoch and Chapelle (2018) and Knoch and Elder (2013) include the test consequences inference, an element not included in the framework previously presented by Chapelle (2008). However, it is included in the AUA discussed above.

Knoch and Macqueen (2020) state that while interpretations and decisions (uses) are linked for practical purposes, the former is often associated with more general aspects, while the latter is often more specific. In LAPP, decisions are workplace-related.

The framework they present is a combination of proposals put forward by Chapelle (2008), Knoch and Chapelle (2018) and Knoch and Elder (2013). For Knoch and Macqueen, the starting point for a validity argument framework is the definition of the Target Language Use (TLU) domain. In language testing, the TLU domain is the 'test construct', defined by Knoch and Macqueen (2020) as "the term used to refer to whatever is measured by an assessment" (p. 39). By describing the domain, it is possible to design test tasks that are representative of the real-world tasks that test-takers are expected to encounter. They present a test development cycle that begins with the needs analysis, which results in the

design statement, test blueprint and specifications, which can be found in Figure 3.4 below. The process continues with the development of the test prototype, which is then pre-tested, evaluation criteria developed and then trialled. Standards are established before the test enters the operational phase.

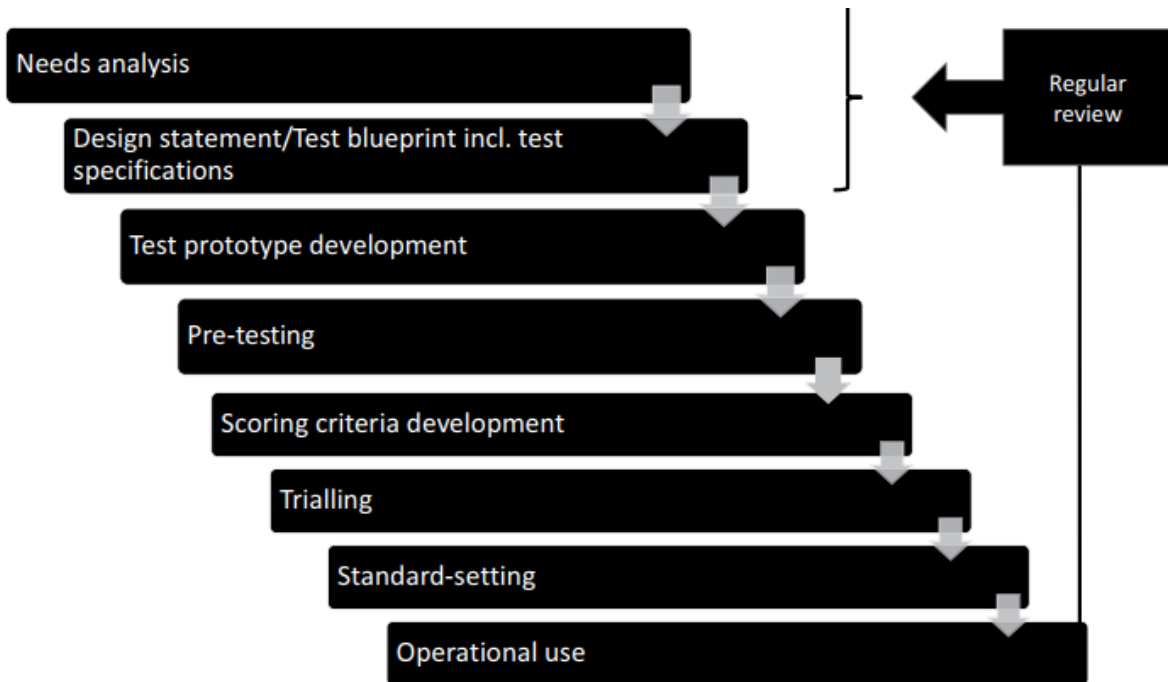


Figure 3.4 LAPP test development cycle. Knoch and Macqueen (2020).

The validation of the developed test is done through a chain of inferences. The LAPP proposal begins with the description of the domain, followed by evaluation, generalization, explanation, extrapolation, decisions, and consequences. This proposal is shown in Figure 3.5 below.

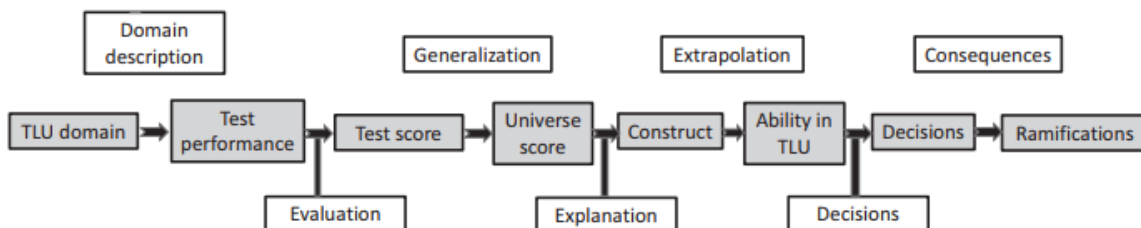


Figure 3.5 Inferences in a validity argument. Knoch and Macqueen (2020).



Knoch and Macqueen state they present a generic proposal because, just as contexts vary one from the other, test purposes and the types of interpretations intended to be claimed based on scores are dissimilar.

### 3.7 Tasks in language testing

Tasks in language assessment have been used since the early 1900s due to the expanded use of task-based language teaching. As in ESP language teaching, tasks present learners with familiar activities in their professional environment (Huhta et al., 2013, p. 9). In language assessment, Wigglesworth (2008) argues that “tasks are used to elicit language... (that)... reflect the kind of real-world activities learners will be expected to perform” (p. 112). In the former situation, learners need to learn how to accomplish tasks in another language. In contrast, in the latter, learners demonstrate they can fulfil tasks pertaining to the context of interest. Besides the expected relation between task-based language teaching and task-based language assessment, using tasks in language assessment will likely result in a more positive ‘washback’ effect on classroom practice (M. Long, 2015). Task-based language assessment is not always related to classroom instruction (M. Long, 2015), as “tasks offer a fundamental, ... foundation for useful language assessment” (Norris, 2016, p. 230). Task-based language assessment is frequently used to certify the abilities of individuals to carry out real-life activities. It measures the extent to which an individual's language ability conforms to “the standards required of a computer programmer, a nurse, an airline pilot, a marksman, and so on” (M. Long, 2015, p. 330). The strength of task-based approaches to language assessment is based on their potential to connect successful test results to effective language use in real-life contexts (R. Ellis et al., 2019). Task-based language assessment (TBLA) seemed to offer “transparent indications of the extent to which language learners can communicate in a target language” (Norris, 2016, p. 232).

All assessments involve the elicitation of data to support a decision-making process (R. Ellis et al., 2019). The degree of estimation required for data interpretation is related to the type of assessment used, as, according to Long (2015), “discrete-point tests of linguistic knowledge reveal little or nothing about the ability to perform real-world tasks” (p. 336). The use of tasks in language assessment “emphasizes the performance of target tasks (as

opposed to the demonstration of knowledge about the language)” (M. H. Long & Doughty, 2009, p. 582).

According to Brown (2010), in language assessment, unlike real-world tasks, the range of language to be used by the test taker is generally restricted to the language related to the tasks through which the specific sample of language use is elicited. The use of tasks for language assessment purposes is organized around the fulfilment of the language use task rather than measuring elements of language such as grammar or vocabulary (Shehadeh, 2012).

Tasks have been playing different roles in task-based language assessment. Norris (2016) discusses how tasks have been used to establish performance standards by establishing an expected level of language ability when carrying out activities within defined contexts. Policymakers or professional bodies, such as the CLB (2012), use these for certification purposes. Tasks within the CLB are set “to reflect contexts that are relevant and meaningful to learners within community, work and study settings” (p. ix). Another language policy document that makes substantial reference to tasks is the CEFR (COE, 2001). Huhta and collaborators (2013) claim that due to “(T)he action-oriented approach adopted by the CEFR...tasks...(are) central to language learning” (p. 9), hence for language assessment.

Tasks used for proficiency assessment purposes aim to predict the degree to which learners possess the language abilities that will be required in a particular domain of language use, such as the academic environment. The International English Language Testing Service (IELTS) Academic (International English Language Testing System (IELTS), 2019) and the Test of English as a Foreign Language Internet-based test (iBT TOEFL) (ETS, 2011) are examples of task-based tests frequently used to make decisions regarding university admission. The IELTS General Training test is used for professional or immigration purposes.

Tasks used “in the certification of profession-related language skills or competencies” (Norris, 2016, p. 235) focus on successful communication within a given domain, replicating tasks and task conditions typical of the workplace of interest. In this sense, a task may consist of a series of stages, such as listening to a recording, followed by a request to explain the recording in detail and express an opinion about the situation described.

Finally, tasks are also used in educational assessment, in the interest of alignment of instruction and assessment, especially in contexts where language teaching aims to develop communicative competence, language use ability and so on (Norris, 2009). The aim of TBLA within the academic setting is to establish a clear link between the language skills required for study purposes or in a future working environment and those taught in the classroom. Students are evaluated on their language performance (Fischer, 2011).

### **3.8 Proposal for a Local Context-Sensitive Language Test (LCSLT)**

The LCSLT begins when the decision to develop a test is made. This decision implies that using an existing test is not possible because its result interpretation is poorly or unrelated to the context of our interest (L. Bachman & Palmer, 2010).

The starting point to establish the intended outcomes in an LCSLT is a needs analysis, as it is in the LAPP model (Figure 3.4 above). However, for the LCSLT, the needs analysis is referred to as a domain analysis because a needs analysis can be carried out without the analysis of the domain. Using the name 'domain analysis,' the analysis of the domain is placed at the centre. This central position does not mean that other types of analysis. i.e., a means analysis cannot be carried out; it just emphasizes the importance of the domain analysis.

The AUA approach for test design begins by looking at the intended outcomes. For an LCSLT, the intended outcomes are directly related to and emerge from the context in which the target language will be used. The intended outcomes are identified through domain analysis. For the present study, a domain analysis can be defined based on the definition discussed in 2.4.2 above. A domain analysis would be the systematic collection and analysis of all the information necessary and available for the definition and validation of the content of a context-sensitive higher education exit language test.

The "second generation NA" discussed at the end of section 2.5 above focused on identifying the language use activities (or tasks) students needed to carry out in a workplace context. This study took the same approach (focusing on identifying the language use activities) for the analysis of the domain. However, as opposed to the "second generation NA" discussed in Huhta et al. (2013), the LCSLT proposal uses work-

related international language descriptors to carry out the domain analysis rather than seeking to compare findings to international language descriptors as Huhta et al. did. This study considers that work-related language descriptors are more closely related to the TLU domain than those for general English purposes. It is contended that participants can relate work-oriented language descriptors to their own perceived or current workplace language needs better than using language descriptors not associated with the job activities they carry in the target language. It is considered that findings can be related and compared to other language requirements when international language descriptors are used to define assessment outcomes. It also provides transparency to all stakeholders regarding what the test takers can do in English in the workplace.

Most English for specific purposes course development published research focuses on what teachers consider candidates require without explicitly referring to the language level in terms of international language descriptors. Relating course or test contents to standardized language descriptors may be relevant for many institutions (Knoch & Macqueen, 2020). Higher education frequently expresses the language level required to enter university or to graduate in terms of the levels of language descriptors (see Appendix Table 1 below), such as the CEFR. A study by Elder and Knoch (2016) reports using the CEFR language descriptors as a reference point to identify the language level required for the National Accreditation Authority for Translators and Interpreters (NAATI) accreditation in Australia. Domain experts were asked to identify the levels to determine the minimum language proficiency for the sub-levels of translators and interpreters.

Knoch and Macqueen (2020) argue that in assessing language for professional purposes (LAPP), “using such general frameworks (as the CEFR) is problematic because LAPPs and the more general language frameworks draw on different stated constructs” (p. 151).

The proposal for the design of a local workplace context-sensitive test uses the equivalent to B1 level work-domain language descriptors from the ALTE (2002) and the Work-Ready (CCLB, 2015) language descriptors from the Canadian Language Benchmarks (CLB). These will be used to identify the language use activities that are relevant for professionals in the local context, as discussed in 4.5.3 below. As both language descriptors were benchmarked to the CEFR (ALTE, 2002; North and Piccardo, 2018), the respective levels (ALTE level 2 and WR level 5) equivalent to the B1 level of the CEFR are used. Using these descriptors allows the study to be carried out at the level established by the university but workplace-oriented. Having test results expressed through international language

descriptors informs all stakeholders about the knowledge, abilities and skills certificate holders can do in a real-life language use situation. Their equivalence to the CEFR makes their interpretation and understanding more straightforward and more transparent for all interested parties.

The design of a test that meets the characteristics of the local context incorporates relevant language use activities and favours a jagged skills profile (see 1.6.1 above). This language profile allows more efficient use of time and resources from all relevant parties, as all efforts concentrate on what is pertinent.

### 3.9 Test task design patterns

The generic template for construct test task development presented by design patterns can be modified to suit a variety of needs (Wei et al., 2008a). Design patterns offer an argument-based approach to test task design that aligns with the argument-based approach to test validation. Although they were initially presented for the assessment of scientific inquiry (Mislevy *et al.*, 2003), design patterns were adapted for language assessment by Wei and collaborators (2008). Design patterns enable the mediation of the features and supporting assessment argument to be assessed. This assessment is done through a scheme that encompasses the elements “that guide(s) task creation and assessment implementation” (Mislevy, Hamel, et al., 2003, p. vi).

The difference between test specifications and design patterns is that the former provides “a detailed blueprint for writing tasks with specified properties that suit the purposes, constraints, and resources of the particular testing context” (Wei et al., 2008b, p. 5) while the latter focus “on assessment arguments at a narrative level and are organized around aspects of language use” (Wei et al., 2008b, p. iv). The aspects of language use, which are the focus of assessment, can be described differently according to the purpose or situation of interest.

A design pattern offers an organisational structure of the aspects that need to be considered when assessing language proficiency. A design pattern is intended to be a non-sophisticated scheme for task design that incorporates the relevant information and justification of the characteristics of the test task. However, some technical aspects of task design are necessary to describe how competences will be assessed. Rather than having two separate documents for test task design, a combination of elements from design

patterns (Wei *et al.*, 2008) and some relevant technical features are combined into the design pattern test-task scheme proposal presented in Table 3.1 below; a design pattern for an LCSLT.

The framework that suits the LCSLT task design requires including information that allows the articulation of the relevant features related to the measurement of the competences for one or more disciplinary groups. Some features of the language assessment design pattern presented by Wei and collaborators remain in the LCSLT design pattern example presented below (Table 3.1); these are indicated by an asterisk (\*).

The *Focus* specifies the type of language to be assessed, whether for general purpose or related to a specific domain of language use. The *Date* indicates when the design pattern was designed. The *Classification* summarizes the focus, the skill(s) involved, the year it was designed and the version. The *Title* identifies and distinguishes each design pattern. The *CEFR reference* box specifies the language level of the competence of the design pattern and the specific language activity addressed. The *Competence* box states the specific competence the design pattern aims to assess. The type of competence defines whether the task will be discipline-specific or it can be used for more than one disciplinary group. The *Input* and *Output* section has two components: mode and language. The *input mode* specifies how the language is presented to students. It could be written, spoken or through a computer-mediated audio or video recording. The *language* box indicates the language in which the task is presented and the language expected to be used by test takers when recording their responses. The *Interaction pattern* section specifies if the task is carried out individually or in pairs or trios. This specification applies to oral interaction assessment only as the rest of the tasks are responded to individually.

The *Target audience* indicates which disciplinary groups the task is aimed at. Task designers can refer to this data during the task validation process (to seek feedback from expert insiders).

The *Summary* box presents a general outline of situations of language assessment that test takers may be presented with and the knowledge, skills, and abilities (KSA) they are expected to demonstrate. These KSAs could be “to read or listen to a text” and the type of action they are expected to carry out. Test takers may be required to interact with an oral examiner, record their responses on the computer, fill in a form or choose an image.

The *Rationale* box states the argument that justifies the relation between the test task presented in the design pattern and the language required within the TLU domain.

The *Focal KSAs* specify the main knowledge, skills and abilities test takers must demonstrate. These could be understanding written instructions to operate discipline-specific equipment or spoken directions to reach a destination. Discipline-specific lexis may be specified here. KSAs could be either discipline-specific or general English. When other KSAs, such as additional discipline-related lexis, are required to fulfil the task, they are specified in the *Additional KSAs* box. In such a case, it is necessary to specify why this lexis is expected to be known by test takers. The presence of additional KSAs should not compromise the validity of the task. An effort should be made to reduce or eliminate this possibility whenever this is identified.

The *Characteristic features* define the specific aspects of the task that aim to elicit the expected performance or exhibit the use of discipline-related lexis or the KSAs of interest. To the extent possible, tasks should emulate real-life language use activities. The type of stimulus material is stated here. While direct measurement of competences is preferred, it is necessary to consider that practicality may call for the inclusion of objectively marked items to be used. The way tasks are marked is indicated in the *Type of measurement* box.

*Potential work products* are the specific performance expected to be observed, what the test taker says or does, whether the test-taker completed a diagram or answered a multiple-choice item. The *Potential observations* refer to the expected output mode and detail what test takers are expected to do to demonstrate fulfilment of the specified language competence—the specific aspects of performance required for competence assessment.

The Potential rubrics refer to the way the task is assessed. Whether the task is objectively or subjectively marked is indicated here. Potential rubrics may provide a link to the marking criteria for language produced by test takers, either oral or written. These could also indicate when an answer is considered correct, incorrect, or incomplete (if half point(s) may be awarded). This section indicates the quality of the inferences desired.

Finally, the *References* box presents the sources of the input materials used, whether in-house materials, links to internet webpages or any other source used.

The argumentative nature of design patterns favours aspects of the construct to be assessed, the context, the task type, performance, and assessment criteria to be organized. The narrative filling the slots of each attribute builds into the justification of the assessment argument of each task and the test in general.

Design pattern for language assessment					
Focus	Type of language to be assessed	Date	Month/ year	Classification	Focus, skill(s) involved, year of development, version
<b>Title *</b>	The label with which the design pattern is identified.				
<b>CEFR reference</b>	Level of the CEFR and general label to language function				
<b>Competence</b>	Competence from the CEFR				
<b>Input/output mode</b>			<b>Input</b>	<b>Output</b>	
	<b>Mode</b>		Written, audio, video, and spoken	Written, spoken	
	<b>Language</b>		English	English/Spanish	
<b>Interaction pattern</b>	Is the task answered individually, or is it in pairs or trios? (for oral tasks)				
<b>Target audience</b>	Relevant disciplinary group(s).				
<b>Summary *</b>	Task type, the KSAs they are expected to demonstrate				
Rationale *	Basic justification of how the task presented is relevant for the target audience				
Focal KSAs *	The language-related KSAs. Discipline-specific lexis used (if applies)				
Additional KSAs *	Other knowledge, skills or abilities required to complete the task.				
Characteristic features *	Specific characteristics of tasks. Characteristics of the input material.				
Type of measurement	Is the construct (competence) being measured directly or indirectly?				
Potential work products *	Item type, output format, expected performance.				
Potential observations *	Expected output. What are test takers required to do to demonstrate the specified language competence? Behaviour or performance				
Potential rubrics *	Answer key or rubrics to assess task.				
References *	Sources of materials used as input.				

Table 3.1 Design pattern attributes and definition, adapted from (Wei et al., 2008)



The model presented above for LCLSLT design patterns may be modified to fit the specific needs of any assessment situation, just as the original design pattern proposal was adapted to meet the needs of an LCLSLT. The LCLSLT design pattern proposal may incorporate other relevant attributes, as those that are not required may be eliminated. The design pattern format aims to meet the needs of the specific assessment situation.

### 3.10 Summary

This chapter discussed the relationship between a test and the purpose which motivates its development. Tests are developed to meet the demands of a specific situation involving particular language use needs within a specific context considering an identified group of test takers. The purpose of such tests is to make inferences about the language ability of individuals to use the language assessed in situations outside the testing session. The extent to which test tasks emulate the target language use situation tasks is the extent to which inferences can predict language performance in non-test situations.

Given that language use varies with the context, a test may be aimed at assessing language for specific purposes, such as language for academic purposes or work purposes. A test that aims to measure language for more than one purpose will necessarily reduce the language samples that may be collected for each purpose, reducing the relation between test tasks and the specific target language use situation.

Tests used for purposes other than those initially intended for may prove unsuitable to make relevant predictions of language use in the target language use situation.

The argument-based approach to test validation provides a framework for analyzing the relation between test tasks and the target context of language use. It also enables the analysis of the effect test development and test use have on individuals and other stakeholders. However, a framework that meets the needs of a local context requires the inclusion of other elements. The proposal for an LCLSLT is considered to meet these needs. It is also considered user-friendly for the non-expert in language assessment development, as could be the case for many teachers worldwide who need to develop tests relevant to making high-stakes decisions.

Design patterns organize relevant information required to build test tasks, connecting stimulus material with expected performance and target language use situations within an argument approach to test development and validation.

## 4 Research Methodology

### 4.1 Introduction

This chapter describes the research methodology used to address the purpose of the present study. First, the research design, “the methodological structure of a study” (Rose et al., 2019, p. 2), which was adopted to answer the research questions, is presented, followed by the data collection methods. It then describes the research participants and the sample size, after which the selection is determined. The following sections discuss the design of the data collection tools, ethical considerations and the role of the researcher.

The data analysis process for each instrument is described. Insights from the analysed data will answer the research questions and provide evidence that will be used later in Chapter 7 in the argument-based approach to the test validation proposal for the local professional context.

The chapter ends with a summary of the aspects covered.

### 4.2 Research design

The research design for the present study will represent the approach taken (Rose et al., 2019) to identify the English language needs of professionals in the local working environment. It is, therefore, necessary to consider the local context in which the language is needed. The case study research method is considered an appropriate approach because, according to Stake (1995), the case study allows for the understanding of the events taking place in significant circumstances, thoroughly and in the context in which they are set. A case study can help to understand “the particularity and complexity of a single case” (Stake, 1995, p. xi) or multiple cases. Rose et al. (2019) and Duff (2020) note that case study research in applied linguistics differs from case studies in the social sciences. While in social science, a case refers to a person, in applied linguistics, this can be “a class, a curriculum, an institution, a speech community, a piece of text or a collection of text types” (Rose et al., 2019, p. 7). A case study in applied linguistics allows researchers to analyse and illustrate different angles of a highly contextualized phenomenon (Duff, 2020). This research looks at the local

professional working environment within the city in which it is located. By considering the entity under study holistically, it will be possible for the researcher of the present study to acquire a new understanding of some topics (P. A. Duff & Anderson, 2015), such as the language needs of the local professional working context addressed in this study. An attractive aspect of case study research is that it is possible to consider the many aspects influencing the entity of interest (ibid). This characteristic, however, may also be seen as a disadvantage. Its highly contextualized nature could make generalization difficult unless the transparency of the description of methods and analyses allows its replicability (Rose et al., 2019).

This study presents a Main Research Objective (MRO) from which four research questions were derived. Data analysis results will be used to explore the extent to which an argument-based approach for language test validation provides a flexible framework for designing a local context-sensitive university exit test. The MRO is specified as follows:

**MRO.** What are the specifications of a high-stakes university exit test which is sensitive to the local context?

The MRO requires the participation of three respondents: a) test takers, b) students and c) professionals in the workplace. The data collected from these three groups of participants contribute to a better understanding of the language needs in the target language use situation: the local working environment.

The first research question (RQ1) refers to the face validity of the test used. Test takers are asked for their opinion on the test used. They are asked if they think any changes should be made and, if given the opportunity, to decide which domain they want the test to focus on. RQ1 reads:

RQ1. From the point of view of test takers, what is the face validity of the current certification test in terms of content, test preparation, timing, and difficulty?

A closed-ended questionnaire (Q1) was designed (see 4.5 below) to answer RQ1. It was applied at the end of four testing sessions over a two-year period. Test sessions are held twice a year in late May and November. Some students were also interviewed after completing the questionnaire. The purpose of this brief semi-structured interview

(see 1.1.1 below) was to ask them which domain they thought the test should focus on and the justification for their choice.

The data collected from the test takers provided a picture of the face validity of the test used.

The second research question (RQ2) refers to the language needs of the target context of language use. It is necessary to understand the language requirements of the work environments that most university graduates are likely to join. For this purpose, two groups of informants were considered relevant: students and professionals. The second research question reads:

RQ2. To what extent are work-related language needs expectations of the students aligned with current professional language needs within the workplace?

The answer to RQ2 requires comparing the results collected using two questionnaires: Questionnaire 2a (Q2a) and Questionnaire 2b (Q2b). These questionnaires aimed at identifying the most frequently used work-related competences. The content of Q2a (see 4.5.3.1 below) and Q2b (see 4.5.3.2) is the same. Their difference lies in the group to which they are applied. Q2a is applied to students in their last year of studies, and Q2b is applied to professionals in the workplace. The result of their application provides data from a different perspective based on their knowledge of the workplace at the time the questionnaire was applied.

On the one hand, the students were asked about their expectations regarding the language required in their future work environment. Responses from students may be considered to be based more on expectations than on experience. However, since participant students are in their last year of studies, they have already spent one or two semesters (depending on the discipline) in the workplace for their practicum (time required to spend in the workplace as part of the requirements of their study programme). However, their expectations may not match current language needs in the workplace due to the type of activities they were assigned during their practicum. As digital natives, students have a different view of the world than today's professionals. On the other hand, professionals are the best informants about current language needs in the workplace.

While the data collected from students may be more related to expectations or perceptions about the future, workplace experts provide data on the current real-world needs within the work environment.

The language competences that are more relevant in the workplace are identified by analysing data collected through closed-ended questionnaires. However, this data is insufficient to design a test. More detail regarding the use of the competences is required. This need raises a third research question. Research question 3 (RQ3) is:

RQ3. What are some of the typical domain-related tasks that employees use?

To answer RQ3, workplace professionals (from the same disciplines as students and test takers), where possible, were asked to answer the questionnaire in the presence of the researcher. The aim was to collect examples of activities or situations where professionals need to use English in the workplace. This instrument is referred to as Questionnaire 3 (Q3) and Supplementary Interview (see 4.5.3.3 below). After obtaining consent from participants, the session was audio recorded. Very few participants did not consent to their voices being recorded. In such cases, they answered the questionnaire out loud while the researcher took notes. Whether recorded or not, participants provided specific examples of language use when an option other than 'never' was selected. The hard data from this tool was added to the quantitative data collected in Q2b, and the samples of workplace language use were analysed qualitatively.

This data is used to inform the design of a high-stakes university-exit language test that is sensitive to the local work context. The need to design a test gives rise to a fourth research question, which aims to identify an approach to language testing that makes it possible to include relevant aspects for a context-sensitive test. This fourth research question (RQ4) is expressed as follows:

RQ4. To what extent does an argument-based approach to test validation offer a flexible framework to address specific testing needs?

This research question is answered by reviewing the literature and considering collected data that provide insights into the context and the language expectations of the test-taker population, their characteristics and opinions.

This study has been divided into three phases. The first two phases involve data collection, while the third explores using or adapting an existing framework for test validation. These phases are summarized in Table 4.1 below.

The aim of Phase I was to understand the perceptions of graduate test takers regarding the final test. This phase consisted of applying Questionnaire 1 (Q1) and a short semi-structured interview (between May 2015 and December 2016).

<b><i>Research design phases</i></b>	<b><i>Instruments applied</i></b>	<b><i>No. of participants</i></b>	<b><i>Group of informants (and place where the instrument was used)</i></b>
Phase I	Questionnaire 1 (Q1)	886	test takers (at the exit point after taking the exit test used)
	Voluntary comments (in Q1)	216	
	Short semi-structured interview	231	
Phase II	Questionnaire 2a (Q2a)	706	students in the last semesters of their undergraduate studies (on the school premises)
	Questionnaire 2b (Q2b)	612	professionals (in the workplace)
	Q3 + interview	93	professionals (in the workplace)
Phase III	Explore existing frameworks for test design and validation to be used or adapted to design a university exit test that is sensitive to the local context.		

Table 4.1 Overview of research design

Phase II addresses the language needs in the local workplace and consists of two questionnaires: Questionnaire 2a (Q2a) and Questionnaire 2b (Q2b). The first questionnaire of Phase II, Q2a, aimed to identify the future job-related language needs of students about to finish university studies. Q2a was applied between February and May 2016. The second questionnaire, Q2b, was applied to professionals at their

workplace in the local area where university graduates are likely to find employment. Q2b was applied between July 2015 and January 2017. Also, within Phase II, Q3 was applied. Q3 aimed to identify specific situations in which professionals of different disciplines need to use English in the workplace. Just as Q2b, Q3 was applied in the local region where university graduates are likely to find employment. Q3 was applied between July 2015 and January 2017. These participants provided examples of language use activities in the workplace during a semi-structured interview. Data collected and analysed in Phase II reveals the language needs within the local working environment.

The data collected aims to learn more about the local context. On the one hand, it is about test taker participants and their opinions on the current test practice. On the other hand, about the language needs of future and current specialists in the local work environment.

Finally, Phase III looks at the results of the analysis of data collected and explores an argument-based approach for test validation to determine whether it provides the flexibility required to design a test that is sensitive to the local context.

An overview of the research design, the instruments used, and the number of participants in each phase is given in Table 4.1 above.

### **4.3 Data collection methods**

Duff and Anderson (2015) position the case study within qualitative, interpretative approaches. This suggests that case study research designs are paired with qualitative data collection methods. Rose et al. (2019) argue that “data collection methods can be used within almost any research design, creating multiple possible combinations” (p. 12) in the search for data that provide insights into the area of interest. Therefore, case study research designs can draw on qualitative and quantitative data collection methods, or a combination of these methods, known as mixed methods research (MMR) which leverages the strengths of both qualitative and quantitative research approaches (Creswell, 2009).

Data collected with qualitative instruments are likely to be open-ended so that unanticipated data may emerge. On the other hand, data collected with quantitative



instruments provide answers to given aspects. Mixed methods arose from the idea that existing data collection methods, qualitative and quantitative, introduced their own biases and shortcomings to a study (Creswell & Creswell, 2018). Mixed methods research (MMR) is currently viewed “as a genre of inquiry that intentionally and systematically connects qualitative and quantitative methods to address substantive questions” (Meixner & Hathcoat, 2019, p. 52). This study falls under a so-called Fixed Mixed Methods Design since using both qualitative and quantitative data collection methods was planned from the beginning (Creswell & Plano Clark, 2011).

One of the benefits of a quantitative approach is the ability to include large numbers of participants in the study. The large amount of data collected is analysed using statistical analysis methods. This analysis focuses on understanding the behaviour of the variables involved. The software used to analyse the collected data is IBM SPSS (Statistical Package for the Social Sciences) (Dörnyei, 2007).

Some of the data collection tools used in this study are Likert questionnaires because they can generate quantifiable data (Dörnyei and Taguchi, 2009). Results from this type of questionnaire are considered reliable because they are not influenced by the researcher’s interpretation (Nunan, 1992). Quantitative data are considered “hard”, objective and focused on finding answers or drawing conclusions (*ibid.*). Their use in this study helped identify facts and characteristics of the relatively large populations that participated in it.

Even though quantitative analysis allows the identification of facts, it is not possible to understand the circumstances or situations behind the facts. Hard data does not explain, justify, or reveal the beliefs behind an answer. Qualitative data collection tools were used to fill the gap to understand the nature of the responses: an open-ended question at the end of a closed-ended questionnaire, a short semi-structured interview, and a supplementary interview. Considering that qualitative research “assumes that all knowledge is relative” (Nunan, 1992, p. 3), the results are influenced by several elements, making them partially true or subjective. Some examples of this were found when data from the interviews were analysed and compared to hard data gathered from the same group of participants.

A limitation of the qualitative approach is that it is not possible to include large numbers of participants, as is the case with quantitative research (Nunan, 1992). The non-

numerical data generated is considered soft compared to the hard data generated in quantitative research. However, Richards (2003) does not consider qualitative research to be soft, arguing that qualitative research requires developing appropriate skills, which are largely neglected in TESOL.

A better understanding of the data provided was gained using an interview supplementing the closed-ended questionnaire (Q3) with one of the participant groups (workplace professionals) (Dörnyei and Taguchi, 2009; Dörnyei, 2007). During this supplementary interview, expert insiders provided examples of the specific situations in which English was used or needed for their professional activities. The responses to the questionnaire fed the quantitative data, while the examples provided data that gave meaning to the situation.

The qualitative data gathered provided a “rich and sensitive description of events and participant perspectives” (Dörnyei & Taguchi, 2009, p. 10), which made it possible to learn about the context in which these participants interact. It was also possible to collect explanations or justifications that helped to understand previous answers and provide a more comprehensive view of the context (Cohen *et al.*, 2011; Dörnyei, 2007; West, 1994).

All collected qualitative data was analyzed using non-statistical methods, such as NVivo, a computer software package for qualitative data analysis from QSR International.

#### **4.4 Research participants, sample selection and size**

The needs analysis approach taken in this study could be viewed as a combination of viewpoints described by Brown (see 2.4.2 above). From the so-called democratic point of view, stakeholders are asked which domain they consider the test should focus on. They are also asked about the language use activities they think they will need to use more often in their later professional practice. The diagnostic perspective deals with the elements that test takers are likely to encounter in the future situation of language use, i.e. in the professional environment (J. D. Brown, 2016).

This study identified more than one target group of research participants (Dörnyei and Taguchi, 2009; Dörnyei, 2007; Nunan, 1992). Having more than one group of

informants makes it possible to confirm results, uncover unexpected perceptions, or enrich the understanding of the results (Dörnyei, 2007; Nunan, 1992; West, 1994), thereby providing valuable information for the study. Even if one of the groups is seen as the primary source of information, all the data collected provides rich and relevant information that feeds into the bigger picture of language use and language needs in the local professional work environment. This study's population is designed to examine consists of state university students, university test takers, professionals (Dörnyei and Taguchi, 2009; Bachman, 2004) or expert insiders. Knoch and Macqueen (2020) refer to the latter as domain insiders. These terms are used interchangeably to refer to this group of informants. Using *domain insiders* reminds the reader that this group has knowledge specifically related to real-world language needs within the domain of interest. On the other hand, using the term *workplace professionals* contextualizes the term with this particular study. These groups will be described after the sample selection procedure has been discussed.

#### 4.4.1 Sample selection process

Given the characteristics of the target populations, the sample was selected using non-probability (Cohen et al., 2011), cluster and quota (Dörnyei, 2007) sampling. Considering that the target populations are widely dispersed in the city of Durango, clusters of individuals were accessed within their study or working environments.

Questionnaires were applied to students in their classrooms when access to the school facilities was possible. Some Schools required that the questionnaires be left and applied by their personnel. Questionnaires were collected a few days after they were left at the school premises. Test takers were reached at the test session premise's exit point after finishing the test. Finally, workplace professionals were accessed at their working environments. Few questionnaires were applied online to domain insiders. For this purpose, the link was given to the participant or a company representative who asked professionals to answer the questionnaire.

Quota sampling observes the subgroups within the target population are proportionally represented in the overall sample (Dörnyei, 2007). Q2a was applied to students from three disciplinary areas. Q2b and Q3 were applied to professionals of the same disciplinary areas, trying to keep the same proportion of students who answered Q2a

from the three disciplinary groups. All samples aimed at being representative of the target population's characteristics (Cohen et al., 2011; Dörnyei and Taguchi, 2009; Bachman, 2004); however, this was not always possible. Gathering data from some disciplines was more difficult than others, especially when applying Q2b and Q3.

Convenience of opportunity (Dörnyei, 2007) was also taken into account. The institutions where the questionnaires and interviews were applied to workplace professionals were selected randomly, trying to have the same number of participants from state and private institutions. Q2b and Q3 were applied in those institutions where permission was granted.

#### 4.4.1.1 Test takers

Q1 was the first instrument applied. It was applied to test takers after finishing the in-house built certification test. In their last semester at the university, test takers were students from three disciplinary groups: health sciences, social sciences, and physical and natural sciences. The questionnaire was applied during established dates and times after test completion. Participants in a test session depended on enrolment.

- University test takers are those participants responding to Q1 after the test session. This group of participants includes students from the 6<sup>th</sup> semester of the School of Medicine (this per request of the School of Medicine), students from the 8<sup>th</sup> semester (except for Medicine, all programmes last eight semesters) and undergraduates (from any discipline), who are retaking the test for the second or third time. Most university students need to fulfil the language requirement to receive their degree.

This group of participants was found at the test session premises. A large table and chairs were placed at the exit point where the questionnaire was applied to students who agreed to answer it.

Due to the logistics of the test session, test takers were finishing the test in pairs, two or three pairs, almost simultaneously. On their way out, the researcher talked to each pair or group of students, explaining the purpose of the questionnaire and asking them to answer it. Test takers were told their participation was voluntary and ensured their answers were anonymous. After they completed the questionnaire, they were asked if

they agreed to answer a short-semi structured interview. The researcher asked permission to audio record the interview and ensured only their discipline would be asked and their name would not be required.

A total of 886 students responded to Q1 over a two-year period: four test sessions at the end of semesters A and B in 2015 and 2016. The name of the study program of test takers was the only factual information requested. Other factual information such as age and gender were not requested as it was considered that test takers would be more willing to answer the questionnaire if their identity remained as anonymous as possible. Despite this, some participants were a little concerned about the possibility of linking their Q1 responses to their test; they thought their scores could be affected by expressing their opinions about the test. Even though 886 questionnaires were collected, the disciplinary area of 300 participants is unknown. Some test takers decided not to include this information, which was respected.

The logistics of the test administration played an essential role in the number of test takers that could be interviewed. While the researcher was interviewing a participant or a group of participants, other participants who finished answering the questionnaire left the premises. Therefore, it was not possible to interview all the participants. It was possible to interview only 231 participants, representing 26% of the total Q1 respondents. Had all the participants been ready to leave simultaneously, very few participants could have been interviewed. Based on the tone of voice, it is thought that 148 participants were female and 83 were male. As can be seen, the number of female participants doubles the number of male test takers.

Participants from the different study programmes were grouped into three major areas of knowledge: social sciences, physical sciences, and health sciences. However, Table 4.2 below shows a fourth group of participants: Unknown discipline. These participants decided not to include the name of their study programme when answering Q1.

Test takers studying law, political sciences, psychology, human communication therapy and social work are included in the numbers reported for the social sciences area. The pharmacy-biological chemist, biotechnological chemist, and materials sciences engineer programs are included in the physical sciences area. Finally, the study programs for medicine, nursing and obstetrics, dentistry and nutrition are included within the health sciences. The difference in the size of groups of participants among

disciplinary areas is due to career popularity and school compliance with the requirement. Some careers are more in demand than others. Typically, the BA in Law from the School of Law & Political Science has the largest student population in the university. The number of test takers from each disciplinary area can also be related to the degree to which their School requires them to comply with this institutional requirement.

<i>Disciplinary area</i>		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Unknown discipline	300	33,9	33,9	33,9
	2 Social sciences	380	42,9	42,9	76,7
	3 Physical sciences	30	3,4	3,4	80,1
	4 Health sciences	176	19,9	19,9	100,0
	Total	886	100,0	100,0	

Table 4.2 Number of participants grouped by disciplinary area

A very positive response from test takers was observed, given that answering the questionnaire was not obligatory, and it is assumed that most might have been tired and ready to leave. There were only eight who apologized because they had to run to work. They are assumed to be related to one of the healthcare disciplines (medics, nurses, or dentists), as students in these areas are dressed in white.

#### 4.4.1.2 University students

The second group of participants is university students. These are described below.

- University students are participants of different disciplines in their last year at the university (whenever possible). Even though most programmes last eight semesters, most participants were between the 6<sup>th</sup> and 7<sup>th</sup> semesters. Participants were from these semesters because some disciplines do not have 8<sup>th</sup>-semester students taking regular classes; they are mainly doing fieldwork.

Even though their opinions may not be considered as accurate as the ones provided by professionals, they do have expectations of what the workplace may require. These students belong to a different generation than most workplace professionals in the current study; therefore, they have a different view of the world. These participants belong to a generation called digital natives (Prensky, 2001), characterised by technology and globalization. Therefore, their view of their future professional activity cannot be detached from these aspects.

This group of participants were approached through their schools. Questionnaires were applied in the classrooms or laboratory premises.

A total of 706 students from several schools within the university aged between 18 and 43 years old, with an average age of 22.3, answered Q2a: 443 female and 262 male respondents. The average age reveals most students finished high school and continued studying. A smaller number of participants returned to school after (presumably) working for several years.

The initial aim of the sample was between 50 and 100 per school; however, the number of students per group varies according to the career. Some careers have very few students; in others, groups are rather large. Considering 30 as the average number of students in the last semesters, the aim was modified, targeting the application of the questionnaire to two groups per school. Some groups were smaller, so permission was requested to apply the questionnaire to more groups. However, in some Schools, this meant the questionnaire would be applied to students in semesters below fifth, so this was not done.

The average socioeconomic profile of the university student population ranges from middle to low class. In general, more than one family member works to contribute to the family's income. Several students work to support themselves or even a family of their own. This data was obtained through personal communication with the Statistics Department within the university.

#### 4.4.1.3 Workplace professionals

The third and last group of participants are workplace professionals.

- Workplace professionals are considered to be the main source of information given that they are the ones who can provide the best account of the current language needs within the local workplace. Underemployed professionals were omitted. Contributions from expert informants within the TLU context are valuable since they not only possess direct experience but may share characteristics with potential future language users within the same TLU context (Tarantino, 1988). Knoch and Macqueen (2020) refer to this group of participants as domain insiders.

These participants were found in many different locations. There were no restrictions concerning age or gender; questionnaires were applied to all workplace professionals available within a premise. It was sought that data gathered from local workplaces represented both the private and the public sectors.

Six hundred twelve questionnaires were applied to workplace professionals: 289 male and 323 female respondents. Other professionals were asked to provide examples of language use while answering the questionnaire. These 93 participants participated in a recorded interview while answering the questionnaire.

The main reason for reaching 411 respondents in public institutions was the convenience of opportunity. Accessing individuals in the private sector (201 in total) was either more difficult or impractical. Within the private sector, there are fewer professionals per workplace visited, so it was necessary to visit more workplaces.

Regarding age range and gender, it was found interesting that there were more female professionals within the youngest age range group. In contrast, more men were observed within the third age range (44 and above) in both state and private sectors. Male and female professionals were almost the same within the 33 and 43 age range group. It is also worth pointing out that the highest number of professionals is in the youngest age group.

### **4.5 Design of the data-gathering instruments**

Two questionnaires were designed for the present study: Q1 and Q2 (which is also Q3, as will be detailed in 4.5.3.3 below). One short semi-structured interview was conducted as a complement after responding to Q1. A supplementary interview was



conducted with some participants while responding to Q3. All data were collected in Spanish; the English version is provided in the Appendixes specified in each section below.

#### 4.5.1 Questionnaire 1

The brief Consent Form was on a slip of paper stapled in front of Questionnaire 1 (Q1). The only factual information asked test takers to provide was their career, although it was not obligatory to include that data. The purpose of this was to ensure anonymity. Test takers needed to be sure it would not be possible to relate their answers in Q1 to the test. It was considered of paramount importance that test takers had no doubt their participation would not affect them in any way. Three hundred participants decided not to include information about their careers. The Consent Form was in Spanish to ensure test takers understood the information. The translated version can be found in Appendix II below.

The questionnaire was designed in Spanish to ensure that misunderstandings did not influence the answers. Test-takers were also believed to be more willing to participate if it were in their native language. Some even said something like “It’s in Spanish, right?” or “Only if it’s in Spanish!” when asked to answer Q1.

Questionnaire 1 (Q1) aimed to identify the face validity of the certification test used, addressing several aspects of it. Q1 is a Likert-type questionnaire with four options to choose from. Participants answered Q1 after completing the exit test.

Q1 is divided into five sections (see Appendix III below for the entire questionnaire), each addressing an aspect of face validity. The purpose of each section is discussed below.

Section 1 (items 2 to 18) aims to determine what competences test-takers think have been assessed by the test they just took, which can be seen in Table 4.3 below.

Items 2 and 3 aimed to determine whether test-takers thought the language measured was related to the workplace or academic environment. It was important to know whether test-takers felt the test measured English for general purposes or was domain-related. Items 4 through 18 were taken from the 15 language descriptors provided by the university's English department (in charge of test development). These

<b>In MY opinion, the B1 English language university exit test I just took</b>	Disagree 1	Quite disagree 2	Quite agree 3	Agree 4
2. included activities (in written form) that are similar to what I will need to do in my professional life				
3. tested the language abilities that are required to study for a master's degree in an English-speaking country				
<b>In MY opinion, the B1 English language university exit test I just took assessed my ability to understand</b>	Disagree 1	Quite disagree 2	Quite agree 3	Agree 4
4. the main points of clear standard speech on familiar matters regularly encountered <b>in workplaces</b>				
5. the main points of clear standard speech on familiar matters regularly encountered <b>in school</b>				
6. the main points of clear standard speech on familiar matters regularly encountered <b>in leisure activities</b>				
7. the main point of many radio or TV programs on current affairs or topics of <b>personal interest</b>				
8. simple technical information, such as operating instructions for everyday equipment				
9. the description of events, feelings and wishes in personal letters				
<b>In MY opinion, the B1 English language university exit test I just took assessed my ability to</b>	Disagree 1	Quite disagree 2	Quite agree 3	Agree 4
10. deal with common situations likely to arise whilst travelling in an area where English is spoken				
11. enter unprepared into conversation on topics that are familiar or <b>personal interest</b>				
12. enter unprepared into conversation on topics that are familiar to <b>professional interest</b>				
13. enter unprepared into conversation on topics that are familiar and are <b>related to everyday life</b>				
14. describe experiences and events, my dreams, hopes or ambitious				
15. briefly give reasons and explanations for opinions and plans				
16. narrate a story in oral or relate a plot of a book or film and describe my reactions				
17. write a simple connected text on topics that are familiar or of personal interest				
18. write personal letters describing experiences and impressions				

Table 4.3 Q1, Section 1. What test takers think the test assessed

language descriptors were used to develop the test, though the exact ones used were not provided.

Two groups distinguish between items 4 to 18; items 4 to 9 relate to receptive skills, and items 10 to 18 relate to productive skills.

Section 2, shown in Table 4.4 below, had two purposes. The first, items 27 through 33, was to identify the competences that test-takers think the test should assess. Second, to determine how relevant test takers consider each of the four skills to be in their future professional job. This information was addressed in item 34.

<b>In MY opinion, the B1 English language university exit test I just took NEEDS</b>		Disagree 1	Quite disagree 2	Quite agree 3	Agree 4
27.	the world of entertainment (music, movies, videogames)				
28.	travelling for business				
29.	doing a postgraduate in an English-speaking country				
30.	to assess English related to talking about the things that I like to do, eat, etc.				
31.	the world of work (useful for my future professional life)				
32.	talking about my family				
33.	travelling for pleasure				
34.	<b>I think a job related to my disciplinary area requires me to</b>	Very rarely 1	With very little frequency 2	Rather frequently 3	Frequently 4
	a) speak <b>in English</b>				
	b) understand what is said to me <b>in English</b>				
	c) write <b>in English</b>				
	d) read <b>in English</b>				

Table 4.4 Q1 Section 2. What test takers think the test needs to include and the expected need to use the four skills within a discipline-related job.

The answers to items 27 to 33 show what the test takers are interested in. It is considered that the need to use English in a professional context can vary not only based on the nature of the subject but also influenced by the type of institution or workplace that test-takers believe they are likely to join. It is considered that test-takers expectations of the need to use each of the four skills in the workplace will vary by speciality.

Section 3 (items 19 to 21 and 35) addresses students' opinions on the usefulness of the courses they have taken in preparing for the exit test. Test-takers were asked their opinion on the courses they attended in items 19 and 20, whether they helped them prepare for the test and whether there were practice activities that prepared them for those found in the test. Item 21 aims to learn to what extent the courses attended included activities that test-takers believe are similar to those they will likely encounter in their future job. The elements in Section 3 are shown in Table 4.5 below.

<b>In MY opinion, the English classes I took before taking the B1 English language university exit test</b>		Disagree 1	Quite disagree 2	Quite agree 3	Agree 4
19.	in general, helped me prepare for the test				
20.	provided opportunities to practice the type of activities included in the test				
21.	included activities that are similar to the type of activities I may need to perform in English in my future job				
35.	Where did you study English? (Please tick as necessary)	<input type="checkbox"/> School of Languages	<input type="checkbox"/> Harmon Hall	<input type="checkbox"/> FADER Language Center	
		<input type="checkbox"/> FECA Center of Languages	<input type="checkbox"/> Inglés Individual	<input type="checkbox"/> ITD Language Center	
		<input type="checkbox"/> Centro de Idiomas del Guadiana	<input type="checkbox"/> Golden English	<input type="checkbox"/> Other (please specify)	

Table 4.5 Q1, Section 3, preparation for the test.

While all students are required to take the test offered by the university, learning the language can be done inside or outside the university. Therefore, it is necessary to know whether test-takers preparing for the test took a course in any of the options offered by the university, studied alone or attended a private language school. This data was collected through item 35.

It was considered relevant to learn what test takers think about the usefulness of the courses taken to prepare for the test. Since a proposal for a new test is likely to have a washback effect, it was considered relevant to know what test-takers' opinion is about the relationship between the test used and the courses offered at the university. If the

current test is replaced, the content of the university's courses would need to be adjusted.

The aim of Section 4, which includes items 1, and 22, to 26, is to examine how test-takers perceive the overall difficulty of the test. These aspects give a picture of the opinion of the test-taker population regarding the difficulty of the test (item 1), the language in which the test instructions are given (item 22), the types of items included in the test (items 23 and 24), the time allotted for the test (item 25) and the number of items in the test (item 26). These items can be seen in Table 4.6 below. Test-takers' opinions on the format and administration of the test can help to understand reactions or responses. A context-sensitive test knows the test-taker population. This does not mean that the test must necessarily be designed to meet all test-taker expectations, i.e., preference for more multiple-choice items or open questions. Being familiar with a general mindset may help provide information about test characteristics.

<b>In MY opinion, the B1 English language university exit test I just took NEEDS</b>	Disagree 1	Quite disagree 2	Quite agree 3	Agree 4
1. was easy to answer				
22. instructions in Spanish				
23. more open-ended questions				
24. to be multiple-choice only				
25. more time to be completed				
26. to be longer				

Table 4.6 Q1, Section 4. General perceptions about the test

Ensuring anonymity to test-takers was crucial to incentivize them to participate.

Therefore, the only factual information requested was contained in item 36, although it was left open not to provide it. They were only asked to provide the name of the school and the degree they were studying. This item is shown in Table 4.7 below. This data is usually requested at the beginning of a questionnaire. On this occasion, it was considered that if it had been at the top, test takers would have felt insecure about answering the questionnaire, as they may have thought there might be a way to link

their answers to their test. For this reason, item 36 was placed at the end of the questionnaire, expecting that after completing the questionnaire, they might no longer feel uncertain about providing this data.

36.	School:		Discipline:	
Is there anything you would like to comment on relating to the test you just took or the type of test you think you should take? (Please continue in the back.)				

Table 4.7 Q1, Section 5 Factual information from test takers.

The final section of Questionnaire 1 (Q1) was a space for test-takers to write any comments they wished to make regarding the test. Open-ended questions at the end of an objectively answered questionnaire allow participants “to express their opinions more freely...” or otherwise, they “...may find it frustrating to be completely limited to choosing from ready-made options.”(Dörnyei & Taguchi, 2009, p. 37). These voluntary responses, viewed as “additional bonuses” (Dörnyei & Taguchi, 2009, p. 48), provided insight into perceptions of the face validity of the test as well as other aspects related to the testing process.

It was considered important to allow test-takers the opportunity to express ideas, thoughts or feelings about the test or test experience, should they wish to do so. Even if the space was small, they were invited to use the back of the sheet if desired.

#### 4.5.2 Semi-structured interview

When the test-takers submitted in Q1, some were asked if they would agree to a short anonymous audio-recorded interview. All test takers asked to participate agreed to do so.

The purpose of the brief semi-structured interview was to ask test-takers which area they would choose the test to focus on (from three options suggested by the researcher) if allowed to do so and to justify their answer. They were also asked if they thought English was important or necessary for professionals in their field living in Durango and why they thought that. The script for the semi-structured interview can be found in Appendix IV below.

Although the gender of the participants in Q1 is unknown, it is possible to know the gender of the test-takers who participated in the interview from their voice, so the appropriate pronoun is used when necessary.

Two-hundred and thirty-one interviews were conducted, some individually, others in pairs or small groups, and the average interview length ranged from 3 to 8 minutes.

Due to the amount of data, the researcher and two research assistants transcribed the interviews. They were transcribed according to standard procedures, and ambiguities have been highlighted and not used in the dataset. The researcher revised transcriptions made by assistants before being uploaded to NVivo. Transcription conventions can be found in Appendix VIII below.

#### 4.5.3 Questionnaire 2 (Q2)

Questionnaire 2 (Q2) aims to identify students' expected language needs and professionals' current language needs in the local workplace. An English for General Professional Purposes (EGPP) approach is adopted given that, just as the IELTS (Davies, 2008), designing an English for Professional Purposes (EPP) test for each discipline is found unfeasible and impractical. A local context-sensitive general work domain test is considered useful and more practical for the local university context and future language users. A test of such nature considers work-related competences common to several disciplines. In contrast with a general-purpose English test, which has not considered the working environment language needs in its design, an EGPP approach aims to measure language and skills related to the discipline of test takers. As the data comes from the local context, the language and skills represent the language use activities in the local workplace.

The test needs to be at the B1 level of the CEFR (COE, 2001), as the university established this level. Several documents were considered for work-related language descriptors at the B1 level. These are listed below:

- The Common European Framework of Reference for Languages (COE, 2001)
- The American Council on the Teaching of Foreign Languages (ACTFL) (American Council on the Teaching of Foreign Languages, 2012)

- Association of Language Testers in Europe (ALTE) (ALTE, 2002)
- Work Ready Tools Non-Regulated Skilled and Semi-Skilled Occupations (CCLB, n.d.)
- The Canadian Language Benchmarks (CLB) (CCLB, 2012)
- Can Do Statements for Employment (CCLB, 2016)

Work domain language descriptors were found in the Association of Language Testers in Europe (ALTE) CAN DO Project (ALTE, 2002) and the Work Ready Tools Non-Regulated Skilled and Semi-Skilled Occupations (WRT NRS SSO) (CCLB, n.d.), a complementary document of the Canadian Language Benchmarks (CCLB, 2012). The language descriptors of both documents were benchmarked against the CEFR language descriptors (ALTE, 2002; Pépin, 2008).

The ALTE identifies three categories of main interest for learners of a language: social and tourism (Category A), work (Category B) and study (Category C). It presents language descriptors in six levels: ALTE Breakthrough, ALTE 1, ALTE 2, ALTE 3, ALTE 4 and ALTE 5. They have their corresponding CEFR levels: A1, A2, B1, B2, C1, and C2. The ALTE 2 level is presented to correspond to the B1 level of the CEFR. The ALTE Category B (work) CAN DO statements at the ALTE 2 level were identified and organized in a list.

Before looking at the WRT NRS SSO language descriptors, it was necessary to identify which level from the Canadian Language Benchmarks (CLB) corresponded to the B1 CEFR level. This was done through the CENNI document (Certificado Nacional de Nivel de Idioma-National Certificate of Language Level) (DGAIR, 2011) (see Appendix V below). According to the CENNI, level 5 of the CLB corresponds to the B1 level of the CEFR. Since the language descriptors within the WRT NRS SSO emerged from the CLB, the level 5 Can Do statements from the WRT NRS SSO were organized in a list.

The work-related language descriptors extracted from the ALTE Category B work domain and those taken from the WRT NRS SSO were compared. A list of 28 competences resulted from merging both lists of work-related language descriptors. A



Likert scale questionnaire was built using them. Numbers 1 to 4 were the options with the following labels: 1 No, almost never; 2 Sometimes; 3 Frequently; 4 Very Frequently.

The questionnaire was translated into non-technical Spanish, so the statements were easy to understand by non-language specialists. During the piloting stage of the questionnaire, some participants highlighted the fact that some of the language descriptors included examples of the language statement being described. Participants commented that these examples helped them understand the meaning of the statements, and it was easier to relate the competences with activities they may carry out using English in their workplace. They also said they would like to have a reference in frequency. As a result of these comments, language descriptors were checked, and examples were elaborated or added. Percentages in the options to choose from were also added. The second version of Q2 was piloted again. Participants said that the examples provided in the competences helped them see how these could be related to their workplace activities, and the percentages helped them to decide which box to tick. The English version of Q2a (applied to students as explained in 4.4.1.2 above) can be found in Appendix VI below.

Questionnaire 2 was applied to two different groups of participants. The difference between them is the introductory paragraph, where consent from participants is requested. Another difference is the factual information collected. The rest of the instrument is the same for both groups, detailed below.

#### 4.5.3.1 Questionnaire 2a (Q2a)

Questionnaire 2a (Q2a) was applied to graduating students. The purpose of Q2a was to identify what their expectations are in terms of work-related language needs. Knowledge of the target test taker population is considered an important aspect of test design (O'Sullivan, 2012), especially because they are a significant element in deciding language needs (West, 1994). Learners bring to the picture their needs and wants, which should not be overlooked, as they are the ones who are going to be affected by the decisions made on test content (L. F. Bachman & Palmer, 1996).

Q2a was applied to a total of 706 students of three disciplinary groups: health sciences, social sciences, and physical and natural sciences. Students were in the last year of their studies.

#### 4.5.3.2 Questionnaire 2b (Q2b)

Questionnaire 2b (Q2b) was applied to workplace professionals.

Applying Q2b to professionals aimed to identify current work-related language use activities and language needs. Two formats of Q2b were prepared: online and paper-based. The English version of the latter can be found in Appendix VII below. This mixed-mode strategy was considered to allow flexibility in the questionnaire application. Smart Survey was used to deliver the online version of Q2b.

When Q2b and Q3 were applied, participants were told that the purpose of applying the questionnaire was not to see if they (the participants) knew English. It was explained that the focus of the questionnaire was on the position occupied by people, regardless of whether individuals may have needed help to carry out some of the activities required in the workplace.

Q2b was applied to 612 professionals found in 132 workplaces.

#### 4.5.3.3 Questionnaire 3

##### Competencies questionnaire (Q3) + supplementary interview

Current language use and needs in the workplace were identified through the use of Q2b. However, this is limited in the type of information collected. Even though we know the frequency with which competences are used in the workplace, this information does not say much about the specific situations in which the language is used. More detailed information is required to design test tasks; this will contribute to designing test items that better reflect the actual language use activities carried out by professionals.

This data was collected by asking participants to answer Q3 in the researcher's presence. The content of Q2b and Q3 does not vary. The questionnaire was relabelled to distinguish data obtained by answering a questionnaire (Q2b) and answering a questionnaire and providing examples of language use activities during an interview (Q3). This data aims to answer SQ3 above. The instrument used to generate professional interview data will be called Q3 data.

Upon requesting participants' permission to audio-record the session, they were asked to read aloud each competence and to state their choice regarding the frequency with which each of them is required in their workplace. The options given to participants to choose from are shown in Table 4.8 below:

<b>1</b> <b>No, almost never</b> <b>0-25%</b>	<b>1</b> <b>Very rarely</b> <b>26- 50%</b>	<b>2</b> <b>Sometimes</b> <b>51-75%</b>	<b>4</b> <b>Frequently</b> <b>76-100%</b>
---	--	---	---

Table 4.8 Q2b Likert scale options participants chose according to the frequency that competences are required in the workplace.

Every time a participant selected an option different from “No, almost never”, they were asked to provide examples of the type of activities or situations they were required to carry out and the situations in which these took place.

Through this means, it was possible to learn about the type of language use activities required within the workplace and the type of language required to fulfil them.

Due to the amount of data, transcriptions were carried out by the researcher and two research assistants. They were transcribed according to standard procedures. Unclear sections were highlighted and not used in the data set. Transcription conventions can be found in Appendix VIII below.

Overall, 93 interviews were conducted in 71 work settings. Workplace professionals provided specific examples of language use in their workplace. The interviews took between 9 and 16 minutes.

#### 4.5.4 Summarizing research design

Figure 4.1 below summarizes the research design. Four sub-questions need to be answered as they will inform the proposal that will provide the answer to the Main Research Objective (MRO). The answer to Research Question 1 (RQ1) will be obtained by analysing data gathered through Questionnaire 1 and a semi-structured interview applied to test takers; this is Phase I.

In Phase II, the answer to Research Question 2 will be the result of the comparison of data gathered through Questionnaire 2a (Q2a) applied to students in their last year of studies and Questionnaire 2b (Q2b) applied to professionals in the workplace.

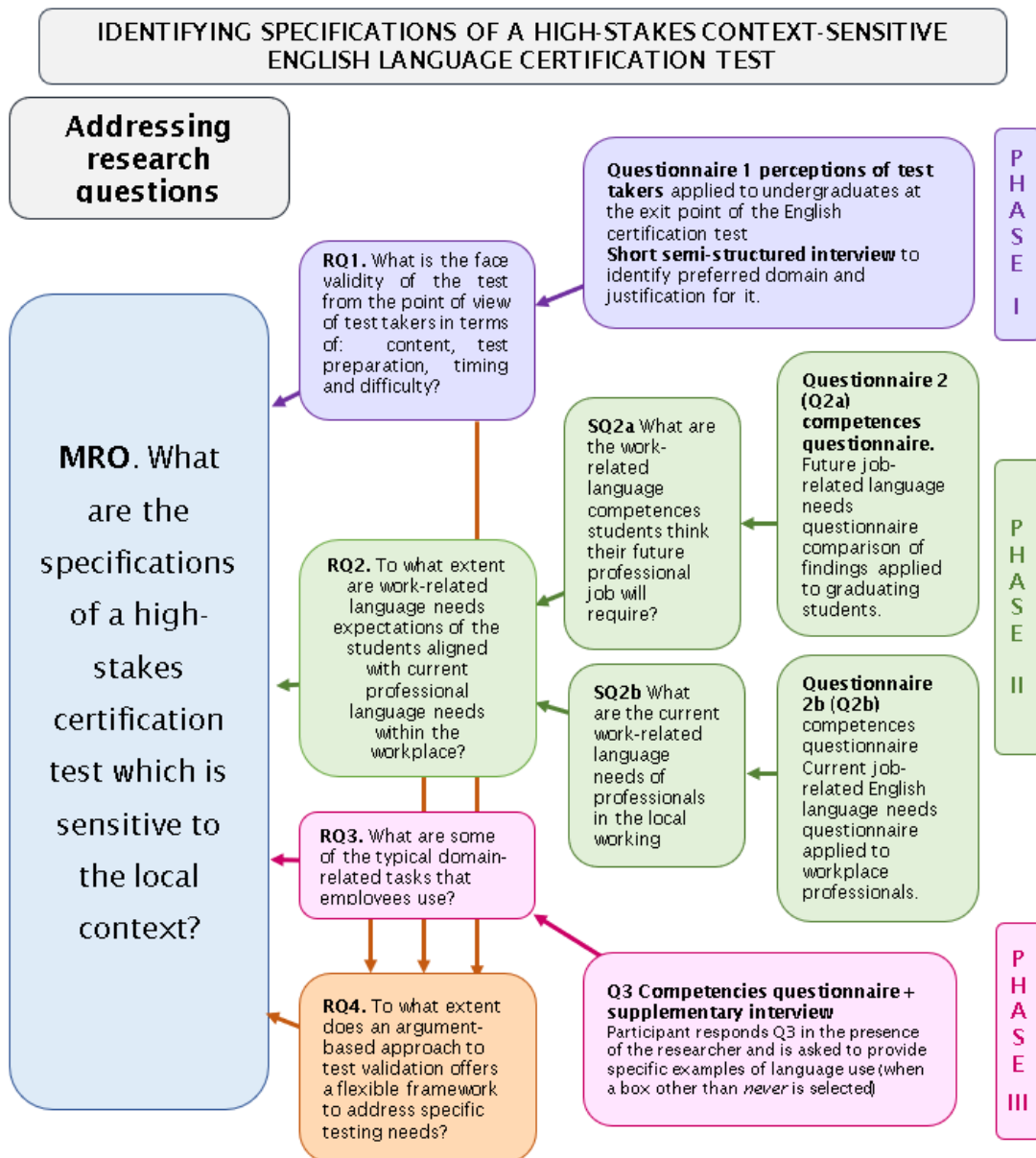


Figure 4.1 Summary of research design

Responses to Research Questions 1 and 2 inform and provide evidence for building a proposal for an assessment use argument for test validation, which is sensitive to the

local context. In Phase III, the answer to Research Question 3 (RQ3) is obtained from the data gathered through the interview while applying Q3. Specific examples of tasks within the context of language use are collected in this Phase. These findings inform a proposal for domain-related test task design patterns, which are an element of the overall proposal of a local context-sensitive test.

Finally, the Main Research Objective (MRO) answer comes from a literature review examining the data collected and the expected outcomes.

## **4.6 Ethics**

Working with respondents as human beings taking part in a research study involves ethical issues (Dörnyei and Taguchi, 2009). Ethics is concerned “with conducting research in a moral and responsible way” (Bruns, 2010, p. 34), where participants need to be well informed and explained the nature of their participation and ensured anonymity.

For the present study, the required documentation was submitted before the Ethics and Research Governance Online (ERGO) at the University of Southampton to comply with established ethical requirements. Under Submission ID 17618, the project was approved by the Ethics Committee, allowing the data-gathering phase of the study to begin.

All questionnaires contained a consent paragraph, which participants read before answering the questionnaire. A consent paragraph for Q1 can be seen in Appendix II and for Q2a and Q2b in Appendix VI and Appendix VII, respectively. When interviews took place, participants were asked for permission to audio-record the conversation. The script used to ask participants (test takers) for permission to record the interview can be found in Appendix IX below.

## **4.7 The role of the researcher**

As a former participant of the test design team, the researcher has a comprehensive view of the aspects related to the test design and application process, as well as some students' reactions when the results were unfavourable. Being the study of the researcher's interest, it is undeniable that hypothesis and expectations unconsciously

emerge. However, an inherent part of the role of a researcher is to identify and address any potential bias emerging from him or herself. Another essential aspect to consider is acknowledging and discussing any trait that may counter the expected results.

However, research studies emerge from the interest in a topic on which the researcher has hypothesized or has specific preconceived ideas. My interest in the present study emerged from being surprised to hear former test-takers complaining about the irrelevance and uselessness of the English certification test for their professional activity. It seemed impossible to believe that every professional in their workplace did not need English. In the interest of offering a test that was perceived to be relevant for professional practice, the quest seemed, at that time, to be focused firstly on establishing why English was required in the professional local working environment. As the study progressed, it was found that the aim was not to justify that every professional required English but to understand the characteristics of local English language use within the working environment. This aim entailed the adjustment of preconceived ideas. It was first acknowledged that not all professionals needed English in their current jobs. It was later learned that language requirements differed among disciplines and even within the same discipline. In this sense, my role as a researcher required the modification of the preconceived idea that everybody needed a balance of the four skills of English.

Among the threats to research validity that may affect the present study could have been the participant's desire to meet the researcher's expectations (Dörnyei, 2007), especially during the interviews. This is not considered a significant threat as, in the case of students, the researcher was a stranger who would have no influence over test scores in any way. Regarding the interviews with workplace professionals, participants were told that any response was relevant to the researcher as the purpose was to understand the context. Stating there were no specific expectations aimed at reducing social desirability bias (*ibid.*).

A strength of using a mixed-methods approach to research, as the present study has, allows qualitative data to be supported by quantitative data. This strengthens the validity of the results and may reduce the possibility of researcher bias (Dörnyei, 2007).

## 4.8 Data analysis

The data generation process may begin, in many cases, during interviews or data transcription, as it is then when the researcher has the first contact with the data (Duff, 2008) and starts to notice salient aspects of the data generated (Duff, 2020). A mixed-methods approach requires analysing the data gathered to be addressed differently. All quantitative data were analysed using the statistical analysis program SPSS. The qualitative data manager NVivo 10 was used for data collected during the interviews.

### 4.8.1 Preparing data for analysis

All quantitative data were fed into a spreadsheet and checked for errors while entered. Unanswered items were also identified and highlighted in the spreadsheet. Two responses to Q1s were eliminated because almost half of the items were left unanswered. Even though the answered items are still data that can inform the results, it was considered that these questionnaires were answered in a careless, hasty way. The reliability of the data of these two questionnaires was in doubt. Some of the questionnaires had a few unanswered items, which could be attributed to an involuntary omission when answering the questionnaire. According to Dörnyei (2007), it is not uncommon for a researcher to find that a few values are missing; however, the rest of the data is still useful.

Qualitative data from Q1 (open-ended question at the end of Q1) was fed to NVivo. Interviews were transcribed into Word and read to check for errors or misunderstandings. Highlighted portions representing unclear data were identified. In those cases, the audio recording was listened to once more, trying to check if it was possible to understand the segment. There were only three cases where this was required. After listening a few times to the section, it was possible to understand what was said, and the transcription was completed. Data was fed to NVivo for its analysis.

### 4.8.2 Analysis of quantitative data

Data collected in Q1, Q2a, Q2b and Q3 were analysed using SPSS. A code was assigned to each discipline area in which they were combined. For Q2 (a and b) and

Q3, where other factual information was collected, data such as gender or the age group of the participants were also coded.

For Q1, the data were analyzed and compared using mean values and standard deviation. The data were first divided into two major groups: known and unknown disciplines and mean values and standard deviations, which were analyzed to see if a pattern or clear difference was found. The same procedure was used for the responses of the different participants (see Table 4.2 above) of the three identified discipline groups (607) and the unknown discipline group (300), also looking for a response pattern or a clear difference between the groups.

For Q2, mean values and standard deviations were used to compare groups and to identify the language competences that are more frequently used or likely to be used.

#### 4.8.2.1 Questionnaire 1

Analysis of the data collected in Q1 will provide insights into test-takers' perceptions of the certification test. This was achieved from the different sections of Q1.

The internal consistency of Questionnaire 1 was assessed by measuring Cronbach's alpha coefficient. The analysis results in Table 4.9 below for Q1 indicate that all items in Q1 measure the test-takers' perception of the exit test.

Cronbach's Alpha	N of Items
.918	37

Table 4.9 Reliability Statistics for Q1

Two to four letters were used to identify the disciplinary area of respondents. Questionnaires answered by participants who chose not to provide the program they are enrolled in are marked as UD (Unknown Discipline). The questionnaires were numbered according to the order in which they were used. The gender of the participants who answered Q1 is unknown. For this paper, when the gender of a participant is unknown, and the use of a pronoun is required, one of the pronouns (he or she) will be used randomly, trying to keep a balance between their use.



As detailed in 4.4.1.1 above, 300 participants chose not to state the study program they belong to; therefore, of 886 completed questionnaires, these data are unknown for about 1/3 of the total data collected.

The type of data collected was identified in SPSS according to its characteristics. Each competence has been labelled TTQ to refer to the *test-takers' questionnaire*, followed by the number in which they appear in Q1, i.e., TTQ4 refers to item 4 in Q1.

The five sections included in Q1 (see 4.5.1 above) were analysed separately by comparing the mean of the discipline groups. The results can be seen in Appendix X below.

The 17 items in Section 1 (Table 4.3 above) aimed to determine what test-takers believed they were being assessed on the test they took. The items were divided into three groups for their analysis: a) related to the work domain (TTQ2, 4 and 12), b) related to the academic domain (TTQ3 and 5), and c) related to personal/general purposes (TTQ6 to 11 and 13 to 18).

The analysis outcome in Section 2 (Table 4.4 above) provided data on the competences that test-takers felt should be assessed. Some of these competences relate to specific domains of language use. It also provided data on the skills test-takers will need for their future discipline-related jobs.

The data analysis results collected in Section 3 (Table 4.5 above) captured what test-takers think about their courses before the test. It also summarizes the places where learners prepared for the test.

The data from Section 4 (Table 4.6 above) reflects the test-takers' overall perceptions of the test.

Factual information, such as the school and study program of the test participants, was recorded using the items in Section 5 (Table 4.7 above), providing a profile of the test-taker population. Finally, at the end of the questionnaire, the space provided for test-takers to write comments or opinions about the test they took or the type of test they thought they should take provided interesting data supporting the opinions of test-takers on various aspects of the test, the test session, or the type of test they would like to take.

## 4.8.2.2 Questionnaire 2a

Q2a aimed to identify the job-related language competences that students believe they will need to use more frequently in their future job. Students selected the frequency with which they expected to use each of the 28 work-related language competences.

The questionnaire has two sections. The first section of the questionnaire collected factual data to create a participant profile: their gender, age, major and semester. The questionnaires were identified by a number assigned based on the order of application.

The work-related competences of 706 student-answered questionnaires from different degree programs offered at the university were analysed by comparing their means and standard deviation to compare responses from the three identified discipline groups (social, physical, and natural and health sciences).

The 28 competences were organized by the skill or skills concerned. Some of these require one skill, while most competences involve combining two skills, as shown in Table 4.10 below.

Skill(s) involved	Competence Number						
Speaking	1	2	3	8	9	10	
and Listening	11	12	13	15	16		
Listening and Speaking	4	5	14				
Listening	6	7					
Reading	17	18	19	20	21	22	
Writing	23	24	25	26	27	28	

Table 4.10 Competences in Q2 according to the skill(s) required for their fulfilment.

As can be seen, there are more competences (to be used in that order) in the Speaking and Listening group than in either the Listening and Speaking group or the Listening group. The speaking and listening group competences assume that the person initiates an interaction with one or more speakers through a request or an utterance. The initiator of the interaction must understand the answer of the interlocutor. Speaking

skills are not considered to happen in isolation within these work-related competences, so interaction is assumed. The number of competences related to reading is the same as in the writing group.

#### 4.8.2.3 Questionnaire 2b

The same procedure used to analyse the data from Q2a was followed to analyse the data collected in Q2b.

As in Q2a, the first section of the questionnaire collected factual data to create the participant profile: their gender, age group, place of work and field of study. The questionnaires were identified by a number assigned according to the order of application.

As with the analysis of the Q2a data, the competences were grouped by the skill or skills concerned (see Table 4.10 above).

#### 4.8.2.4 Comparison of results -- Q2a and Q2b

The mean values of the three discipline groups were compared. The mean values of students and professionals from the health sciences were compared. The same procedure was carried out for the other two discipline groups. An average mean was obtained from each comparison. The highest and lowest scores for each competence were identified, as Appendix XI below shows.

The highest mean values of each discipline group per competence were determined and compared. The competences that did not report a mean value at or above the total mean were discarded. These were discarded because a mean below the total mean indicates that these competences are not very important to any group of participants. This comparison showed that some competences were equally important for students and professionals in the same discipline group or across discipline groups. It was found that only a few competences were relevant only to one group of participants. This comparison resulted in a list of competences and their importance to the participant groups.

### 4.8.3 Analysis of qualitative data

Qualitative research is often criticized for not being rigorous (Dörnyei, 2007). However, even though the criteria to which quantitative research is subject cannot apply to qualitative research, some principles "should be applied rigorously to the data" (Braun & Clarke, 2006a, p. 26) when doing qualitative research. In this study, I followed the 6-step guide presented by Braun and Clarke (2006) to do a thematic analysis, as will be discussed below.

Three data sets of a qualitative nature were collected. The two data sets collected as part of Phase I were: a) responses to the open-ended question at the end of Q1 and b) semi-structured interviews with test takers. These were independently analysed using thematic analysis to "identif(y), analyz(e), and report patterns (themes) within data" (Braun & Clarke, 2006b, p. 6). The third data set collected as part of Phase II were examples of the specific language use activities carried out by professionals when using English in the workplace.

The first step is to get familiar with the data. Concerning the interviews, the researcher had the first opportunity to learn about the data during its generation process.

Interviews were transcribed, creating a Word file for each data item. During this process, the researcher had a second opportunity to further familiarize with these data sets. Responses to the open-ended question in Q1 were also transcribed and saved as an individual Word file, allowing the researcher to have a first encounter with this data set.

The nature of the data sets required them to be addressed differently, as discussed below. The data set for Phase I was analysed using the qualitative data analysis tool NVivo.

#### 4.8.3.1 The open-ended question at the end of Q1

Once data was uploaded to NVivo, it was read several times. The 215 voluntary comments were distributed as follows: Unknown Discipline (UD) = 89, Health Sciences (HSc) = 35, Social Sciences (SSc) = 86, and Physical and Natural Sciences (PNSc) = 5. Comments from the questionnaires were transcribed to Word documents; this was the first opportunity for the researcher to familiarize with this data. After data had been

transcribed, they were reread, resulting in the elimination of five responses. These were discarded because they were thought not to provide data regarding the thoughts or opinions about the test or the test session (they were “thank you” or “no”). After being cleaned, a total of 210 comments were uploaded to NVivo.

The open-ended question in Q1 did not address a particular topic; it was a space for test-takers to express, if desired, anything they wanted regarding the test or the test session. There was no anticipation regarding the type of responses that would be collected; therefore, a data-driven thematic analysis method was adopted: an inductive approach to analyse the data (Braun & Clarke, 2006b). Braun and Clarke (2006) state that an inductive approach requires the data to be read and re-read to identify patterns in the data without looking for data related to a specific theme. Data was read and re-read to identify codes that represented interesting features. During this stage, it was identified that some comments were repeated. The initial analysis of the data revealed that many of the responses were a rejection of the test. More negative comments were identified, and the first codes were created. The data review within the negative and positive themes revealed several topics were addressed within the negative comments. Codes were created to represent the different types of comments. Data that was relevant to each code was collated. Code generation is the second step Braun and Clarke (2006a) propose as part of doing good thematic analysis. The third step is theme generation. Data coded was read and re-read, looking for themes or topics representing the data.

An example of this is *time*. Time was a word that appeared several times: some wanted more for the writing part of the test, and others more time in the listening or speaking sections. A larger group requested more time without specifying a section within the test. These codes have been grouped under the *More time* theme.

Some of these codes were collated to create a theme. The themes or topics that are relevant or interesting and contribute to participants' understanding were used. Figure 4.2 below shows the themes and their organization. As seen above, 176 of the 210 comments represented dissatisfaction or disapproval, while only 34 were positive about the test.

The fourth step in Braun and Clarke's proposal is theme review. Data within each theme was reviewed to verify all extracts belonged to the same group and that the

theme's name represented the extracts. Some theme names were modified, looking for a better representation of the data. The fifth step is to define the themes. Data within each theme had to be revisited to generate their definition. The sixth and final step is to produce the report. This step, however, did not take place soon after completing the fifth step. The data was left to rest for a few weeks. When the file was revisited, data within each theme was re-read, and names of themes and descriptions were also revisited to confirm their belonging and relevance. During this iterative process, some data were moved to another theme, and new themes or child themes were created. The process ended when there were no modifications to the data, the codes or the themes.

Name	Files	References
Characteristics	26	28
Compulsory	36	47
Contents	15	15
Difficulty	31	31
Difficult	20	20
Easy	11	11
Evaluators	13	13
Logistics	25	25
More time	48	49
In general	38	39
Listening	2	2
Speaking	3	3
Writing	5	5
Negative comment	176	176
Positive	34	34
Preparation for test	20	20
Relevancy	30	32
Not relevant	30	32
Relevant	0	0
Speaking	12	12
Test taker reliability	6	7

Figure 4.2 Themes and codes for voluntary comments to the open-ended question at the end of Q1.

## 4.8.3.2 Semi-structured interviews for test takers

The 231 interviews represent 26% of the questionnaires applied. The interviews gave the researcher an initial overview of the data. During the transcription, the researcher gained further insight into the aspects mentioned by participants, allowing the researcher to refresh the content of the data collected during the interviews.

Some interviews were individual; however, some were in pairs or small groups, according to the availability of test-takers. The data obtained were initially labelled with two to six letters related to the disciplinary area reported by the participants. The disciplinary letters were followed by an interview number, which was assigned in the order in which they occurred. Each disciplinary group was numbered separately. Finally, a letter indicating the gender of the participant was added. Thus, the eighth interviewed student from the social sciences group, a male student, was coded as SSc8M. Examples are shown in Figure 4.3 below.

Data source	Participant Major	Number randomly assigned to questionnaire or interview	Gender	
			Female=F Male=M	Code used
	Health Sciences	10	F	HSc10F
Semi-structured interview	Social Sciences	1	F	SS1F
	Physical and Natural Sciences	10	M	PNs10M

Figure 4.3 Examples of participant coding

When the data was fed into NVivo, responses related to the preferred domain were grouped into a theme called Domain. Child nodes were created for each domain: social, study and work. A third node was created because some participants said they did not want to study only one of the domains; they considered the three equally important. Child nodes for each domain were also created to nest the arguments given by participants when justifying their selection.

The data were read multiple times to identify themes or patterns meaningfully related to the research questions (Braun & Clarke, 2006) or relevant to the study. Initial themes or codes were generated representing comments on the same idea or topic. As themes

and codes were created, they were defined and described. After all data had been coded, the file was left unseen for a few weeks. When data was revisited, the definition and description of the themes were reviewed, and some modifications were made. The data coded to each theme was revisited to check for belonging and representativeness. Some codes were eliminated from a theme or moved to another one. Some new themes were created, or their name was modified in a several-week interactive process. The process ended when no more changes were made.

### 4.8.3.3 Supplementary interviews

For workplace professionals (Q3 + interview)

Data collected during the interview that accompanied the answer to Q3 provided specific examples of workplace language use by professionals from the disciplines included in the present study. The number of Q3 applied is the same number of supplementary interviews conducted (see Table 4.1 above).

Interview data were identified using the letter E (from the word *entrevista*, meaning interview in Spanish), followed by the first two to three letters that refer to the participants' disciplinary area. The letter "E" was preferred to "I" to avoid possible confusion with the number 1. The number corresponding to the order in which they were conducted was also used. In this way, a healthcare practitioner who was participant No. 561 was identified as HScE561.

These data were transcribed and organized according to the competence they referred to. The examples of language use given by participants on how they use each competence were grouped by competence, allowing the researcher to see if there were similarities in how professionals of different disciplinary areas use the language in the workplace. Data were read multiple times to understand better the specific types of activities performed by professionals in the workplace. This data was useful in the design process of test tasks (see Appendix XII below).

This data was used to learn more about the specific situations, interlocutors, types of texts, required information, etc., that participants needed English to perform.



## 4.9 Summary

The mixed methods approach chosen for this primary research case study was justified. The research design introduced the Main Research Objective (MRO) that the study aims to answer and the research questions that need to be addressed to answer the MRO. Next, the data collection process and the description and number of participants from three different informant groups were described. The design of the data collection instruments was then discussed. The Ethics and Role of the researcher sections followed this. After that, the limitations of the study were discussed. Finally, the data analysis process was described.

## 5 Understanding the test-taker population

### 5.1 Introduction Phase I results

Phase I aims to examine test-takers' perceptions of the university-leaving test. The results and analysis of the data collected in this phase (Table 4.1 above) are presented in answering Research Question 1 (RQ1) at the end of the chapter.

After completing the test, almost all participants agreed to answer Questionnaire 1 (Q1), collecting data from 886 test participants. Around 26% (231) also participated in a short semi-structured interview. Questionnaire 1 is mainly quantitative, but interesting qualitative data was also collected as the questionnaire allowed participants to comment at the end of it. Participants could write anything they wanted to share about the test or its procedure. Two-hundred and fifteen (24.3%) voluntary comments on various topics were received. The latter, together with the interview responses, not only contributed to the understanding of the results of the quantitative data analysis but also provided data reflecting the opinion of participants on aspects of the test and the test session that were not included in the questionnaire or the interview. These spontaneous comments reflect the thoughts and feelings of the test-takers.

The first section of this chapter allows students to comment on aspects not included in the questionnaire or the interview. The "Student Voices" section (5.2 below) draws on voluntary comments at the end of Q1 identifying issues of interest to them. Some comments made during the short semi-structured interview on the same topics are also recorded and merged with comments from the questionnaire.

Section 5.3 and its subsections consider the mean scores of the three identified disciplinary areas on the opinion of test takers regarding test content. These are compared according to the sections identified in 4.5.1 above within Q1.

### 5.2 Student voices

*"It is wrong that we have to take it to get our degree."* Q1SSc2

*“It shouldn’t be obligatory.”* Q1SSc412

*“It should not be a degree requirement.”* Q1SSc440

The comments above illustrate test-takers' opinions on the mandatory nature of the test, an issue not addressed in the questionnaire. The opinions regarding the obligatoriness of English in undergraduate programmes may vary among stakeholders, i.e. students, teachers, and employers. Upon finding that employers of multinational companies locally based considered university graduates did not have the language proficiency level required by their companies, a study was conducted to address this situation. University authorities, teachers and students from the university in central México where the study was conducted were asked what actions they would suggest so university graduates could meet the expected proficiency level. Interestingly, having “English as part of the requirements in the BA and BSc curricula” (Garcia-Ponce, 2020, p. 154) was mentioned only by university authorities.

The mandatory nature of university board decisions related to English was one of the themes identified in the voluntary written comments. There were 36 references with a total of 47 comments. This is shown in Figure 5.1 below. Some comments expressed

○ Compulsory	36	47
○ Against	26	33
○ Obstacle	14	15
○ In favour	3	3
○ Part of programme	12	12

Figure 5.1 Theme and code compulsory nature of the test taken.

disapproval of the test (33), and 15 indicated that it was a barrier to their professional development. Only three comments favoured the mandatory nature of the test, and 12 comments argued that English should be a compulsory subject for all majors. They

consider students should not be left to decide when to enrol in English courses or how to fit them into their schedule.

Of the 47 comments under the mandatory code, 22 came from participants in the Social Sciences (SSc) group, two from Health Sciences (HSc), 11 from participants in the Unknown Discipline (UD) group and one from the PNSc group.

The following two excerpts exemplify feelings that the test hinders their professional development.

*“That it should not be an obstacle to get the degree.”* Q1UD7

*“That it should not be a degree requirement, thanks to it, many former students haven’t gotten their degree yet.”* Q1UD169

Other 14 participants also view the test as an obstacle to their career completion. All these comments are believed to be from former students who have taken the test at least once but have not yet passed it. An example of this situation is shown below:

*“Many former students did not pass the test, and it is urgent for us to get our degree.”* Q1UD54

The urgent need to pass the test was also mentioned in the interview by a participant from the social sciences group SSc22:

*“It is two years since I finished school, and I cannot get my degree... I already have my test; my CENEVAL\* I have already passed it.”* SSc22

\*CENEVAL: The National Degree Certification Test

This participant states that the university-leaving test prevents her from completing her degree. Three other students, also from the SSc group, commented on the same situation as expressed by SSc26 during the interview:

*“It has been a lot of time, and we cannot get our degree; they should give us the option of getting our degree and then be able to do the test because ... really, sometimes in our jobs, we need the degree, and we cannot get it because of this regulation.” SSc26*

SSc26 spoke on her behalf and of her colleagues SSc25 and SSc27 as they were interviewed together (they nodded in agreement or agreed). They are dissatisfied with the language requirements as they will not let them proceed with the graduation process. Rather than considering passing the test as an asset for professional development, test takers undermine its relevance and question its compulsory nature. Test takers' perception of the test may be a powerful demotivator to study the language and pass the test as their “attitudes towards English play a major role in learning the language” (Altasan, 2016, p. 1135).

Even though test takers are aware of the high-stakes nature of the test, they feel discouraged. Their comments resemble feelings reported by Bai (2020), where test-takers' previous unpleasant test experiences (not passing the test) may be “discouraging rather than encouraging test takers to learn” (p. 12).

Keblawi (2022) accounts for several studies on motivating and demotivating factors. Most of the studies reviewed report tests could be a motivating or demotivating factor for language learners. His research found tests could demotivate learners *before*, *during* and *after* taking the test. Finding the motivation to study to pass the test presents a more significant challenge for those who have already failed to pass the test.

Students' attitudes, perceptions and feelings towards the compulsory nature of the test contrast those of an Indonesian study on the use of a proficiency test as a graduation requirement. Regardless of the discipline, survey responses reveal students *strongly agree* (SA) (38.2%) or *agree* (A) (52.1%) that the proficiency test motivated them to study. A similar percentage was observed when students were asked if the test benefited them (A= 39.7% and SA= 53.3%). Although a little lower, students' perception regarding the test not being an obstacle to following academic activities: 28% SA and 48.6% A (that the test is not an obstacle for them). The positive tendency was not maintained when participants were asked about the test's difficulty. The

highest percentage was for a neutral (N) response (neither agree nor disagree), with 64.1% of participants saying they neither agreed nor disagreed that the test was easy to respond to. However, a slight tendency towards positive perceptions on the test was reported: 17.6% A and 6.1% SA. They were also asked about the test preparation courses. For Participants SA (34.5%) and A (42.8%), the test preparation courses were important; this is aligned with Rohman's (2019) comment regarding test-takers' opinions on test preparation courses. The contrasting results are likely related to the context in which they were generated. The historical and geopolitical contexts of the countries are dissimilar. Even though the British rule of Indonesia dates back to the early 1800s, English may have had a more substantial influence in the region where Indonesia is located. On the other hand, the increasing presence of English in México dates back to the early 90s decade with the signing of the NAFTA (National Free Trade Agreement between México, the United States of America and Canada). The 2022 EPI reports English proficiency in Indonesia is higher than in México, occupying 81 (low proficiency) and 88 (very low proficiency) positions, respectively (among 111 countries ranked) (Education First, 2022).

Opposite opinions were also expressed, albeit in much smaller numbers, as exemplified by the following comments:

*"I think it is a good idea... that of an exam, and I think it is excellent that it is compulsory."* Q1SSc482

*"That it is fine that it is a requirement."* Q1UD228

*"Of great importance to be a better-qualified professional."* Q1SSc554

These participants consider the test to be a positive requirement. The latter sees advantages in this for his professional practice. As Bai (2020) reported, the comments above reflect a test-unrelated motivation for learning English that closely matches the L2 Ideal Self in Dörnyei's L2 Motivational Self System (2009). These test takers may be willing to study hard to achieve their ideal self-image and to become professionals with better opportunities due to their English proficiency.

Other comments in favour of the test relate to the need to link the test and the discipline, as shown below:

*“I think it would be good if there were a higher and specialized level according to the degree you are in, for example, legal English, English for business, medical English, etc., this would open doors to all of them.”* Q1SSc572

This participant does not explicitly agree to the test as a prerequisite for career completion; however, his comment suggests that he does. He believes that the level and content of the courses, and consequently, of the university-leaving test, is limited. Having specialized content for students of different disciplines would be an advantage. Other comments relating to specialized content do not have the same positive tone as the above. An example of this is shown below:

*“I think that if it is to get the degree...where are the topics of our discipline?”*  
Q1UD20

*“It should be offered separately for each discipline and focus on the language inherent to each discipline, so it has real value for professionals.”* Q1SSc561

The first comment questions the relationship between the content of the test and the purpose of taking the test. Q1UD20 suggests that if passing an English test is a graduation requirement, this test should be discipline-related. The second comment also favours having subject-specific tests and considers the general English test they took irrelevant to her profession. After analysing the proficiency level of individuals in 111 countries and regions worldwide, the 2022 issue of the English Proficiency Index (EPI) (Education First, 2022) presents a series of recommendations. It is suggested that besides including English language requirements for all university majors, having subject classes taught in English and learning relevant vocabulary related to each discipline would benefit students. Many universities in México and other non-English speaking countries have included an English language requirement. However, in many universities, this requirement is not subject-specific (see Appendix I). According to the 2022 EPI issue, México occupies the 88<sup>th</sup> position among the countries participating in

this study. A *very low* proficiency index was identified, corresponding “to the lower half of CEFR level B1 and A2” (p. 33). This low average level suggests that subject classes taught in English are pretty unfeasible, especially in some areas of the country where contact with English outside the classroom is scarce. However, contact with the language relevant to disciplines can take place in the language classroom, where at least discipline-related vocabulary can be learned and practised. Within this scenario, a discipline-related exit test would be the natural way to assess the language learned in the classroom. It may provide intrinsic motivation because the language assessed is related to their field of study.

The above three test-taker comments (Q1SSc572, Q1UD20 and Q1SSc561) relate to the content validity of the test because they felt that a test containing discipline-related content (L. F. Bachman & Palmer, 1996) would be relevant to them.

Most of the above comments do not favour the university-leaving language requirement. Many are from test-takers who have taken the test more than once and failed to pass. It is therefore not surprising that they do not have a positive attitude or opinion about the language requirements. Their negative experience taking the test may be hindering their willingness to engage in studying the language, hence, to pass the test (Dörnyei, 2019b). These comments give a first insight into the face validity of the test from the perspective of some test participants.

### 5.3 Test content

Two aspects of the test were included in the questionnaire: a) what test-takers believe the test (they just took) assessed (Section 1 in Q1, see Table 4.3 above) and b) what test-takers believe the test should assess (Section 2 in Q1, see Table 4.4 above). In this section, the test-takers were also asked what language skills they expected to need in their future discipline-related job. The opinion of test takers about what the test measures is discussed first, followed by what they think the test should measure. Finally, the results of the analysis of the skills they expect to need in their future work are discussed.

For the analysis of these data, participants were divided into four groups, as indicated in Table 4.4 above. Data and results from the unknown discipline group were retained because they helped to understand the overall mean per item in the questionnaire. Still,



they are not included in the comparison between groups. The latter is because these data do not help much in understanding the similarities or differences between known discipline participants.

### 5.3.1 Section 1. What test takers think the test assessed

This section looks at the opinion of test takers about what was assessed in the test they just took. Competences in this section were divided into three groups to facilitate analysis. This analysis shows whether they think academic, work-related, or personal/general language was assessed. The items were also grouped according to the skills they contained: receptive or productive.

#### 5.3.1.1 Workplace or academic-related language

##### Items 2 to 5 and 12

On items 2 through 5, test-takers were asked whether the test measured language related to work (TTQ2, TTQ4 and TTQ12) or the academic domain (TTQ3 and TTQ5). Item TTQ2 referred expressly to written language. TTQ4 and TTQ5 relate to oral language in work and academic settings, respectively. Item TTQ3 relates to language skills for study purposes. A comparison of the means for items 2 to 5 between the different groups reveals some differences, as seen in Appendix Table 2Appendix X below.

The PNSc group reports the highest mean scores on two (TTQ2 & TTQ4) of the three items related to language at work. The highest mean for TTQ12 is from the SSc group. The lowest mean for the workplace language items comes from the HSc group. For the items on academic language (TTQ3 & TTQ5), the highest mean values come from the PNSc and SSc groups, respectively, and the lowest mean values come from the HSc group.

The highest mean values, ranging from 3.07 to 2.63, indicating that the participants overall range between quite agree (3) and quite disagree (2) values, with a slight tendency at the quite agree end. The lowest mean scores (highlighted in lighter grey) between 2.51 and 2.78 indicate that participants generally do not really *quite disagree* (2) that these competences were measured, and there is a slight tendency towards the *quite agree* (3) end.

Although SSc participants have the highest mean score in either of the two competences related to language for study purposes, it was one test-taker from this group (Q1sSc373) who (at the end of Q1) provided a written statement on the test content and language for academic purposes, as exemplified below:

*“It (the test) is very simple and basic; it is not enough for postgraduate studies in an English-speaking country.” Q1SSc373*

This social sciences student feels that the test did not measure the language required within the academic domain, which would be required for studying abroad. Although he cannot specify the content of a language test for study purposes, he believes that such a test should be much more challenging than the level of difficulty offered by the current test. This comment questions the face validity of the test; as for him, it does not match the expectations of an exit test. He considers the purpose of an exit test to be to ensure candidates are prepared to study abroad (Phillips et al., 2020).

The highest standard deviation values (highlighted in darker grey) are from the SSc group (1.070 and 1.048) for TTQ2 and TTQ3, respectively, from the PNSc for TTQ4 (1.081) and the HSc for TTQ5 (0.996). The lowest means (highlighted in light grey) are from the PNSc group for TTQ2 (1.020) and TTQ5 (0.968), from the SSc group for TTQ4 (1.005) and from the HSc group for TTQ5 (1.008). The standard deviation indicates the degree of disagreement among members of the same group. These results suggest that there seems to be an association between the disciplinary group and the competence for the standard deviation values.

A comparison of the total means of TTQ2 to TTQ5 and TTQ12 shows that participants overall are more likely to agree that the test covers both study-related language (TTQ3 and TTQ5) and work-related language (TTQ2, TTQ4 and TTQ12). Almost all (except for the PNScs for TTQ3) range between *quite agree(ing)* (3) and *quite disagree(ing)* (2) that they encountered activities representing work or study-related language in the test they just took.

5.4.1.2 Receptive skills, social/general purpose domain

Items 6 to 9

Items 6 to 9 of the receptive skills for personal or general purposes also show differences in the means between the groups.

Items TTQ6 and TTQ7 relate to listening comprehension. The PNSc group reports the highest mean scores (2.93 and 2.87, respectively) on both items, while the lowest mean scores (2.81 and 2.67, respectively) are from the HSc group. These values indicate that the PNSc group was more likely to agree that the test assessed their ability to understand clear standard speech in unfamiliar leisure environments (TTQ6) and the main points in TV programs on current or familiar topics (TTQ7). The lowest mean scores are from the HSc group in the four competences (TTQ6, TTQ7, TTQ8 and TTQ9), with scores ranging from 2.67 to 2.81 for the receptive skills in this group, as shown in Appendix Table 3 below. Accordingly, the HSc group was the least likely to agree that the test measured these competences.

Items TTQ8 and TTQ9 relate to understanding written instructions, such as manuals, or describing events, feelings or wishes of personal interest, respectively. The highest mean for TTQ8 is from the PNSc group (2.80), while the SSc group reports the highest mean for TTQ9 (2.81). The lowest mean values (2.69 and 2.75, respectively) come from the HSc group.

The range of the mean values indicates that the answers range between *quite agree* (3) and *quite disagree* (2) but closer to the *quite agree* end.

The highest standard deviation values for listening competences come from the physical and natural sciences (TTQ6, 1.202 and TTQ7, 1.106) and reading competences from the social sciences (TTQ8, 1.05 and 9, 1.006). The lowest standard deviations are in the SSc group for TTQ6 and TTQ7. For TTQ8, this value is from the HSc, and for TTQ9, this value is reported from the PNSc group.

5.4.1.3 Productive skills, social/general purpose domain

Items 10 to 18

Almost all competences in this group relate to familiar topics: oral language required when travelling (TTQ10), ability to spontaneously engage in conversations on topics that are familiar or of personal interest (TTQ11), ability to spontaneously engage in everyday conversations or familiar topics (TTQ13), ability to describe personal aspects (TTQ14), to provide reasons or explanations for plans (TTQ15) and ability to tell a story or describe reactions to movies or plots (TTQ16). Among these speaking competences, the PNSc group reports the highest mean for all. The lowest mean comes from the HSc group for all items except TTQ16, where the low value is the same for the HSc and SSc groups. These results can be seen in Appendix Table 4 below.

While the highest mean values represent higher agreement with the assessed competences, the lowest mean values mean lower agreement.

Regarding the writing-related competences, the same pattern can be observed for the competences in this section. TTQ17, writing simple texts on familiar topics, and TTQ18, writing personal letters describing experiences, give the highest mean scores from the PNSc group and the lowest mean scores from the HSc group.

Importantly, all of the highest mean productive skills scores are above 3, *quite agree*. Although the highest score (3.20) is closer to the *quite agree* end than the *agree* (4) end, these high mean scores are higher than most in the previous sections.

The highest standard deviation values for TTQ17 are from the HSc and the SSc groups. The lowest standard deviation values are from the PNSc group.

The overall results show that a more significant number of test takers agree that the productive competences related to the social/personal purpose domain were measured, as these competences have higher total mean values than the school or work-related competences or the receptive skills competences analyzed above. The PNSc group has most of the highest mean scores, while the HSc group has the lowest mean scores in more competences. The highest mean values represent respondents' agreement that the selected competences were assessed in the test they took. The

lowest mean values represent the opinion of the respondents between *quite agree* and *quite disagree* that the competences were measured in the test that was just taken.

A male health science participant’s comment, HSc85's, illustrates test-takers' perception that the certification test was in the personal domain. When asked during the brief semi-structured interview to select one of the three domain options (social, academic or work) on which to focus the test, he explained:

*“Well...talk about the family and all that, as we just did...hum, I just took it and found it very comprehensive.”* HSc85

HSc85 indirectly chooses the social domain when he says he would choose a test like the one he just took. It was probably during the oral exam component that HSc85 had to *talk about family*, which he is likely to associate with the social domain. Interestingly, HSc85’s expectation of what the test is expected to measure differs from Q1SSc373’s above, for whom the test does not meet expectations.

Many participants felt that the social/general-purpose domain (including the tourist domain) was assessed. Even if some participants consider the work and study domains were also assessed, the mean values for the competences for assessing these domains are smaller overall than those from the social/general purpose domain.

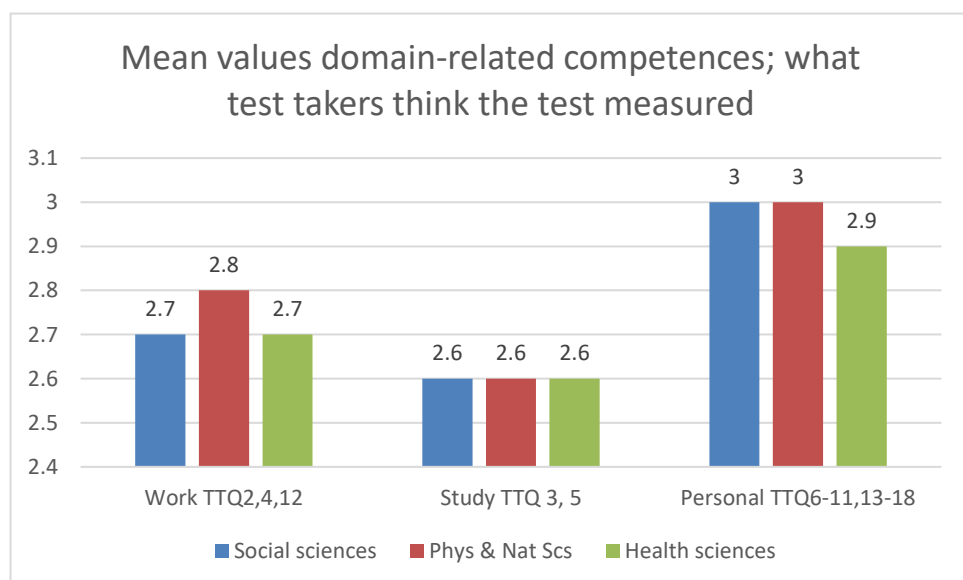


Figure 5.2 What test takers think the school exit test assessed

The mean value of the domain-related competencies per disciplinary group shows that SSc and PnSc participants have the same opinion regarding personal domain competences in the test they took. The HSc group is not as unanimous as the other groups in that they encountered the personal domain competences listed in the university-leaving test. The PnSc group has a stronger opinion that the work-related competences are part of the university-leaving test they just took. In contrast, the SSc and HSc groups share their opinion about the encounter with the test tasks for work-related competences. These results can be seen in Figure 5.2 above. Finally, the three groups share the same opinion about encountering study-related competences in the university-leaving test. These answers indicate the lowest mean, although they tend slightly towards the quite agree end.

### 5.3.2 Section 2: What test takers think should be assessed

#### Items 27 to 34

This section looks at the competences test-takers believe the test should assess and the skills they feel will be used more frequently in their future discipline-related job. Findings from qualitative data are integrated with quantitative data when the former supports, explains or appears to contradict the latter.

For practical purposes, items are classified and analysed according to their content. Items related to the academic (TTQ29) and workplace (TTQ28 and TTQ31) domains are compared, and those related to personal interests (TTQ27, TTQ30, TTQ33) or family (TTQ32) are analysed together.

The PnSc group is more approving of a test that includes work-related language (TTQ31, mean value = 3.23) and travelling for business purposes (TTQ28, mean value=3.07) than the other two groups. These results can be seen in Appendix Table 5 below. However, the SSc group is more inclined to take a test that assesses academic language (TTQ29, mean value= 3.08). The lowest mean scores for these three competences come from the HSc group, ranging from 2.84 to 3.09. The HSc group does not seem too enthusiastic about a university-leaving test assessing work or academic language.

Language related to entertainment (TTQ27) is also highly valued by the PNSc (3.27), and the HSc group seems the least interested in this content (2.93). The SSc group is the group most interested (3.25) in language, which allows them to talk about themselves and the things they like doing, eating, etc. (TTQ30), and the HSc is the least interested group (3.15). Talking about family (TTQ32) and travelling for pleasure (TTQ33) seem to be more relevant for the PNSc group (3.30 and 3.37 respectively) than the other groups, especially the HSc group, with the lowest mean values (3.08 and 3.14 respectively).

The highest mean values in Section 2 are above Section 3. Responses range between the *quite agree* (3) and the *agree* (4) range but clearly lean towards the former.

The lowest standard deviation value reported in four competences from the PNSc group shows a more homogeneous distribution of responses, with values between 0.809 and 0.980 (TTQ27, TTQ28, TTQ32 and TTQ33), with the other three low standard deviation values coming from the HSc with values between 0.922 and 0.949 (TTQ29, TTQ30, and TTQ31). The lowest standard deviation was for the competence of travelling for pleasure (TTQ33). On the other hand, the least homogeneous distribution of responses comes from the PNSc group for the competence of talking about oneself (TTQ30, 1.073). Five high standard deviation values come from the SSc group on competences related to studying abroad (TTQ29, 1.019), business travelling (TTQ28, 1.016), the language for the workplace (TTQ31, 1.001), travelling for pleasure (TTQ33, 0.948) and talking about the family (TTQ32, 0.928). The remaining high standard deviation value is from the HSc group for the competence related to entertainment (TTQ27, 0.922).

A summary of the mean scores of the competences test-takers think the test should measure can be seen in Figure 5.3 above. The total mean was calculated for C27, C28, C30, C32 and C33, as they all represent the same domain. As can be seen, the SSc and the PNSc groups have similar opinions on domain-related competences. The three groups seem to consider the social and work domains more relevant than the study ones. The PNSc and the SSc groups seem to place as much importance on the social domain as on the work domain, with a slightly increased interest in the work domain on the part of the PNSc group. For HSc participants, the work domain is seen as more interesting than the social domain. The study domain seems to be the one where participants are less interested.

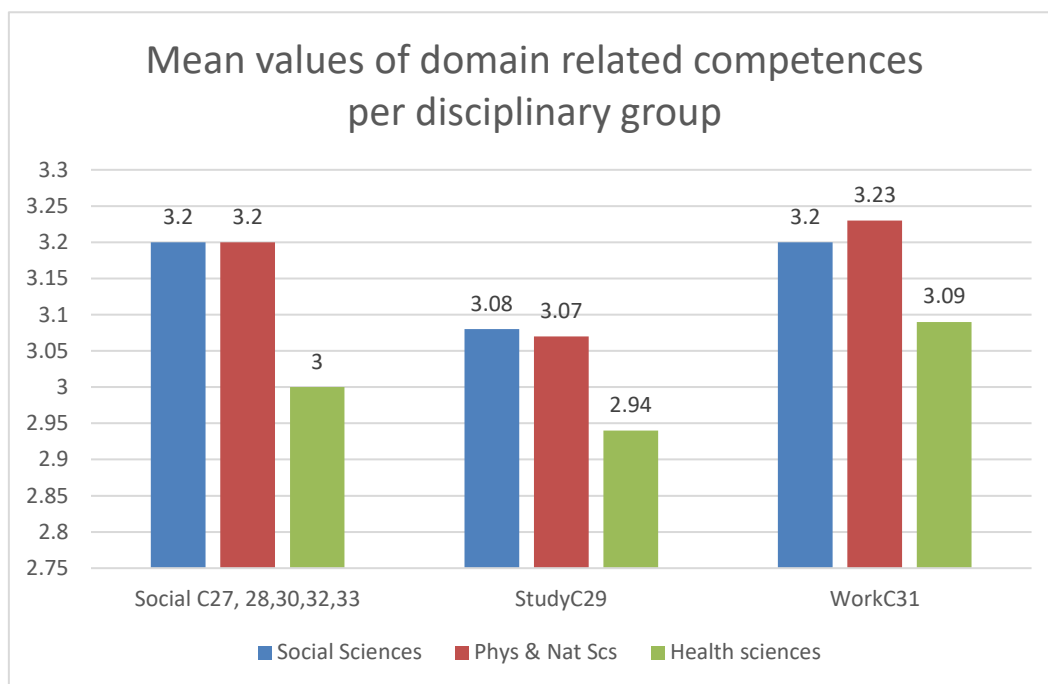


Figure 5.3 What test takers think the test should measure

When asked which language domain they would prefer to include in the test, participants' comments indicated that they were undecided as they were interested in more than one domain. The following excerpts make this clear:

*“Only one?”* (SSc55),

and

*“The three of them”.* (HSc103)

Response from SSc55 does not say if she is interested in two or three domains, but one domain does not seem enough. This feeling is shared by HSc103, who initially refused to choose one of them. The comments from these test takers seem to reveal interest; there is a positive attitude toward learning English. A positive attitude towards the language has been related to a positive attitude towards the test (Keblawi, 2022). Within the same idea of not being limited to one domain, another male participant, HSc85, emphasized his interest in learning English thoroughly:



*“Yes, everything...everything because it is a language that is being instilled a lot here and...we need to speak it, all topics, family, hum...work, friends, hum...there is a lot, it is being instilled a lot here, and I think it is good to learn it 100%”. HSc85*

The excerpts above show the importance that English has within the local community. Whether it is parents, local authorities, teachers, or all, this comment confirms the emphasis on learning English. However, this does not seem related to any particular context. Whether the comments reveal interest in more than one domain or they have not considered that English could be learned for a specific purpose, participants seem interested in using English in all contexts in which they may engage. The comment made by HSc85 reveals high motivation towards the language. The phrase “...it is...being instilled a lot here...” reveals the belief of the external expectations of the attributes an individual is expected to possess. Repeating the same phrase a few words after reinforces the understanding of the ought-to L2 self (Dörnyei & Ushioda, 2021).

Socializing is seen as an essential activity by some participants, but they do not see it in a general context. When they think of socializing, they consider it within the work environment, as expressed by male test-taker HScs13, who said:

*“...I think all of them are important, but the one that prepares us specifically for social harmony situations related to our discipline...at the job...I think should be the best option.” HScs13*

For HSc13, the social aspect within the work environment seems the most relevant.

Participants expressing interest in more than one domain (36 of 215) were asked to select one during the interview. However, when answering Q1, there were no restrictions on the choices.

Mean results could also be explained by looking at other types of participant comments. While some test-takers do not seem to think one domain is sufficient,

others believe language cannot be broken down into domains. Female participant HSc6F explained:

*“...you can’t teach exclusively one of the three because in one, while you are learning one, you are teaching the others without having the intention.”* HSc6F

This participant believes that language cannot be divided into the domains mentioned. For her, language is an indivisible system. Two other examples of comments against the compartmentalization of language domains were made by HSc test-takers (HSc5M and HSc1M, respectively) below. Both see the social domain as all-encompassing.

*“...the social encompasses all the relationships, not only with family but with friends, superiors, etc.”* HSc5M

*“It (the social domain) includes everything. Includes from work and relating to others, with other people...”* HSc1M

The social domain is not only seen as the overarching domain encompassing all types of social interaction but as the underlying domain that provides the essential elements to use the language in every other domain. Two female test takers from the SSc groups SSc24F and SSc75F (both below) express how they think the social domain plays a central role in language learning.

*“The social domain is vast, and knowing how to communicate in a basic way...that will allow us to apply or use it in our work domain, but it is necessary to know it, because (...) how am I going to communicate with someone in the work domain if I do not know how to communicate in the social domain?”* SSc24F

*“...(the social domain) is where everything starts, you start developing and if you know how to express yourself in society, to understand cinema and that, you can later understand books and from there it*

*would take you to the work domain and interact with your colleagues and talk about your job or a topic in their work domain.” SSc75F*

Apparently, mastering the social domain leads to developing the ability to use the language in another domain. It seems to be believed that once mastery of the social domain has been achieved, the ability to use the language in the academic and professional domains comes naturally. Participants seem to view the work and academic domains as general domain language “with technical vocabulary thrown in” (Douglas, 2000, p. 1), and learning the language needed to socialize is all that is required, as the specific uses of the language “will take care of themselves” (ibid. pg. 1).

Some test-takers believe that learning a domain is the starting point for learning the other two or that they are included in learning a domain. However, not everyone thinks that the underlying domain is the social domain. Other test takers think it is another domain which should be learned first. The following excerpts reveal how they consider one domain comprehensive or develop the others. A female student, SSc82F, explains how she considers the work domain will lead to the development of the social domain:

*“I also consider that the third option (work domain) (is the one I prefer), because from there the social starts to break down and then we can...like get closer to people and talk to them.” SSc82F*

SSc82F above likely thinks that the social domain is related to the language required by a professional in the social sciences and not to personal socialization. She may feel that the work domain will provide her with technical language (e.g., vocabulary) and the specific purpose language ability required to interact within her professional practice. A similar interpretation may be given to the comment made by male student HSc48M:

*“...because I think that interacting... when teaching you to interact with patients and the people you work with, goes hand in hand with the social... because you need to socialize at work too...then, I think it would be*

*like...when you study for the work domain, you would be studying indirectly for the social domain.” HSc48M*

This comment reveals an awareness of some of the language activities his future work as a health care provider will require: interacting with patients (Douglas, 2000, p. 94) and colleagues. His priority seems to be the work domain. He might consider that once he can interact with patients, he will be prepared to interact personally. While greeting and leave-taking may be a competence used in both the social and health domains, the nature of doctor-patient interaction (asking fact-finding questions to reveal the condition's origin) is specific to this domain. Talking about personal preferences is a competence found in the social domain and is unlikely to be commonly observed in interactions within the health care provider's office. Some language competences are domain-specific.

Participants seemed unfamiliar with the specific language features of the three domains they had to choose from. An example of this can be seen in the following excerpt, where a female SSc student (SSc80F) stated:

*“Ummh... I would choose the work domain because once you start reading articles and all... it is easier to learn the social and the academic...” SSc80F*

The above comment seems to reveal unawareness of the competences related to the domains. SSc80F considers reading articles to be part of the work domain. Even though reading articles may be part of the activities undertaken in some work situations, this activity is more often related to the academic domain (Douglas, 2000). The commentary also reveals misunderstandings regarding the subskills required to read articles and the linguistic features of the social domain. SSc80F thinks the work domain is the foundation for developing social and academic domains.

The academic domain is also considered the basis for developing other domains. The following excerpt exemplifies how female student HSc110F believes that domains are related and evolve from others:

*“...and if you learn (English) in the academic domain, you will also learn it in the work domain, and it will help you develop professionally (in your work) and then...uummh...the social... umm...comes hand in hand, I think the social will derive from the other domain...”* HSc110F

There were another four comments from participants in the HSc group (HSc106, HSc23, HSc53 and HSc54) under the same tenor. They also believe that learning the language in the academic domain will lead to the development of the work and social domains.

Test takers may think that language cannot be divided into domains because they are unfamiliar with the boundaries between them. On the other hand, participants may be aware that some competences are useful in more than one domain (COE, 2001). Although social interaction also occurs within the academic and work domains, the nature of social interaction within each can have specific characteristics. The ability to use the language to socialize at a personal level will not automatically translate into socializing within the academic or work domains. They may also assume that the academic and work domains are similar since they require discipline-specific vocabulary. This argument does not seem to consider that language use varies according to the domain of use (Douglas, 2000).

While it is true that socializing is an activity that is expected to take place in several domains, e.g., school or workplace, the extent to which this occurs varies greatly. This variation does not only depend on the discipline but on the type of activities that the context requires. It is possible that when voicing their thoughts, many of these test takers were thinking of themselves in an international setting; they were thinking of their *Ideal-Self* (Dörnyei & Ushioda, 2021). Studying abroad or working for a multinational company would undoubtedly require more extensive language command than living in an environment where most day-to-day social interactions take place in the native language of the country (Douglas, 2000). In such situations, the presence and need of the target language are very different.

In summary, many comments above indicate that participants are unaware of the characteristics of the three language domains on which they expressed their opinions. The inseparability of domains (e.g. HSc6F above), the perceived innate ability to

transfer language knowledge from one domain to another (e.g. SSc24 above), the overarching characteristic of a domain (e.g. HSc5M) and the belief that when one domain is learned, another will take care of itself (e.g. HSc110F, HSc48M above) illustrate this. On the other hand, it could be argued that the participants do not limit themselves as users of the language to the professional domain but see themselves as the whole person who is part of society. Whether employed or aspiring to be employed, they are and will always be “a family member, a stakeholder in local and national elections and, of course, a colleague in the workplace” (Huhta et al., 2013, p. 15). They see that they use English in more than one context, and while they may (or may not) find English useful or not for professional practice, it will be required for other purposes.

All respondents who initially supported the selection of more than one domain were asked to select only one and justify their choice. In general, comments from test-takers regarding their preferred domain vary. The excerpts above show that although one domain is chosen, they are also interested in the other domains. Their choice is because they believe gaining control in one domain is necessary to develop the other domains. This reasoning may also help explain the origin of the high overall mean values seen in Appendix Table 6 below.

The interview responses on the preferred domain do not entirely confirm the quantitative analysis findings in Figure 5.3 above. According to the domain selection in the interview, HSc participants prefer the study domain, followed by the work domain. Fewer participants chose the social domain, as shown in Figure 5.4 below, which shows the percentage of discipline participants coding in each domain. Among the social science participants, the social domain was chosen by almost as many participants as those who preferred the work domain. The number of SSc participants who prefer the study domain is about half that of the other two domains. Finally, responses from the PNSc, with a much smaller number of participants, show that more of them are interested in the work domain, followed by the study domain. The social domain was not chosen by PNSc participants.

It could be argued that the quantitative results representing mean values cannot be compared to the qualitative results since these represent percentages. However, it can be contended that Figure 5.3 and Figure 5.4 represent the tendency of the data they analysed. They would have been very similar if the preference for the selected domain

had been retained in both instruments. There is no significant difference between the SSc group's qualitative and quantitative result behaviour. The questionnaire results show equal interest in the social and work domains, while the interview results show a slight preference for the work domain. The study domain does not appear as relevant or important for SSc respondents as the other domains.

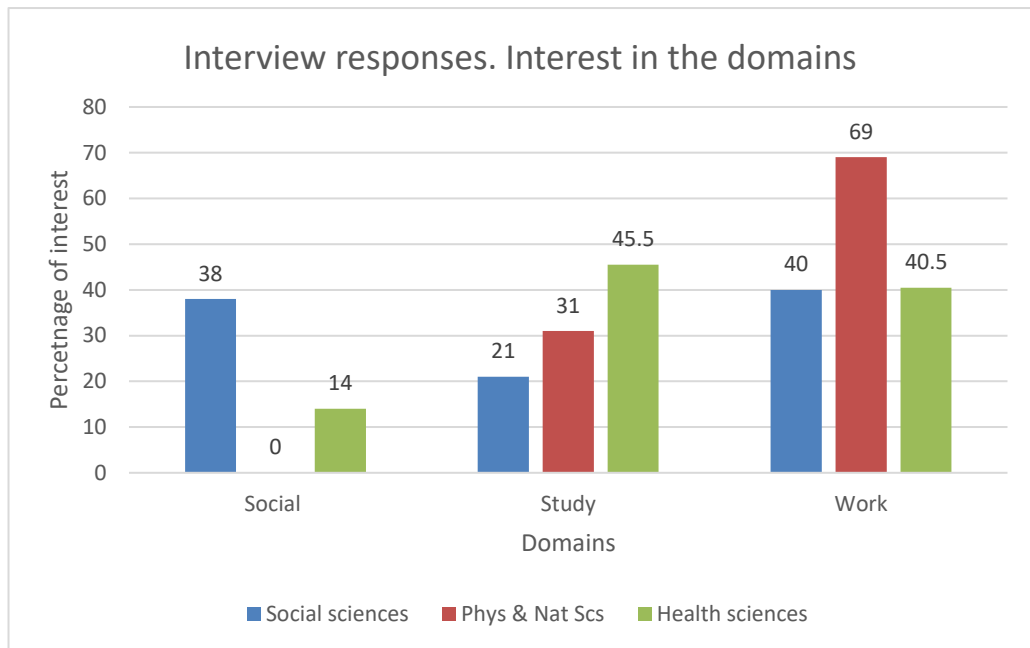


Figure 5.4 Interview domain preference per disciplinary group

The work domain is still relevant for HSc participants, but in contrast to the questionnaire results, more HSc respondents are interested in the study domain than other domains. The study domain seems to be less important than the other domains. A smaller number of HSc participants state that they are interested in the social domain.

During the interview, the relatively small PNSc group reported that they were more interested in the work domain, followed by the study domain. No participants stated that they were interested in the social domain. These results differ significantly from those obtained from questionnaire data.

The work domain has the largest or second largest number of participants (among the three domains) who want the university-leaving test to focus on that domain. Further variations in data results are reported from the social and study domains.

Overall, the interview results show that more participants are interested in the work domain (100), followed by the study domain (81). A few participants stated that they were interested in the social domain.

Several aspects may have contributed to this discrepancy. First and foremost, the questionnaire allowed respondents to select any competence they were interested in, while the interviewer asked them to select just one skill. Interview responses such as “*the three (of them)*” (HSc 103M) show interest in more than one domain. Several interviewees may have been interested in more than one domain but did not express this as they were requested to select only one.

Another aspect that may have contributed to the results obtained is awareness. It may be that the test-takers did not think thoroughly when choosing the competences during the questionnaire. However, they may have been more reflective in their responses during the interview, as they had to verbalize and justify their choices. Additionally, participants may have reflected on their peers’ responses during the group interviews. They may have reconsidered their opinion after listening to peer comments. When asked for their opinion, on more than one occasion, some participants limited their responses to saying they agreed with their peer’s opinions.

While most students chose a domain and justified their response, others could not reconcile their wants and needs (West, 1994). Although they acknowledge the importance of a domain because of its relationship to their future professional life, they would rather have the test focus on another domain because of their interest in it. An example is the comment of a female student, HSc26F, who expressed an interest in the social domain. However, she acknowledged the importance of the academic domain in her professional life, as seen below:

*“Well, I would choose the social because...well, the academic is important and much more in our discipline, but...I think that to learn a language... it is first social.” HScs26F*

She seems to have put aside the domain that will be useful in her professional life due to her interest in the social domain. While aware of the need, this does not seem to be



what she wants. As West (1994) noted, needs and wants sometimes conflict; test-takers may choose what they want rather than need.

### 5.3.2.1 The importance of the work domain

Regarding the percentages within the domains, the PNSc group reports a higher percentage of participants coding in the working domain, while the percentage for the HSc and the SSc is almost the same.

Interviewees made several comments about the importance of being able to use English for work-related activities. Some of them consider that the work domain will be an asset for their professional development, as PNSc4 put it:

*“To be well prepared for our degree”.*

Another participant spoke about the possibility of needing English in a professional environment; PNSc5 said:

*“English is very used in any job and more in my degree.”* PNSc5

English proficiency is also considered a door opener, as expressed by HSc50M below:

*“...because the opportunities where we can work are widened.”*

When choosing the work domain, some students, like SSc47F, justified their answer by referring to its usefulness;

*“...because it would be more useful”.* SSc47F

Others, like HSc3M, mentioned the kind of activities they would need work-related English for in the near future:

*“...either talk in English with our patients, understand texts, and well, being able to translate articles downloaded from the internet, they frequently come in English”.* HSc3M

Some already working participants said they did not think they would need English in the workplace and were surprised when they did. SSc61F gave an example:

*“Hum, because I have encountered cases... I thought I would not need to (use English) in my career, cases in English. But I have had several in which I have (needed English)”.* SSc61F

Other participants also felt that being able to speak English would give them an advantage over non-English speaking professionals in their same discipline. The 2022 English Proficiency Index (Education First, 2022) states that speakers of other languages who can communicate in English “are better equipped to collaborate internationally with partners and within their own organization” (Education First, 2022, p. 5). The comments made by these students reveal they acknowledge English as an asset for their future professional activity. The investment made by organizations, governments and individuals in English instruction has been found to be of benefit to those involved (Education First, 2022).

Within the recommendations based on the data analysed, the 2022 English Proficiency Index (Education First, 2022) argues that by “recognizing the essential role of the workplace in driving English acquisition” (p. 29), country leaders can take advantage of the global lingua franca to establish connections across borders and access the international arena.

#### 5.3.2.2 The value of the academic domain

For the academic domain, the highest percentage of participants comes from the HSc group, followed by the PNSc group. Although the SSc group has a more significant number of participants than the latter, the percentage of cases coding in this domain is the lowest, as seen in Figure 5.4 above.

There were two main reasons for participants to choose the academic domain as the focus of the university-leaving test.

The first reason was the desire for higher education. Participants consider proficiency in academic English to be a necessary tool, as HSc91 expressed:

*“Because, what’s next, that is, the specialisation (English) it is a must, well, from my point of view, it’s a must to do postgraduate studies”.* HSc91

Although most said they would study locally, studying abroad was also mentioned. This reduced number of students thinking of pursuing higher education in an English-speaking country could explain why very few mentioned the need to give academic presentations in English, although understanding articles was often mentioned, as HSc22M illustrates:

*“To be able to understand the articles that come in English more easily”.*  
HSc22M

HSc20M below mentioned not only the need to understand articles in English but also the need to understand spoken production, as students sometimes attend academic congresses. This is shown below:

*“Well, because in this and all (degrees), it is necessary to...it is necessary to be updated, and the recent articles are only available in English; besides, many conferences and academic events are in English, and we need the level to understand them”.* HSc20

The choice of the academic domain was not only related to their needs as students but was also seen as necessary for their later professional life, as expressed by HSc9F:

*“...it is necessary to prepare oneself, reading constantly and all those things”.* HSc9F

## Chapter 5 Understanding the test taker population

Most participants coding at the academic node, saying they need to understand written academic texts, are from the HSc group. In terms of percentages, the PNSc group also had a high percentage. Both the HSc and PNSc groups feel they need to stay current. They need English because most new information is shared in English. The comments above could be examples of what the English Proficiency Index (Education First, 2022) published in their 2022 issue. As English is more commonly used for international information exchange, it is “a key component for accessing knowledge and expertise” (p. 5). Reading discipline-related articles, as mentioned by the four participants above (HSc91, HSc22M, HSc20M, and HSc9F), is necessary to learn about the latest discipline-related innovations. However, reading comprehension is not enough to keep up to date as through video-streaming, it is possible to attend international conferences that take place in distant locations (Education First, 2022).

Many SSc students find that they do not need academic English as their studies are conducted primarily in Spanish due to the nature of their degree. Q1SSc26 illustrates this belief:

*“...it is unnecessary for many degrees. I am studying to be a lawyer, and my degree is based on Mexican law; therefore, I do not see why we have this English level”. Q1SSc26*

Although Q1SSc26 does not need English to study Mexican law, he might be surprised, as was SSc61 (above), and find that English is required for a job-related activity.

### 5.3.2.3 The interest in the social domain

The SSc group had the most significant percentage of participants coding in the social domain, followed by the HSc group—no participants from the PNSc group coded in this domain (Figure 5.4above).

Some participants indicated that they chose the social domain because they feel English is not required in their discipline. Hence, they would learn it for travel purposes, as SSc64M specifies:

*“(..) it is necessary, but in other jobs, not in my case, it is not so necessary, as would if I travelled or interacted with people from other places.” SSc64M*

Several participants who expressed interest in the social domain consider they need English for travel. Participants do not expressly state they would like to visit English-speaking countries. As the 2022 English Proficiency Index states, these students consider English the language that enables them to communicate with speakers of other languages when travelling to non-Spanish speaking places (Education First, 2022).

Other participants consider the social domain as the one they will use for work-related activities, as exemplified by SSc91F and SSc89F, respectively, below:

*“Well, in my area of psychology, it is more of human resources and an important position; for example, it actually needs to handle English. All the companies, in fact, handle English at the recruiter level of diagnostic test applicator; then, in management, an important position really needs a good level of English.” SSc91F*

*“Yes, I think it is the most important for our degree because we are in constant contact with individuals within society, and I think it would be the most important.” SSc89F*

It could be argued that SSc91F and SSc89F consider the work domain the technical language required in degrees with little human interaction. Although they feel their degrees require a lot of human interaction, they think the social domain is the one they need to use within the local work environment. They believe that the social domain will provide the necessary language to carry out discipline-specific activities, as in the case of SSc91, interviewing individuals for recruitment purposes.

Other participants see the social domain as more motivating for them, as HSc99F expressed

*“...because it is more appealing, more than academic that you must do it compulsory, it is more interesting to learn to understand cinema in English or being able to interact with... (participant got interrupted by another test taker)”*. HSc99F

Not all students consider the work or study domains to be the most important. Others are more interested in using English to communicate with people beyond borders or to enjoy international media (Education First, 2022). It could be thought that HSc99F is not aware of the usefulness of English for her studies and future professional life. However, when asked if she considers English important for her career, she said

*“Oh no! In my career, it is basic... English is very, very basic, ...all the manuals for specialization are in English; in the same way, if you want to join a private university or do a postgraduate degree, they have English as an entry requirement.”* HSc99F

Even though HSc99F is aware of the importance of English for job-related purposes, she is more interested in the social domain. The two HSc99F excerpts above illustrate the tension observed between participants' needs and wants (Hutchinson & Waters, 1987; West, 1994). More than one participant may share this preference.

#### 5.3.2.4 The relevance of language skills

The oral interview to test takers did not specifically ask participants to say which skills were more important for the workplace. However, 37 participants referred to the use or relevance of the skills in the workplace. These results will be addressed below in the discussion of the findings for each skill.

Item 34 has four sub-questions. Each asks participants the extent to which they think any of the four skills will be required in their future professional job. According to the total mean, listening, reading, writing, and speaking are the order of importance of these skill-related items. This is shown in Figure 5.5 below. The highest mean scores for listening (TTQ34b = 2.91) and speaking (TTQ34a = 2.58) by disciplinary area come from the SSc group, as shown in Appendix Table 6 below. The mean values for the

writing (TTQ34c = 2.67) and reading (TTQ34d = 2.93) items come from the PNSc group. The HSc has the lowest mean value for all these four items.

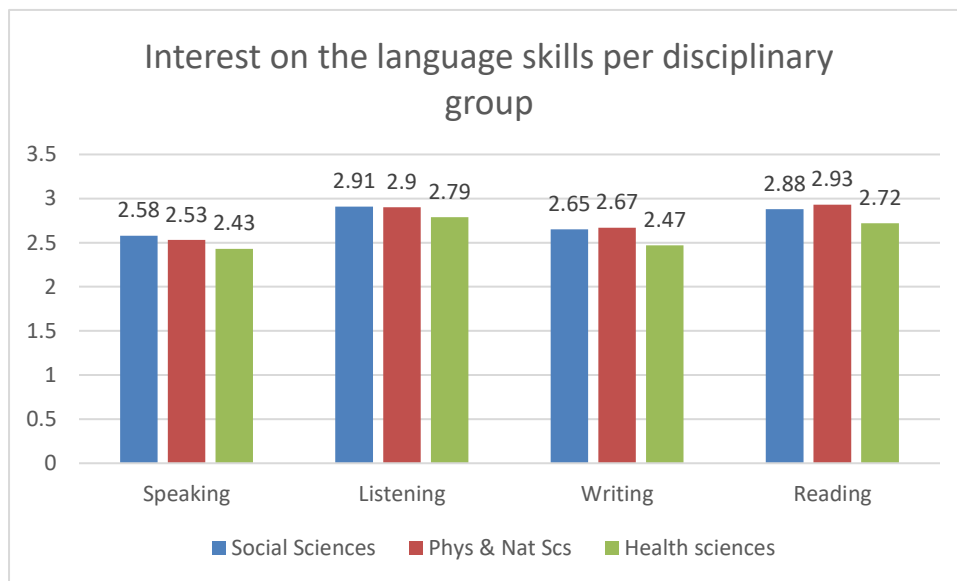


Figure 5.5 Interest in the language skills per disciplinary group. Q1 responses

The highest standard deviation value is reported by the HSc group for speaking (1.184), followed by listening (1.038) and writing (1.175), and by the SSc group, the skill for reading (1.147). The lowest standard deviation values are from PNSc for speaking (1.167), listening (0.995) and reading (1.081), and from SSc for writing (1.141). Almost all standard deviation values are above 1, showing significant differences between the answers.

During the interview, some participants mentioned that they considered more than one skill necessary or valuable for certain activities inherent in their professional practice. This belief is exemplified in the following exchange with an HScs student (HSc85). This participant said he lives in the northern part of the state of Durango, where communities of Mennonites who came from Canada settled. Their presence in Durango is minimal, as they only come to sell some of their products, like cheese and baked goods, and purchase goods and services they cannot find near their community. HSc85 said that many Mennonites speak English and have difficulty communicating in Spanish.

- Researcher: *“As a dental surgeon, what would you need to use English for? ...to speak? To understand what is being said or written... or to write it? What do you think you would need to do?”*
- HSc85: *“Well, write it because, as I just said, almost every patient we have requires us to explain what we are going to do; if we are giving them a prescription, we need to write... well written, so they understand it, and there will be no mistake because we handle medication, so then...for them (Mennonite patients) English is easier, and write it to them (after orally giving treatment instructions) that would be good”.*
- Researcher: *“So, you would need to know how to write a prescription, and you would need to be able to talk to the patient? And read?”*
- HSc85: *“Yes, yes and read it too, because in the field of dentistry many of the...hum what’s its name? The...the books and all of that...what’s its name? ...the articles, many come in English”.*

As a practitioner, this participant was currently confronted with the need to communicate with Mennonites in English, so he was aware of the needs in his professional practice. HSc85 recognized the need for oral interaction, reading and writing in his dental work. However, not all participants were as explicit or aware of the demands of their profession. The following exchange with two female HSc participants, HSc28F and HSc29F, interviewed together illustrates a different perception of the demands of the profession as a dentist:

- Researcher: *“Do you think English is important for medics here in Durango?”*



HSc28F: *“Here in Durango? I, I don’t think so. Not here in Durango. But...because here there are almost no...hum, situations like...taking care of foreigners or...something like that. And... there aren’t emergencies where you need to use English here in Durango. (...)”*

HSc29F: *“Well, I think we do, because well, (...) the best books, the best bibliography...the best articles, the...updates, really, well...are in English, so...it would be good that we learned to interpret that”.*

HSc28F: *“Yeah! Uh-huh!”*

It seems that HSc28F only reflected on the need to speak English during her professional practice without acknowledging that the use of a language is not limited to oral interaction. However, her colleague, HSc29F, did not mention the need to speak English, although she acknowledges the need to understand academic material in their field. Following HSc29F, HSc28F seems to agree that English would be relevant to her, although perhaps not as much for oral interaction as she initially said. Both participants do not consider speaking to be necessary in their professional practice.

During the interviews, it was common for participants to think of only one skill when considering the need to use English. During group interviews, students sometimes changed their initial responses after listening to colleagues expressing a different opinion, as illustrated by the abovementioned exchanges with HSc28F and HSc29F.

Questionnaire data shows participants in the three groups assumed that they would most likely need to receive information (reading and listening ranked first and second, respectively). Interviewees were not explicitly asked about the skills they felt they would need in their future professional roles. Still, voluntary comments helped the researcher understand the data results from the questionnaire.

Extracts of the participants’ contributions to each skill are presented individually below.

5.3.2.4.1 Understanding written English is necessary.

Questionnaire results show that understanding written English is the most relevant skill for future job-related activities for the PNSc group (mean = 2.93), the second most important for the HSc (mean= 2.72) and SSc (mean = 2.88) groups (Figure 5.5 above). A closer look at the data shows that the SSc group chose *agree* (4) more often than the other disciplines, as shown in

Figure 5.6 below. The number of participants who chose option 4 represents 42.8% of the domain group. For the HSc group, the number of participants selecting *agree* represents 32.5% and for the PNSc group, this option represents 36% of the group participants.

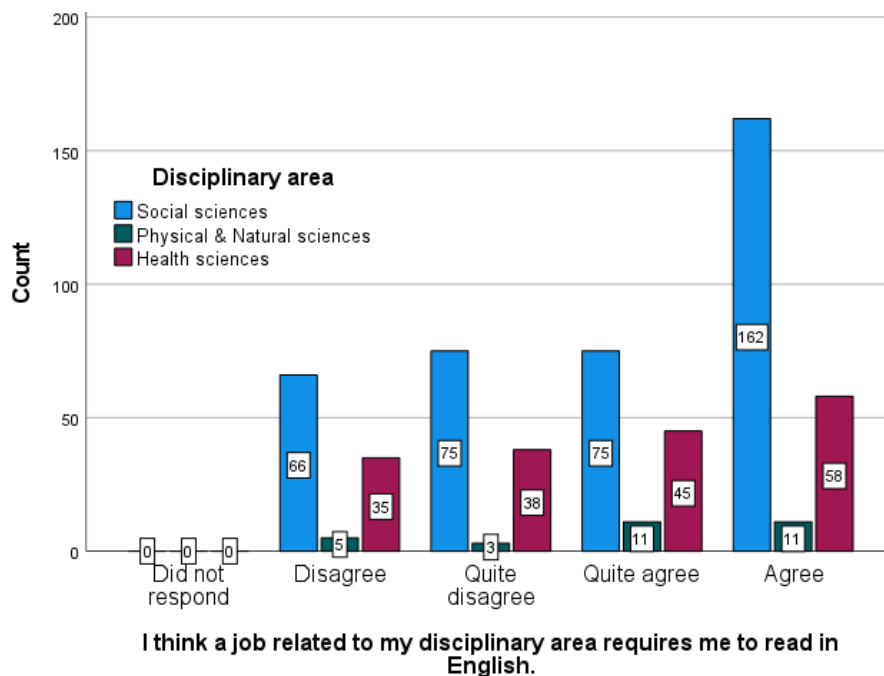


Figure 5.6 Number of participants per option selection for C34d (reading)

Regarding the interview, among the 37 mentions related to skills, reading was mentioned 19 times related to the workplace. This value was the highest among the skills. These results contrast those obtained in a study conducted at a university in a highly industrialized state in central México (Garcia-Ponce, 2020). Even though the study did not distinguish students by their disciplinary area, the overall result reported that 61.3% of the nearly 140 students participating in that study said reading was

necessary for the workplace. This skill was the second most important for these students in central México.

Questionnaire results reveal overall reading is the most important skill. Comments made by participants during the interview show their opinion on the need to understand discipline-related texts in English. The following excerpts from PNSc2 and PNSc13 each relate to the information being published in English:

*“...because generally all the texts that we.... um... well look at, that we, ...um,... get articles or something, all come in English.”* PNSc2

*“Well, it is that, for us, almost all the information we receive is in English.”*  
PNSc13

Participants state that more information is available in English than in their mother tongue. Others comment that updated information in their areas is also published in English, as SSc43M and HSc1M put it:

*“...there is a lot of information, for example, that one can find in English, and in fact, it is like the most updated.”* SSc43M

*“...the most advanced things, almost all of them are in English, articles, books, all of that is in English.”* HSc1M

The participants from the three groups consider reading discipline-related information important during their school days and in their future professional activities.

Other types of texts the participants mentioned that they may need to read as part of their work are those that are part of international companies' documentation. Handbooks, especially in the health sector, and foreign legal documents and medicines were also mentioned.

Other participants think they might work for an international company where they might encounter a situation where they need to read documents other than scientific articles. This is illustrated by SSc87F, she said:

*“How are we going to limit ourselves if we get a proposal that is in English? We need to be able to cope with it.”* SSc87F

SSc87F wants to be prepared should the situation arise, although she does not know what type of document she might need to understand. Other participants mentioned the type of documents they encountered as professionals in their field, such as those mentioned below. Participant SSc61F stated:

*“For example, the title (deed) of a foreign truck.”* SSc61F

It is not uncommon for American vehicles to operate in Durango. There may be situations where legal documents from these vehicles are required. SSc95M said reading is necessary because his current job involves translating legal documents. He said:

*“Because, for example, in my degree, there are many legal translations that are needed in my job, for example, the seizure that comes from foreign courts, then, it is necessary to have a good command, not basic, more thorough, of English, for the translations.”* SSc95M

SSc54 M also mentioned the need to translate legal documents, so English was important to him as a professional.

#### 5.3.2.4.2 Understanding others is important.

Understanding what others say is more important for the SSc group (2.91), followed by the PNSc (2.90) group. The HSc (2.79) group occupies the third position, as shown in Figure 5.5 above. The mean scores indicate that overall, test-takers ranged between

*quite disagree* and *quite agree*, with a tendency towards *quite agree*, particularly the SSc and PNSc groups, as their mean is closer to 3 than the value from the HSc group (see Appendix Table 6).

Analysis of the responses shown in

Figure 5.7 below shows that more than 43% of PNSc participants chose number 3 (*quite agree*), and 30% chose number 4 (*agree*). Option 3 also had the highest percentage of responses (33%) from the HSc group, followed by option 4 (29%). The behaviour of the SSc group is different, as option 4 had the highest percentage of responses (37%), followed by option 3 (29.4%). Overall, listening is relevant for most students in all disciplinary areas.

Regarding the interview mentions of the skills, among the 37 spontaneous comments, only eight referred to the relevance of listening (not connected to speaking) in the workplace. This skill occupies the third position. For the students participating in the study conducted in a central México university, listening occupies the fourth place in terms of importance. Among the almost 140 participants, 45.2% said English was relevant for the workplace (Garcia-Ponce, 2020).

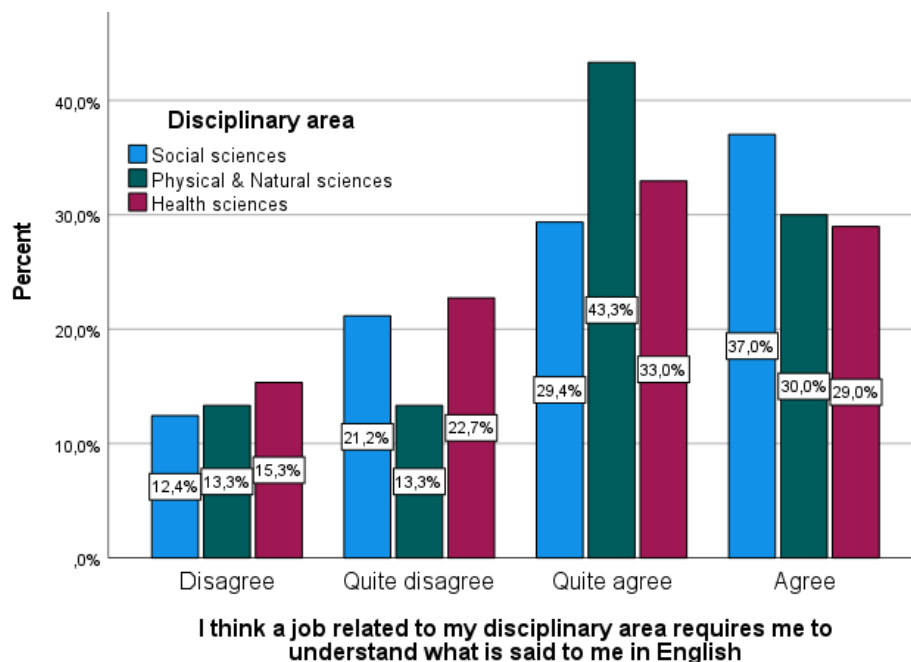


Figure 5.7 Number of participants per option selection for C34c (listening)

## Chapter 5 Understanding the test taker population

Participants' comments on the need to understand what others are saying come from various sources. Some of them from the HSc group feel that they may need to interact with English-speaking patients, as HSc89 stated:

*"...to understand what their problems are."* HSc89

A patient in a doctor's office must state the reason for his visit, and the doctor must understand the problem to present a solution. SSc88F made a similar comment:

*"...it could happen to us that we could be working at a place and someone who speaks English arrives... or be able to understand, be able to explain."* SSc88F

Both excerpts indicate that listening is viewed as part of a spoken interaction exchange rather than considering listening alone. However, there were other comments related to the need to understand spoken production, such as a conference presentation by an international speaker, as expressed by HSc20:

*"...and many academic events and conferences are in English, and we need to understand them".* HSc20

Most of the above examples show that the listening skill was not chosen solely for the sake of the skill but as part of the oral exchange. Most comments related to listening relate to part of a conversation rather than a situation where no interaction is expected. Therefore, it could be assumed that most participants expect to have to understand what others are saying in an oral exchange instead of understanding a monologue.

### 5.3.2.4.3 Writing in English in the workplace

Responses to the questionnaire ranked writing third for the three groups of participants in terms of the level of importance of the skill in the future workplace. Mean values are between 2.47, response option 2 (*quite disagree*) for the HSc and 2.67, response option 3 (*quite agree*) for the PNSc (Figure 5.5 above).

The results for each answer option show that 33.3% of PnSc participants selected option 4 (*agree*), closely followed by the SSc group with 32.3% of the answers. The HSc group had 27.3% of the responses to option 4, as shown in Figure 5.8 below. The percentage of participants from the HSc selecting option 4 is identical to those who chose option 1 (*disagree*). Overall, HSc participants show a stronger tendency to *disagree* (51.7%) than to *agree* (47.8%). The opposite trend is observed in data from the PnSc and the SSc groups. The latter groups show a tendency towards the *agree* end with 53.3% and 53.5%, the remaining percentages of 46.7% and 46.5% towards the *disagree* end.

Only five participants mentioned the need to write in English during the interview. These results contrast those obtained in the study conducted in central México (first mentioned in 5.3.2.4.1 above) (Garcia-Ponce, 2020). For those university students, writing occupies the third place in importance for the workplace. The writing skill was selected by 48.9% of the participants regarding its relevance in the workplace.

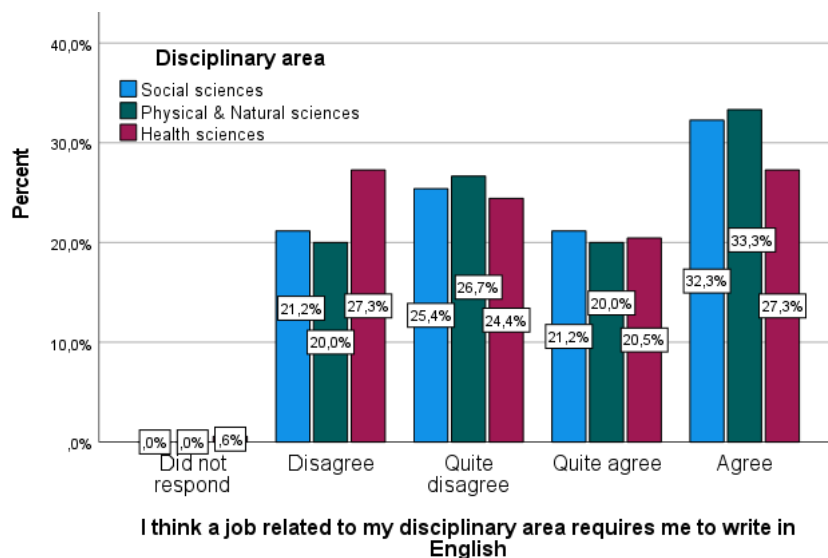


Figure 5.8 Number of participants per option selection for C34c (writing)

Participants' responses to the questionnaire may have been related to the need to write in English at school, not necessarily at work, as the following excerpt from HSc55F exemplifies:

*“For example, in the national test, there is an English part in it.”* HSc55F

Some disciplines within the healthcare provider community must take a discipline-related national test as part of the university-leaving requirements. HSc55 shows that one of the test components is in English. HSc55 understands that this test may require him to write in English, although it is unclear what type of writing is required.

When considering the possibility of having English-speaking patients, participant HSc85M commented that writing a receipt for the patient might be necessary.

Other participants, such as HSc17, feel that the ability to write in English would allow them to publish their findings in international journals, provided they get engaged in research, as the example below shows:

*“I think that knowing that I could publish in English, that is, that I carry out research and publish in foreign countries’ journals”.* HSc17

Other participants who expressed a desire to get involved in research also belong to the HSc group.

#### 5.3.2.4.4 Oral interaction

Being able to express thoughts and ideas ranks fourth on mean results in the three groups of participants. A closer look at the responses for each option shows that the HSc group has more than 31% of the responses on number 1 (*disagree*), as shown in Figure 5.9 below. Many HSc respondents believe they will not need to speak English at work. On the other hand, 26% believe they will need to speak English at work. Overall, more HSc participants, 51.8%, *disagree* or *quite disagree* with the need to use English, while 48.3% *agree* or *quite agree* on the need to speak English in the workplace.

The opposite behaviour is observed in the SSc group, as 31% of participants feel they will need to speak English for professional purposes (option 4, *agree*), and 25.4% are at the opposite extreme (*disagree*). Overall, most SSc participants (53%) are more inclined towards options 3 and 4 (*quite agree* and *agree*), and a smaller number



(46.8%) tend towards numbers 2 and 1, with a higher percentage choosing number 1 (*disagree* – 25.4%).

The PNSc group shows a different behaviour. The percentage of participants who chose options 4 (*agree*), 3 (*quite agree*) and 1 (*disagree*) is the same: 26.7. The remaining 20% chose *quite disagree* (number 2).

An overall tendency towards any option cannot be claimed as responses are distributed along the axis, with higher peaks at both ends (*disagree* and *agree*). However, this was the second skill most mentioned by students during the interview. Among the 37 spontaneous mentions, 17 students said speaking was relevant in the workplace. These results contrast those obtained in the abovementioned study conducted in a highly industrialized state in central México (Garcia-Ponce, 2020). For these university students, the overall result reported that 88.3% of the participants reported that high speaking skills were relevant in the workplace. This skill was the most important for them.

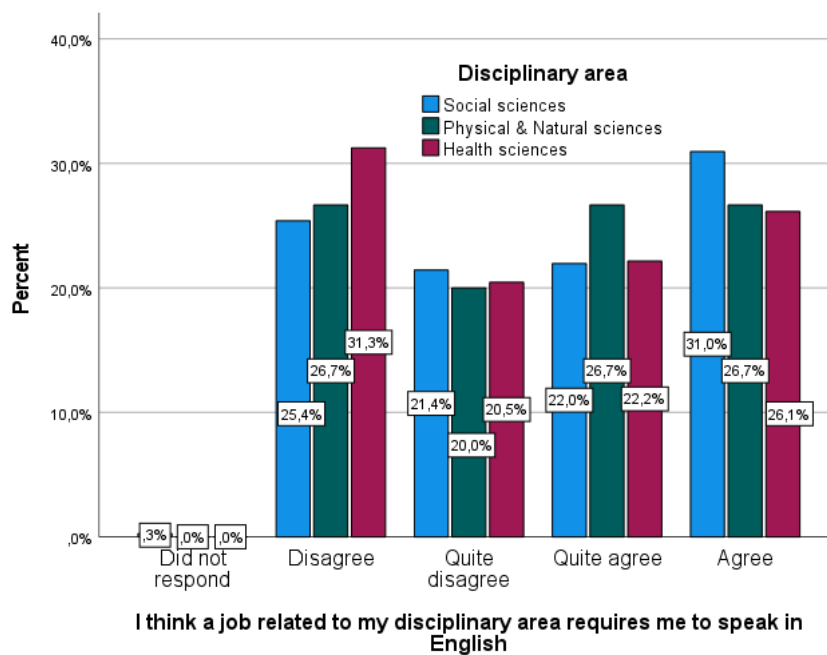


Figure 5.9 Number of participants per option selection for C34c (speaking)

In addition to the reference by SSc88F above to the possibility that an English speaker will come to her workplace and she will need to interact with that person, other

participants also commented on the need to speak in the workplace. Participant PNSc2 said:

*“...I believe that also speaking because generally all the conferences we attend are in English, and then well, there are...there are researchers that...that are very well known and if we do not know English, we can't interact...then, well...imagine!”* PNSc2

This participant finds it essential to be able to interact with researchers and speakers at academic events. Another 14 comments on the same topic related mainly to academic events, although some expressed the need to interact with colleagues about discipline-related issues.

Another group of comments related to someone having experienced or knowing someone who needed to speak English for professional reasons. An example of this is the exchange shown below between the researcher and female student SSc59:

SSc59: *“Because many people come from out of town. For example, in my job, people come from out of town, then yes. In my job, nobody has a good command of English, then, more or less understand them a little what they say...or at least what they want...”*

Researcher: *“And where do you work?”*

SSc59: *“...it is the urban vice-chancellorship, then yes, suddenly they arrive...tourists or people who bought land here, then it is necessary...to understand them a little... about that.*

There were 20 further comments from participants who felt the need or knew someone who needed to use English for work. HSc24F spoke about a medic she and her partner know who speaks English frequently; she said:

*“...yes, yes, well, a medic we know has many patients from out of town, maybe because he is an excellent medic and it is easy for them to talk to him.” HSc24F*

While not all participants have experienced the need to speak English in the work environment, knowing about other professionals doing so increases awareness of its usefulness.

A third group of comments considers speaking English for professional purposes important, but not in Durango. They think it is required in big cities or abroad. The excerpt below SSc36F exemplifies this belief:

*“Mmm...not here in Durango necessarily, but in other places that they send me, maybe out of town, hum...well...I could not communicate, I would not know what to say, and I may blow out the business, hum...because I did not know how to use the language appropriately.” SSc36F*

This group of 13 participants might feel that speaking English in the workplace is not required in Durango, although they acknowledge that it will be beneficial elsewhere.

The last group of comments are from participants who said they do not know situations or people in their profession where speaking was required. However, they believed there was a high probability that they would need to speak English in Durango for professional reasons. HSc89F commented on what she thinks could happen in the future regarding her degree:

*“Yes, they could come from out of town to have their dental work done here.” HSc89F*

The increasing number of foreign contract factories or outsourcing companies bring foreigners who may not be very competent Spanish speakers. They may need services from professionals from different disciplines. These participants feel they need to be

prepared to use English to interact with speakers of other languages to provide the required services.

Overall, the responses to the interviews show that some test-takers believe that English will be required in their future professional activities. Perhaps they have already experienced the need or know someone who does. They may think they need it outside of Durango or for academic purposes in the local area. In any case, speaking English is seen as a necessary skill in a professional context.

### 5.3.3 Section 3. Preparing for the test

#### Items 19 to 21

The participants' opinions on the courses taken to prepare for the leaving test range between *quite disagree* (2) and *quite agree* (3), as shown in Appendix Table 7 below. Participants from the PNSc group had a better opinion of the leaving test preparation classes, while they were not so relevant for the HSc group. The minimum mean for TTQ19 (the courses I took helped me prepare for the test) was 2.41 from the HSc group and the highest from the PNSc group, with a mean of 2.77 for the same item.

Participants in the PNSc have a better opinion (mean = 2.63) than those in the HSc (mean = 2.43) about the classroom opportunities to practice the activity types they found in the test (TTQ20). Participants in the PNSc group agreed more (mean = 2.73) than those in the HSc (mean = 2.48) about the similarity of activities between those included in the classroom and those found in the exit test.

Respondents' opinions on the opportunities they had in university courses to carry out activities that helped them prepare for the test (TTQ21) ranged from *quite disagree* to *quite agree*. The HSc group disagrees the most, reporting the lowest mean (2.48), while the PNSc group reports a tendency toward *quite agreeing* (2.73). This group also reports a minor difference between opinions, with a standard deviation of 0.868. The group with the highest degree of disagreement among participants is the SSc, with a standard deviation of 1.166.

All means from the SSc group are between the highest from the PNSc group and the lowest from the HSc group.

For items in Section 3, all the highest mean values are from the PNSc group, and the lowest is from the HSc group. The lowest level of disagreement for the three items in this section comes from the PNSc group. The highest standard deviation values are from the SSc for TTQ19 and TTQ21 and the HSc for TTQ20.

Exam preparation was not an issue addressed in the interview; however, this issue was raised in some voluntary written comments at the end of Q1. Of the 232 voluntary comments, 18 related to test preparation and eight specifically to the courses offered by the university. In general, these comments indicate that participants feel that the courses they took at university did not help them succeed on the test, as exemplified in the excerpt from Q1SSc303 below:

*“The course I took that the university offered is unrelated to this test”.*

Q1SSc303

Other participants felt that the courses should only focus on the exam content and not include other things that will not be included in the test. They seem unsatisfied with the relationship between the language courses and the test content.

#### 5.3.4 Section 4: General opinions about the current test

Items 1, and 22 to 26

The mean scores of the items in Section 4 have the smallest and highest scores among all responses in the questionnaire, ranging from 1.93 (between *disagree* and *quite disagree*) to 3.20 (between *quite agree* and *agree*), as shown in Appendix Table 8 below.

With a mean = 2.73, the PNSc group considered the test easy to answer (TTQ1). The lowest mean (2.59) was for the HSc group. All standard deviation values for the known discipline groups are more significant than 1, with the highest being from the SSc group (1.051) and the lowest being from the PNSc group (1.015).

This issue arose in some participants' voluntary written comments at the end of Q1. There were 31 comments on the difficulty of the test; 20 of them said they found the test difficult, while 11 said they found it easy. Below is an example made by a

participant whose discipline is unknown (Q1DU118); this is one of the few comments made in English.

*“It is very simple”.* Q1DU118

Other comments mentioning the difficulty of the test do not seem to have a positive connotation. Q1HSc604’s comment clarifies this:

*“Well, it was short; I don’t think it evaluates many things”.* Q1HSc604

The comment does not explicitly state that the test was considered easy, but the comment regarding its length may be interpreted that way. A lengthy test is often associated with being demanding and vice versa. These comments suggest the test did not have the complexity or length a test of this nature is expected to have. For these test takers, the face validity of the test may be low because the test did not match their expectations of a high-stakes test (Phillips et al., 2020). These seemingly superficial perceptions of validity may influence test takers’ attitudes toward the test (Keblawi, 2022).

Most test-takers do not find it necessary to use their native language (Spanish) for test instructions. All TTQ22 mean scores ranged from quite *agree* and *quite disagree*, with the highest mean reported by the PNSc group (2.47) and the lowest by the SSc group (2.26). However, the standard deviations for this item are the highest in the entire questionnaire. These high standard deviation values show that opinions on this aspect vary widely among participants, particularly those of the SSc, as this group reports the highest value. Although the standard deviation value of the HSc group (1.159) is also above 1, this group has the slightest difference in opinion among the three groups. This topic was not part of the interview; however, at the end of Q1, a voluntary comment referred to using the mother tongue for instructions. Participant Q1HSc775 wrote:

*“I also think they should put instructions in Spanish or agree to clarify doubts about an instruction.”* Q1HSc775

Only Q1HSc775 commented in this respect, but it may represent participants who chose option 3 or 4, *quite agree* or *agree* (on using Spanish for test instructions). The comment by Q1HSc775 suggests that instructions can be confusing. Using test-takers' native language to deliver test instructions (Bachman and Palmer, 1996) is recommended in low-level testing situations (Underhill, 1987). A group of a test-taker population like the one taking part in this study (at the B1 level of the CEFR-an independent language user) (British Council Global, 2019) is expected to understand test task instructions, assuming they are well written. It is assumed that the weakest test-takers are the most likely to misinterpret test instructions (Hughes, 2003).

The suggestion that the test contained more open-ended questions (TTQ23) was not very popular. Mean values varied between *disagree* and *quite agree* but closer to *quite disagree*. The PNSc group most disagrees with the increase in open-ended questions. Their mean value (1.97) lies between the *disagree* and *quite disagree* options. The PNSc group reports the lowest value among the three groups. Even if the SSc group is not very happy with increasing the number of open-ended questions, their mean value of 2.21, between the *quite disagree* and *quite agree* options, is the highest for this item. The standard deviation values are also quite significant, ranging from 1.202 for the SSc group as the highest and 1.089 for the PNSc group as the lowest. These high standard deviation values indicate more significant disagreement among SSc participants regarding the increase in open-ended questions than among the PNSc participants.

This topic was not part of the interview, but there was a voluntary written comment on open-ended questions. Participant Q1HSc817 wrote:

*"I liked it very much because it was the first time they made open-ended questions".* Q1HSc817

There were other comments on questions, such as a desire for more variety in the types of questions used, but only one on open-ended questions. This test taker may seem to be motivated by the test. Keblawi reports some test takers "enjoy doing tests or even feel happy about tests" (2022, p. 703). No comments were made on the multiple-choice questions.

Increasing the number of multiple-choice items (TTQ24) was more popular than increasing open-ended questions, particularly in the PNSc group, with the highest

mean = 3.20. This group was the most dissatisfied with the idea of increasing open-ended questions. The HSc group has the lowest mean value, 3.00. Even though they do not fully support increasing the number of multiple-choice questions, they do not dislike the idea. This group reports the highest standard deviation, 1.131, showing more variation in responses among healthcare test-takers than the other two groups. The group most favouring an increase in the number of multiple-choice items is the one reporting the least within-group variation for this item, the PNSc reporting a standard deviation of 0.961.

Many participants considered the time allotted for the test (TTQ25) insufficient. The mean values for this item are between 3.20 and 3.14. The SSc group is more in favour of increasing the time it takes to complete the test. The lowest mean from the HSc group is also in the range of *quite agree* and *agree*, indicating that more than half of the test takers felt that the time allotted to take the test was insufficient. All standard deviation values are above one and range from 1.057 for the SSc to 1.097 for the HSc group, showing that, while many participants would like more time to complete the test, there are also participants who do not share the same opinion.

The mean values do not indicate whether participants felt more time was needed for the entire test or specific sections. Appropriate time allocation to answer the test was another topic not raised during the interview. However, it was mentioned quite frequently in the voluntary written comments. This topic had the highest number of codes. Figure 5.10 below shows that most comments regarding more time to take the test were made in general terms. A few comments were specific to the testing component they would like more time to complete. More participants considered the time allocated for the written part insufficient, but speaking and listening were also mentioned and asked for more time.

<input type="radio"/>	More time	48	49
<input type="radio"/>	In gene	38	39
<input type="radio"/>	Listenin	2	2
<input type="radio"/>	Speakin	3	3
<input type="radio"/>	Writing	5	5



Figure 5.10 Voluntary written comments coding at “More time”.

Comments on time came mainly from the HSc and SSc participant groups. The comment by Q1HSc336 was:

*“...too little time.”* Q1HSc336

This comment does not indicate whether more time is necessary for all test components; it just states that time was insufficient. Other comments were more specific, as Q1SS356 wrote:

*“...more time is required for the written test.”* Q1SS356

The listening section is included in the written component of the test. Participant Q1SSc716 revealed in his comment that he had difficulties with the listening section of the test:

*“...more time or pause in the listening”.* Q1SSc716

Perhaps the time between the small conversations of the listening component of the test is considered insufficient to process the information and select an answer. The speaking component was also mentioned; Q1HSc378 wrote:

*“...more time for the oral test.”* Q1HSc378

As noted by the comment above, requests for more time did not preclude the written component, although most comments were not specific. A study was conducted in the Midwest USA to explore the consequences of using a locally designed 3-hour English placement test (EPT) for non-native-English speaking students. Even though it is a compulsory test, it is a post-entry pencil-and-paper test. The results of the EPT are

## Chapter 5 Understanding the test taker population

used to determine which course(s), if at all, students should enrol on. Though the results of the EPT may require students to enrol in non-credit-bearing courses, it does not determine university admission. Four of the eight participant students mentioned time-related issues when interviewed. Fifty per cent of participants consider “they could have done better in an untimed situation”. (Li, 2021, p. 305).

The length of the test (TTQ26) is another aspect that was not very popular among participants. The PNSc group is the only one with a mean above 2 (*quite disagree*). The highest mean value was 2.07, while the HSc group, with a mean of 1.93, was somewhat divided about increasing the number of items in the test. As for the standard deviation, the highest value, with 1.23, came from the PNSc group and the lowest, with 1.034, from the HSc group.

Test length was also an issue that participants commented on. Some specifically stated that they consider it short, as expressed by Q1HSc604 above. The length of the speaking component was also mentioned. Q1HSc441 wrote:

*“I think more questions and more opportunities in the oral test.”* Q1HSc441

According to Q1HSc441 above, the number of questions in the written component is insufficient, and the oral component does not seem to provide enough opportunities for expression. Another comment about the oral component suggests there is a limit on the amount of language that can be sampled. Q1SSc602 explained:

*“The oral test should be longer because, with three activities, it is difficult to evaluate the English that each person knows.”* Q1SSc602

Q1SSc602 might feel that the oral component did not allow him to demonstrate the language he can produce. He feels more opportunities could provide a sample that better represents his oral ability.

Test takers feeling unhappy about the amount of time allocated to answer the test or certain parts of the test may also feel demotivated by the test. If test takers felt uneasy when they were taking the test because they found they needed more time to think and

respond, that feeling may have affected their attitude towards the test. They may not have made the same effort when answering the test (Keblawi, 2022).

## 5.4 Summary

This chapter looked at the results of qualitative and quantitative data from Q1 and the responses to a short semi-structured interview. This data was collected from test takers upon finishing the school exit test. Findings provide answers to Research Question 1 (RQ1). RQ1 reads:

**SQ1a.** What is the face validity of the test from the point of view of test-takers in terms of content, test preparation, timing and difficulty?

A comparison of the domains of the competences in Section 1 (what test takers think the test measured) and Section 2 (what they think it should measure) suggests that the current school exit test is not fully meeting the expectations of test takers. The mean values from Section 2 are generally higher than those from Section 1, indicating that more participants favour the work and study domains being considered in the design of a test than those who think these domains were tested. Higher mean values in Section 2 than in Section 1 could indirectly represent students' desire that the test included the work and academic domains. Overall, quantitative results suggest that the test taker's expectations are not fully met. The interest of most test takers differs from the actual contents of the test.

Voluntary comments at the end of Q1 reflect displeasure from some test takers on the compulsory nature of the test. Some comments also refer to the content of the test; they think the test should be discipline-related. Some comments claim the test has had a negative impact on their professional development. Not being able to pass it after more than one attempt has not allowed them to award their degree, placing their career development on hold.

During the interview, participants chose which domain they would like the test to focus on, thinking of their interests and, many of them, in their future professional lives. Results reveal a more significant interest in the study domain, a domain not considered strongly present in the school exit test by many participants. There is also great interest in the test to be focused on the work domain. The interview responses support the interest in these domains. It is worth noticing that even though quantitative results

reveal that SSc participants are equally interested in the social domain and the work domain, interview responses reveal many SSc participants consider the social domain to be the one that will allow them to interact within the workplace. They consider that because their discipline is social science and they need to interact with people, the social domain will allow them to do it.

Quantitative data results show many test takers are interested in being able to use English for socializing purposes. For them, socializing occurs within all the three domains they were asked to select from. This perception may explain why some test takers said the test measured competences related to the work domain. However, some of the comments made by test takers concerning the domains (as their inseparability, one comprising the others or one emerging from another) reveal that many test takers are unfamiliar with the competences inherent to each domain. This argument may be a more substantial explanation of why some test takers consider the competences related to the academic and work domains were assessed. This perception could also explain why some test takers consider that their classes offered opportunities to practice language use activities like the ones they may encounter in their future working environment.

The short semi-structured interview allowed participants to tell their “individual stories” (Dörnyei, 2007) concerning the domain of their choice. A different scenario from the above was presented through interview responses, as participants were allowed to express their opinions and justify their answers (*ibid.*). These revealed that, if allowed to decide, some of them would have the three domains in the test. However, if one had to be chosen, the previous preferred choice may become the last domain of interest. A more significant interest in the study domain was found, followed by the work domain for the three groups of participants. This preference indicates a mismatch between what is currently being assessed and what test takers are interested in. They consider the study domain more beneficial than the general-purpose content of the current test. Some test takers questioned the content of the test, not only in terms of the domain but also in terms of its comprehensiveness. A few comments about the whole test and the speaking component consider the test needs to broaden its content to yield more reliable results of test takers’ language ability.

The latter questions the face validity of the test in terms of the relevance of its current content. This perception is because it does not address the language required for study purposes or the local working environment where test takers are likely to use it.

In general, test takers consider the language classes taken helped them, to a certain degree, to prepare for the test; however, it was not enough. Some test takers expressed the opposite on the open-ended question in Q1. These comments represent those participants who selected *quite disagree* or *disagree* when asked about the courses they took and their relationship with the test.

Students prefer tests that do not require much language production and are mainly based on recognition items. However, it must be acknowledged that this is not a unanimous opinion.

In general, the test was perceived to be of average difficulty. However, some comments made by test takers reveal instructions are perceived to be confusing. Even though (in general) they do not think the instructions should be in Spanish, some of them think instructions need to be clearer, especially regarding the listening component. An aspect that did not only report a high mean value but revealed the feelings test takers have towards the test was the time allocated for its completion. Many test takers considered more time was needed to answer the test.

Several comments regarding time reveal many test takers consider the time allocated to respond insufficient. They mainly refer to the written component.

A difference in opinion regarding the domain the test should focus on may generate an adverse psychological reaction in the learner, who may not be able to perform at his best, affecting his overall test result (Brown, 2004).

The test-taking session *per se* is already an event prone to generate stress and/or anxiety. Face validity has an impact on attitudes and motivation towards a test (Heaton, 1975). Some test takers may arrive to the testing session with adverse feelings due to its compulsory nature and/or the domain it focuses on (L. F. Bachman, 1990). Not having enough time to complete the tasks, having problems with instructions, or facing item types they find confusing will naturally increase the already unfavourable psychological state. It is not only the individual test results of the test taker who is going

through this emotional stage that is affected, but in general, test results will not reflect the actual language proficiency of the test taker population (Hughes, 2003).

Test designers have overlooked the relevance of face validity due to its subjective nature, which is based on the eye of the layperson (Bachman, 1990). Even though not frequently considered, test takers may sometimes be able to provide a knowledgeable opinion about a test (Davies et al., 1999), and it should be regarded that low face validity is considered to have a negative effect on the performance of test takers (Sato & Ikeda, 2015). Failing to consider the psychological effects generated by the test may lead to undesirable consequences for the test and the institution that offers it (Davies et al., 1999).

Test takers said they thought they would need English in their future career-related jobs; however, the specific kind of language required in the workplace in the local environment is unknown. The result of data collection on this aspect is addressed in the following chapter.

## 6 Local context language needs

### 6.1 Introduction Phase II

This chapter accounts for the findings regarding the relevant work-related language competencies in the local environment. These were identified through the analysis and comparison of Questionnaire 2 data applied to two groups of participants: students (Q2a) and professionals (Q2b). Specific examples of language use activities for the disciplines this study is concerned with, which were gathered through semi-structured interviews with professionals (Q3) while responding to Q2a, are also presented. Other examples of current language use activities were provided by test takers, such as HSc3, HSc85, SSc61, SSc95 and others above, during the short semi-structured interview applied upon taking the school language exit test.

Data from both groups, students and professionals, was organized according to their disciplinary area: Health Sciences (HSc), Physical and Natural Sciences (PNSc), and Social Sciences (SSc), as identified in Table 4.2 above. The mean value results of the Q2 data analysis are presented below. The competences comprised in Q2a (students) and Q2b (professionals) were grouped by skill(s) (Table 4.10 above) and compared in 6.2 below.

The nature of the origin of the items used in the instrument applied may influence the responses of the target population. The competences comprised in the questionnaire were designed for international/intercultural environments, such as Europe (ALTE, 2002) and Canada (CCLB, 2012). Given that México belongs to the Expanding Circle (Figure 1.1 above), English is not commonly used for everyday matters as it is in countries within the Inner Circle. Some of the competences in the questionnaire pertain to activities carried out in the country's mother tongue, as will be exemplified below with some of the comments from professionals. Carrying out these competences using a foreign language may happen in exceptional circumstances.

### 6.2 Work-related language competences

The working environment grouped by the skills involved

This section compares the mean value results of the work-related language competences collected using Q2a (students) and Q2b (professionals). Competences are grouped according to the skills involved. Results for each group of participants are compared.

The tables in Appendix XI, containing results analysis organized by areas, show the mean values from student data are higher than those from professional data. Mean values from students and professionals of the same disciplinary area will be compared. The total mean from each group is used as the cutting point to determine the most relevant competences for each group.

Mean values below the corresponding total mean are considered not to be so important or relevant for respondents due to the frequency of use represented by mean value results.

Mean values above the total mean for its corresponding group are considered relevant to be included in the design of an exit test due to the frequency of use reported by participants.

Even though high mean values from professionals may be considered more relevant because they represent current language needs within the workplace, high mean values from student data cannot be underestimated. Jin (2023) argues the relevance of considering test takers' voices for language assessment practices (including test development) and policies as they provide valuable insights. Her contention is based on the studies published in the Virtual Special Issue (VSI) published in September 2022. Chancová (in 2.5.2 above) also argues not to disregard pre-service students' needs and wants. While it may be argued that many students lack job experience, their responses represent the vision of the future. Another difference between student data and professional data is language preparedness. Proficiency in English may have been an accessory for some current professionals during their student time.

As was argued in Chapter 1 above, technology and globalization have impacted the role of English proficiency for professional practice worldwide (Education First, 2016). It is currently possible for almost anybody to access oral and written specialized, updated field-specific material on the internet. However, this information can only be used if individuals can understand the language in which it is commonly shared,



English (Education First, 2020). Students know the importance of understanding written English, as exemplified in 5.3.2.4.1 above. A glance at the EF reports between 2011 and 2022 shows a continuous gap among generations regarding English proficiency. The generations between 18 and 25 report higher proficiency levels than those over 41. These results suggest that the probability that students use English is higher than the probability of adults. The country's low/very low proficiency index makes the language barrier for adults bigger than for the average university age group (Education First, 2011; 2012; 2013; 2015; 2016; 2017; 2018; 2019; 2020; 2021; 2022)

Once competences with mean values above the total mean were identified, they were analysed in the light of the comments made by professionals regarding how they are used for job-related purposes. Their comments reveal how competences are interpreted and how they relate them to their professional activity. Some of the comments made by interviewees exemplify how competences are misunderstood, sometimes generalizing them or considering a competence to be the same as another one that requires a different skill or language functions. Interviewees' comments shed light on the relevance of competences for the local workplace. In the analysis of data and decision-making process of determining the competences that will be part of the exit test, the context in which competences are used cannot be disregarded. It cannot be ignored that the study takes place in a country where English as a foreign language reports a very low proficiency level (Education First, 2020; 2021; 2022) and a state occupying the third position above the lowest proficiency level in México (1.6.5 above). All these factors are considered to retain or discard the competences (among those with mean values above the total mean) that will be part of the school exit test.

The results of the above data analysis per disciplinary area will be compared to define if one test is relevant to the three groups. It may be found that a different test is needed for each disciplinary area or that there could be shared content along with specialized components for the disciplinary areas.

### 6.2.1 Speaking and listening-related language competences

This section comprises 11 competences, representing 39.3% of the competences in Q2. This section is the most numerous in the instrument applied.

Data collected from students and professionals were analyzed according to the disciplinary group. Results among disciplinary groups were compared, finding that students and professionals from the three groups agreed with each other on most of their choices, as seen in Appendix Table 9 below.

Asking questions of a fact-finding nature (C3), asking for clarification when something is not clear (C8), and greeting a visitor and engaging in a limited conversation for a short time (C12) are equally important. Both students and professionals from the three disciplinary areas report mean values for these competences above their total mean values. Interactive communication is one of the desired skills mentioned by employers in the study conducted in central México (first mentioned in 2.4.1 above). The human resources executives from multinational companies in the local area claimed university graduates had limited speaking skills and struggled to communicate to the extent that there had been misunderstandings with foreign employers (Garcia-Ponce, 2020).

Professional SScE554 mentioned having to ask questions about the problems with the equipment he uses when giving an example of how he may need to use C3 in the workplace. He stated

*“The equipment I use is manufactured and sold outside the country (México), and I have had to, well, ...had to communicate in English so they can understand what I need or get an explanation to a failure on the equipment or see how I can get the piece or the replacement part I need to fix the equipment delivered”*. SScE554

A similar type of situation was expressed by HScE582 when he said that

*“Sometimes, in this case, I would say yes, sometimes. When one buys equipment or other things that one uses, and the company is not meeting the delivery date, one needs to get in touch, and we must call, and here we need to speak in English”*. HScE582

Asking for clarification (C8) was mentioned to be needed when they attend conferences about new developments in their field, as stated by HScE582.

*“When we attend conferences, as they are in English if one has a doubt, one needs to ask in English”*. HScE582

This competence was also reported to be relevant for Chilean dentists (Olusiji et al., 2022). Speakers of other languages may have investments in Durango, as explained by PNScE635 below when he had had to interact with them during their visit.

*“I work with mining companies, sometimes the owners come, and we interact with them... a little... not only about work”*. PNScE635

Even though he does not specify the language the visitors speak, PNScE635 had already mentioned that the mining companies they work with are Canadian. The owners spoke English (rather than French). Besides, as this comment was an example of the situations in which he had to greet a visitor and engage in a limited conversation (C12), it is implied that the conversation takes place in English.

Students do not seem to consider they will need to offer advice to clients within their job area (C1) as professionals of the three discipline groups do (see Appendix Table 9 ). Within health science, HScE582 said

*“When providing service to a patient, it is important to explain the differences between the treatments offered, their advantages and disadvantages, and the cost”*. HScE582

This language use activity was also found relevant for Chilean dentists. They need to interact with their English-speaking patients (Olusiji et al., 2022). However, for Rumanian dentists, this competence was found to be less likely to be required (Mihaela et al., 2020).

Within the social sciences, SScE568 commented:

*“When foreign clients arrive at the establishment, we immediately identify them and, almost always, I go to greet them and explain to them what we offer them”. SScE568*

SScE568 explained that she would care for foreign customers, as most of her personnel are not fluent in English. Even though she said her English proficiency was not very high, she said it was better than the language abilities of most of her personnel.

For Iranian managers, assistant administrators or company expert engineers working in a large Iranian company, this competence is relevant when interacting with foreign guests or colleagues (Chalak, 2019).

Students of SSc, as well as students and professionals of PNSc, find requesting routine services related to their own area of work (C2) a relevant competence. This competence may have been conceived to be relevant within a multicultural environment, i.e., a company worker requesting a service (in English) to another worker within the same company. However, examples of the use of this competence reveal it was understood to refer to purchasing services or goods from international providers. An example from the social sciences refers to requesting legal documents, as SScE590 stated:

*“...we had to request information about her migratory status, request her passport and even had contact with her family in the United States...and we had to request papers to be sent to us”. SScE590*

Social sciences professionals and Health sciences participants (students and professionals) report mean values below their total mean values.

Even though SSc professionals did not report a mean value above the total mean, C2 will also be considered for this disciplinary group because the mean value from students was above its total mean.

Participating in job-related meetings and seminars (C9) seems important to all participants except PNSc professionals, who report a mean value below its total mean.

As in C2 above, C9 will also be considered for the PNSc group because of the high mean value recorded from student data. The examples participants gave reveal they thought they would need to use English in seminars they may attend. Participating in other job-related meetings would be done in their mother tongue. An example of this is stated by HScE642 below:

*“Yes, it has happened when I have doubts in a conference or similar; there are times I go directly to the speaker, and I ask him, more or less in English, and he responds likewise”.* HScE642

Saying “more or less in English” was how HScE642 wanted to say he considered his English proficiency not to be very advanced, though he managed to ask questions and understand the response given.

Expressing opinions in simple terms, for example, “I don’t agree” (C10), is considered relevant for PNSc students and professionals from the SSc and HSc. Students from these two disciplines report mean values below the total mean. The examples of language use situations provided were also related to attending conferences. Still, some participants from the SSc also expressed using this competence in dealings with wholesalers who are also non-native English speakers, as exemplified by SScE560 below:

*“In fairs and exhibitions, many Koreans, Chinese and Gringoes\*, and you place your order and get to the stand, and they do not speak Spanish, or they do not speak it well, so it’s easier for them and me to use English”.*  
SScE560

*\* Gringoes is a colloquial expression to refer to people from the United States of America*

English is also used to communicate with speakers of other languages.

Providing instructions and directions on how to get to a place (C13) seems relevant only for PNSc professionals. Therefore, this competence will be considered only for this group. An example of the use of this is provided by PNScE635 below:

*“We participate in species conservation programs, and sometimes we have experts from the USA, and they ask questions, sometimes when we are in the woods or things about the city, and we tell them”.* PNScE635

HScE572 provides an example from another context of use:

*“Sometimes we need to tell them how to get to the office”.* HScE572

Making phone calls requesting information or services (C15) reports mean values above the total mean only from HSc students. Therefore, it will only be considered for this disciplinary group. HScE582 provided an example of the use of this competence:

*“Sometimes, when we need supplies, we have to call...”* HScE582

Providing information or opinions, such as when planning a company event (C11), making job-related phone calls, leaving a phone message, and placing a service or materials order (C16), do not seem as important as others. C11 and C16 report mean values below the total mean for all groups of participants. These competences will not be considered for any disciplinary group in the school exit test.

## 6.2.2 Listening and speaking-related language competences

Among the three competences in this section, only dealing with predictable requests from a visitor (C14) reports mean values above the total mean for all participants from the three disciplines so that C14 will be part of the test content. All participants who gave examples of the need to deal with predictable visitor requests mentioned giving directions. Some participants from the SSc and the PNSc mentioned they might need to give directions to get to a place within the company. Participants from the HSc said sometimes they need to give directions to get to the facility. Patients or clients have also asked participants from the HSc or SSc for directions to get to another place from the participant’s location. Several participants thought C14 was the same as C13, with their answer *“Yes, the same”*.

Taking a routine order (C4) reports a mean value above the total mean only from the PNSc students. The few examples of language use of this competence reveal it was interpreted as for requesting goods or services, not as being providers, as exemplified by HScE582 below:

*“Well, this is similar to a previous one; it is talking about when a product is being bought. In my case, it will be to place the order, do the paperwork, and if it does not arrive on time, we need to request a refund”.* HScE582

None of the professionals interviewed expressed being providers of goods for international buyers; therefore, they do not require this competence. This competence will not be considered for the test of any of the disciplinary groups.

Receiving phone calls, taking messages, and making notes (C5) are reported to be relevant for SSc professionals and HSc students and professionals. Competence 5 will be considered for the SSc and the HSc areas. There are few examples of how this competence is used, but they specifically state or imply that notes will not likely be taken in English, as another Spanish speaker will most likely use them. An example is provided by HScE582 below:

HScE582: *It has happened that a patient calls to make an appointment, then I write his name on the agenda and the problem he has or the ongoing treatment if it's the case.”*

Researcher: *“What language do you use to make these notes?”*

HScE582: *“Spanish, of course; why would I use English?”*

Whether notes taken are for the use of the person receiving the phone call or will be passed on to somebody else, these notes are most likely to be taken in Spanish, except for proper names. This situation may be different if participants work in multinational companies, where individuals may be required to take notes that will be shared with another non-Spanish speaker (Garcia-Ponce, 2020).

Looking at the total means of the S&L and the L&S competences, it could be suggested that respondents may consider the need for the L&S competences to be less frequent than the S&L competences (see 6.2.1 above).

### 6.2.3 Listening-related language competences

Understanding spoken directions, such as safety procedures (C6) and following a simple presentation or understanding explanations regarding a product (C7) are the two competences that involve only listening. This section is the smallest within the instrument.

The mean values reported for all groups reveal that students and professionals of the three disciplinary areas agree that C7 is relevant for them and C6 is not, as Appendix Table 11 below shows. The former got mean values above the total mean, while the latter reported mean values below the total mean. This result reveals only C7 will be considered for test content.

Regarding C6, some participants said Spanish was used to understand spoken directions, such as safety procedures. Others said they might need to understand the safety procedures of their equipment, but those are in written form, not spoken as C6 specifies.

Competence 7 was related to understanding the information in academic conferences or product presentations. Participant SScE554 said that even though sometimes the presenters bring a translator with them, it is better to understand what the speaker is saying:

*“All the design and embroidery courses and training we take are in English, they are very technical, they are not offered in Spanish... the teachers and the technicians speak English... they regularly have a translator, but a lot of the explanation is missed if you don't know English”. SScE554*

Likely, the translator is not specialized in the discipline of her courses; therefore, SScE554 considers it better not to rely on the translator.



#### 6.2.4 Reading-related language competences

This group, as the writing-related group, comprises six competences.

Recognizing and understanding, at least partially, the general meaning of a non-routine letter (C17) is reported to be relevant for professionals of the PNSc and HSc groups; however, not for those within the SSc group. These results can be observed in Appendix Table 12 below. Concerning the mean values from students means above the total mean are reported from the SSc and HSc groups. This competence is quite relevant for students and professionals from the HSc group (both mean values are above 2). Only students (SSc) or professionals (PNSc) consider it relevant for the other groups.

When providing examples of situations in which C17 is used, some interviewees referred to understanding other types of documents, like assembly or operating manuals (which is C21), not necessarily letters. Others said they do not receive letters on the post but receive emails related to products, events, or services.

Examples of the use of C17 reveal participants were thinking about other types of documents rather than understanding the general meaning of a non-routine letter within their own area of work. Communication within and between companies in México takes place in the country's mother tongue. Communication in English with foreign companies does not occur every day. Based on these arguments, C17 will not be considered for any participant group.

All participants from the three disciplinary areas agree that understanding straightforward factual information on routine job-related faxes, memos, emails, etc. (C18) and obtaining basic relevant information, such as product specifications, professional or commercial leaflets, advertisements, internet pages, etc. (C19) is relevant, as all mean values are above the total mean. These competences will be considered for the three disciplinary groups.

Regarding C18, participants referred to receiving advertisements about products or services in the email. They also mentioned they receive emails with invitations to courses, conferences, or events. Reading academic articles was also mentioned in the comments regarding C18, such as SScE583 expressed:

*“Sometimes you start reading an article you are interested in, but then you don’t understand...”*. SScE583

SScE583 may be able to understand the primary or general idea of discipline-related articles, but apparently, she cannot understand the details of it. HScE637 presents a different scenario:

*“Many companies send me emails in English, especially about products or medication that is about to be released or was recently released”*.

HScE637

Her comment suggests she can understand the information being received but does not make reference to understanding routine job-related documentation.

When providing examples of how C18 is used, most participants gave examples of the situations in which C19 is used. The examples provided suggest it is most likely that C19 is the competence interviewees use rather than C18. As argued on using C17 above, all internal communication within companies will likely occur in Spanish, even within international companies. No comments concerning the use of C18 referred to routine documents. The lack of reference on how professionals use C18 supports not including C18 among the competences for developing the test. Competence 19 will be considered for all groups of participants.

The same argument regarding C17 and C18 is relevant for C20, which refers to recognizing and understanding, at least partially, the general meaning of a routine letter within own work area. This competence (C20) is only reported as relevant for PNSc students. The few comments made by interviewees when referring to using this competence refer to sporadic communication regarding an event, updates on user’s manuals or recent releases on medication to treat an illness. No participant expressed using English regularly for job-related purposes, as communication within the company, such as routine letters or memos, takes place in Spanish. Competence 17 will not be considered for the design of the test.

Understanding general written instructions, for example, installation, functioning or maintenance manuals (C21), was found relevant by all participants from the three groups. This competence will be considered for the three disciplinary groups.

Many interviewees referred to understanding installation or functioning instructions for office equipment, such as printers or copy machines. On the other hand, other interviewees mentioned an office within their institution in charge of installing and explaining how office equipment operates. SScE593 explained:

*“No, almost never. If there were a situation with any office equipment, people from the Systems department would come and fix it”.* SScE593

Other interviewees mentioned they needed to understand the installation and use of specialized equipment, such as those used for healthcare or analysing materials or substances. PNScE620 stated the need to understand manuals:

*“70% of our products are foreign, and all the manuals are in English; it is not unusual that they come in English”.* PNScE620

Participant HScE562 expressed the need to understand instructions. She is self-employed; therefore, she does not have a department in charge of installing and explaining to her how to use the equipment:

*“Yes, because they send you something, for example, simply put...many of the indications of (use of) the materials are in English...also the equipment we buy; however, many of them also have instructions in Spanish”.*

HScE562

Some interviewees said they occasionally look at the instructions (whether in English or Spanish) because they are already familiar with how the equipment works. They would do so only with new equipment they are not familiar with.

The last item within the reading competence section is another one where all participants agree on its importance. All participants agree that using several pieces of

information, such as a work schedule to see who should be working or several reports on a process to identify the source of a problem (C22), is not so relevant for them. All mean values are below the relevant total mean. Therefore, C22 will not be considered for any of the disciplinary groups.

Qualitative data results reported that understanding academic journals is, perhaps, the most essential skill. Given that the questionnaire applied did not contain a competence related to this language use activity, it was not reflected in the quantitative findings. However, it was a language use activity frequently mentioned by domain insiders and test takers during their respective interviews. Many participants from the two groups and the three disciplinary areas acknowledged the most recent information in their fields of interest is shared in English. Using English for sharing developments in the professional world is done by both native and non-native speakers of English. Understanding new advancements is one area many consider very important to be a *good* practitioner.

### 6.2.5 Writing-related language competences

The writing-related language competences are the last group of items in the questionnaire. Among the six competences within this group, few were found relevant for participants.

Writing short reports, faxes, emails, and memos on basic everyday job-related matters (C23) is reported to be relevant for students from the three disciplinary groups; however, it is not the same for professionals. These results can be seen in Appendix Table 13 below. PNSc professionals do not seem to consider this competence to be relevant. Very few comments were made on this topic. Sometimes, it seemed like participants wanted to justify the need for that competence, but the examples provided were unrelated, as is exemplified below by SScE568 and HScE599, respectively:

*“Well, sometimes, not much, really. I don’t know if it applies here; as I said before, we are thinking of having a menu in English...we think it is necessary”.* SScE568

*“Well, maybe sometimes. If we do, as I said, it would be when we receive information (through email), and we just reply”.* HScE599

The comment made by SScE568 seems to try to please the interviewer with the response. She mentions having plans to write a menu in English; however, even if she does this, that will be a one-time activity. Moreover, the text she refers to belongs to a different genre than the one this competence refers to (everyday written interaction). The comment made by HScE599 also seems to refer to a different type of genre, perhaps more related to C24. However, it would require other language functions (accepting, refusing, clarifying, etc., rather than requesting and reporting) than those implied in C23. Even though the competence includes email as one of the means used for sending written reports, the activity mentioned by HScE599 is to reply or follow up on communication received, not on writing reports. These, if required to be done by professionals, would most likely be carried out in Spanish, as stated by HScE637:

*“No, I don’t do it in English, not at all”.* HScE637

The response given by HScE637 reveals when she needs to write reports, these are done in her mother tongue. This response is aligned with the results found in a study aimed at identifying the language needs of a dentist in Chile. Most writing would be done in Spanish, except when emailing non-Spanish-speaking colleagues in foreign countries (Olusiji et al., 2022).

Even though the mean values for this competence are above the mean value, it is considered that participants misunderstood what this competence entails; therefore, they marked it as required. Communication among employees in a language other than the mother tongue is unlikely to occur in México. If it took place, it would likely be among senior staff of an international company where a non-Spanish speaker is involved.

These findings contrast with one of the competences international medical graduates seeking professional employment in Australia are required to perform when taking the Occupational English Test (OET). One of the test tasks in the OET requires non-English speakers to write a referral letter (Davidson, 2022), where the health

professional “asks another health provider or health service to treat or diagnose a patient for a particular condition” (Healthdirect, 2022).

The above argument and absence of relevant examples of language use situations of C23 by interviewees leads to excluding this competence from those to be considered for the design of the test. These results contrast those obtained from interviewing employers of multinational companies based in central México. Concerning writing, all nine employers participating in the study said they needed university graduates to have written communication. Within writing, all needed professionals to write emails and have problem-solving interactions in written form; 3 needed employees to do persuasive writing, and 2 required written reports (Garcia-Ponce, 2020).

On the other hand, some HSc professionals report having English-speaking patients. Part of the process of attending to patients is to provide indications on how to take the medication prescribed. Only HScE582 commented on the need to prescribe in English; he stated:

*“Reports and these others, no, not really, the only thing... I don't know if it applies here... the prescription we write to the patient, the indications one gives them so they can follow the treatment. It's easier for them if it's in English, so they don't get confused.”. HScE582*

Most Q2 respondents linked C23 to other activities and expressed not needing English to fulfil that competence or not requiring the use of that competence at all. Only HScE582 linked this competence to writing a medical prescription in English. Some test takers from the HSc also mentioned it might be possible they had English-speaking patients. This situation was commented on by HSc85 above; he said he currently needed to write prescriptions in English so that patients could understand indications following the healthcare provider's appointment. Writing indications to follow the healthcare treatment is considered part of the process of attending to patients. In the interest of meeting this language use activity, the previous C23 competence will be restated as follows:

- C23. Write short, simple indications to follow a healthcare treatment, where most are comprised of abbreviations, fixed expressions, or formulaic language.

As argued above, this competence will be included only for the HSc group and dropped for the other two groups.

All participants agree that writing requests for goods, services, etc., on a range of routine matters (C24) is a relevant competence for their job-related activities. All mean values were reported above their respective total mean.

Despite the high mean values reported by all groups, the examples of language use situations provided were limited. Most of the examples of language use situations came from the HSc group. Two examples, HScE575 and HScE630 are provided below:

*“It would sometimes be when we are going to buy equipment, we check prices, sometimes it’s better to buy things here, sometimes not”.* HScE575

*“...yes, to buy equipment”.* HScE630

Another participant mentioned requesting goods but not always needing English, even though the order was for an American company. HScE563 stated:

*“...I use imported goods, all of them, 98%, so...most of the documents I send are in English; however, when they arrive in the United States, there are a lot of people who answer in Spanish because there are a lot of people who are paisanos\*”.* HScE563

This situation is not always the case; it may depend on each company. If a company sees México as a potential buyer, it may appoint Spanish-speaking employees to assist Mexican/Latin American customers. HScE637’s comment below provides an example of orders placed to an international company without using English:

*“No, I don’t do it in English. I do everything in Spanish, and they accept it that way”.* HScE637

The comment made by HScE637 implies the requests are placed to an English-speaking company when he says, “...*they accept it that way*”. This comment would not be necessary if the speaker referred to placing an order to a Spanish-speaking company.

Other interviewees said the language use activity represented by C24 is not within their job description, as stated by SScE614:

*“This is not something I do”.* SScE614

What SScE614 was probably trying to say is that there are people in other departments who take care of ordering the supplies needed.

Despite some interviewees stating they do not carry out the language use activity C24 represents, it is considered that being able to carry it out is relevant for professionals. Language should not be a limitation for professionals to acquire the goods or services required for their professional activity (Education First, 2014b). On these grounds, C24 will be included in the competences for test design.

The first competence in which all students and professionals have a different opinion regarding its relevance is making notes on routine matters, such as taking/placing orders (C25). This competence is reported to be relevant for all professionals and not so much for students of the three disciplines (mean values from students are below their respective total mean), as seen in Appendix Table 13 . The mean value from the PNSc is the same as the total mean for the writing-related competences. The difference between the HSc mean value and the total mean is just 0.01. The difference between the mean value from the SSc and its total mean is the largest; though it is not so great, it is 0.03 above the total mean. This slight difference between the total mean and the one from each group of professionals may be related to the fact that participants provided no examples of language use situations. Six participants responded they would not make notes in English, and two said it was the same as C24: requesting for goods, as exemplified in the comment made by SScE612 below:



*“Yes, sometimes, it’s the same as the previous one when one places orders for goods”. SScE612*

This comment suggests a misunderstanding of the difference between C24 and C25. Many participants who checked this competence may have thought it was the same as the previous competence (requesting goods or services). While C24 is addressed to English-speaking companies and may require the order to be placed in English, C25 refers to making notes in English about routine matters. The language used in C25 may depend on whether the notes are for personal use or another employee. If they are for personal use, they may not be written in a language other than the mother tongue. Based on qualitative (absence of examples of language use situations provided by interviewees) and quantitative (mean values the same or almost the same as the total mean, therefore on the verge of being considered or not) data, this competence will not be taken into account for the design of the test.

Filling out a form, such as an incident/accident report form, such as a medical record (C26), is considered relevant only for SSc students. There was only one example from HScE637 regarding the use of this competence; she stated:

*“Yes, this I have done in English. Um, I work for an insurance company, and yes, they have um...it is necessary to fill out some reports in English, but not very often; it happens when an incident occurs in a foreign country”.*  
HScE637

The rest of the professionals interviewed said they did not need to do that kind of report or, if required, would do them in Spanish. Hence, this competence will not be considered for any of the groups.

All students and professionals agree that writing reports of events or incidents involving products, machines, or personnel (workers, clients) (C27) is barely relevant. All mean values for C27 are reported below their total mean, revealing participants do not think this competence is required. This competence will not be considered for any of the disciplinary groups.

The last competence of the questionnaire, C28, reports the same behaviour seen for C25 above; all professionals agree, and all students disagree on its importance. The exact opposite opinion on the need to make notes for personal use on non-routine aspects such as a client's requirement or simple characteristics of a new product or service (C28) is expressed by the two groups of participants, as all mean values from student data are reported below their respective total means.

As with C25, the small number of comments made by interviewees regarding C28 specifies that if they needed to take notes, they would make them in their mother tongue. They also said that if notes were taken in English, they would likely be copying information displayed in the equipment used (which is in English). No examples of language use situations of C28 were provided. As stated before, it is unlikely that somebody would make notes for personal use in a language other than the mother tongue.

This competence will also be discarded from the group of competences to be included in the school exit test.

### **6.3 Most relevant competences within the workplace**

The analysis of mean values and comments provided by professionals and students made it possible to identify the competences relevant to the local workplace environment. The number of competences was reduced from twenty-eight to seventeen. The result of the analysis shows similarities and differences in the language needs of the three groups. This difference in language needs implies developing three tests, one per discipline group. The competences that are relevant for each disciplinary group are presented below.

#### **6.3.1 Speaking and listening**

The group of speaking and listening competences are shown in Table 6.1 below. The check (✓) shows the relevance of each competence and, consequently, the test where it will be included. It can be seen that some competences are relevant to more than one disciplinary group. Offering advice to clients (C1), asking for clarification (C8), expressing opinions (C9), participating in job-related meetings (C10) and greeting a

visitor (C12) are relevant to the three disciplinary groups. Requesting routine services within one's own area of work (C2) and making questions of a fact-finding nature (C3) are relevant to two groups (SSc and PNSc) and not relevant to HSc participants. Finally, providing instructions and directions (C13) is only relevant for the PNSc group and making phone calls requesting information and services (C15) is only relevant for HSc participants.

Speaking and listening competences				
	Competence	SSc	PNSc	HSc
C1.	offer advice to clients within their own job area on simple matters.	√	√	√
C2.	Request routine services related to your own area of work.	√	√	—
C3.	ask questions of a fact-finding nature.	√	√	—
C8.	ask for clarification when something is not clear (e.g. "Can you repeat that please?", "Is it ...or ...?", "So, the.... is/has /needs...")	√	√	√
C9.	express opinions in simple terms, for example, "I don't agree" "I think..." provided the question/issue has been put clearly and simply.	√	√	√
C10.	participate in job-related meetings and seminars (e.g. making suggestions or asking questions: "Why don't we...", "Can we...", "Is it possible to...").	√	√	√
C12.	greet a visitor and engage in a limited conversation for a short time, for example, enquiring about a visitor's journey, hotel, etc. (e.g., "How was your ...", "Did you like...?", "Is/Was your .... ?").	√	√	√
C13.	provide instructions and directions, such as giving a client clear direction to the business/company. (e.g., "Take a taxi to. ...", "Go straight...", "Walk two blocks, then turn...").	—	√	—
C15.	make phone calls requesting information or services (e.g., "Do you have...", "I would like...", "How much/many...?")	—	—	√

Table 6.1 Most relevant speaking and listening competences per disciplinary group

Some competences are relevant for more than one disciplinary group. Offering advice to clients (C1), asking for clarification (C8), expressing opinions (C9), participating in job-related meetings (C10) and greeting a visitor (C12) are relevant to the three disciplinary groups. Requesting routine services within one's own area of work (C2) and making questions of a fact-finding nature (C3) is relevant to two groups (SSc and

HSc) and not relevant to HSc participants. Finally, providing instructions and directions (C13) is only relevant for the PNSc group and making phone calls requesting information and services (C15) is only relevant for HSc participants.

Competences 9 and 10 were referred to be relevant when learning more about their field of interest. Participants (HScE582, HScE642, HSc20) said they would use these competences when attending courses, seminars, conferences, or any other activity in which they could interact with experts on the topics addressed. Competence 1 was related to letting the client know about the characteristics of the options available in products or services (SSc88student SSc59, HSc24, HScE582). Competences 12 and 13 are seen as more related to socializing with foreigners (PNScE635), whether with company visitors, clients, patients, or other participants in the learning events they may attend. Competence 8 and 15 were reported to be relevant when acquiring goods or services from international companies, getting information, or asking about characteristics, terms, and conditions of what they are interested in acquiring (HScE582).

### 6.3.2 Listening and speaking, and listening-only competences

The listening and speaking competences and the listening-only competence that are relevant for the disciplinary groups are shown in Table 6.2 below. In speaking and listening competencies, receiving phone calls (C5) is relevant for SSc and HSc, but not for PNSc. Dealing with predictable requests from visitors (C14) is relevant for the three groups of participants. Receiving phone calls was related to patients making an appointment on the telephone or as a follow-up to a product or service acquired from an international company (HScE582). Participants related competence 14 to receiving foreign company visitors (HScE572, PNScE635). Some participants reported they have sometimes been appointed to pick visitors up at the airport and attend to them while in Durango.

Only one competence from the listening-only group was found relevant to participants; this is related to keeping up to date in the discipline of interest. All groups are interested in being able to follow a simple presentation, demonstration or explanation

related to their own area of professional practice (C7). This competence also relates to understanding the latest information within their practice field.

Listening and speaking competences						
	Competence	SSc		PNSc		HSc
C5.	receive phone calls (e.g., take messages/make notes, etc.).	√				√
C14.	deal with predictable requests from a visitor, for example, "Can you arrange a taxi to the airport?", "Can I use a projector in the presentation?", "Where can I....?"	√		√		√
Listening competence						
	Competence	SSc		PNSc		HSc
C7.	follow a simple presentation/ demonstration and understand explanations concerning a product or topic within your own area of expertise.	√		√		√

Table 6.2 Most relevant listening and speaking and listening only competences per disciplinary group

Only one competence from the listening-only group was found relevant to participants; this is related to keeping up to date in the discipline of interest. All groups are interested in being able to follow a simple presentation, demonstration or explanation related to their own area of professional practice (C7). C7 is also related to understanding the latest information within their practice field.

### 6.3.3 Reading competences

One of the reading competences found to be very relevant for the three groups of participants is related to acquiring goods or services for professional practice (C19). This is shown in Table 6.3 below. Participants mentioned that new equipment is not acquired as frequently as supplies are. This competence was also mentioned as relevant to learning about events or products related to the field of interest. They may get information about new products or training and information-sharing gatherings.

For some professionals, understanding installation or operating manuals may be an advantage over waiting until somebody from another department addresses the situation, as could be the case with office equipment.

For others, finding their way around the equipment may be an essential aspect within their professional practice, as was expressed by some practitioners of the health HScE562 and physical and natural sciences PNScE620. Other practitioners also mentioned the importance of understanding labels, intake instructions and medication or supplement dosage they instruct their patients to take.

For others, finding their way around the equipment may be an essential aspect of their professional practice, as was expressed by some practitioners of the health HScE562 and physical and natural sciences PNScE620. Other practitioners also mentioned the importance of understanding labels, intake instructions and medication or supplement dosage they instruct g their patients to take.

Reading competences						
	Competence	SSc		PNSc		HSc
C19.	obtain basic relevant information (product specifications, professional or commercial leaflets, advertisements, internet pages, etc.)	√		√		√
C21.	understand general written instructions (installation, functioning or maintenance manuals).	√		√		√
C29.	understand, given sufficient time, most information of a factual nature that (s) he is likely to come across during his/her studies.	√		√		√

Table 6.3 Most relevant reading competences per disciplinary group

Reading was reported to be an essential skill for test takers of the PNSc and the second for the other two disciplines. There were many comments related to understanding written discipline-related texts (see 5.3.2.4.1 above). Some test takers, such as HSc3M and SSc80, referred to understanding field-related texts as part of the activities they consider to be part of their future professional practice. The comments made by test takers regarding their interest in the academic domain (see 5.3.2.2

above) support the relevance this language use activity has for them. Professionals also frequently referred to the need to stay updated in their field. Some of them mentioned the ability to understand articles or other field-related publications as an example of the language use situation of one of the reading competences (even though that type of activity was not in the competence mentioned).

Considering the above, it is considered that adding a competence from the Study domain from the ALTE CAN DOs (ALTE, 2002) responds to the many references and comments regarding the relevance of understanding journal articles or other field-related written publications. It is assumed that most of those interested in this competence would like to be able to understand academic articles fully. This level of proficiency is achieved at C1 or C2 levels of the ALTE CAN DOs (ALTE, 2002). Given that the school exit level that students are required to demonstrate is B1, including competences in the test above this proficiency level would be against this university regulation. Modifying the UEP (Reyes Fierro, 2008) specifications would require the approval of both the Chancellor of the university and the Board of Directors. An alternative to modifying the UEP would be to use an ALTE 2 (Study domain) competence. The one that best meets this need reads:

CAN understand, given sufficient time, most information of a factual nature that (s)he is likely to come across during his/her studies.

This competence was added to Table 6.3 above. It will be referred to as competence 29 (C29). This competence will be added to the reading competences that will be included in the school exit test for the three groups.

#### 6.3.4 Writing competences

Competence 23 was dropped as initially stated, given that the type of written products it included are not likely to be required in the local context, as argued in 6.2.5 above. However, in the interest of addressing expressed needs by HSc participants, it has been modified to meet the needs of this group. Given its specificity, it applies only to the HSc group. This is shown in Table 6. 1 below.

Writing competences						
	Competence	SSc		PNSc		HSc
C23.	Write short, simple health care instructions, mostly comprised of abbreviations, fixed expressions or formulaic language.	--		--		√
C24.	write requests for goods, services, etc., on a range of routine matters, but MAY need to get these checked.	√		√		√

Table 6. 1 Most relevant writing competences per disciplinary group

The writing activities all participants reported to be more relevant for their job-related activities are requests for goods or services they require or are familiar with (C24). Interview data revealed professionals used this competence for placing supply orders to international companies on the internet, as discussed in 6.2.5 above. Some said they buy equipment (see HScE575 and HScE630 above); however, this does not happen as frequently as they acquire supplies, as HScE563 above states.

## 6.4 Summary

Results reveal that spoken interaction (speaking and listening) competences are more important than only listening, though it is still required. Reading is also much more relevant than writing; participants report they are required to do little writing.

Even though there are more language use activities related to speaking and listening than those related to reading, the latter is reported to be of great importance in maintaining professional knowledge.

These results are consistent with the findings reported in the study conducted by Cambridge English in 2016 (2016b) in 1.6.1 above.



## **7 Considerations for a local context-sensitive test.**

### **7.1 Introduction**

Data analysis results reveal the need for a more work-related exit test. Results from Questionnaire 1 (in Chapter 5 above) disclose test takers consider a general English test is not related to their profession: They would like the exit test to be related to the workplace so they can be prepared to carry out the language use activities they may encounter in their job. Questionnaire 2 results (in Chapter 6 above) show the language competences university students and professionals selected as the most relevant for the workplace. The semi-structured interview results provided specific examples of language use activities in the workplace for the different disciplinary groups of participants. The design of a more work-related exit test would consider the local workplace context language needs; it would be a local context-sensitive language test (from now on, LCSLT). The argument-based approach for test validation was suitable for validating an LCSLT.

The argument-based approach was found to be appropriate and flexible, as discussed in 7.2 below.

Knoch and Macqueen (2020) claim the target language use (TLU) domain is the centre of language assessment activities. It is argued that tests designed considering the real-world language needs of test takers will bring beneficial consequences to them and other interested parties.

An LCSLT can bring beneficial consequences to its primary stakeholders, test takers, because it considers the characteristics of the language used. The LCSLT content draws from identifying and analysing the language use activities required in the local context of interest.

The test development process proposal of the LCSLT draws initially from Bachman and Palmer (2010), then incorporates some elements from the Knoch and Macqueen (2020) and Bachman and Damböck (2018) proposals to the elements that the LCSLT presents. This proposal is considered in 7.3 below.

The validation proposal of the LCSLT draws primarily on the AUA (Assessment Use Argument) presented by Bachman and Palmer (2010); however, it also incorporates elements from the LAPP approach. This proposal is discussed in Section 7.4 below.

## 7.2 The argument-based approach to validation

The argument-based approach (ABA) to test validation focuses on evaluating the existing evidence to justify using a test in a particular situation (Sireci, 2013), which “addresses the limitations in the traditional frames of validations (content, criterion-related, and construct validities)” (Im & Cheng, 2019, p. 2). Even though it was developed for the educational measurement community (Xi, 2008), reviews on the use of ABA by Im and Cheng (2019) and Dursum and Li (2021) report this approach has increasingly been used in many research studies.

Im and Cheng (2019) reviewed eight of thirty-three empirical studies found using an ABA in five language testing journals and the ProQuest Dissertation and Thesis database. These were published between 1992 and 2016. For their part, Dursum and Li (2021) found seventy publications, journal articles, reports, and thesis dissertations between 2000 and 2018. Thirty-five publications were found in five language testing journals. They also identified twenty-five publications in the ProQuest Dissertation and Thesis Global database and ten in the Educational Testing Service’s (ETS) Research Report Series. They claim that upon the appearance of the first ones in 2005, the number of publications gradually increased, with an increase observed after 2011. Their analysis reveals the ABA provides a flexible framework that may be used in various contexts and for different purposes using the methodological approaches that best suit the research purposes of the authors. Most of the studies were conducted in North America (56), mainly in the United States of America (USA) (50). Five corresponding authors were affiliated with institutions in Canada, and only one was carried out in México. The publications whose corresponding authors had affiliations in Asia (6), the Middle East (2), Australia (4), and Europe (2) indicate that the ABA has been found relevant beyond its North American origin.

Im and Cheng report the six dissertations reviewed in their study conducted research using the ABA in high-stakes testing contexts. Dursum and Li report a broader use of the ABA. Besides investigating the validation of standardized tests developed by

professional companies, the ABA has been used in projects addressing institutional and domain-specific tests or tasks of recent development. Most of these studies were part of dissertation projects, some technology-related. Some publications report on placement tests, rating scales and classroom-based assessments. They do not report, however, on the use of the ABA in the development of a new test.

Both Im and Cheng (2019) and Dursum and Li (2021) report that most of the studies did not go through the whole evaluation process of all the inferences or claims. According to Dursum and Li (2021), the number of dissertations addressing a maximum of six inferences is more significant than the number of journals or research reports. They contend this may be related to the nature of the publication venue. The most frequently studied is the explanation inference, and the least addressed is the domain definition inference (twelve studies). The former can be established if the evidence gathered “indicate(s) that the test scores reflect the construct that the test is intended to assess” (2021, p. 56). The latter can be established by determining the link between test task performance and the expected target domain of language use performance.

Dursun and Li (2021) report publications used quantitative and qualitative research methods, such as document analysis, observations, needs analysis, statistical analysis, expert judgement, interviews, surveys and conversation analysis.

More recent reports, journal articles or conference papers found using an ABA support the trend reported by Dursum and Li (2021) regarding the affiliation of corresponding authors. Eight studies were retrieved, though one of them (Fechter et al., 2021) is biology-focused rather than language-testing related. Three of the seven studies are research reports: two from American professional testing companies (Stone & Wylie, 2019; J. Schmidgall et al., 2021) and one from a British testing company (Xu et al., 2020). One conference paper and three journal articles were also identified. The corresponding author of the former (Pardo-Ballester, 2020) and two of the latter (Yan & Staples, 2019) (Suzumura, 2022) have an American affiliation. The third corresponding journal article author has a Chinese affiliation (Chen, 2022). Six of the eight publications found between 2019 and 2022 come from institutions based in the USA. Four of the seven language testing-related publications are technology-related. None of these studies is concerned with test or task development. Their inquiries pertain to exploring one or more inferences of an existing test to test-taker

performance. However, some similarities to the current study were found in earlier reported studies.

Liu (2014), professor Lyle F. Bachman's former Ph.D. student, used the CEFR (COE, 2001a) CAN DO statements to design a domain-related task-based reading test. The results of applying this test were compared to those of the high-intermediate reading General English Proficiency Test (GEPT) (Roever & Pan, 2008). The purpose was to learn the relationship between general English tasks and work domain-related tasks. The study concluded that the information the GEPT reading scores provided was insufficient to generalize to the target language use domain; "more task-based and workplace specific" (2014, p. iii) items may provide more meaningful or generalizable scores. As the present study, Liu's dissertation used international language descriptors to address the target language use domain, and both studies favour the use of task-based test tasks. Moreover, both studies contend test tasks need to reflect real-world language use for meaningfulness and generalizability.

Only one study focusing on a high-stakes exit test was identified among those reported by Im and Cheng (2019), Dursun and Li (2021) and the ones found between 2019 and 2022. He and Min (2017) developed and validated a computer adaptive test. They focused on the third claim, "interpretations of assessment records" of Bachman and Palmer's (2010) assessment use argument (AUA), particularly on the "meaningfulness" quality. This high-stakes university exit test was developed to assess non-English major Chinese students' reading and listening proficiency in English. The present study and the one by He and Min are concerned with a high-stakes English language exit test; however, they do not state how the construct to be assessed was determined. They did not conduct a needs or domain analysis or consider the local language needs of test takers. Furthermore, they did not distinguish between disciplinary areas of test takers and were only concerned with testing the receptive skills.

In contrast, more similarities were found with the study by Deygers et al. (2018) in northern Belgium. Even though this study focused on assessing the use of university entrance tests rather than exit tests, as the present study does, both are high-stakes. Both use an argument-based approach, though the Deygers and collaborators' study draws on Kane's (2013) interpretation-use-argument (IUA) (see 3.6.1 above). The present study draws mainly on Bachman and Palmer's AUA (see 3.6.2 above) and Knotch and Macqueen's (2020) LAPP approach (3.6.3 above) to propose the LCSLT

discussed below. Both studies consider a language level related to international language descriptors: the B2 (COE, 2001a) for the Flemish entrance tests and the B1 (ibid) for the present study. While the Belgian entrance tests were already in use, this study presents a test proposal. Both studies are concerned with the four skills and language for specific purposes, academic for Flemish and workplace for English, and both consider students of different disciplines. However, the Belgian study only reports overall results rather than discipline-related ones. The absence of this information in the journal article could be due to publication venue constraints or the limited number of participants in the study. Even though Deygers and collaborators do not apply a needs or domain analysis, they consider experts (subject teachers) and test takers. While data from expert insiders (professionals) in this study and test takers are considered, the former group of participants are deemed to be better informed of real life language needs than the latter. In contrast, in the Belgian study, teachers' opinions are valuable, but test takers' comments reflect real-life language needs. An example of this situation reported by Deygers and collaborators is that teachers assume first-year students are familiar with specific academic vocabulary. Still, a comment from a Spanish-speaking student reveals the opposite.

The studies described above and a quick look at the titles of the studies analyzed by Dursun and Li reveal the flexibility of the argument-based approach (ABA).

Researchers in most publications do not focus on investigating all the claims in an argument-based approach. In contrast, they seek to investigate the relevant claims for their purpose and context. Moreover, they do not necessarily adopt any of the claims concerning an ABA; they may adapt the claims to fit their research interests and resources.

The following sections discuss how some ABA proposals are combined and adapted to fit the needs of an LCSLT.

### **7.3 Test development process**

From initial planning to operational use

The steps in the test development process are described, starting with initial planning, then identification of local language use needs, then the development of the design statement, the blueprint and finally, design patterns. This process is discussed below.

### 7.3.1 Initial planning

Bachman and Palmer (2010) argue a decision maker may need to answer a series of questions during the initial planning stage to decide whether a test is necessary or whether the required information could be gathered through other means, such as classroom observations or student work.

If a test is required before embarking on a test development project, it is advisable to consider using an existing test (L. Bachman & Palmer, 2010).

Clearly identifying the test taker population and the situation(s) in which the language will be used (testing needs) provides the test designer with the information necessary to verify the suitability of existing tests for the purpose and the interest group. For this research, existing tests were considered but discarded as they were designed for international students aiming to work or study in an English-speaking country. The target population and context of use of the revised tests differ from those involved in this study.

A test is developed when there is a weak or no correspondence between existing tests and the test needs (test takers and the TLU situation). Weak or no correspondence was found between existing tests and the testing needs of this research. The LCLST proposal assumes the need for a test has been identified and emphasizes the importance of identifying the test takers and the TLU domain of interest. The identification stage is the starting point in the test development process, as the test developer can make informed decisions in the following stages. More importantly, considering these factors is vital to the intended outcome; the consequences benefit the test-takers and stakeholders.

The initial stage of the LCLST is shown in Figure 7.1 below. The initial stage considers Bachman and Palmer's proposal (in blue) to ponder using an existing test. However, the LCLST (in white) believes it is only possible to consider whether an existing test can be used once the test-takers and the target language use domain of interest have been identified. If this information is not known, the grounds on which the decision to use a test or not may not be valid.

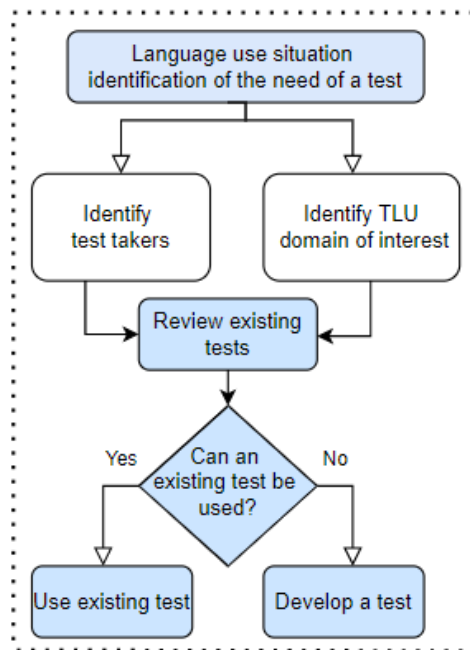


Figure 7.1 Initial planning stage, adapted from Bachman and Palmer

This research project began with the premise that there was no correspondence between existing tests and local language needs in the workplace. However, until the data collected was analysed, this premise was no longer an assumption; no correspondence was found with existing tests. The test (see 1.7.8) used during data collection (see 4.4.1.1 above) was considered irrelevant by the test-takers.

### 7.3.2 Identifying the local language use needs

Once an existing test has been discarded, the development stage begins by analyzing the TLU domain under the real-life approach to communicative language testing (Liao et al., 2023). This analysis is a step not included in Bachman and Palmer's (2010) diagram (see Figure 3.2 above). Identifying the TLU domain is also the starting point in the diagram presented by Knoch and Macqueen's (2020) diagram (see Figure 3.4 above), although they refer to this as a needs analysis. They argue that needs analysis involves analysis of the domain. However, it is claimed that this stage needs to be called "domain analysis" to ensure that the domain analysis takes place instead of conducting other types of needs analysis, such as *wants* or *perceived needs* analysis. Labelling this stage as "domain analysis" does not suggest conducting other types of needs analysis would not be beneficial. Still, it does ensure that the analysis of the

domain is taking place and gives it the prominent place it holds in the LCSLT development process.

In an LCSLT, domain analysis considers both test-takers and relevant informants in the target language use situation. Relevant informants can be expert insiders, teachers, employers or receiving institutions. Who the relevant informants are depends on the purpose of the test development purpose and the people who can help identify the language use activities related to the context of interest. In an LCSLT, the domain analysis is adapted to the specific context of interest. Suppose the test aims to identify people's language use activities in the tourist industry. In that case, the relevant informants will be found in the specific establishments of the relevant tourist-related activities (i.e. restaurants, hotels, museums, amusement parks, etc.). The target language use situation for the domain analysis will be tourist-related. The expert insiders may be waiters, managers, guides, hosts, receptionists, etc.). The entire diagram for the LCSLT development process is shown in Figure 7.2 below.

Domain analysis shows the target language use (TLU) activities or situations of the domain of interest. The LCSLT suggests using international language descriptors to perform the domain analysis. These could come from the CEFR (COE, 2001), the ALTE (ALTE, 2002) or the CLB (CCLB, 2012). The former and the latter can be used to identify the target language use activities of the context of interest that is not domain-specific. The ALTE and some of the domain-specific language descriptors developed based on the CLB can be used to identify the target language use activities related to specific contexts of language use, such as study, work, leisure or travel purposes.

An advantage of using international language descriptors is that results can be expressed and related to them. The well-known international language descriptors allow interested parties to better understand and compare the language requirement level to other language requirements or test results.

In the absence of a syllabus of instruction, as is the case, the result of the domain analysis becomes the basis for defining the construct to be assessed (L. Bachman & Palmer, 2010).

If there is only one group of participants, a set of language-use activities will be selected to be included in the test (L. F. Bachman & Palmer, 1996). The present study identified three groups of participants, so it was necessary to analyse their language



use activities in search of similarities and differences. A comparison of the data analysis outcomes of the Health Sciences, the Social Sciences and the Physical and Natural Sciences revealed most language use activities are relevant across the three groups of participants (see 6.3 above). Some activities are relevant for only one or two of the groups of participants. Some of the language use activities require the use of specialized language.

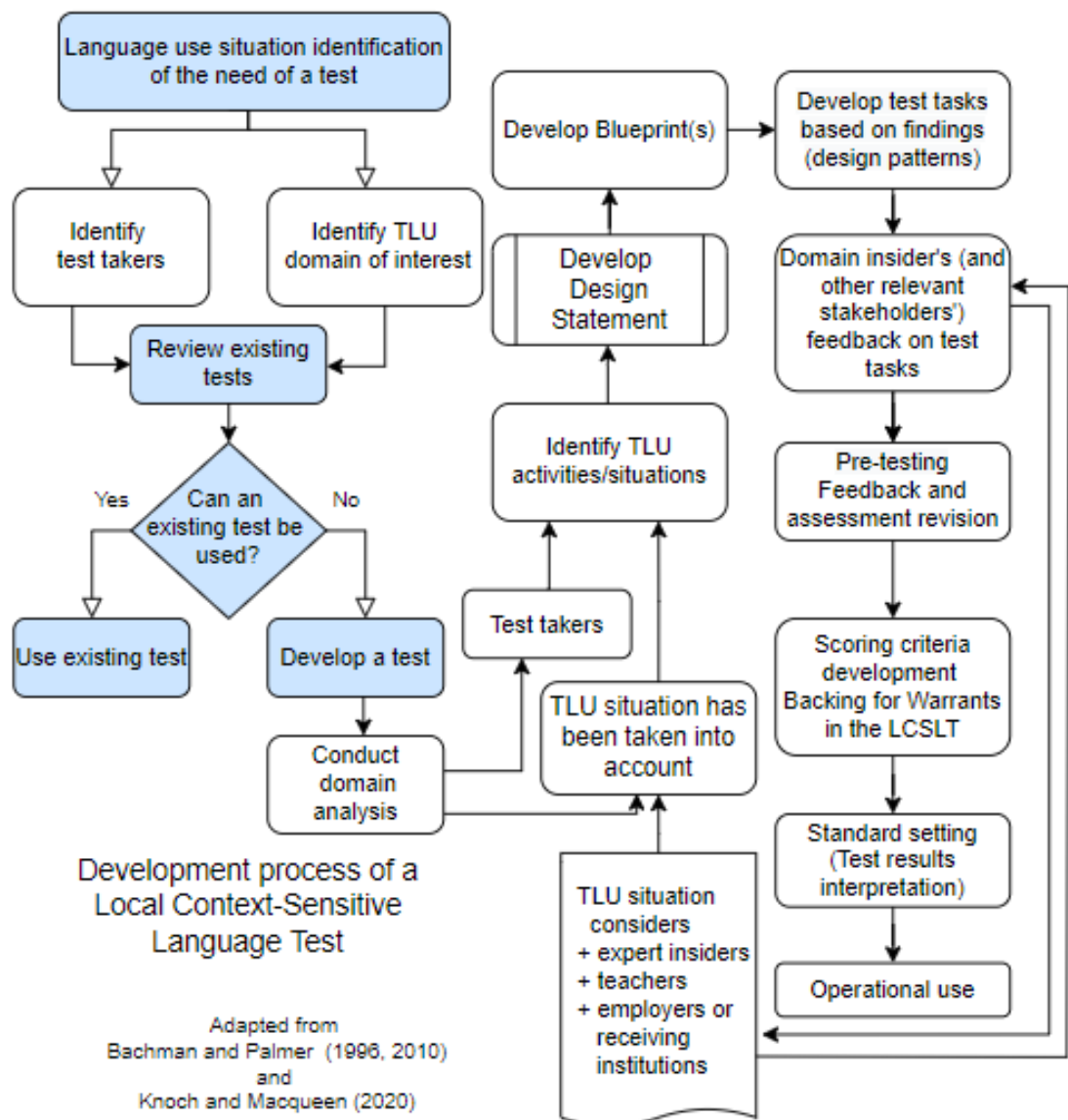


Figure 7.2 Development process of a Local Context-Sensitive Language Test

The results of the data analysis showed that all participants need to use competences that include listening, speaking and reading skills. However, most participants do not need to write in English. The skill requirement is discussed further below.

### 7.3.2.1 Domain analysis: identifying TLU activities

Future test takers, test takers and domain insiders were involved in identifying the local language needs in the workplace. The latter provided examples of language use activities; a few more came from test-takers. These were found in the public (state or federal) and private (employees or self-employed) sectors.

Test takers, as the end users of the university-leaving test, were an essential source of information (O'Sullivan, 2012, p. 79). They were given a voice to express what was relevant to them. Their opinion on several test-related aspects was obtained; one of them is the domain they think the test should focus on. Some chose the academic domain because they felt it would give them the tools to understand subject-related texts. Whether oral or written, they value understanding the language of science, not only during their academic life but also for their later professional practice. They believe that staying up-to-date in their field is characteristic of qualified professionals.

A less general, more subject-related language test was the result of asking the test-takers about the relevance of the social, academic and work-related domains.

Participants' comments confirm the importance of the correspondence between test content and the use the test is put to (see discussion in 3.5 above). A test serves its purpose when it measures the knowledge, abilities or skills required in the context in which the test-takers will use the language. Most participants think the current university-leaving test (1.7.8 above) does not measure the language they need after graduation. Comments such as those from participants Q1UD20 and Q1SSc561 above exemplify this. In addition, test-taker comments justifying their choice of the domain (academic or work) for the university-leaving test focus and reinforce the need to change the content of the current test. The language use activities chosen by students and professionals and the specific examples of language use situations for the different disciplines provided by the latter and some test-takers provide additional support for a more discipline-focused test.

Mean scores from Q2 (see Appendix XI below) indicate that English is not used for everyday communication. As discussed in 1.6.5 above, daily communication for work-related activities in the local context of the present study takes place in the native language of the professionals. English is only used for certain types of activities. Their frequency may vary.

The results also show that some skills are required more frequently than others. The highest mean ( $M=3.16$ ) reported for reading competence is consistent with international research findings (see 1.6.1 above) that reading was identified as the most essential skill in countries where English is not an official language. Although speaking is the second most important skill according to international research, participants in this study ranked understanding what others are saying as the second most important skill. The increased importance of listening could be related to the increasing opportunity to access international online conferences and videos on almost any topic.

A comparison of the most relevant competences shows more similarities than differences between disciplinary groups in terms of the types of language use activities considered relevant to them. The following sections will discuss whether the selected competences are subject-specific or not. From this, it follows whether the tasks to be created can be used for more than one disciplinary group or must be designed to target a specific area of knowledge.

#### 7.3.2.2 Speaking and listening competences

This group's competences imply oral interaction with at least one other language user. They also imply that the language user initiates the interaction, understands what their interlocutor replies to, and can respond accordingly.

Competences 3, 8 to 10 and 12 are relevant for the three groups. Competences 1 and 2 are only relevant for the SSc and the PNSc groups. Competence 13 is only relevant for the PNSc and C15 for the HSc group. This relevance is shown in Table 7.1 above.

Competences 3 and 8 to 10 are not considered to be discipline-specific, so the tasks designed could be the same for the relevant groups. Competences 1 and 2 are discipline-specific (light grey boxes in Table 7.1 above), so they need to be addressed separately for each discipline. Test tasks will have discipline-related content. Even

## Chapter 7 Considerations for a context-sensitive tests

though Competences 13 and 15 (dark highlight below) are not discipline-specific, a separate task needs to be designed for each. The separate tasks are designed because they are only relevant to a disciplinary group of test takers.

Speaking and listening competences		SSc		PNSc		HSc
1	offer advice to clients within their job area on simple matters.	√		√		
2	request routine services related to one's own area of work.	√		√		
3	ask questions of a fact-finding nature.	√		√		√
8	ask for clarification when something is not clear (e.g., "Can you repeat that please?", "Is it ... or ...?", "So, the...is/has/needs...")	√		√		√
9	participate in job-related meetings and seminars (e.g. by making suggestions or asking questions: "Why don't we...", "Can we...", "Is it possible...?")	√		√		√
10	express opinions in simple terms, for example, "I don't agree" "I think..." provided the question/issue has been put clearly and simply.	√		√		√
12	greet a visitor and engage in a limited conversation for a short time, for example, enquiring about a visitor's journey, hotel, etc. (e.g., "How was your ...", "Did you like...?", "Is/Was your .... ?").	√		√		√
13	provide instructions and directions, such as giving a client clear direction to the business/company. (e.g., "Take a taxi to. ...", "Go straight...", "Walk two blocks, then...").			√		
15	make phone calls requesting information or services (e.g., "Do you have...", "I would like...", "How much/many...?")					√

Table 7.1 Most frequently used speaking and listening competences for disciplinary groups.

### 7.3.2.3 Listening & speaking and listening competences

There are two types of competences in this section. The listening and speaking competences imply interaction, although the language user is not the initiator, as in the

Chapter 7 Considerations for a context-sensitive tests group of competences in section 7.3.2.2 above. Listening competence does not imply a verbal response to an interlocutor.

Competence 5, from the Listening and Speaking competences group, is relevant for both the SSc and the HSc groups but not for PNSc participants, as shown in Table 7.2 below. The task for assessing this competence can be the same for both groups, as it is not subject-specific.

Listening and speaking competences		SSc		PNSc		HSc
5	receive phone calls (e.g., take messages/make notes, etc.).	√				√
14	deal with predictable requests from a visitor, for example, “Can you arrange a taxi to the airport?”, “Can I use a projector in the presentation?”, “Where can I....?”	√		√		√
Listening competences		SSc		PNSc		HSc
7	follow a simple presentation/demonstration and understand explanations concerning a product or topic within your expertise.	√		√		√

Table 7.2 Most frequently used listening and speaking and listening competences for disciplinary groups.

Competence 14 is also not subject-specific and is relevant for the three groups. It could be assessed with C12 (section 7.3.2.2 above), where the incoming visitor expresses some requests. The task to assess these two competencies can be used for the three groups of participants.

Competence 7 is also relevant for the three groups; however, it is discipline-specific, so a separate task is required for each group.

#### 7.3.2.4 Reading competences

The reading competences listed in Table 7.3 below are relevant to the three groups of participants. Competences 19 and 21 could be discipline-specific or more general in

## Chapter 7 Considerations for a context-sensitive tests

content, but Competence 29 needs to be more discipline-oriented, as the competence clearly states. For reasons of practicality, the items designed for C19 and C21 will not be discipline-specific so the items will be used for the three disciplinary groups. On the other hand, C29 will be discipline-specific, and a separate task will be required for each disciplinary group.

Reading competences		SSc	PNSc	HSc
19	obtain basic relevant information (product specifications, professional or commercial leaflets, advertisements, internet pages, etc.)	√	√	√
21	understand general written instructions (installation, functioning or maintenance manuals).	√	√	√
29	Understanding discipline-related texts such as articles from journals, summaries, etc.	√	√	√

Table 7.3 Most frequently used reading competences for disciplinary groups.

### 7.3.2.5 Writing competences

Participants in this study do not need to write much in English. A writing competence is only relevant for the HSc group, as Table 7.4 below shows. This competence has been adapted from its original to meet the writing needs of this disciplinary group (see 6.2.5 above). A discipline-related task will be designed for this competence.

Writing competences		SSc	PNSc	HSc
23	write short, simple indications to follow a healthcare treatment, may make extensive use of commonly used abbreviations, fixed expressions, or formulaic language.			√
24	write requests for goods, services, etc., on various routine matters, but MAY need to get these checked.	√	√	√

Table 7.4 Most frequently used writing competences for disciplinary groups.

The other writing competence, C24, was reportedly used when writing emails to make requests for goods or services. This activity was found to be relevant for the three

Chapter 7 Considerations for a context-sensitive tests disciplinary groups. The test task could be designed as a non-discipline-specific task so that the same task can be used with the three groups.

### 7.3.3 Design statement

Knoch and Macqueen (2019) conceive the design statement as a description of the test purpose, the test-takers, the target domain, the construct and the available resources. On the other hand, Bachman and Palmer (2010) see the design statement as the pathway that guides the following stages (test development and validation) and defines how the results are to be interpreted. For them, the design statement contains selected elements of the first three claims (see 3.6.2 above) and their warrants. Arguments (claims) express the test characteristics believed to be favourable to the testing situation, context, target population and language needs. These three aspects are at the heart of an LCSLT. This approach to test validation is perceived to be more user-friendly. It allows test designers to engage in a spiral revision of task characteristics and the construct to be assessed to ensure the intended characteristics are addressed.

Bachman and Palmer's proposal for a design statement considers aspects that are relevant to an LCSLT. Therefore, the elements to be included in the design statement of an LCSLT are:

- the description of the purpose of the test
- the description of the test-takers
- what is the target domain
- the construct to be assessed
- an estimate of the resources required

Additionally, the Major Claim (beneficial consequences) is included in the Design statement of the LCSLT.

***Design statement.*** The purpose of the LCSLT is to measure the extent to which test takers (university students who have graduated from one of the following disciplinary areas within the institution: health sciences, physical and natural sciences or social sciences) have the language skills and knowledge to perform some of the work-related tasks the B1 level of the CEFR which are likely to be more common in the local professional workplace. The consequences of using the LCSCT and the decisions based on the test results will benefit students, teachers, the university, and the

## Chapter 7 Considerations for a context-sensitive tests

community in which they live. The construct to be assessed is shown in Table 7.5 below.

Construct to be assessed.			
Skill(s)	Competence(s)		Task Type
S&L	1	offer advice to clients within their job area on simple matters (e.g., "This model works faster, but is more expensive").	Record a voice message offering advice to a <i>client/patient</i> in response to recorded input provided.
S&L	2	request routine services related to own work area, such as a document or package to be sent (e.g., Can you send/type/call...)	Record a voice message requesting routine services in a familiar work-related matter based on written input provided.
S&L	3	ask questions of a fact-finding nature, for example, establishing what is wrong with a machine/process/situation, where something or somebody is, the frequency of an event, etc.	Live role play (pairs). Engage in an imaginary phone call where a client asks about an order that has not arrived.
S&L	8	ask for clarification when something is not clear (e.g. "Can you repeat that please?", "Is it .. or ...?", "So, the ...is/has/needs...")	Record the questions that would be made to the presenter of a topic or product based on the input material read.
S&L	9	express opinions in simple terms, for example, "I don't agree" "I think..." provided the question/issue has been put clearly and simply.	Live role play (pairs). Making suggestions, agreeing, disagreeing, and asking for clarification.
S&L	10	participate in job-related meetings and seminars (e.g., by making suggestions or asking questions: "Why don't we...," "Can we...," "Is it possible to...").	
S&L	12	greet a visitor and engage in a limited conversation for a short time, for example, enquiring about a visitor's journey, hotel, etc. (e.g., "How was your ...", "Did you like...?", "Is/Was your .... ?").	Live role play (pairs)—professional and visitor—greeting, engaging in a simple and short conversation. Deal with visitors' requests.
L&S	14	deal with predictable requests from a visitor, for example, "Can you arrange a taxi to the airport?", "Can I use a projector in the presentation?", "Where can I....?"	
S&L	13	provide instructions and directions, such as giving a client clear direction to the business/company. (e.g., "Take a taxi to. ...", "Go straight...", "Walk two blocks, then turn...").	Record a voice message giving instructions or directions based on the recorded input provided.
S&L	15	make phone calls requesting information or services (e.g., "Do you have...," "I would like...," "How much/many...?")	Record a voice message requesting information about products and/or services based on written input provided.



## Chapter 7 Considerations for a context-sensitive tests

L&S	5	receive phone calls (e.g., take messages/make notes: who called, what time, etc.)	Listen to a recording (a phone call, a voice message) and take messages to give to somebody else.
L	7	follow a simple presentation/demonstration and understand explanations concerning a product or topic within your expertise.	Listen to a recording of the presentation/demonstration of a product or topic within your area. Demonstrate understanding by choosing the appropriate options provided.
R	19	obtain basic relevant information (product specifications, professional or commercial leaflets, advertisements, internet pages, etc.)	Read a professional/commercial leaflet and obtain basic relevant information.
R	21	understand general written instructions (for example, installation, functioning or maintenance manuals)	Read instructions and demonstrate understanding.
R	29	understand, given sufficient time, most information of a factual nature that (s) he is likely to come across during their studies.	Read a discipline-related text to get factual information.
W	23	write short, simple indications to follow a healthcare treatment, may make extensive use of commonly used abbreviations, fixed expressions, or formulaic language.	Write short, simple healthcare indications based on input provided in Spanish.
W	24	write requests for goods, services, etc., on a range of routine matters, but MAY need to get these checked.	Write a request for goods and/or services based on input provided in Spanish.

Table 7.5 Design Statement: Competences and task types

The LCSLT needs to measure the intended construct, which is the language required in the workplace, as shown in Table 7.5 above. The test must provide tasks that enable evaluators to observe and evaluate test-takers' language performance. Test-takers' performance should allow evaluators to make relevant predictions of test-takers' performance in real-world situations. Test tasks should emulate the characteristics of real-world tasks as much as possible (L. Bachman & Palmer, 2010), so it is essential to consider the real-world output mode. In a multicultural environment where English is the means of communication, output in English is essential, as it may be necessary to discuss the content of the text with a colleague. In a monolingual environment where English is the means of obtaining information but not communication, the outcome of the reading comprehension activity, if shared orally, will not be in English. As discussed in 1.6.1 above, the reader will undoubtedly use their native language to

share or discuss the content of the reading passage with a colleague. Using the mother tongue to refer to the content of input in English was also mentioned by HScE582 above. Based on this, having the questions and/or the outcome of the reading task expressed in the test taker's native language is considered a better reflection of the real use of the language.

A more significant number of speaking and listening competences are less practical to assess directly than assessing listening or reading competences. The former requires several trained oral evaluators (pairs of evaluators would be preferable due to the high-stakes nature of the test) per pair of test-takers. For practical reasons, decisions must be made about how many competences can be assessed directly and compromises made to assess the other competences semi-directly or indirectly.

The oral interaction task type is preferred to assess speaking and listening competences. Whenever possible, role-play-type tasks will be preferred to assess oral competences. However, real-life language assessment has certain limitations for practical reasons (L. Bachman & Palmer, 2010). Where compromises are required, these will be sought to be the minimum, i.e., before deciding to assess a speaking and listening task using a written task, an attempt will be made to use a speaking or listening-only task. A computer-mediated task could assess speaking or listening-only competences, where participants record their responses to the oral prompts provided.

The nature of some of the speaking and listening competences in Table 7.5 above, C3, C9, C10, C12 and C14, make them more suitable for live pair assessment. The other speaking and listening competences, C1, C2, C8, C13 and C15, could be assessed indirectly for practical reasons. Their nature determines whether the input entries are made in written or recorded form; test-takers will need to listen to a recording to answer C1 and C13, as the nature of both implies interaction. On the other hand, C2 and C15 involve making a phone call. Although not the same, both competences can be partially fulfilled by pretending to leave a phone message by recording the message on a device. The input could be in writing as they are likely to read the information in a brochure, or it was a request from someone within the company. Finally, participants said they needed C8 when listening to a product or service conference presentation. These are likely to be supported by visuals. This situation will be best recreated with video footage. While it may be possible to have a computer-mediated task, the input could also be in printed form. However, C8 can be assessed together with C7, which is

Chapter 7 Considerations for a context-sensitive tests a listening competence only (i.e., following a presentation) and does not require an oral response from the participant. The same video recording can assess C7 and C8 if a computer-mediated task is used. Participants can show they understand the presentation by checking boxes with options or by making notes on the discipline-specific topic being presented or demonstrated (to fulfil C7). Then, they can record the questions they would ask the presenter (to fulfil C8).

There are two listening and speaking competences: C14 and C5. Competence 14 is directly assessed, together with C12, in a live role-play. Competence 5 would be assessed semi-directly. Test-takers will listen to a recording to take a message and then have to record the messages for someone else. Although both skills are required, an interlocutor is not part of the interaction mode. Even though this modification is based on practicality, it is considered that the test task interaction mode is likely to take place in the real world. Listening to a recorded message and taking notes are likely real-world tasks, but in the context of interest, conveying the message to a third party is more likely to take place in Spanish than in English.

Three reading competences, C19, C21 and C29, do not require a written or oral response to an interlocutor. For all of them, the test-taker will be required to read the documents corresponding to each. For C19, understanding basic relevant information, such as product specifications, could be demonstrated by writing answers in Spanish (C19). Using the test-takers' native language is a better emulation of real-world use of the language since that is the type of interaction that test-takers are likely to have with this type of text. Another important reason for using the native language of test-takers to demonstrate understanding is that when obtaining basic relevant information, using synonyms may not always be possible. Using the exact words in the questions and the input text may not show an understanding of the content, only the ability to identify words in a text. By using the correct words in Spanish to demonstrate understanding, test-takers may also be able to demonstrate knowledge of cognates and false cognates, which can lead to confusion and embarrassment if they are unaware.

Understanding of general written instructions could be demonstrated by matching or labelling pictures according to the instructions. These tasks may be related to how a person might interact with this type of text. Instructions or user manuals are often accompanied by illustrations intended to help the user of the product or machine understand how to use it.

## Chapter 7 Considerations for a context-sensitive tests

The last reading competence, C29, understanding discipline-related texts such as journal articles, could only be assessed using English. Although it is likely that, as with the other reading competences, the knowledge gained from reading an article or similar will be discussed with colleagues in Spanish (rather than English), it is believed that the use of English will require test-takers they show understanding of the content of the text with other English words, i.e., when it is paraphrased. Test-takers would be required to choose the correct answer from the options given.

Finally, there are two writing competences, one of which, C23, applies only to the health sciences group, and the other one, C24, is relevant to the three disciplinary groups. Fulfilling C23, participants could be asked to write short, simple indications of a health treatment from the input provided in Spanish. Providing the input in the test takers' native language better reflects real-life use of the language. Healthcare providers must use English to write the information generated in their native language. Using English to write the healthcare treatment would help the English-speaking patient understand the instructions to follow the treatment. The same is believed about C24. Indications for writing a request for goods or services are likely to be received in the test taker's native language, either from a supervisor, a colleague, or from personal notes. Although the products' names are likely in English, the rest of the information is likely in Spanish.

It is essential to include an estimate of the human and physical resources that are expected to be required for the institution's administrators to obtain them. If resources are insufficient, changes must be made before the test development process begins (L. Bachman & Palmer, 2010).

### 7.3.4 Blueprint

After the design statement is formulated, the blueprint is developed.

The blueprint below, Table 7.6, describes the general structure of the tests (L. Bachman & Palmer, 2010) for the three disciplinary groups. An example of the blueprint for health sciences is presented later in Table 7.7 below—design patterns, discussed in 7.3.5 below, present test task specifications.

The blueprint contains four columns: task number, competences, task type and specificity. The *Task number* helps identify and refer to the tasks but does not indicate

the order of appearance or importance. The *competence(s)* box contains the number of the competence(s) each task aims to assess. Task designers can refer to Table 7.5 above to verify the content of the competences. The information in the *task type* column describes the type of task that is considered appropriate and relevant for assessing each competence or combination of competences selected for assessment (see 6.3 above).

The *specificity* column tells the item writer the amount of specialized language required for each task. Domain analysis results (6.3 above) show that some competences are common in the three groups of participants; these are referred to as *Common3* competences. Tasks marked as *Common3* are designed to be used with the three groups of participants. Of the competences relevant to the three groups, only a few require discipline-specific content. These are each marked as *Unique for each*, which means that a specific task needs to be designed for each group. This label distinguishes them from those unique to a disciplinary group, referred to as *UniqueX*, where X refers to the disciplinary group to which it is relevant, i.e., *UniqueHSc*. This label indicates that the test task will be designed for only one disciplinary group. Competences common to two groups are referred to as *CommonXY*, where X and Y refer to the disciplinary groups to which the competence is relevant, i.e., *CommonHScSSc*. This label indicates that a This indicates test task must be designed for these two disciplinary groups. This classification indicates the level of specificity that tasks require and the disciplinary groups with which the task is used.

The spoken interaction sections of the test are designed to take place in live role-playing tasks. The other tests (reading, listening, spoken production and writing) are designed to be computer-mediated, allowing test-takers to listen to the audio and record their responses when required.

Table 7.6 below shows the general overview of the LCLSLT blueprint, while Table 7.7 below shows an example of the blueprint for a disciplinary group (health sciences). The blueprint for the Health Sciences group contains information specific to this group of participants. A similar blueprint for each disciplinary group would be part of the complete LCLSLT document. This blueprint contains 13 tasks, while the general blueprint contains 15. The difference in the number of competences is because the general blueprint contains the tasks for the three groups of participants. In contrast, the

Chapter 7 Considerations for a context-sensitive tests

health sciences blueprint does not contain information on tasks irrelevant to this group, which helps test task designers focus on the content of the test for that disciplinary group.

LCSLT general blueprint			
Task No.	Comp.	Task Type	Specificity
1.	3, 8	Live role play (pairs). Engage in an imaginary phone call where a client asks about an order that has not arrived.	<i>Common3</i>
2.	9, 10,	Live role play (pairs). Making suggestions, agreeing, disagreeing, and asking for clarification.	
3.	12, 14	Live role play (pairs). Greet a visitor and engage in a simple and short conversation. Deal with visitors' requests.	
4.	5	Listen to a recording (a phone call, a voice message) and take messages, write, or record the message for somebody else.	<i>CommonSSc &amp; HSc</i>
5.	2	Record a voice message requesting routine services in a familiar work-related matter.	<i>CommonSSc &amp; PNSc</i>
6.	1	Record a voice message in response to questions posed by a <b>client</b> offering advice.	<i>CommonSSc &amp; PNSc</i>
		Record a voice message in response to questions posed by a <b>patient</b> offering advice.	<i>UniqueHSc</i>
7.	13	Record a voice message giving instructions or directions based on written input provided.	<i>UniquePNSc</i>
8.	15	Record a voice message requesting information about products and/or services based on written input provided.	<i>UniqueHSc</i>
9.	7	Listen to a recording of the presentation/demonstration of a product or topic within their area. Demonstrate understanding by selecting the correct response.	<i>Unique for each</i>
10.	19	Read a professional/commercial leaflet and obtain basic relevant information.	<i>Common3</i>
11.	21	Read instructions and demonstrate general understanding by selecting the correct response.	<i>Common3</i>
12.	29	Read a discipline-related text to obtain factual information.	<i>Unique for each</i>
13.	23	Write short, simple healthcare indications based on the written input provided.	<i>UniqueHSc</i>
14.	24	Write requests for goods and services based on written input provided.	<i>Common3</i>

Table 7.6 LCSLT general blueprint

The specifications for the test task design are presented in the LCSLT design patterns discussed below.

### 7.3.5 From the LCSLT Design patterns through Operational use

The organizing principle of design patterns provides a framework for developing context-sensitive test tasks for an LCSLT.

The first element in Table 7.8 below, the Focus box, indicates the type of language to be assessed. The following example aims to measure competence 21, understanding written instructions. This competence was considered relevant for the three disciplinary groups and classified as *Common3* (Table 7.6 above). A task will be designed and applied to the three disciplinary groups for practical reasons. A corpus on the language of instructions (installation, function or maintenance manuals) can be used to identify the language more frequently used in those types of documents that may be most relevant to the three disciplinary groups. The Date and the Version help track their development.

The title “understanding manuals” not only identifies the design pattern but provides a quick reference to the aim of the competence. The CEFR level and reference indicate the level and activity type. The *competence box* shows the competence number in brackets (see Table 7.8) and the construct, i.e., the work-related competence statement to be assessed.

Competence 21 relates to reading, so the assessment of C21 requires input in the form of written text or images. Test-takers are expected to demonstrate an understanding of the content of written text and/or images presented as specified in the input/output mode. The interaction pattern shows that this task is answered individually.

The Version box indicates that this task is classified as *Common3*. This box also indicates the skills involved, namely reading (R), the year the design pattern was developed (21) and its version (1).

## Chapter 7 Considerations for a context-sensitive tests

LCSLT Health Sciences blueprint				
No.	Skill(s)	Competences	Task Type	Specificity
1.	S&L	3	Live role play (pairs or trios): client and seller. Engage in an imaginary phone call where a client asks about an order.	<i>Common 3</i>
2.	S&L	9, 10,	Live role play (pairs or trios). Co-workers. Making suggestions, agreeing, disagreeing, and asking for clarification.	
3.	S&L	12, 14	Live role play (pairs or trios)—professional and visitor. Greet a visitor and engage in a simple and short conversation. Deal with visitors' requests.	
4.	S&L	8	Record the questions that would be made to the presenter of a topic or product based on the input material read.	
5.	L&S	5	Listen to a recording (a phone call, a voice message) and take messages, recording the message for somebody else.	<i>Common SSc &amp; HSc</i>
6.	S	1	Record a voice message in response to questions posed by a <b>patient</b> offering advice.	<i>UniqueHSc</i>
7.	S	15	Record a voice message requesting information about products and/or services based on written input provided.	<i>UniqueHSc</i>
8.	L	7	Listen to a recording of the presentation/demonstration of a product or topic within their area.	<i>Unique for each</i>
9.	R	19	Read a professional/commercial leaflet and obtain basic relevant information.	<i>Common 3</i>
10.	R	21	Read instructions and demonstrate general understanding by selecting the correct response.	<i>Common 3</i>
11.	R	29	Read a discipline-related text to obtain factual information.	<i>Unique for each</i>
12.	W	23	Write short, simple healthcare indications based on the written input provided.	<i>UniqueHSc</i>
13.	W	24	Write requests for goods and services based on written input provided.	<i>Common 3</i>

Table 7.7 LCSLT Blueprint for the Health Sciences

The information in the target audience box includes the names of the relevant disciplinary groups and the areas the design pattern intends to assess. This information reminds task designers which expert insiders to consult for task validation (Table 7.3 above).



## Chapter 7 Considerations for a context-sensitive tests

An overview of the assessment situation can be found in the summary box. For C21, test takers will be presented with written input material related to the competence to be assessed. The text would include general written instructions for equipment, products or processes. Depending on the selected content, images may be included. Test takers are asked to demonstrate an understanding of the task content by selecting, checking, organizing, matching or numbering the text and/or images.

The rationale section justifies the type of task used by indicating how it relates to the competence being assessed and the real-world work-related language use needs of test-takers. This argument supports the relevance aspect within the beneficial consequences (see 7.4.2) of the design pattern for the target audience.

The knowledge, abilities or skills (KSAs) this task is designed to elicit from test-takers is that they understand the general language related to instructions. Focal KSAs state the need to use discipline-related content. Discipline-related tasks may require additional KSAs. This section is marked NA (not applicable) when not required.

The characteristic features show how the task aims at eliciting the expected performance. For C21, test-takers are asked to read the input text provided and show understanding of the content by selecting, labelling, matching, numbering, or organizing the text and/or images provided. The reading skill to which C21 refers is measured directly as the test task asks the test-taker to read the input material.

The Potential work products state the four options that can be used so test-takers show an understanding of the written text related to general instructions. These give task designers some flexibility in the type of task that can be developed. Test-takers might be asked to match pictures and text or to label parts of a picture (fill in the blanks). Other possibilities include numbering a series of pictures according to the order given in a written text or organising a series of steps in a procedure. Since there are four options for the potential work products, there are also four options for the expected correct answers. These potential observations provide evidence for the KSAs.

The potential observations are based on the potential rubrics. For C21, the potential rubrics box indicates that these can be found in the Answer Key.

The References box for this design pattern suggests using a language corpus to identify the language related to the competence of interest.

LCSLT task design specifications					
Focus	Common3	Date	August 2021	Version	Com3R21V1
<b>Title *</b>	Understanding manuals				
<b>CEFR reference</b>	B1 reading instructions				
<b>Competence</b>	(21) Understand general written instructions (for example, installation, functioning or maintenance manuals)				
<b>Input/output</b>		<b>Input</b>	<b>Output</b>		
	<b>Mode</b>	written	written		
	<b>Language</b>	English	English		
<b>Interaction pattern</b>	Individual				
<b>Target audience</b>	Health sciences: medics, nutritionists, nurses, dentists Physical and natural sciences: chemists, forestry engineers Social sciences: lawyers, social workers, psychologists, human communication therapists				
<b>Summary *</b>	In this design pattern, test takers are presented with general written instructions on equipment, products and/or procedures. Images may be included. Test takers are asked to carry out an action based on the reading (and/or images if included).				
<b>Rationale *</b>	In real-life work-related situations, professionals may encounter circumstances in which they need to understand and follow written instructions for the installation and or operation of equipment or the steps to be followed concerning understanding the instructions of a product or procedure coming from abroad.				
<b>Focal KSAs *</b>	The understanding of the general language related to instructions.				
<b>Additional KSAs *</b>	NA				
<b>Characteristic features *</b>	Written instructions for: Instructions (installation, functioning or maintenance manuals) are presented to test takers. Test takers read the input material. Test takers must demonstrate understanding by selecting, labelling, organizing, matching, numbering, etc., the input material.				
<b>Type of measurement</b>	Direct measurement.				
<b>Potential work products</b>	Responses reveal their ability to understand and/or follow general instructions (installation, functioning or maintenance manuals). 1. Matching instructions to pictures (with an extra element in one of the columns) or 2. Sequencing steps in a procedure (using numbers), or 3. Filling in blanks with labels given based on reading input or 4. Selecting the pictures the texts refer to (checking boxes).				
<b>Potential observations *</b>	1. Correct matching of instructions to pictures 2. Correct sequence of steps in a procedure 3. Correct labels in blanks 4. Correct selection of pictures				
<b>Potential rubrics *</b>	Answer key (link to Answer key).				
<b>References *</b>	Corpus of the language of instructions as a reference to select the texts and images used.				

Table 7.8 *Common3* design pattern to assess reading; understanding instructions

An example of a design pattern for spoken interaction competence is shown in Table 7.9 below. This design pattern is also a *Common3* task, so the task can be used with

the three disciplinary groups. Competence 3, “Can ask questions of a fact-finding nature,” and Competence 8, “Can ask for clarification when something is unclear, ” are assessed together using a role-play format. The role-play format aims to resemble, to the extent possible, the type of oral exchange that may occur in the workplace when asking about goods or services ordered. The design of this task type is based on comments made by some participants, such as SScE554, who expressed he needed to get in touch with a company in a foreign country and may be required to explain or clarify the situation with the equipment. HScE582 mentions the need to contact a company that has not met the programmed delivery date. HScE582HScE582 adds that buying supplies would be another situation when he may need to talk in English to a company representative. A role-playing task type aims at engaging test-takers in an interactive exchange with a peer, in which one pretends to be the client and the other the salesperson in a company. An example of a test task can be seen in Appendix XII below. This task aims to elicit the language pertaining to this specific situation, allowing oral evaluators to observe test takers’ performance of what they can do with the language. Performance observation will allow oral evaluators to predict test takers’ ability to engage in situations where they need to check on goods or services ordered in non-test situations.

Each oral interaction test task refers to the competence it aims to measure, making it explicit to oral evaluators what the purpose of the task is. Test takers are given cards with information to use in a role-play where one is the client and the other is the salesperson in a company. To complete the task, participants must ask questions and ask for clarification, as the cards intentionally contain incorrect information that they must correct.

Once developed, test tasks are shared with relevant stakeholders, mainly with domain insiders, to receive feedback on how the test task resembles tasks encountered in the real work environment. Their comments will be reviewed and analyzed to determine feasibility. Changes will be made provided they remain within the task's limits, difficulty level, linguistic demands on participants, and practicality set by the testing institution, in this case, the university authority.

## Chapter 7 Considerations for a context-sensitive tests

Attributes	Definition
Focus	Common3   Date   Sept. 2021   Version   Common3.SI.21.V1
Title *	Making questions: fact-finding and clarification
CEFR reference	B1 spoken interaction
Competence(s)	(3) CAN ask questions of a fact-finding nature. (8) CAN ask for clarification when something is not clear
Input/output mode	Written (keywords & visuals), spoken/spoken
Interaction pattern	Pairs or trios
Target audience	<b>Health sciences:</b> medics, nutritionists, nurses, dentists <b>Social sciences:</b> Administrators, accountants, social workers, psychologists, human communications therapists, lawyers <b>Physical and Natural Sciences:</b> Chemistry, Forestry
Summary *	In this task, test takers are asked to imagine a job-related situation in which they must engage in oral interaction to clarify information about an order placed with a company. One of the test takers pretends to be the client, and the other is the salesperson in a company. They both need to ask questions of a fact-finding nature and ask for clarification.
Rationale *	In real-life work-related situations, test takers may need to call a company to clarify an order that has not arrived or has a problem.
Focal KSAs *	Making questions of a fact-finding nature. Responding to questions of a fact-finding nature. Ask for clarification/repetition. Understand when clarification/repetition is being requested and respond accordingly.
Additional KSAs *	1. Greeting and leave-taking.
Characteristic features *	2. Sets of cards with <ul style="list-style-type: none"> <li>• keywords to make questions</li> <li>• Information to respond to questions</li> </ul> Test takers engage in oral interaction. Student A asks questions using the keywords in the role card, and Student B responds to them using the information in the role card.  When they finish, roles are exchanged (Student B makes the questions and Student A answers) using a second (similar) set of cards.
Type of measurement	Direct measurement.
Potential work products	A conversation between two participants.
Potential observations *	1. Test takers make grammatically appropriate questions using the keywords provided. 2. Test takers understand questions made 3. Test takers produce grammatically appropriate responses using the information provided.
Potential rubrics *	Link to the rubrics designed for oral interaction task assessment.
References *	Comments made by SScE554 and HScE582 are examples of how professionals in the disciplinary area use this competence.

Table 7.9 *Common3* design pattern to assess spoken interaction, making questions of fact-finding nature.

The test tasks that have been processed are then compiled, and preliminary tests are carried out. The results are analyzed, and relevant changes are made before developing the scoring criteria. Warrants for the claims made in the Design Statement will be revised to provide backing.

The final stage before the Operational use stage is to set the standard. The standard is set by specifying how the test results will be interpreted. Once the test has been applied, the results will be revised to assess whether changes need to be made. New versions of the test (new test tasks) must go through the same stages: verification by stakeholders (especially domain insiders), pre-testing and feedback, revising of the scoring criteria and warrants to provide backing and finally, specification of the meaning of test results of the new version of the tasks.

## **7.4 Justifying the use of an LCSLT**

The validation of the local context-sensitive language test is discussed below.

### **7.4.1 Introduction**

After deciding to develop a test, the starting point of an LCSLT is to analyse the language used in the domain of interest, as shown in Figure 7.2 above. A difference between the LCSLT and the AUA (above) is that in the former, domain analysis is explicitly included between the decision to design a test and the development of the design statement. The domain analysis must also include test takers, expert insiders, and other relevant stakeholders. The data collected informs the Design Statement and the development of test tasks; hence, it can be argued that there is a correspondence between the TLU situation and the test applied. It is, therefore, conceivable that the consequences for stakeholders are beneficial. Beneficial consequences are an overarching argument in the LCSLT. This argument is another difference between the LCSLT and the AUA. The consequences will be beneficial mainly because of the correspondence between the test tasks and the real-life target language use domain. The beneficial consequences are based on the claims made on the characteristics, quality and strength of the three pillar elements in the LCSLT: interpretations, decisions and assessment records. These are shown in Figure 7.3 below. These

claims are supported by backings, which are statements describing their characteristics.

On the other hand, the LCSLT differs from the LAPP (Figure 3.4 above) in that the latter starts with the needs analysis, while the LCSLT, as AUA, starts with identifying a need for a test. The test-takers and the TLU domain of interest are then identified. Identifying these two elements is relevant to the following stage: the possibility of using an existing test. Using an existing test cannot be pondered if the test taker population and the TLU domain have not been clearly identified.

#### 7.4.2 The Local Context-Sensitive Certification Test (LCST)

The Design Statement would be:

Overarching argument: Consequences are beneficial. The consequences of using the LCST and the decisions that will be made based on test results will benefit students, teachers, the university and the community in which they live.

Claim 1. Interpretations. Results interpretation of the LCSLT allows the university English program to determine test takers' ability to use the language in the workplace because they are:

Backing 1A. Useful. Test results are useful in predicting the future ability of test takers to perform the most commonly required language use activities in the workplace.

Backing 1B. Relevant. Test takers find test content relevant for their future language use needs because it was developed considering the language use activities reported to be the most frequently needed in the workplace by professionals living in the community where test takers are likely to work (see 6.3 above).

Backing 1C. Positive washback. The test brings a positive washback effect to the language classes offered to students within the university, as teachers prepare students for a test related to their future activities in the workplace. Test content and task types perceived as relevant are likely to have a positive washback effect on learners (see examples of test takers' comments 1SSc572 and SSc47F above).

Backing 1D. Face validity. A high degree of face validity is attributed to the test since the content and tasks are related to the students' perceived future language use; therefore, they will consider it authentic. Test takers are expected to have a positive attitude towards it. The LCSLT test incorporated end-user perceptions of the test content and format (see the summary of test takers' comments regarding test content in 5.4 above).

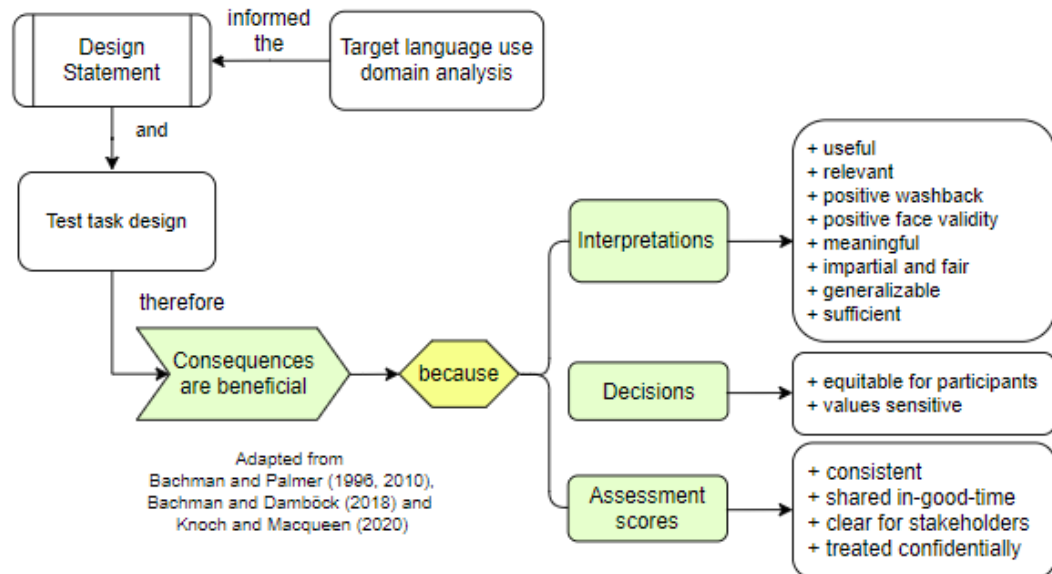


Figure 7.3 Justification of a local context-sensitive school-leaving English language test

Backing 1E. Meaningful. The interpretations of language ability are meaningful concerning the language competences more commonly used in the local workplace. The abilities assessed are based on the competences reported by professionals and test-takers relevant to the workplace (see 7.3.3 above).

Backing 1F. Impartial and fair. Test takers' language ability interpretations are impartial and fair to all test takers. Test content does not include topics considered offensive to test takers. All test takers are treated equally throughout the certification test administration process. All test takers have equal access to information about the competences to be assessed, the test format and the procedure. All test takers have equal opportunity to demonstrate their job-related language knowledge and skills. All test takers have the same opportunity to

## Chapter 7 Considerations for a context-sensitive tests

register for the test administration session of their choice. All tests and test taker performances are marked equally regardless of their discipline.

Backing 1G. Generalizable. Interpretations about test takers' language ability are generalizable to the work environment in which test takers are expected to use the language. There are correspondences between the most commonly used language competences in the workplace and test content and test tasks. Task characteristics (input, format, type of interaction, expected response, and expected response mode) have been defined to emulate, as closely as possible, those taking place in the real world.

Backing 1H. Sufficient. The language performance of test-takers in the LCSLT is sufficient to make the intended decisions. The language, skills, task types and interaction modes assessed by the LCSLT are representative of the language competences most frequently used in the local workplace.

Claim 2. Decisions. The decisions made reflect the regulations and values of the current Institutional Development Plan of the university.

Backing 2A. Equitable for participants. The decisions made are based solely on the interpretation of the results and do not relate to the disciplinary group of test takers.

Backing 2B. Values sensitive. By using the LCSLT to assess students' language ability, the university will provide professionals ready to interact with the globalized world and contribute to the development of their community.

Claim 3. Assessment scores. Assessment scores support the decision-making process due to their consistency. They also support stakeholders through timely sharing, clarity and confidentiality.

Backing 3A. Consistent. Assessment scores are consistent across assessment tasks, administrative procedures and grading. The criteria and grading process for assessing the objectively and subjectively graded tasks correspond closely to the ones used by university English teachers at the corresponding level (B1 of the CEFR). Test Administrators follow the same procedures for each session. Oral evaluators follow the script and time frames for each activity. Performance



Chapter 7 Considerations for a context-sensitive tests evaluators (oral and written) do not favour any test taker. Two oral evaluators evaluate pairs of students. Written performance tasks are assessed double-blind.

Backing 3B. Shared in-good-time. Assessment scores will be shared within the set time frame. The time is set, considering that at least two assessors must mark written performance tasks.

Backing 3C. Clear for stakeholders. Assessment scores are delivered using a number and a description of what the test taker can do with the language at the assessed level. These are reader-friendly for test takers and stakeholders, so their meaning is clear.

Backing 3D. Treated confidentially. Assessment records are not shared publicly. These are delivered to the stakeholders either personally or by email.

## **7.5 Main Research Objective (MRO)**

The study aimed to identify the specifications for a high-stakes university exit test sensitive to the workplace's local language use needs. The domain analysis considered test-takers as well and domain insiders as sources to inform the contents of a test that is likely to better meet the demands of the local workplace than using a general purposes test. The data analysis showed the local workplace's language needs, which were used to inform the design of test tasks. The validation of the use of test task results as an exit requirement for the workplace was found in a proposal for a Local Context-Sensitive Language Test (LCSLT), which was developed based on two argument-based approaches to test validation: Bachman and Palmer's AUA (2010) and Knoch and Macqueen's LAPP (2020) frameworks.

It is considered that the argument-based approach provides a flexible and adaptable framework for assessment validation purposes in different scenarios.

## 8 Conclusions

### 8.1 Introduction

English has become an essential tool for speakers of other languages professionals as they can communicate and exchange information, opening doors to a variety of opportunities (BCEI, 2015; Borjian, 2015; Education First, 2016, p. 10; Ferguson et al., 2011; Mexicanos Primero, 2015). University graduates will be ready to be part of the global information exchange community if they can use the language that enables them to do so. A university exit language test that considers the workplace's language requirements can show how well graduates are prepared to join the international global community.

Although the importance of the relationship between test tasks and the target language use situation has been outlined (L. F. Bachman & Palmer, 1996), many international tests are used for purposes other than those for which they were designed.

While tests developed at a local university may not have the rigour and expertise of international testing houses, their advantage is that they may be more tailored to the community's needs. An in-house tailored test will be useful to the community provided they are designed following the best testing practices and to the best of the abilities of those involved in their design. Individuals looking to study abroad or pursue an international career will find that they may have already completed part of the journey in preparation for an international exam.

Professional comments show how English can open doors but also restrict career progression. Limited language skills can prevent someone from advancing and climbing in their profession. Opportunities such as specialized training or better job positions may be tied to English proficiency. The following comment made by a professional from the social sciences (SScE583) exemplified this:

*“We live in a world that is becoming increasingly globalized, and it (English) is no longer a plus or an extra, because, simply stated, if you want access to a better job, he who has English has more possibilities, giving them more confidence. The one who has English has a better chance to do specialized studies. I have relatives that are good at English, and they are fluid. My*

*niece is currently in the United Kingdom because the language did not limit her.”*

The speaker discussed how English proficiency in the workplace can open doors for potential employees for job enhancement or academic opportunities. Her concluding comment shows that English could also limit professionals in their career advancement. The phrase “*because she was not restricted by language*” implies that English restricted other people, although it is unclear whether she refers to herself or others she knows.

## 8.2 Summary of findings

This study uncovered aspects directly related to the research questions and revealed other interesting data.

Research question 1 (RQ1) asks about the face validity of the current university-exit test from the test-takers' perspective. The results revealed a variety of opinions, beliefs and perceptions about the university exit test's language, content and format. Some participants believe that language cannot be divided into domains. These perceptions show that test-takers are unfamiliar with the differences between domains and the subskills that each one entails. This belief is not surprising given that they did not have English for specific purposes classes. All English courses offered at the university are for general purposes. Another belief revealed by the study is that a domain arises naturally after being proficient at one.

The results also show that many test takers are unaware of the English proficiency required in their profession. While some test-takers, such as law students, said English was not required in their profession, other law students provided examples of current job-related English language needs in their field. Likewise, some test-takers believe that English is not required in a professional context in Durango. However, they believe it is necessary if they wish to live in another state within the country or abroad. In part, this may be because many test-takers seem to have English in mind concerning spoken interaction. Some responses show participants think there are no foreigners in their profession to talk to; therefore, they think they do not need English for work purposes. Some of them changed their opinion when a colleague mentioned that it was

relevant for them to understand journal articles that shared the latest information in their fields.

The data collected to answer research question 2 (RQ2) showed that students and professionals agree on the relevance of many work-related language competences but disagree on others. The total mean was the cutting point for selecting the competences perceived as more relevant for students and professionals. These would be the constructs that would be assessed in the university exit test that was proposed.

The data show that reading and spoken interaction are the skills that are more frequently required or expected to be needed in the workplace. Listening skill is also needed, but not as frequently or for various language activities such as reading and spoken interaction. Writing is the least needed skill. These results are an example of the irregular skills profile reported to characterize language needs in the local work environment in different settings, as discussed in 1.6.1 above (Aliakbari and Boghayeri, 2014; Cambridge English, 2016b; Didiot-Cook *et al.*, 2000; Langues et employabilité and Erasmus+, 2015; Shamsudin *et al.*, 2013).

Research question 3 (RQ3) aimed to identify employees' typical domain-related tasks. These were provided not only by expert insiders while responding to Q3 but also by test-takers who had already joined the workforce. These helped to understand how competences are used in the workplace by providing specific examples of language use situations. The data also indicated that some language use activities might be similar for some disciplines within the local work environment. Genre and lexis can vary from one context to another, while professional activities within the same discipline can have different language needs. The language needs of a professional may be pretty frequent, while another professional of the same discipline may need to use English occasionally.

The suitability of an argument-based approach to test validation for the design of a specific needs test was the purpose of Research Question 4 (RQ4). A combination of the frameworks presented by Bachman and Damböck (2018), Bachman and Palmer (2010), and Knoch and Macqueen ((2020) were used in the design of the Local Context-Sensitive Language Test (LCSLT) proposal. Initially presented by Wei *et al.* (2008b), design patterns for language assessment were also used as the basis for the LCSLT framework design patterns.

### 8.3 Contribution to the research area

This section discusses the possible research contributions this study can make, particularly in testing English for professional purposes, face validity (H. D. Brown, 2004) and the relevance of the test context (L. F. Bachman & Palmer, 1996). Studies in the field of testing validate past and current standardized testing practices. Still, empirical research that considers contextualized language needs is lacking, an area where the present research could prove helpful.

This case study can contribute to discussing the delicate nature of language test validity and instrument design, as it argues that contextual needs must be considered. While the literature shows that English language use continues to spread globally (Ferguson et al., 2001; Education First, 2016a, p. 10), and language needs vary according to the context of use (Cambridge English, 2016b; Education First, 2018) the practice of language testing does not fully demonstrate the latter (Fang, 2017).

The LCSLT proposal may be relevant to institutions using language tests as an exit requirement but also interested in preparing graduates to join the international workforce. Using test tasks that resemble, to the extent possible, the language required in the target language use domain may be an indicator of test takers' ability to use the language in non-testing workplace situations. Test results that consider the language required in the workplace. The LCSLT framework is transferable to any context where the test developer can clearly identify the target language use situation and the specific types of language use activities. These are identified through a domain analysis that considers the relevant informants. Test tasks are designed to try to emulate real-world language use activities. Test results that are a good predictor of language use in non-testing situations consider real-world language tasks' demands.

No literature was found related to exploring this phenomenon in Spanish-speaking countries. Based on the localized language needs of future test-takers and potential future employees, this proposed framework for designing a context-sensitive English language exit test may aid in designing language tests that prepare university graduates to use the language in the workplace, meeting the demands of employers (Garcia-Ponce, 2020).

A context-sensitive language test may have high face validity among laypersons, mainly test takers. High face validity may have an impact on test takers' motivation. Higher test-taking motivation may contribute to other types of validity because test

takers may have better attitudes toward the test and may be willing to perform at their best (Phillips et al., 2020).

Although the scope of this case study is limited, it can be replicated in other contexts and languages. Findings from this and other studies may contribute to the research gap on language assessment relevant to the local language needs of professionals in their workplace.

Publications often concentrate on reporting statistical analysis of their investigations but do not provide specific examples of language use activities carried out by professionals in the disciplines. Neither many examples are provided on assessing the language competences relevant to the workplace.

Most studies on languages for specific purposes focus on the academic domain; few focus on the workplace. Those focusing on the workplace aim at a company's specific discipline language needs. This study adopted a general workplace purpose approach. The LCSLT framework is transferable to the needs of a large company with different departments where language needs may be shared in some areas and differ in others. It can also be adapted to target the needs of a specific discipline.

## 8.4 Implications

Most changes in curricula start with course development. A change in curricula starting with a test will likely bring many implications.

The use of a new test will likely generate a washback effect. Students cannot pass a test they have not prepared for or are familiar with. Teaching should be modified. This change implies teachers are in favour of such a change and willing to change their English for General Purposes (EGP) teaching programme to English for general working purposes (EGWP). While teachers may feel a burden to change their familiar EGP to new EGWP content, this change is expected to bring beneficial consequences for teachers when they see a change in students' motivation to attend classes and engage in class activities. Students' increase in motivation is expected based on their responses concerning their interest in language for academic purposes or the workplace.

Another implication is that teachers should be familiar with the test format and task types so students know what to expect when they take the test. At least a bank of task

types for the whole test should be available online for teachers to use. Having sample material online is also relevant for students who do not enrol in the UEP classes and decide to prepare independently.

Due to the characteristics of the context in which the LCSLT is developed, there will be three tests, one per disciplinary area. The three of them, however, will have a common core. The competences that are equally relevant for the three groups will feed task design that will be the same in the three tests. Therefore, each disciplinary test will have a common core component; the rest will be disciplinary specific for each group. This difference implies that the discipline-specific tasks will have their own marking criteria. An essential aspect to consider is the number of items and their difficulty level. These different marking criteria also imply training raters. Due to the high-stakes nature of the test, two raters are required, both in marking the speaking component and for the writing tasks. Efforts should be made to have the same number of items in the three tests and that the difficulty level does not vary significantly among tests. Test takers of a disciplinary group could feel at a disadvantage if they perceive their test is more linguistically demanding than the other two groups' tests.

## 8.5 Constraints and limitations

Every research has some inherent and incidental constraints and limitations. One of the limitations of the present study, shared in social research, was the data collection.

When the research design was planned, English teachers and employers were part of the informant group. Their responses to Q2 would have enriched the knowledge gained about future test takers and the workplace. However, due to administrative circumstances, it was not possible to interview teachers. It was also not possible to interview employers. Only two of them agreed to be interviewed. Their answers were quite similar. They checked all the boxes as 'very frequently'. When I asked if I could have some examples of the specific type of activities they carried out, they responded, "Just as it is described in the statements, that's what they have to do". Their justification for ticking all the boxes as 'very frequently' was that they thought of all employees, so some would use some of the competences, while others would have different language needs. According to one of the employers, all competencies are generally relevant to the company. It is considered that these managers did not really understand their employees' language needs.

Other employers said they were swamped and did not have time for the interview, but permission was granted to interview some of the company's employees. Others said they did not need English in their job, so there was no reason for the interview. It is believed that some of these employers thought their English level would be assessed (possibly indirectly), so they did not want to embarrass themselves.

Another limitation that can significantly impact research is sample size and selection. Ideally, an even number of questionnaires or interviews from the different disciplines offered at UJED would have been used. However, the disciplines within the university have different numbers of participants. Some disciplines, such as law or medicine, had two or three groups per semester. Other disciplines, like chemistry, had only a small group of students. This difference in group sizes affected the number of questionnaires used. Although taking the university exit test is mandatory for the degree, some schools have not complied with this requirement. As a result, very few students from some disciplines, such as chemistry or forestry, took the exit test during the four test sessions. Therefore, the number of students from some disciplines who answered Q1 was very limited.

Access to professionals in the workplace was another limitation. Many locations were visited during the data collection, requesting access to expert insiders. There was not always a positive answer. Access to facilities or personnel was not always allowed. Also, at the individual level, some people were unwilling to participate. It was always explained that the purpose of the questionnaire or interview was to identify the language needs of their professional position and not their level of English. Despite this initial comment, their unwillingness to participate sometimes revealed that they felt their answers would reflect their limited proficiency.

Another limitation concerns the quality of information reported by professionals. More accurate data would have been collected if it had been possible to observe professionals using the language. Shadowing professionals at the workplace would provide richer and more precise data on language use. Alternatives to shadowing professionals may be to ask them to use their mobile devices to record themselves when engaged in spoken interaction. They can also provide samples of the non-sensitive documents or written exchanges they use. Additionally, they can keep journals, keeping records of their use of English, challenging situations, and how they solved them. Focus groups with specific informants (Garcia-Ponce, 2020), such as healthcare-related professionals, can be conducted to identify or validate findings.



Other constraints and limitations that emerged were related to the type of instruments used. Questionnaires and semi-structured interviews have their own inherent limitations. Questionnaires do not allow clarification of ideas. On the other hand, semi-structured interviews offer a limited range of negotiation of meaning.

Professionals and the interviewer may have had a 'power dynamic'. Some participants may have felt intimidated by the 'language expert' who would judge them based on their self-perceived limited knowledge of English. Whilst this was never intentional, and it was stated that the focus was on the activities carried out and not on assessing their language proficiency, it cannot be assured that this dynamic was not present among all participants.

This framework promotes the use of international language descriptors to identify the language use activities in the context of interest, as others, like Huhta et al., use the CEF Professional Profiles (CEFPP) in a "second generation needs analysis" (2013, p. 14), which is task-based. Zela (2018) comments explicitly on using the CEFPP to identify the language needs of Albanian business companies. Liu (2014) used the CEFR CAN Dos to design a reading test, and the Occupational Language Analysis (OLA) is suggested to identify the language needs "of an occupation by relating an ESP (Essential Skills Profiles) to the Canadian Language Benchmarks" (ATESL, nd. p. S1-9). The ESP and OLAs provide detailed information on the language required to complete tasks in many occupations. On the other hand, needs analysis can be based on interviews (Changpueng & Pattanapichet, 2015) (Garcia-Ponce, 2020) (Chalak, 2019), focus groups (Garcia-Ponce, 2020), questionnaires (Chalak, 2019) (Changpueng & Pattanapichet, 2015), and observations (Chalak, 2019) to identify the language needs of the context of interest. The use of international language descriptors may be seen as restrictive or unrelated. It may be considered that richer data, more context-related, can be obtained through other approaches to needs analysis, as mentioned above. However, as Harsch (2018) discussed, several studies have used the CEFR to empirically research how the CEFR can support university entry requirements in countries where this Framework's levels are used as reference. The use of language descriptors like the CEFR (COE, 2001c) or the CLB (CCLB, 2012) to identify language use activities in specific contexts may provide a starting point for identifying language needs. Moreover, they provide a common ground to establish comparison among language learning, teaching and assessment materials. It is essential to acknowledge, however, that any set of language descriptors has been developed based on specific contexts, language users and language use purposes,

and these should be pondered to determine the extent to which they can be used. More frequently than not, these should be adapted to the particular contexts of interest.

Regardless of how well a test task may capture the essence of the real-world task or how authentic the task may be, a test is always a test. A test taker's performance in a test replicating a real-world task will never accurately predict how the test-taker will perform in a non-testing situation (McNamara, 2000). Therefore, test results will always be estimates predictions. They will never be true, accurate representations of real-world language ability.

## **8.6 Future directions**

The ever-increasing use of technology and social media for everyday information and communication requires language skills related to these means of communication. The Companion Volume for the CEFR (COE, 2018) has addressed the issue of online communication for general purposes. Research on these for specific purposes is required.

Specific examples of language use in the workplace could be enriched. Expert insiders can also validate them. Before designing test tasks, expert insiders of each disciplinary area can be asked to mark the extent to which the examples of language use activities reflect their workplace language needs. This validation process would better reflect the professional language use needs.

Research on test-taking motivation may have relevant implications on test validity because of the way motivation may affect test performance (Knekta, 2017). Research on test-taking motivation may also aid in better understanding the test-taker population, thus informing test developers on relevant issues to consider.

## Bibliography

- Acedo Dominguez, G., & Edwards Rokowski, P. (2002). Bridging the gap between English for Academic and Occupational Purposes. *English for Specific Purposes World*, 1(2).
- Alderson, J. C. (2000). *Assessing Reading*. Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511732935>
- Alderson, J. C. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing - LANG TEST*, 22(3), 301–320. <https://doi.org/10.1191/0265532205lt310oa>
- Alderson, J. C. (2009). Air safety, language assessment policy, and policy implementation: The case of Aviation English. *Annual Review of Applied Linguistics*, 29, 168–187. <https://doi.org/10.1017/S0267190509090138>
- Alderson, J. C., & Clapham, C. (1992). Applied Linguistics and Language Testing: A Case Study of the ELTS Test. *Applied Linguistics*, 13(2), 149–167.  
<https://doi.org/10.1093/applin/13.2.149>
- Al-Hoorie, A. H. (2017). Sixty Years of Language Motivation Research: Looking Back and Looking Forward. *SAGE Open*, 7(1), 2158244017701976.  
<https://doi.org/10.1177/2158244017701976>
- Aliakbari, M., & Boghayeri, M. (2014). A Needs Analysis Approach to ESP Design in Iranian Context. *Procedia - Social and Behavioral Sciences*, 98, 175–181.  
<https://doi.org/10.1016/j.sbspro.2014.03.404>

- Alsamadani, H. A. (2017). Needs Analysis in ESP Context: Saudi Engineering Students as a Case Study. *Advances in Language and Literary Studies*, 8(6), 58.  
<https://doi.org/10.7575/aiac.all.v.8n.6p.58>
- Altasan, A. (2016). Motivational orientations and their effect on English Language Learning: A Study in EFL Saudi Context. *American Journal of Educational Research*, 4, 1131–1137. <https://doi.org/10.12691/education-4-16-2>
- ALTE. (2002). The ALTE CAN DO Project—*Association of Language Testers in Europe*.
- AltoNivel. (2015, April 28). *¿Hablas inglés? Felicidades, eres minoría en México*. Alto Nivel. <https://www.altonivel.com.mx/liderazgo/management/50607-bajo-nivel-de-ingles-reduce-competitividad-de-mexico/>
- American Council on the Teaching of Foreign Languages. (2012). *ACTFL Proficiency Guidelines 2012*. American Council on the Teaching of Foreign Languages.
- Arkoudis, S., Hawthorne, P. L., Baik, C., Hawthorne, G., O'Loughlin, K., Bexley, E., & Leach, D. (2009). *The Impact of English Language Proficiency and Workplace Readiness on the Employment Outcomes of Tertiary International Students*. Australian Government. <http://hdl.voced.edu.au/10707/142064>
- Ashton, K. (2008). The languages ladder and asset languages: A new assessment framework for languages in England. In *Multilingual Europe: Diversity and Learning* (p. 175). Trentham Books.  
[https://www.researchgate.net/publication/262915625\\_The\\_languages\\_ladder\\_and\\_asset\\_languages\\_A\\_new\\_assessment\\_framework\\_for\\_languages\\_in\\_England](https://www.researchgate.net/publication/262915625_The_languages_ladder_and_asset_languages_A_new_assessment_framework_for_languages_in_England)

- ATESL. (nd). *Determining Needs*. Government of Alberta.  
[https://www.atesl.ca/documents/1449/Determining\\_Needs.pdf](https://www.atesl.ca/documents/1449/Determining_Needs.pdf)
- Bachman, L., & Damböck, B. (2018). *Language Assessment for Classroom Teachers: Assessment for Teachers*.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.
- Bachman, L. F. (2004). *Statistical Analyses for Language Assessment*. Cambridge University Press.
- Bachman, L. F. (2005). Building and Supporting a Case for Test Use. *Language Assessment Quarterly*, 2(1), 1–34. [https://doi.org/10.1207/s15434311laq0201\\_1](https://doi.org/10.1207/s15434311laq0201_1)
- Bachman, L. F. (2015). Justifying the Use of Language Assessments: Linking Test Performance with Consequences. *JLTA Journal*, 18, 3–22.  
[https://doi.org/10.20622/jltajournal.18.0\\_3](https://doi.org/10.20622/jltajournal.18.0_3)
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford University Press.
- Bachman, L. F., & Purpura, J. E. (2008). Language Assessments: Gate-Keepers or Door-Openers? In B. Spolsky & F. M. Hult (Eds.), *The Handbook of Educational Linguistics* (pp. 456–468). Blackwell Publishing Ltd.  
<https://doi.org/10.1002/9780470694138.ch32>
- Bachman, L., & Palmer, A. (2010). *Language Assessment in Practice*. OUP Oxford.
- Bai, Y. (2020). The relationship of test takers' learning motivation, attitudes towards the actual test use and test performance of the College English Test in China.

*Language Testing in Asia*, 10(1), 10. <https://doi.org/10.1186/s40468-020-00108-z>

Barjesteh, H., & Shakeri, F. (2013). CONSIDERING THE ISSUES OF LANGUAGE FOR SPECIFIC PURPOSES AT IRANIAN UNIVERSITY: ITS GENESIS/PROBLEMS AND SUGGESTIONS. *Indian Journal of Fundamental and Applied Life Sciences*, 3(3), 540–552.

Barnard, R., & Zemach, D. (2003). Materials for Specific Purposes. In *Developing Materials for Language Teaching* (pp. 306–323). Continuum.

Basturkmen, H. (2010a). *Developing Courses in English for Specific Purposes*. Palgrave Macmillan.  
[https://www.academia.edu/33544413/Developing\\_Courses\\_in\\_English\\_for\\_Specific\\_Purposes](https://www.academia.edu/33544413/Developing_Courses_in_English_for_Specific_Purposes)

Basturkmen, H. (2010b). *Developing Courses in English for Specific Purposes | H. Basturkmen | Palgrave Macmillan*. Palgrave Macmillan.  
<http://www.palgrave.com/gp/book/9780230227972>

Basturkmen, H., & Elder, C. (2004). The Practice of LSP. In *The Handbook of Applied Linguistics* (pp. 672–694). Blackwell Publishing Ltd.

Baumert, J., & Demmrich, A. (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education*, 16(3), 441–462.

BCEI. (2015). *English in Mexico: An examination of policy, perceptions and influencing factors*. British Council Education Intelligence.

- Beacco, J.-C., & Byram, M. (Eds.). (2016). *Guide for the development and implementation of curricula for plurilingual and intercultural education*. Council of Europe.
- Belcher, D. (2009). What ESP Is and Can Be: An Introduction. In *English for Specific Purposes in Theory and Practice* (p. 20). University of Michigan Press.
- Belcher, D. D. (2006). English for Specific Purposes: Teaching to Perceived Needs and Imagined Futures in Worlds of Work, Study and Everyday Life. *TESOL QUARTERLY*, 40(1), 133–156.
- Benavent, G. T., & Sánchez-Reyes, S. (2015). Target Situation as a Key Element for ESP (Law Enforcement) Syllabus Design. *Procedia - Social and Behavioral Sciences*, 173, 143–148. <https://doi.org/10.1016/j.sbspro.2015.02.044>
- Berwick, R. (1989). Needs assessment in language programming: From theory to practice. In R. K. Johnson (Ed.), *The Second Language Curriculum*. Cambridge University Press.
- Borjian, A. (2015). Learning English in Mexico: Perspectives From Mexican Teachers of English. *The CATESOL Journal*, 27(1), 163–173.
- Braine, G. (2001). Twenty years of needs analyses: Reflections on a personal journey. In *Research Perspectives on English for Academic Purposes* (pp. 195–207). Cambridge University Press.
- Branden, K. van den. (2006). Introduction: Task-based language teaching in a nutshell. In K. van den Branden (Ed.), *Task-Based Language Education* (pp. 1–16). Cambridge University Press. <https://doi.org/10.1017/CBO9780511667282.002>

- Braun, V., & Clarke, V. (2006a). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.  
<https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2006b). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77–101.  
<https://doi.org/10.1191/1478088706qp063oa>
- Brindley, G. P. (1989). The role of needs analysis in adult ESL programme design. In R. K. Johnson (Ed.), *The Second Language Curriculum*. Cambridge University Press. <https://www.cambridge.org/core/books/the-second-language-curriculum/B450FA5EB84A25C8553E4A469C3C3C7F>
- British Council, & Aptis. (2016). *Empower your business with Aptis*. British Council.
- British Council Global. (2019). *B1 English level | British Council*.  
<https://www.britishcouncil.es/en/english/levels/b1>
- Brown, H. D. (2004). *Language Assessment—Principles and Classroom Practice* (1st edition). Pearson ESL.
- Brown, J. D. (2009). Foreign and Second Language Needs Analysis. In *The Handbook of Language Teaching* (pp. 267–293). John Wiley & Sons, Ltd.  
<https://doi.org/10.1002/9781444315783.ch16>
- Brown, J. D. (2016). *Introducing Needs Analysis and English for Specific Purposes* (1st edition). Routledge.
- Brunfaut, T. (2014). Language for Specific Purposes: Current and Future Issues. *Language Assessment Quarterly*, 11(2), 216–225.  
<https://doi.org/10.1080/15434303.2014.902060>



- Bruns, A. (2010). *Doing Action Research in English Language Teaching A Guide for Practitioners*. Taylor & Francis.
- BUAP. (2019). *Opciones de Acreditación*. Benemérita Universidad Autónoma de Puebla.
- Cambridge English. (2016a). *English at Work: Global analysis of language skills in the workplace Full report*. Cambridge English Language Assessment, Quacquarelli Symonds.
- Cambridge English. (2016b). *English at Work: Global analysis of language skills in the workplace*. Cambridge English Language Assessment.
- Cambridge English Language Assessment. (2016). *Cambridge English First*. UCLES.
- CCLB. (n.d.). *Work Ready Tools Non-Regulated Skilled and Semi-Skilled Occupations*. Centre for Canadian Language Benchmarks, Centre des niveaux de compétence linguistique canadiens.
- CCLB. (2006). *The First Year of Official CELBAN Administration in Canada (March 2005—April 2006)* (p. 19) [CELAS Centre Report]. CCLB. <https://cpb-ca-c1.wpmucdn.com/www.rrc.ca/dist/6/47/files/2016/06/Final-Official-CELBAN-report-First-year-March-2005-to-April-2006.pdf>
- CCLB. (2012). Canadian Language Benchmarks English as a Second Language for Adults. *Centre for Canadian Language Benchmarks*, 235.
- CCLB. (2015). *Work Ready Resources for Counselling, Hiring, and Working with Internationally Trained Individuals*. Centre for Canadian Language Benchmarks, Centre des niveaux de compétence linguistique canadiens.

- CCLB. (2016). *Can Do Statements for Employment*. Centre for Canadian Language Benchmarks, Centre des niveaux de compétence linguistique canadiens.
- Chairat, P. (2016). Communication Skills Used by English Major Graduates in the Workplace: A Case Study. *International Journal of Languages, Literature and Linguistics*, 2(1), 38–43.
- Chalak, A. (2019). English Needs Analysis in the Workplace: The Case of Engineers in Esfahan Steel Company. *International Journal of Foreign Language Teaching & Research*.  
[https://www.academia.edu/50944305/English\\_Needs\\_Analysis\\_in\\_the\\_Workplace\\_The\\_Case\\_of\\_Engineers\\_in\\_Esfahan\\_Steel\\_Company](https://www.academia.edu/50944305/English_Needs_Analysis_in_the_Workplace_The_Case_of_Engineers_in_Esfahan_Steel_Company)
- Chambers, F. (1980). A re-evaluation of needs analysis in ESP. *The ESP Journal*, 1(1), 25–33. [https://doi.org/10.1016/0272-2380\(80\)90007-4](https://doi.org/10.1016/0272-2380(80)90007-4)
- Changpueng, P., & Pattanapichet, F. (2015). An Analysis of English in the Workplace: The Needs of Engineers in Writing in English. *Silpakorn University Journal of Social Sciences, Humanities, and Arts*.  
<https://www.semanticscholar.org/paper/An-Analysis-of-English-in-the-Workplace%3A-The-Needs-Changpueng-Pattanapichet/54939e0916ef1fb47d860a729c3f995577cd1837>
- Chapelle, C. A. (2008). The TOEFL validity argument. In *Building a Validity Argument for the Test of English as a Foreign Language* (pp. 319–352). Routledge.
- Chapelle, C. A. (2011). Validity argument for language assessment: The framework is simple.... *Language Testing*, 29(1), 19–27.  
<https://doi.org/10.1177/0265532211417211>

- Chapelle, C. A. (2021). *Argument-Based Validation in Testing and Assessment*. SAGE Publications.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. (2010). Does an Argument-Based Approach to Validity Make a Difference? *Educational Measurement: Issues and Practice*, 29(1), 3–13. <https://doi.org/10.1111/j.1745-3992.2009.00165.x>
- Chapelle, C. A., Grabe, W., & Berns, M. (1997). *Communicative Language Proficiency: Definition and Implications for TOEFL 2000* (Monograph Series). ETS. [https://www.ets.org/research/policy\\_research\\_reports/publications/report/1997/hzeu](https://www.ets.org/research/policy_research_reports/publications/report/1997/hzeu)
- Chapelle, C. A., & Lee, H. (2021). Understanding Argument-Based Validity in Language Testing. In *Validity Argument in Language Testing*. Cambridge University Press.
- Chapelle, C. A., & Voss, E. (2013). Evaluation of Language Tests Through Validation Research. In *The Companion to Language Assessment* (pp. 1079–1097). American Cancer Society. <https://doi.org/10.1002/9781118411360.wbcla110>
- Chapelle, C. A., & Voss, E. (2021). Introduction to Validity Argument in Language Testing and Assessment. In *Argument-Based Validation in Testing and Assessment* (Vol. 184, p. 16). Iowa State University.
- Chapman, M. (2004). Voices in the Field: An Interview with Kazuhiko Saito. *Shiken: JALT Testing and Evaluation SIG Newsletter*, 8(2), 8–11.
- Charge, N., & Taylor, L. B. (1997). Recent developments in IELTS. *ELT Journal*, 51(4), 374–380. <https://doi.org/10.1093/elt/51.4.374>

- Chen, T. (2022). An Argument-Based Validation of an Asynchronous Written Interaction Task. *Frontiers in Psychology, 13*.  
<https://www.frontiersin.org/articles/10.3389/fpsyg.2022.889488>
- Chovancová, B. (2014). Needs Analysis and Esp Course Design: Self-Perception of Language Needs Among Pre-Service Students—*Studies in Logic, Grammar and Rhetoric, 38*(1), 43.
- COE. (2001a). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Council of Europe Cambridge University Press.
- COE. (2001b). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Council of Europe Cambridge University Press.
- COE. (2001c). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Council of Europe.
- COE. (2018). *Common European Framework of Reference for Languages: Learning, teaching, assessment Companion Volume with New Descriptors*. Council of Europe.
- COE, & ALTE. (2016). *Language tests for access, integration and citizenship: An outline for policymakers*. Council of Europe, Association of Language Testers in Europe.
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research Methods in Education* (7 edition). Routledge.
- College Board. (2008). *Puerto Rico and Latin America Office*. College Board.
- College Board. (2015a). *ELASH 1 Guía del estudiante*. The College Board.

- College Board. (2015b). *Guía del estudiante ELASH II*. College Board.  
<https://latam.collegeboard.org/publicaciones/guia-del-estudiante-elash-ii/>
- College Board. (2018). *Informe final de la alineación del ELASH con el Marco Común Europeo*. College Board. <https://latam.collegeboard.org/publicaciones/informe-final-de-la-alineacion-del-elash-con-el-marco-comun-europeo/>
- Congreso Constituyente. (1993). Artículo 3o Gratuidad de la Educación. In *Constitución Política de los Estados Unidos Mexicanos*. UNAM.
- Consulta Mitovsky. (2013). *MEXICANOS Y LOS IDIOMAS EXTRANJEROS*.  
<http://www.consulta.mx/index.php/estudios-e-investigaciones/mexico-opina/item/823-mexicanos-y-los-idiomas-extranjeros>
- Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: A practical guide to Kane's framework. *Medical Education, 49*(6), 560–575. <https://doi.org/10.1111/medu.12678>
- Council of Europe. (n.d.). *Council of Europe / Language Policy Unit (Strasbourg)—Project LIAM*. Council of Europe. [www.coe.int/lang-migrants](http://www.coe.int/lang-migrants)
- Coxhead, A., & Nation, P. (2001). The Specialized vocabulary of English for academic purposes. In *Research Perspectives on English for Academic Purposes* (pp. 252–267). Cambridge University Press.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed). Sage Publications.
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design* (Fifth). SAGE Publications.  
<https://uk.sagepub.com/en-gb/eur/research-design/book255675>

- Creswell, J. W., & Plano Clark, V. L. (2011). Choosing a mixed methods. In *Designing and Conducting Mixed Methods Research* (pp. 53–106). SAGE Publications.
- Cronquist, K., & Fiszbein, A. (2017). *El aprendizaje del inglés en América Latina*. Pearson.
- Crystal, D. (2003). *English as a global language* (Second edition). Cambridge University Press.
- Davidson, S. (2022). The domain expert perspective: A qualitative study into the views expressed in a standard-setting exercise on a language for specific purposes (LSP) test for health professionals. *Language Testing*, 39(1), 117–141.
- Davies, A. (2001). The logic of testing Languages for Specific Purposes. *Language Testing*, 18.
- Davies, A. (2008). *Assessing Academic English: Testing English Proficiency 1950-1989 - The IELTS Solution* (1 edition). Cambridge University Press.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of Language Testing: Studies in Language Testing 7*. Cambridge University Press.
- Dawadi, S. (2021). Factors Influencing Washback of a High-stakes English-as-a-Foreign- Language Test. *TESL-EJ: The Electronic Journal for English as a Second Language*, 25(3), 1–16.
- de Chazal, E. (2012). THE GENERAL–SPECIFIC DEBATE IN EAP: WHICH CASE IS THE MOST CONVINCING FOR MOST CONTEXTS? *Journal of Second Language Teaching & Research*, 2(1), 135–148.

- De Wit, H., Jaramillo, I. C., Gacel-Ávila, J., & Knight, J. (2005). *Educación Superior en América Latina*. Banco Mundial.
- Deygers, B., Van den Branden, K., & Van Gorp, K. (2018). University entrance language tests: A matter of justice. *Language Testing*, 35(4), 449–476.  
<https://doi.org/10.1177/0265532217706196>
- DGAIR. (2011). *Normas Correspondientes al Control Escolar, así como a la Acreditación y Certificación de Conocimientos y Aptitudes en Lenguas Extranjeras y en el Español como Lengua Extranjera, a que se sujetará la Dirección General de Acreditación, Incorporación y Revalidación*. SEP.  
[https://www.cenni.sep.gob.mx/work/models/cenni/Resource/PDFs/norma\\_cenni.pdf](https://www.cenni.sep.gob.mx/work/models/cenni/Resource/PDFs/norma_cenni.pdf)
- Di Bella, E., & Batista, J. (2007). La enseñanza del inglés instrumental mediante estudios a distancia. Caso: Postgrado de arquitectura y diseño de LUZ. *Laurus Revista de Educación*, 13(23), 413–440.
- Didiot-Cook, H., Gauthier, V., & Scheirlinckx, K. (2000). *Language needs in business, a survey of European multinational companies*. 25.
- Dörnyei, Z. (1994). Motivation and Motivating in the Foreign Language Classroom. *The Modern Language Journal*, 78(3), 273–284.
- Dörnyei, Z. (2005). *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Routledge.  
<https://doi.org/10.4324/9781410613349>
- Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford University Press.
- Dörnyei, Z. (2009). The L2 Motivational Self System. In *Motivation, Language Identity and the L2 Self* (p. 34). Multilingual Matters.

- Dörnyei, Z. (2019a). From Integrative Motivation to Directed Motivational Currents: The Evolution of the Understanding of L2 Motivation over Three Decades. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave Handbook of Motivation for Language Learning* (pp. 39–69). Springer International Publishing. [https://doi.org/10.1007/978-3-030-28380-3\\_3](https://doi.org/10.1007/978-3-030-28380-3_3)
- Dörnyei, Z. (2019b). Towards a better understanding of the L2 Learning Experience, the Cinderella of the L2 Motivational Self System. *Studies in Second Language Learning and Teaching*, 9(1), 19–30. <https://doi.org/10.14746/sslit.2019.9.1.2>
- Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in Second Language Research: Construction, Administration, and Processing* (2 edition). Routledge.
- Dörnyei, Z., & Ushioda, E. (2021). *Teaching and Researching Motivation* (3rd ed.). Routledge. <https://doi.org/10.4324/9781351006743>
- Douglas, D. (2000). *Assessing Languages for Specific Purposes*. Cambridge University Press.
- Dudley-Evans, T. (2000). Genre analysis: A key to a theory of ESP? *Ibérica*, 3–11.
- Dudley-Evans, T., & St John, M. J. S. (1998). *Developments in English for Specific Purposes: A Multi-Disciplinary Approach*. Cambridge University Press.
- Duff, P. (2020). Case study research: Making language learning complexities visible - 2020. *The Routledge Handbook of Research Methods in Applied Linguistics*, 144–153.
- Duff, P. A. (2008). Language Socialization, Higher Education, and Work. In N. H. Hornberger (Ed.), *Encyclopedia of Language and Education* (pp. 2818–2831). Springer US. [https://doi.org/10.1007/978-0-387-30424-3\\_211](https://doi.org/10.1007/978-0-387-30424-3_211)



- Duff, P. A., & Anderson, T. (2015). Case study research. *The Cambridge Guide to Language Research*.  
[https://www.academia.edu/6790647/Duff\\_P\\_and\\_Anderson\\_T\\_2015\\_Case\\_study\\_research\\_In\\_J\\_D\\_Brown\\_and\\_C\\_Coombe\\_Eds\\_The\\_Cambridge\\_Guide\\_to\\_Language\\_Research\\_New\\_York\\_Cambridge\\_University\\_Press](https://www.academia.edu/6790647/Duff_P_and_Anderson_T_2015_Case_study_research_In_J_D_Brown_and_C_Coombe_Eds_The_Cambridge_Guide_to_Language_Research_New_York_Cambridge_University_Press)
- Dursun, A., & Li, Z. (2021). A Systematic Review of Argument-Based Validation Studies in the Field of Language Testing (2000–2018). In C. A. Chapelle & E. Voss (Eds.), *Validity Argument in Language Testing: Case Studies of Validation Research* (pp. 45–70). Cambridge University Press.  
<https://doi.org/10.1017/9781108669849.005>
- Dussel, I. (2005). Las políticas curriculares de la última década en América Latina: Nuevos actores, nuevos problemas. In *Educación de calidad para todos: Iniciativas iberoamericanas* (pp. 93–102). Santillana.
- Education First. (2011). *English Proficiency Index*. Education First Ltd.
- Education First. (2012). *English Proficiency Index* (pp. 1–35). Education First Ltd.  
<http://www.ef.com.mx/epi/>
- Education First. (2014a). *English Proficiency Index*. Education First Ltd.  
<http://www.ef.com.mx/epi/>
- Education First. (2014b). *English Proficiency Index*. Education First Ltd.  
<http://www.ef.com.mx/epi/>
- Education First. (2015a). *EF EPI 2015 HOJA DE DATOS*. Education First Ltd.  
<https://www.ef.fr/sitecore/~/media/centralefcom/epi/downloads/fact-sheets/v5/ef-epi-country-fact-sheet-v5-mx-es.pdf>

- Education First. (2015b). *English Proficiency Index* (pp. 1–48). Education First Ltd.  
<http://www.ef.edu/~/media/centralefcom/epi/downloads/full-reports/v6/ef-epi-2016-english.pdf>
- Education First. (2017). *English Proficiency Index* (pp. 1–48). Education First Ltd.  
<http://www.ef.edu/~/media/centralefcom/epi/downloads/full-reports/v6/ef-epi-2016-english.pdf>
- Education First. (2018). *English Proficiency Index*. Education First Ltd.
- Education First. (2019). *English Proficiency English* (pp. 1–52). Education First Ltd.
- Education First. (2020). *English Proficiency Index* (pp. 1–27). Education First Ltd.
- Education First. (2021). *English Proficiency Index* (pp. 1–40). Education First Ltd.
- Education First. (2022). *EF English Proficiency Index*. <https://www.ef.com/wwen/epi/>
- Education First. (2016). *EF EPI - México*. <http://www.ef.com.mx/epi/regions/latin-america/mexico/>
- EF. (2013). *EF EPI EF English Proficiency Index*. Education First Ltd.
- Elder, C. (2017). Language Assessment in Higher Education. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language Testing and Assessment* (pp. 271–286). Springer International Publishing. [https://doi.org/10.1007/978-3-319-02261-1\\_35](https://doi.org/10.1007/978-3-319-02261-1_35)
- Elder, C., McNamara, T., Woodward-Kron, R., Manias, E., McColl, G., & Webb, G. (2013). *Towards improved healthcare communication Development and validation of language proficiency standards for non-native English speaking health professionals* (p. 43) [Final Report for the Occupational English Test Centre]. University of Melbourne. <http://hdl.handle.net/11343/55166>

- Ellis, M., & Johnson, C. (1994). *Teaching Business English: An Introduction to Business English for Language Teachers, Trainers, and Course Organizers*. Oxford University Press.
- Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (Eds.). (2019). Task-Based Testing and Assessment. In *Task-Based Language Teaching: Theory and Practice* (pp. 241–280). Cambridge University Press.  
<https://doi.org/10.1017/9781108643689.014>
- Enright, A. (2007). *ELPAC A Test for Air Traffic Controllers to meet ICAO Language Proficiency Requirements*. EUROCONTROL.  
[https://www.icao.int/Meetings/AMC/MA/Second%20ICAO%20Aviation%20Language%20Symposium%20\(IALS-2\)/22.Enright.pdf](https://www.icao.int/Meetings/AMC/MA/Second%20ICAO%20Aviation%20Language%20Symposium%20(IALS-2)/22.Enright.pdf)
- ETS. (n.d.). *The TOEFL Test: Find Your Format*. Retrieved 22 February 2019, from [https://www.ets.org/bin/getprogram.cgi?urlSource=toefl&newRegURL=&test=TOEFL&greClosed=new&greClosedCountry=China&browserType=&toeflType=&redirect=&t\\_country1=group\\_Mexico](https://www.ets.org/bin/getprogram.cgi?urlSource=toefl&newRegURL=&test=TOEFL&greClosed=new&greClosedCountry=China&browserType=&toeflType=&redirect=&t_country1=group_Mexico)
- ETS. (2011). TOEFL iBT Research Insight TOEFL Program History. *Educational Testing Service Princeton, NJ*, 6(1), 1–7.
- ETS. (2012). *TOEIC Examinee Handbook Listening and Reading*.  
[http://www.ets.org/Media/Tests/TOEIC/pdf/TOEIC\\_LR\\_examinee\\_handbook.pdf](http://www.ets.org/Media/Tests/TOEIC/pdf/TOEIC_LR_examinee_handbook.pdf)
- EUROCONTROL. (2018). *ELPAC | ICAO Language Proficiency Testing for Air Traffic Controllers and Pilots*. <https://elpac.eurocontrol.int/>
- Export Enterprises SA. (2019, April). *Inversión extranjera en México—Santandertrade.com*. Santander Trade Portal.

<https://es.portal.santandertrade.com/establecerse-extranjero/mexico/inversion-extranjera?>

Facultad de Lenguas. (n.d.). *Política Institucional de Lenguas Extranjeras*. BUAP.

Fang, F. (2017). English as a Lingua Franca: Implications for Pedagogy and Assessment. *TEFLIN Journal: A Publication on the Teaching and Learning of English*, 28(1), 57–70.

Fechter, T., Dai, T., Cromley, J. G., Nelson, F. E., Van Boekel, M., & Du, Y. (2021). Developing a Validity Argument for an Inference-Making and Reasoning Measure for Use in Higher Education. *Frontiers in Education*, 6.  
<https://www.frontiersin.org/articles/10.3389/feduc.2021.727539>

Ferguson, G., Pérez-Llantada, C., & Plo, R. (2011). English as an international language of scientific publication: A study of attitudes. *World Englishes*, 30(1), 41–59. <https://doi.org/10.1111/j.1467-971X.2010.01656.x>

Ferrari Fermín, V. I., & Torrealba M, C. E. (2009). Inglés para matemáticos en la Universidad Simón Bolívar: Hacia un rediseño basado en un análisis de necesidades. *Núcleo*, 21(26), 011–035.

Fischer, J. (Ed.). (2011). *Guidelines for task-based university language testing*. European Centre for Modern Languages / Council of Europe Publ.

Flowerdew, J., & Peacock, M. (2001). Issues in EAP: A preliminary perspective. In *Research Perspectives on English for Academic Purposes* (pp. 8–24). Cambridge University Press.

Flowerdew, L. (2005). Integrating traditional and critical approaches to syllabus design: The ‘what’, the ‘how’ and the ‘why?’ *Journal of English for Academic Purposes*, 4, 135–147.

- Flowerdew, L. (2013). Needs Analysis and Curriculum Development in ESP. In *The Handbook of English for Specific Purposes* (pp. 325–346). Wiley-Blackwell.
- Fulcher, G. (2009). *Motivation and Language Testing*.  
<http://languagetesting.info/features/motivation/mil.html>
- Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment: An Advanced Resource Book*. Taylor & Francis.
- Fulcher, G., & Owen, N. (2016). Dealing with the demands of language testing and assessment. In *The Routledge Handbook of English Language Teaching*. Routledge.
- Garcia-Ponce, E. E. (2020). Needs Analysis to Enhance English Language Proficiency at a Mexican University. *Profile: Issues in Teachers' Professional Development*, 22(2), 145–162.
- Gardner, R. C. (2001). Language Learning Motivation: The Student, the Teacher, and the Researcher. *Texas Papers in Foreign Language Education*, 6(1), 20.
- Gardner, R. C. (2019). The Socio-educational Model of Second Language Acquisition. In *The Palgrave Handbook of Motivation for Language Learning* (pp. 21–37). Palgrave Macmillan.
- Gardner, R. C., & MacIntyre, P. D. (1993). A student's contributions to second-language learning. Part II: Affective variables. *Language Teaching*, 26(1), 1–11.  
<https://doi.org/10.1017/S0261444800000045>
- Gardner, R. C., & Tremblay, P. F. (1994). On Motivation, Research Agendas, and Theoretical Frameworks. *The Modern Language Journal*, 78, 359–368.
- Graddol, D. (2000). *The Future of English?* The British Council.

- Graddol, D. (2006). *English Next*. British Council.
- Graves, K. (2016). Language curriculum design: Possibilities and realities. In *The Routledge Handbook of English Language Teaching* (pp. 97–112). Routledge.  
<https://doi.org/10.4324/9781315676203-15>
- Green, A. (2014). *Exploring Language Assessment and Testing Language in Action*. Routledge.
- Greenberg, I. (2012). ESL Needs Analysis and Assessment in the Workplace. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyloff (Eds.), *The Cambridge Guide to Second Language Assessment* (pp. 178–186). Cambridge University Press.
- Guerrero Cárdenas, E. (2008). La Enseñanza del Inglés Instrumental en los Formadores. *Evaluación e Investigación*, 3(1), 12.
- Haggerty, J., & Fox, J. (2015). Raising the bar: Language testing experience and second language motivation among South Korean young adolescents. *Language Testing in Asia*, 5, 1–16. <https://doi.org/10.1186/s40468-015-0018-0>
- Hamel, R. E., Álvarez López, E., & Carvalhal, T. P. (2016). Language policy and planning: Challenges for Latin American universities. *Current Issues in Language Planning*, 17(3–4), 278–297.  
<https://doi.org/10.1080/14664208.2016.1201208>
- Hamid, M. O., Hardy, I., & Reyes, V. (2019). Test-takers' perspectives on a global test of English: Questions of fairness, justice and validity. *Language Testing in Asia*, 9(1), 16. <https://doi.org/10.1186/s40468-019-0092-9>

- Harlen, W., & Crick, R. (2003). Testing and Motivation for Learning. *Assessment in Education: Principles, Policy & Practice*, 169–207.  
<https://doi.org/10.1080/0969594032000121270>
- Harmer, J. (2001). *The Practice of English Language Teaching* (Third). Longman.
- Harsch, C. (2018). How Suitable Is the CEFR for Setting University Entrance Standards? *Language Assessment Quarterly*, 15(1), 102–108.  
<https://doi.org/10.1080/15434303.2017.1420793>
- He, L., & Min, S. (2017). Development and Validation of a Computer Adaptive EFL Test. *Language Assessment Quarterly*, 14(2), 160–176.  
<https://doi.org/10.1080/15434303.2016.1162793>
- Healthdirect. (2022, October 28). *What is a referral?* [Text/html]. Healthdirect Australia.  
<https://www.healthdirect.gov.au/what-is-a-referral>
- Heaton, J. B. (1975). *Writing English Language Tests*. Longman Group UK Limited.
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. Cambridge, Mass. : Newberry House Publishers.  
<http://archive.org/details/guidetolanguaget0000henn>
- Heredia, B., & Rubio, D. (2015). Inglés y desigualdad social en México. In *Sorry -El aprendizaje del inglés en México*. Mexicanos Primero.
- Hughes, A. (2003). *Testing for Language Teachers* (2nd edition). Cambridge University Press ELT.
- Huhta, M., Vogt, K., Johnson, E., & Tulkki, H. (2013). *Needs Analysis for Language Course Design: A Holistic Approach to ESP*. Cambridge University Press.

- Hutchinson, T., & Waters, A. (1987a). *English for Specific Purposes*. Cambridge University Press.
- Hutchinson, T., & Waters, A. (1987b). *English for Specific Purposes*. Cambridge University Press.
- Hyland, K. (2002). Specificity revisited: How far should we go now? *English for Specific Purposes*, 21, 385–395.
- Hyland, K. (2006). *English for Academic Purposes: An Advanced Resource Book*. Routledge.
- Im, G.-H., & Cheng, L. (2019). The Test of English for International Communication (TOEIC®). *Language Testing*, 36(2), 315–324.  
<https://doi.org/10.1177/0265532219828252>
- IMCO. (2015). *Inglés para la competitividad y la movilidad social Inglés es posible Propuesta de una Agenda Nacional*. Instituto Mexicano para la Competitividad A.C. [http://imco.org.mx/wp-content/uploads/2015/04/2015\\_Documento\\_completo\\_Ingles\\_es\\_posible.pdf](http://imco.org.mx/wp-content/uploads/2015/04/2015_Documento_completo_Ingles_es_posible.pdf)
- IMCO. (2018). *¿Ahora con quién me enojo?* [Economico informativo].
- International English Language Testing System (IELTS). (2019). *IELTS guide for teachers*. British Council.
- International Legal English (ILEC) | British Council*. (n.d.). Retrieved 23 January 2021, from <https://www.britishcouncil.es/en/exam/ilec-international-legal-english-certificate>



- Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). *TOEFL 2000 Framework: A Working Paper* (Monograph Series). Educational Testing Service.
- Javid, C. Z. (2014). EMP Needs of Medical Undergraduates in a Saudi Context. *Kashmir Journal of Language Research*, 14(1), 89–110.
- Jenkins, J., & Leung, C. (2016). Assessing English as a Lingua Franca. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language Testing and Assessment* (pp. 1–15). Springer International Publishing. [https://doi.org/10.1007/978-3-319-02326-7\\_7-1](https://doi.org/10.1007/978-3-319-02326-7_7-1)
- Jin, Y. (2023). Test-taker insights for language assessment policies and practices. *Language Testing*, 40(1), 193–203.  
<https://doi.org/10.1177/02655322221117136>
- Johnson, R. C. (2011). *Assessing the assessments: Using an argument-based validity framework to assess the validity and use of an English placement system in a foreign language context* [Thesis for the degree of Doctor of Philosophy in Applied Linguistics]. Macquarie University.
- Jones, C. (1991). An integrated model for ESP syllabus design. *English for Specific Purposes*, 10(3), 155–172. [https://doi.org/10.1016/0889-4906\(91\)90022-O](https://doi.org/10.1016/0889-4906(91)90022-O)
- Jordan, R. R. (1997). *English for Academic Purposes: A Guide and Resource Book for Teachers*. Cambridge University Press.
- Kachru, B. B. (1992). *The Other Tongue: English Across Cultures*. University of Illinois Press.

- Kane, M. (2011). Validating score interpretations and uses: Messick Lecture, Language Testing Research Colloquium, Cambridge, April 2010. *Language Testing*, 0265532211417210. <https://doi.org/10.1177/0265532211417210>
- Kane, M. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Kane, M., Crooks, T., & Cohen, A. (1999). Validating Measures of Performance. *Educational Measurement: Issues and Practice*, 18(2), 5–17. <https://doi.org/10.1111/j.1745-3992.1999.tb00010.x>
- Kane, M. T. (1990). An Argument-based Approach to Validation. *ACT Research Report Series*, 90–13.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527–535. <https://doi.org/10.1037/0033-2909.112.3.527>
- Kassim, H., & Ali, F. (2010). English communicative events and skills needed at the workplace: Feedback from the industry. *English for Specific Purposes*, 29, 168–182.
- Keblawi, F. (2022). Language Tests, Motivation, and Demotivation: A Complex Relationship. *International Journal of Educational Methodology*, 8, 699–709. <https://doi.org/10.12973/ijem.8.4.699>
- Knekta, E. (2017). *Motivational Aspects of Test-Taking Measuring Test-Taking Motivation in Swedish National Test Contexts*. Umea Universitet.
- Knoch, U., & Chapelle, C. A. (2018). Validation of rating processes within an argument-based framework. *Language Testing*, 35(4), 477–499. <https://doi.org/10.1177/0265532217710049>

- Knoch, U., & Elder, C. (2013). A framework for validating post-entry language assessments (PELAs). *Papers in Language Testing and Assessment*, 2(2), 19.
- Knoch, U., & Macqueen, S. (2016). Language assessment for the workplace. In *Handbook of Second Language Assessment* (pp. 291–308). De Gruyter Mouton.  
[https://www.researchgate.net/publication/302940518\\_Language\\_assessment\\_for\\_the\\_workplace](https://www.researchgate.net/publication/302940518_Language_assessment_for_the_workplace)
- Knoch, U., & Macqueen, S. (2020). *Assessing English for Professional Purposes*. Routledge.
- Knoch, U., May, L., Macqueen, S., Pill, J., & Storch, N. (2016). *Transitioning from university to the workplace: Stakeholder perceptions of academic and professional writing demands* (IELTS Research Report Online Series). IELTS.
- Knoch, U., McNamara, T., & Elder, C. (2017). *Submission to the Australian Government Department of Immigration and Border Protection on the discussion paper 'Strengthening the test for the Australian Citizenship'*. The University of Melbourne.
- Knoch, U., Paltridge, B., Macqueen, S., & Starfield, S. (2019). *Assessing English for Professional Purposes*. Routledge. <https://doi.org/10.4324/9780429340383>
- Lamb, M. (2004). Integrative motivation in a globalizing world. *System*, 32(1), 3–19.  
<https://doi.org/10.1016/j.system.2003.04.002>
- Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema*, 26(1), 10.

- Langues et employabilité, & Erasmus+. (2015). *Analysis of the foreign language skills requirements of French employers summary*. Langues et Employabilité, Erasmus+.
- Lewis, C., & Kingdon, B. (2016). CELBAN: A 10-Year Retrospective Catherine Lewis & Blanche Kingdon. *TESL Canada Journal*, 33(2), 69.  
<https://doi.org/10.18806/tesl.v33i2.1238>
- Li, Z. (2021). Investigating the Consequences of an ESL Placement Test. In C. A. Chapelle & E. Voss (Eds.), *Validity Argument in Language Testing: Case Studies of Validation Research* (pp. 294–316). Cambridge University Press.  
<https://doi.org/10.1017/9781108669849.016>
- Liao, L., Ye, S. X., & Yang, J. (2023). A mini-review of communicative language testing. *Frontiers in Psychology*, 14.  
<https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1058411>
- Liu, H. (2014). *Investigating the Relationships between a Reading Test and Can-do Statements of Performance on Reading Tasks* [UCLA].  
<https://escholarship.org/uc/item/7br4g6pc>
- Long, M. (2015). *Second Language Acquisition and Task-Based Language Teaching*. John Wiley & Sons.
- Long, M. H. (2005). *Second Language Needs Analysis*. Cambridge University Press.
- Long, M. H. (2016). In Defense of Tasks and TBLT: Nonissues and Real Issues. *Annual Review of Applied Linguistics*, 36, 5–33.  
<https://doi.org/10.1017/S0267190515000057>
- Long, M. H., & Doughty, C. J. (2009). *The Handbook of Language Teaching*. Wiley-Blackwell.

[https://www.researchgate.net/profile/Martin\\_Bygate/publication/227987748\\_Teaching\\_and\\_Testing\\_Speaking/links/5734733008ae9f741b27e091.pdf#page=571](https://www.researchgate.net/profile/Martin_Bygate/publication/227987748_Teaching_and_Testing_Speaking/links/5734733008ae9f741b27e091.pdf#page=571)

Louis Cohen, L. M. (2011, April 28). *Research Methods in Education, 7th Edition*.

<http://cw.routledge.com/textbooks/cohen7e/>

Lowenberg, P. H. (2002). Assessing English proficiency in the Expanding Circle. *World Englishes*, 21(3), 431–435. <https://doi.org/10.1111/1467-971X.00261>

Mann, W., & Marshall, C. R. (2010). Building an Assessment Use Argument for sign language: The BSL Nonsense Sign Repetition Test. *International Journal of Bilingual Education and Bilingualism*, 13(2), 243–258.

Mansfield, G., & Poppi, F. (2012). The English as a Foreign Language / Lingua Franca Debate: Sensitising Teachers of English as a Foreign Language Towards Teaching English as a Lingua Franca. *Profile Issues in Teachers' Professional Development*, 14(1), 159–172.

Markus, H., & Nurius, P. (1986). Possible Selves. *American Psychologist*, 41(9), 954–969. <https://doi.org/10.1037/0003-066x.41.9.954>

Mazatlántoday.net. (2019, April 13). *Durango Day Trip from Mazatlan—Road tours to Victoria Durango, Durango State*.

[https://mazatlantoday.net/mazatlan\\_durango\\_day\\_trip.html](https://mazatlantoday.net/mazatlan_durango_day_trip.html)

McNamara, T. (2000). *Language Testing*. OUP Oxford.

McNamara, T. (2001). The challenge of speaking: Research on the testing of speaking for the new TOEFL. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 5(1), 2–3.

- McNamara, T. (2010). The use of language tests in the service of policy: Issues of validity. *Revue Francaise de Linguistique Appliquee*, Vol. XV(1), 7–23.
- McNamara, T., & Roever, C. (2006). *Language Testing: The Social Dimension*. John Wiley & Sons.
- Meixner, C., & Hathcoat, J. D. (2019). The Nature of Mixed Methods Research. In P. Liamputtong (Ed.), *Handbook of Research Methods in Health Social Sciences* (pp. 51–70). Springer Singapore. [https://doi.org/10.1007/978-981-10-5251-4\\_76](https://doi.org/10.1007/978-981-10-5251-4_76)
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35(11), 1012–1027. <https://doi.org/10.1037/0003-066X.35.11.1012>
- Messick, S. (1989). Validity. In *Educational Measurement* (pp. 1–213). Educational Testing Service.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Mexicanos Primero. (2015). *Sorry- El aprendizaje del inglés en México* (1era ed.). Mexicanos Primero.
- Mihaela, M., Goia, L., Muresan, O., Ursa, O., & Uzoni, M. (2020). *ENGLISH FOR DENTISTRY PRACTICAL COURSE: NEEDS ANALYSIS AND FEEDBACK QUESTIONNAIRE* (p. 79).
- Mislevy, R. J., Hamel, L., Fried, R., Gaffney, T., Haertel, G., Hafter, A., Murphy, R., Quellmalz, E., Rosenquist, A., Schank, P., Draney, K., Kennedy, C., Long, K., Wilson, M., Chudowsy, N., Morrison, A. L., Pena, P., Butler Songer, N., &

- Wenk, A. (2003). *Design Patterns for the Assessment of Science Inquiry*. SRI International.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). Focus Article: On the Structure of Educational Assessments. *Measurement: Interdisciplinary Research & Perspective*, 1(1), 3–62. [https://doi.org/10.1207/S15366359MEA0101\\_02](https://doi.org/10.1207/S15366359MEA0101_02)
- Mislevy, R. J., & Yin, C. (2012). Evidence-centered design in language testing. In *The Routledge Handbook of Language Testing* (pp. 208–222). Routledge.
- Mohamed, A. A., Mohd Radzuan, N. R., Kassim, H., Azmat Ali, M. M., & Ali, A. (2014). Conceptualizing English Workplace Communication Needs Of Professional Engineers: The Challenges for English Language Tertiary Educators. *International Journal of Contemporary Business Management*, 1(1), 1–9.
- Moore, P. (2013). Mexican national and international policy on second languages (principally English). *Lengua y Voz*, 3(1), 3–18.
- Moslehifar, M. A., & Ibrahim, N. A. (2012). English Language Oral Communication Needs at the Workplace: Feedback from Human Resource Development (HRD) Trainees. *Procedia - Social and Behavioral Sciences*, 66, 529–536. <https://doi.org/10.1016/j.sbspro.2012.11.297>
- Navarro Leal, M. A., Sánchez Rodríguez, L. I., & Lladó Lárraga, D. M. (2009). La conducción de las universidades públicas en México: Un largo proceso de innovación. In *Universidad, sociedad e innovación, una perspectiva internacional* (p. 24). EDUNTREF.
- Niño-Puello, M. (2013). El inglés y su importancia en la investigación científica: Algunas reflexiones. *Revista Colombiana de Ciencia Animal*, 5(1), 243–254.

- Norris, J. M. (2009). Task-Based Teaching and Testing. In *The Handbook of Language Teaching* (pp. 578–594). John Wiley & Sons, Ltd.  
<https://doi.org/10.1002/9781444315783.ch30>
- Norris, J. M. (2016). Current Uses for Task-Based Language Assessment. *Annual Review of Applied Linguistics*, 36, 230–244.  
<https://doi.org/10.1017/S0267190516000027>
- North, B., Milanovic, M., & Saville, N. (2016). *The CEFR in Practice*. Cambridge University Press.
- North, B., & Piccardo, E. (2018). Aligning the Canadian Language Benchmarks (CLB) to the Common European Framework of Reference (CEFR). *Centre for Canadian Language Benchmarks*, 233.
- Notimex. (2017, July 19). Durango destino turístico emergente, la tierra del cine y de los... *Más México*. <https://mas-mexico.com.mx/durango-destino-turistico-emergente-la-tierra-del-cine-y-de-los-alacranes/>
- Nunan, D. (1992). *Research Methods in Language Learning*. Cambridge University Press.
- Nunan, D. (2004). *Task-based language teaching*. Cambridge University Press.
- OET. (n.d.). Cambridge Boxhill Language Assessment Trust.  
<https://oetmedical.co.uk/wp-content/uploads/2018/07/OET-Preparation-Support-Pack-c.pdf>
- O'Loughlin, K. (2008). Assessment at the Workplace. In N. H. Hornberger (Ed.), *Encyclopedia of Language and Education* (pp. 2214–2225). Springer US.  
[https://doi.org/10.1007/978-0-387-30424-3\\_168](https://doi.org/10.1007/978-0-387-30424-3_168)



- Olusiji, L., Salazar-Aguilar, P., Botto Beytía, A. M., & Boris, A. (2022). Needs Analysis of Chilean Students of Dentistry for Dental English Course Development. *World Journal of English Language*, 12, 162–171.  
<https://doi.org/10.5430/wjel.v12n8p162>
- O’Sullivan, B. (2011). *Language Testing—Theories and Practices* | B. O’Sullivan | Palgrave Macmillan. Palgrave Macmillan.  
<http://www.palgrave.com/la/book/9780230230620>
- O’Sullivan, B. (2012). Assessment Issues in Languages for Specific Purposes. *The Modern Language Journal*, 96, 71–88. <https://doi.org/10.1111/j.1540-4781.2012.01298.x>
- O’Sullivan, B. (2015). *Aptis Test Development Approach* (TR/2015/00; APTIS Technical Reports, p. 18). British Council.  
[https://www.britishcouncil.org/sites/default/files/tech\\_001\\_barry\\_osullivan\\_aptis\\_test\\_-\\_v5\\_0.pdf](https://www.britishcouncil.org/sites/default/files/tech_001_barry_osullivan_aptis_test_-_v5_0.pdf)
- Pardo-Ballester, C. (2020). AN ASSESSMENT USE ARGUMENT FOR SPANISH FOR THE PROFESSIONALS. *Proceedings of INTED2020 Conference*, 7172.  
<https://doi.org/10.21125/inted.2020.1893>
- Pearson Inc. (2020). *Pearson PTE English Test Centers & Fees*. PTE Academic.  
<https://pearsonpte.com/the-test/test-centers-and-fees/>
- Peña Dix, B., & de Mejía, A.-M. (2012). Policy Perspectives from Colombia. In *The Common European Framework of Reference The Globalisation of Language Education Policy*. MPG Books Group.
- Penk, C., Pöhlmann, C., & Roppelt, A. (2014). The role of test-taking motivation for students’ performance in low-stakes assessments: An investigation of school-

track-specific differences. *Large-Scale Assessments in Education*, 2(1), 5.

<https://doi.org/10.1186/s40536-014-0005-4>

Pépin, C. (2008). *Common European Framework of Reference for Languages: Learning, teaching, assessment and Canadian Language Benchmarks 200/Niveaux de compétence linguistique canadiens 2006: An overview of concepts, structures and applications*. Centre for Canadian Language Benchmarks.

Phillips, K. A., Brown, A. L., & Blankson, A. N. (2020). Validity, Issues of. In *The Wiley Encyclopedia of Personality and Individual Differences* (pp. 187–189). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118970843.ch102>

Porto, M. (2014). The Role and Status of English in Spanish-Speaking Argentina and Its Education System. *SAGE Open*, 4(1), 2158244013514059. <https://doi.org/10.1177/2158244013514059>

Powers, D. E., & Powers, A. (2015). The incremental contribution of TOEIC® Listening, Reading, Speaking, and Writing tests to predicting performance on real-life English language tasks. *Language Testing*, 32(2), 151–167. <https://doi.org/10.1177/0265532214551855>

Prensky, M. (2001). Digital Natives, Digital Immigrants Part 1. *On the Horizon*, 9(5), 1–6. <https://doi.org/10.1108/10748120110424816>

Presidencia de la República. (2012). *Puente Baluarte*. <http://calderon.presidencia.gob.mx/el-blog/puente-baluarte/>

Purpura, J. E. (2004). Validating questionnaires to examine personal factors in L2 test performance. In *Studies in Language Testing 18 European Language Testing in a Global Context*. Cambridge University Press.

- Qian, D. D. (2007). Assessing University Students: Searching for an English Language Exit Test. *RELC Journal*, 38(1), 18–37.  
<https://doi.org/10.1177/0033688206076156>
- Ramírez Romero, J. L. (2015). Consideraciones metodológicas y conceptuales relacionadas con la investigación sobre la enseñanza del inglés en las primarias públicas en México. In *La enseñanza del inglés en las primarias públicas mexicanas* (1era ed., pp. 1–22). Pearson.
- Reyes Fierro, M. del C. (2008). *Programa Universitario de Auto-Aprendizaje y Acreditación de Lengua Inglesa: Una Respuesta al Modelo Educativo de la UJED*. Universidad Juárez del Estado de Durango.
- Reyes Fierro, M. del C., Cano Vara, R., & Durán Howard, K. G. (2008). *Programa Universitario de Aprendizaje y Acreditación de Lengua Inglesa PUAALI*. Universidad Juárez del Estado de Durango.
- Reyes Solís, J. R. (2018, April 6). *Inversiones extranjeras en México 2013-2018*. El Universal. <https://www.eluniversal.com.mx/cartera/economia/inversiones-extranjeras-en-mexico-2013-2018>
- Richards, J. C., & Schmidt, R. (2010). *Longman Dictionary of Language Teaching and Applied Linguistics* (4th ed.). Pearson. <https://doi.org/10.4324/9781315833835>
- Richards, K. (2003). *Qualitative Inquiry in TESOL*. Palgrave Macmillan.
- Richerich, R. (1972). *A Model for the Definition of Language Needs of Adults Learning a Modern Language*. Council of Europe.  
<http://files.eric.ed.gov/fulltext/ED082544.pdf>
- Roever, C., & Pan, Y.-C. (2008). Test review: GEPT: General English Proficiency Test. *Language Testing*, 25(3), 403–408. <https://doi.org/10.1177/0265532208090159>

- Rohman, K. A. (2019). The Students Perceptions on a Test of English Proficiency as a Graduation Requirement. *LLT Journal: A Journal on Language and Language Teaching*, 22(2), Article 2. <https://doi.org/10.24071/llt.v22i2.2031>
- Rose, H., Mckinley, J., & Baffoe-Djan, J. (2019). *Data Collection Research Methods in Applied Linguistics* (1st ed.). Bloomsbury Academic.
- Saber es práctico. (2014, September 30). *Superficie de los estados de México*. Saber es práctico. <https://www.saberespractico.com/geografia/superficie-de-los-estados-de-mexico/>
- Sánchez, M. E., & Diez, I. (2014). *La enseñanza del inglés como política pública en América Latina: Estudio comparado*. Instituto de Fomento e Investigación Educativa, A. C.
- Sarudin, I., Mechraoui, A., Noor, Z. M., Ibrahim, E. H. E., Muhamad, A. J., & Malek, F. A. (2013). Stakeholders' Perceptions of an English for Occupational Purposes Course. *World Applied Sciences Journal*, 15.
- Sato, T., & Ikeda, N. (2015). Test-taker perception of what test items measure: A potential impact of face validity on student learning. *Language Testing in Asia*, 5(10), 1–16.
- Savignon, S. J. (2017). Communicative Competence. In *The TESOL Encyclopedia of English Language Teaching*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118784235.eelt0047>
- Schedl, M. (2010). *Background and Goals of the TOEIC® Listening and Reading Test Redesign Project*. 18.
- Schmidgall, J., Cid, J., Carter Grissom, E., & Li, L. (2021). Making the Case for the Quality and Use of a New Language Proficiency Assessment: Validity Argument

- for the Redesigned TOEIC Bridge® Tests. *ETS Research Report Series*, 2021(1), 1–22. <https://doi.org/10.1002/ets2.12335>
- Schmidgall, J. E. (2017). Articulating and Evaluating Validity Arguments for the TOEIC® Tests. *ETS Research Report Series*, 2017(1), 1–9. <https://doi.org/10.1002/ets2.12182>
- Schneider, G., & North, B. (2000). *Fremdsprachen können - was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit ; nationales Forschungsprogramm 33, Wirksamkeit unserer Bildungssysteme*. Rüegger.
- SEED. (2015, July). Entrega certificados Internacionales de Inglés. *Secretaría de Educación Del Estado de Durango*.
- SEP. (n.d.). *Certificación Nacional de Nivel de Idioma Antecedentes*. Secretaría de Educación Pública. <http://www.cenni.sep.gob.mx/pdf/antecedentes2.pdf>
- SEP. (2011). *Programa Nacional de Inglés en Educación Básica: Segunda Lengua: Inglés. Fundamentos curriculares. Preescolar. Primaria. Secundaria. Fase de expansión*. Secretaría de Educación Pública.
- SEP. (2017a). *Planes de Estudio de Referencia del Marco Curricular Común de la Educación Media Superior*. SEP.
- SEP. (2017b). *Programa Nacional de Inglés 2017—2018*. gov.mx. <http://www.gob.mx/sep/documentos/programa-nacional-de-ingles-2017-2018?state=published>
- Serafini, E. J., Lake, J. B., & Long, M. H. (2015). Needs analysis for specialized learner populations: Essential methodological improvements. *English for Specific Purposes*, 40, 11–26. <https://doi.org/10.1016/j.esp.2015.05.002>

- Shamsudin, S., Manan, A. Abd., & Husin, N. (2013). Introductory Engineering Corpus: A Needs Analysis Approach. *Procedia - Social and Behavioral Sciences*, 70, 1288–1294. <https://doi.org/10.1016/j.sbspro.2013.01.189>
- Shehadeh, A. (2012). *Task-Based Language Assessment: Components, Development, and Implementation*.
- Sireci, S. G. (2013). Agreeing on Validity Arguments. *Journal of Educational Measurement*, 50(1), 99–104.
- Stake, R. E. (1995). *The Art of Case Study Research*. SAGE.
- Stevens, B. (2005). What communication skills do employers want? Silicon Valley recruiters respond. *Journal of Employment Counseling*, 42(1), 2–9. <https://doi.org/10.1002/j.2161-1920.2005.tb00893.x>
- Stone, E., & Wylie, E. C. (2019). Building a Validity Argument While Developing and Using an Assessment: A Concurrent Approach for the Winsight® Summative Assessment. *ETS Research Report Series*, 2019(1), 1–20. <https://doi.org/10.1002/ets2.12261>
- Sundre, D. L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology*, 29(1), 6–26. [https://doi.org/10.1016/S0361-476X\(02\)00063-2](https://doi.org/10.1016/S0361-476X(02)00063-2)
- Suzumura, N. (2022). Content Analysis of Test Taker Responses on an AP Japanese Computer-Simulated Conversation Test: A Mixed Methods Approach for A Validity Argument. *Language Assessment Quarterly*, 19(5), 480–502. <https://doi.org/10.1080/15434303.2022.2130326>

- Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
- Tarantino, M. (1988). Italian in-field EST users self-assess their macro- and micro-level needs: A case study. *English for Specific Purposes*, 7(1), 33–53.  
[https://doi.org/10.1016/0889-4906\(88\)90005-1](https://doi.org/10.1016/0889-4906(88)90005-1)
- The 5 Main Economic Activities of Durango*. (n.d.). Life Persona. Retrieved 28 January 2019, from <https://www.lifepersona.com/the-5-main-economic-activities-of-durango>
- The Mazatlan Post. (2018). Durango Mexico, Land of Films, Scorpions and Pancho Villa. *The Mazatlán Post*. <https://themazatlanpost.com/2018/07/16/durango-mexico-land-of-films-scorpions-and-pancho-villa/>
- Tickoo, M. L. (1988). Michael West in India: A centenary salute. *ELT Journal*, 42(4).
- Tinsley, T. (2012). *Briefing note re Asset Languages*. Alcantara Communications.
- TOEFL. (2020). TOEFL® Research Insight Series, Volume 6: TOEFL Program History. *ETS*, 6, 10.
- Toulmin, S. E. (2003). *The Uses of Argument*. Cambridge University Press.
- Tratnik, A. (2008). Key Issues in Testing English for Specific Purposes. *SDUTSJ*, 4(1), 3–13.
- Trim, J. (2007). Modern Languages in the Council of Europe 1954-1997. *Council of Europe Language Policy Division*. [www.coe.int/lang](http://www.coe.int/lang)
- UAA. (2012). *Criterios de Operación del Programa Institucional de Lenguas Extranjeras*. UAA.

- UAEH. (2013). *Programa Institucional de Lenguas*. Universidad Autónoma del Estado de Hidalgo.
- UCLES. (n.d.). *Encuentre un centro de exámenes autorizado | Cambridge English*. Retrieved 22 February 2019, from <https://www.cambridgeenglish.org/latinamerica/find-a-centre/find-an-exam-centre/>
- UCLES. (2017). *IELTS Information to Candidates Introducing IELTS to test takers*. UCLES.
- UCol. (2016). *Programa Universitario de Inglés*. Presentación para planteles, Colima.
- UdG. (2011). *Política Institucional de Idiomas Universidad de Guadalajara*.
- UJED. (2015, February 23). *Escuela de Lenguas capacitará a prestadores de servicio para 'Durango bilingüe'*.
- UJED Escuela de Lenguas. (n.d.). *Preguntas frecuentes*. Escuela de Lenguas UJED. Retrieved 2 February 2019, from <http://escueladelenguas.ujed.mx/puaali/preguntas-frecuentes/>
- UNAM. (2018, November 26). *Noticias: Exámenes de Inglés—Requisito de Idioma para Egreso. UNAM, México*. <http://www.fciencias.unam.mx/comunicacion/detalle/2165>
- Underhill, N. (1987). *Testing Spoken Language: A Handbook of Oral Testing Techniques*. Cambridge University Press.
- Universidad Anáhuac. (2019). *Cubrir el requisito de inglés como segundo idioma | Centro de Lenguas | Universidad Anáhuac Querétaro*.



<https://www.anahuac.mx/queretaro/cubrir-el-requisito-de-ingles-como-segundo-idioma-centro-de-lenguas>

Universidad de Colima. (2016). *Programa Universitario de Inglés*.

Universidad de Córdoba. (2018). *Reglamento para la Acreditación de la Competencia Lingüística para la obtención de los Títulos de Grado*. Boletín Oficial de la Universidad de Córdoba.

Ushioda, E., & Dörnyei, Z. (2012). Motivation. In *The Routledge Handbook of second language acquisition* (pp. 369–409). Routledge.

USTR. (n.d.). *North American Free Trade Agreement (NAFTA)*. Retrieved 23 February 2019, from /trade-agreements/free-trade-agreements/north-american-free-trade-agreement-nafta

Velázquez, P. (2015). El inglés y la oportunidad económica. In *Sorry -El aprendizaje del inglés en México*. Mexicanos Primero.

Vidal, K. (2005). El inglés para fines específicos y la universidad: Un estudio empírico exploratorio. *Dialnet, Extra*(1), 213–231.

Wall, D., & Horák, T. (2011). The Impact Of Changes In The Toefl ®Exam On Teaching In A Sample Of Countries In Europe: Phase 3, The Role Of The Coursebook Phase 4, Describing Change. *ETS Research Report Series, 2011*(2), i–181. <https://doi.org/10.1002/j.2333-8504.2011.tb02277.x>

Wang, H., Choi, I., Schmidgall, J., & Bachman, L. F. (2012). Review of Pearson Test of English Academic: Building an assessment use argument. *Language Testing, 29*(4), 603–619. <https://doi.org/10.1177/0265532212448619>

- Watt, D. L. E., Lake, D., Cabronch, T., & Leonard, K. (2003). *Assessing the English Language Proficiency of International Medical Graduates in their Integration Into Canada's Physician Supply*. Health Canada.  
file:///C:/Users/Cubiculo\_00/Downloads/Assessing-ELP-of-IMGs\_Final\_2003%20(1).pdf
- Weger, H. D. (2013). Examining English Language Learning Motivation of Adult International Learners Studying Abroad in the US. *RELC Journal*, 44(1), 87–101. <https://doi.org/10.1177/0033688212473272>
- Wei, H., Mislevy, R. J., & Kanal, D. (2008a). *An introduction to Design Patterns in Language Assessment*. SRI International.  
[https://padi.sri.com/downloads/TR18\\_Language\\_Design\\_Patterns.pdf](https://padi.sri.com/downloads/TR18_Language_Design_Patterns.pdf)
- Wei, H., Mislevy, R. J., & Kanal, D. (2008b). *An introduction to Design Patterns in Language Assessment*. SRI International.  
[https://padi.sri.com/downloads/TR18\\_Language\\_Design\\_Patterns.pdf](https://padi.sri.com/downloads/TR18_Language_Design_Patterns.pdf)
- Weir, C. J. (2005). *Language Testing and Validation*. Palgrave Macmillan.  
<http://www.palgrave.com/page/detail/language-testing-and-validation-cyril-j-weir%2F%3FK%3D9781403911889>
- Weir, C. J. (2013). Measured constructs: A history of Cambridge English language examinations 1913-2012. *Cambridge English Research Notes*, 51, 2–10.
- West, R. (1994). Needs analysis in language teaching. *Language Teaching*, 27(1), 1–19. <https://doi.org/10.1017/S0261444800007527>
- Wigglesworth, G. (2008). Task and Performance Based Assessment. In N. H. Hornberger (Ed.), *Encyclopedia of Language and Education* (pp. 2251–2262). Springer US. [https://doi.org/10.1007/978-0-387-30424-3\\_171](https://doi.org/10.1007/978-0-387-30424-3_171)

- Woodford, P. E. (1982). *An Introduction to TOEIC: The Initial Validity Study*. Educational Testing Service.
- Woodrow, L. (2017). *Introducing Course Design in English for Specific Purposes* (1 edition). Routledge.
- Wu, R. Y., Yeh, H., Dunlea, J., & Spiby, R. (2016). APTIS–GEPT Test Comparison Study: Looking At Two Tests From Multi-Perspectives Using The Socio-Cognitive Model. *British Council*, 50.
- Xi, X. (2008). Methods of Test Validation. In N. H. Hornberger (Ed.), *Encyclopedia of Language and Education* (pp. 2316–2335). Springer US.  
[https://doi.org/10.1007/978-0-387-30424-3\\_175](https://doi.org/10.1007/978-0-387-30424-3_175)
- Xi, X., & Davies, L. (2016). Quality Factors in Language Assessment. In *Handbook of Second Language Assessment* (Vol. 12). De Gruyter Mouton.
- Xu, J., Brenchley, M., Jones, E., Pinnington, A., & Benjamin, T. (2020). *Linguaskill: Building a validity argument for the Speaking test*.
- Yan, X., & Staples, S. (2019). Fitting MD analysis in an argument-based validity framework for writing assessment: Explanation and generalization inferences for the ECPE. 2, 37, 1–26. <https://doi.org/10.1177/0265532219876226>
- Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing*, 32(2), 199–225.  
<https://doi.org/10.1177/0265532214557113>
- Zela, E. (2018). CEF Professional Profile as an Instrument in Language Needs Assessment for Business Companies. *Journal of Educational and Social Research*, 8(2).  
<https://www.richtmann.org/journal/index.php/jesr/article/view/10238>

## Appendix I Tests accepted to fulfil the school-exit requirement in non-native English-speaking universities.

University	Explicit purpose or level established as a university exit requirement	Tests accepted as a university exit requirement	Reference
<b>México</b>			
University of Colima (UCol)	Independent user B2- of the CEFR	FCE, CAE, CPE, BEC, TOEFL, iBT TOEFL	(Universidad de Colima, 2016)
Autonomous University of Aguascalientes (UAA)	A2	FCE, TOEFL, TOEIC and other locally designed tests	(UAA, 2012)
Juárez University of the State of Durango (UJED)	B1 of the CEFR	ELASH 1, ELASH 2, KET, PET, TOEFL ITP, TOEFL iBT or any other included in the CENNI or the locally designed test	(UJED Escuela de Lenguas, n.d.)
Autonomous University of the State of Hidalgo (UAEH)	...development of graduates that are professionally competitive in a globalized world (pg. 14)	PET, FCE, TOEFL, iBT TOEFL, IELTS,	(UAEH, 2013)
Anahuac University	Not stated	TOEFL iBT 89 points, IELTS Academic 6.5, FCE A or B, CAE A, B or C	(Universidad Anáhuac, 2019)
<b>CEFR</b>			
National Autonomous University of México	B2 reading comp B1 listening comp, oral expression and writing	No data on tests accepted is published; hence, it may be an in-house test	(UNAM, 2018)
<b>Europe</b>			
University of Córdoba, Spain	B1 of the CEFR	APTIS, IELTS, FCE, PTE Academic, iBT TOEFL, and TOEIC among other international tests	(Universidad de Córdoba, 2018, pp. 9–10)

Appendix Table 1 Non-native English-speaking universities and the tests accepted to fulfil the university requirement

## Appendix II Test Takers Consent Form

The text below (information and Consent form) was presented on a separate slip of paper stapled in front of the questionnaire to let participants know the questionnaire was only one sheet of paper. It was considered relevant in determining their willingness to cooperate and answer the questionnaire. It was perceived that if they were told that there was more than one sheet of paper, they may have decided not to participate or did so carelessly.

Dear student:

This questionnaire is part of a PhD research Project. It would be greatly appreciated if you took a moment to answer the following questions. Your answers will help understand students' perceptions of the ECGE you just took and may impact the future content and task types of the ECGE. The questionnaire is anonymous, and the respondent cannot be identified. Your answers will have no impact on the grades you are awarded in the ECGE you just took. Thank you for your collaboration.

# Appendix III Questionnaire 1

## B1 CERTIFICATION TEST CONTENT AND EXPECTATIONS UJED STUDENTS' PERCEPTIONS ON THE TEST TAKEN

		Career:			
		Disagree 1	Quite disagree 2	Quite agree 3	Agree 4
<b>According to MY opinion the B1 English language Certification test I just took</b>					
1.	was easy to answer				
2.	included activities (in written form) that are similar to what I will need to do in my professional life				
3.	tested the language abilities that are required to study a master's degree in an English speaking country				
<b>According to MY opinion the B1 English language Certification test I just took tested my ability to understand</b>					
4.	the main points of clear standard speech on familiar matters regularly encountered in work places				
5.	the main points of clear standard speech on familiar matters regularly encountered in school				
6.	the main points of clear standard speech on familiar matters regularly encountered in leisure activities				
7.	the main point of many radio or TV programs on current affairs or topics of personal interest				
8.	simple technical information, such as operating instructions for everyday equipment				
9.	the description of events, feelings and wishes in personal letters				
<b>According to MY opinion the B1 English language Certification test I just took tested my ability to</b>					
10.	deal with common situations likely to arise whilst traveling in an area where English is spoken				
11.	enter unprepared into conversation on topics that are familiar of personal interest				
12.	enter unprepared into conversation on topics that are familiar of professional interest				
13.	enter unprepared into conversation on topics that are familiar and are related to everyday life				
14.	describe experiences and events, my dreams, hopes or ambitions				
15.	briefly give reasons and explanations for opinions and plans				
16.	narrate a story in oral or relate a plot of a book or film and describe my reactions				
17.	write a simple connected text on topics which are familiar or of personal interest				
18.	write personal letters describing experiences and impressions				
<b>According to MY opinion the English classes I took before taking the B1 English language Certification test</b>					
19.	in general helped me prepare for the test				
20.	provided opportunities to practice the type of activities included in the test				
21.	included activities that are similar to the type of activities I may be needing to perform in English in my future job				
<b>According to MY opinion the B1 English language Certification test I just took NEEDS</b>					
22.	instructions in Spanish				
23.	more open ended questions				
24.	to be multiple-choice only				
25.	more time to be completed				
26.	to be longer				
27.	to measure				
28.	English related				
29.	to				
30.	to				
31.	to				
32.	to				
33.	to				
27.	the world of entertainment (music, movies, videogames)				
28.	travelling for business				
29.	doing a post graduate in an English speaking country				
30.	talking about the things that I like to do, eat, etc.				
31.	the world of work (useful for my future professional life)				
32.	talking about my family				
33.	travelling for pleasure				
34.	I think a job related to my disciplinary area requires for me to	Very rarely 1	With very little frequency 2	Rather frequently 3	Frequently 4
a)	speak in English				
b)	understand what is said to me in English				
c)	write in English				
d)	read in English				
35.	Where did you study English? (please tick as necessary)	School of Languages Language Center	Harmon Hall	FADER Language Center	
		FECA Center of Languages	Inglés Individual	ITD Language Center	
		Centro de Idiomas del Guadiana	Golden English	Otro (please specify)	
36.	School:	Discipline:			
Is there anything that you would like to comment relating the exam you just took or the type of exam you think you should be taking? (Please continue in the back.)					

Thank you for your participation ☺

Appendix Figure 1 Questionnaire 1 (Q1) for test takers

## Appendix IV Semi-structured interview guide

Interview applied to students at the end of the language certification session. Q1 was first applied. As they turned in the answered questionnaire, they were asked if they would agree to a short anonymous interview.

Short, anonymous, semi-structured interview  
applied at the end of the language certification session

Script for the semi-structured interview

R: May I ask you a couple of questions? It's about the same topic, about the test you just took.

R: Is it OK if I record it? It is only audio, and I won't ask for your name, only the school you come from.

R: If you were able to choose, if you could decide which language domain the test would focus on, which would you choose? One is the social domain, which will be useful for you to talk about yourself, the things you like, or to travel. Another domain would be the academic domain; this would be useful to attend a foreign university where the language of instruction is English. For a master's degree, for example, understanding academic articles or lectures would be helpful. The other domain would be the work domain, which would be useful in your future job as a (saying the name of their discipline, e.g. chemist), you may need to understand specific material or instructions of equipment, attend conferences or talk to colleagues in English in an international company or at a conference. Which one would you prefer?

R: Why?

Further questions were not very frequent, mainly used to clarify or ask them to elaborate on their response.





## Appendix VI Questionnaire 2a

Dear student:

This questionnaire is part of a PhD research Project. It would be greatly appreciated if you took a moment to answer the following questions. Your answers will help us understand English's importance in the future professional life of UJED graduates. By writing down your name in the line below, you consent to use the information you provide for research purposes. Confidentiality of participants will be kept at all times and the information gathered will be stored in a password-protected computer and used responsibly and discretely for research purposes only. Participants will not be affected in any way. Thank you for your collaboration.

Name: _____		Date: _____	
School: _____		Degree: _____	
Semester: _____		Age: _____	
Gender: F M	E-mail: _____		

<b>Information related to English and job-related activities</b>					
(Select the option that best represents your current or near future needs in terms of frequency of use from 0 to 100%.)					
	<b>As a graduate, I think I will need to be able to:</b>	<b>1 No, almost never 0-25%</b>	<b>2 Some times 26- 50%</b>	<b>3 Frequently 51-75%</b>	<b>4 Very frequently 76-100%</b>
1.	offer advice to clients within their own job area on simple matters (e.g. "This model works faster, but is more expensive").				
2.	request routine services related to own work area, such as a document or package to be sent (e.g. Can you send/type/call...)				
3.	ask questions of a fact-finding nature, for example, establishing what is wrong with a machine/process/situation, establishing where something or somebody is, the frequency of an event, etc.				
4.	take a routine order, provided that this is restricted to matters such as quantity, delivery date, etc. (e.g. can understand and write down reference codes containing letters and numbers)				
5.	receive phone calls (e.g. take messages/make notes: who called, what time, etc.)				
6.	understand spoken directions, such as following safety procedures at work (e.g. Don't....Be careful...Watch...)				
7.	follow a simple presentation/demonstration and understand explanations with reference to a product or topic within your own area of expertise.				
8.	ask for clarification when something is not clear (e.g. "Can you repeat that please?", "Is it .. or ...?", "So, the....is/has/needs...")				
9.	participate in job-related meetings and seminars (e.g. making suggestions or asking				

	questions: "Why don't we..." "Can we..." "Is it possible to...").				
10.	express opinions in simple terms, for example, "I don't agree", "I think..." provided the question/issue has been put clearly and simply.				
11.	provide information or opinions, such as when planning a company event (e.g. "We could..." "There is..." "I think...").				
12.	greet a visitor and engage in a limited conversation for a short time, for example, enquiring about a visitor's journey, hotel, etc. (e.g. "How was your ...", "Did you like...?", "Is/Was your ....?").				
13.	provide instructions and directions, such as giving clients clear directions to the business/company. (e.g. "Take a taxi to. ...", "Go straight...", "walk two blocks, then turn...").				
14.	deal with predictable requests from a visitor, for example, "Can you arrange a taxi to the airport?", "Can I use a projector in the presentation?", "Where can I....?"				
15.	make phone calls requesting information or services (e.g. "Do you have...", "I would like...", "How much/many...?")				
16.	make job-related phone calls, leave a phone message, place a service or materials order (e.g. "I need...", "Can I speak to...", "I'm looking for...", "I")				
17.	recognize and understand at least partially the general meaning of a non-routine letter within own work area.				
18.	understand straightforward factual information on routine job-related faxes, memos, emails, etc.				
19.	obtain basic relevant information (product specifications, professional or commercial leaflets, advertisements, internet pages, etc.)				
20.	recognize and understand at least partially the general meaning of a routine letter within own work area.				
21.	understand general written instructions (for example, installation, functioning or maintenance manuals)				
22.	use several pieces of information, such as a work schedule, to see who should be working on several reports on a process to identify the source of a problem.				
23.	write short reports, faxes, emails, and memos on basic everyday job-related matters (e.g. simple factual information related to a recent event, the time and place of an activity, people attending, and topics to be addressed).				
24.	write requests for goods, services, etc., on a range of routine matters, but MAY need to get these checked.				
25.	make notes on routine matters, such as taking/placing orders.				
26.	fill out a form, such as an incident/accident report form or a medical record.				

27.	report events or incidents involving products, machines or personnel (workers, clients), i.e. a workplace accident, a complaint, or a failure in a process.				
28.	make notes for personal use on non-routine aspects such as a client's requirement and basic characteristics of a new product or service.				
29.	other (please specify) _____				

Adapted from ALTE [http://www.apinex.org/mcre/mre\\_cvc\\_13.pdf](http://www.apinex.org/mcre/mre_cvc_13.pdf)

**MARK THE OPTION THAT BEST REFLECTS YOUR OPINION**

The activities in English above mentioned are, in general _____ considering the activities required in my business/company.			
a. very complicated	b. more or less appropriate	c. appropriate	d. very simple
I agree to be contacted _____ if more information is required (circle all the options that apply).			
a. on the phone	b. personally	c. electronically (email/Skype)	

If you feel you have been put at risk or have any concerns related to this study, you may contact the Administrator of the Ethics Committee, Humanities, University of Southampton, SO17 1BJ, UK. Phone: +44 (0)23 8059 4663, Email: [I.Ghose@soton.ac.uk](mailto:I.Ghose@soton.ac.uk)

Thank you for your participation

Appendix Figure 3 Questionnaire 2a (Q2a) for students

## Appendix VII Questionnaire 2b

### ENGLISH REQUIREMENT ON GRADUATES

Dear graduate:

This questionnaire is part of a PhD research project. It would be greatly appreciated if you took a moment to answer the following questions. The purpose of this research is to find out the importance of English in the life of professionals in Durango. By writing your name and information in the following lines, you consent to the information being used for the abovementioned purposes. Confidentiality of participants will be kept at all times and the information gathered will be stored in a password-protected computer and used responsibly and discretely for research purposes only. Participants will not be affected in any way. You also acknowledge that your participation is voluntary, and you can withdraw at any time without any of your rights being affected. Thank you for your collaboration.

Name:		Place of work:	
Date:	School(latest studies):	Gender: F M	
Career:		Position or activity:	
Year of graduation of latest studies:	Age: 22 to 32 33-43 44 or more		

Information related to English and job-related activities					
(Select the option that best represents your current or near future needs.)					
As a graduate, I think I will need to be able to:		No, almost never 0-25%	Very rarely 26- 50%	Some times 51-75%	Frequently 76-100%
1.	offer advice to clients within their own job area on simple matters (e.g. "This model works faster, but is more expensive").				
2.	request routine services related to own work area, such as a document or package to be sent (e.g. Can you send/type/call...)				
3.	ask questions of a fact-finding nature, for example, establishing what is wrong with a machine/process/ situation, establishing where something or somebody is, the frequency of an event, etc.				
4.	take a routine order, provided that this is restricted to matters such as quantity, delivery date, etc. (e.g. can understand and write down reference codes containing letters and numbers)				
5.	receive phone calls (e.g. take messages/make notes: who called, what time, etc.)				
6.	understand spoken directions, such as following safety procedures at work (e.g. Don't...Be careful...Watch...)				
7.	follow a simple presentation/demonstration and understand explanations with reference to a product or topic within your own area of expertise.				
8.	ask for clarification when something is not clear (e.g. "Can you repeat that please?", "Is it .. or ...?", "So, the....is/has/needs...")				
9.	participate in job-related meetings and seminars (e.g. making suggestions or asking questions: "Why don't we...", "Can we...", "Is it possible to...").				
10.	express opinions in simple terms, for example, "I don't agree," "I think..." provided the question/issue has been put clearly and simply.				

11.	provide information or opinions, such as when planning a company event (e.g. "We could... ", "There is...", "I think...").				
12.	greet a visitor and engage in a limited conversation for a short time, for example, enquiring about a visitor's journey, hotel, etc. (e.g. "How was your ...", "Did you like...?", "Is/Was your .... ?").				
13.	provide instructions and directions, such as giving clients clear directions to the business/company. (e.g. "Take a taxi to. ...", "Go straight...", "walk two blocks, then turn...").				
14.	deal with predictable requests from a visitor, for example, "Can you arrange a taxi to the airport?", "Can I use a projector in the presentation?", "Where can I....?"				
15.	make phone calls requesting information or services (e.g. "Do you have...", "I would like...", "How much/many...?")				
16.	make job-related phone calls, leave a phone message, place a service or materials order (e.g. "I need...", "Can I speak to...", "I'm looking for...", "I")				
17.	recognize and understand at least partially the general meaning of a non-routine letter within own work area.				
18.	understand straightforward factual information on routine job-related faxes, memos, emails, etc.				
19.	obtain basic relevant information (product specifications, professional or commercial leaflets, advertisements, internet pages, etc.)				
20.	recognize and understand at least partially the general meaning of a routine letter within own work area.				
21.	understand general written instructions (for example, installation, functioning or maintenance manuals)				
22.	use several pieces of information, such as a work schedule, to see who should be working on several reports on a process to identify the source of a problem.				
23.	write short reports, faxes, emails, and memos on basic everyday job-related matters (e.g. simple factual information related to a recent event, the time and place of an activity, people attending, and topics to be addressed).				
24.	write requests for goods, services, etc., on a range of routine matters, but MAY need to get these checked.				
25.	make notes on routine matters, such as taking/placing orders.				
26.	fill out a form, such as an incident/accident report form or a medical record.				
27.	report events or incidents involving products, machines or personnel (workers, clients), i.e. a workplace accident, a complaint, or a failure in a process.				
28.	make notes for personal use on non-routine aspects such as clients' requirements and basic characteristics of a new product or service.				
29.	other ( please specify) _____				

Adapted from ALTE [http://www.apinex.org/mcre/mre\\_cvc\\_13.pdf](http://www.apinex.org/mcre/mre_cvc_13.pdf)

**MARK THE OPTION THAT BEST REFLECTS YOUR OPINION**

The activities in English above mentioned are, in general _____ considering the activities required in my business/company.			
a. very complicated	b. more or less appropriate	c. appropriate	d. very simple
I agree to be contacted _____ if more information is required (circle all the options that apply).			
a. on the phone	b. personally	c. electronically (email/Skype)	

If you feel you have been put at risk or have any concerns related to this study, you may contact the Administrator of the Ethics Committee, Humanities, University of Southampton, SO17 1BJ, UK. Phone: +44 (0)23 8059 4663, Email: [I.Ghose@soton.ac.uk](mailto:I.Ghose@soton.ac.uk).

Thank you for your participation

Appendix Figure 4 Questionnaire 2b (Q2b) for domain insiders

## Appendix VIII Transcription conventions

(...) indicates part of the conversation has been omitted because it was not considered relevant

*“italics”* indicates it is the translation of a participant’s comment.

... indicates a short pause, as when people are thinking about what to say next

(not italics) indicates information placed by the researcher for clarification purposes

## Appendix IX Short semi-structured interview example in Spanish and English version

Two female students from the human communication therapy study programme HCT9 and HCT 10

Ex.No.	Translation to English	Part.	Original in Spanish
1.	What School are you from?	Int	¿De qué escuela eres?
2.	<i>From the School of Psychology and Human Communication Therapy</i>	HCT 9	De la facultad de psicología y terapia de la comunicación humana
3.	<i>From the same School</i>	HCT 10	De la misma facultad
4.	From which study programme?	Int	¿De qué especialidad?
5.	<i>From clinical environment therapy</i>	HCT 9	De terapia de ambientes clínicos
6.	And you?	Int	¿Y tú?
7.	<i>The same, from therapy, from the seventh</i>	HCT 10	Igual, de terapia, de séptimo
8.	If you were able to choose, if you could decide which language domain the test would focus on, which would you choose? One is the social domain, which will be useful (see Appendix IIa for the complete introduction of the interview in English)	Int	Si pudieras escoger, si tu pudieras decidir el ámbito en el que se enfocar el examen, ¿cuál escogerías? Uno sería el ámbito social que te sería útil para hablar de ti, las cosas que te gustan o para viajar. Otro ámbito sería el académico, que te fuera útil para estudiar en una...
9.	<i>Specific situations in the work domain</i>	HCT 10	Situaciones específicas en el área de trabajo
10.	<i>Yes, in the work domain</i>	HCT	Si, en el área de trabajo

11.	Why?	Int	¿Por qué?
12.	<i>Because it would be more useful for us and it would meet the purpose of having to take this test to graduate from this program</i>	HCT 10	Porque nos serviría más y se cumpliría el objetivo de tener que hacer este examen para egresar de la carrera
13.	<i>Besides, we would have the opportunity to work, for example, abroad, in the United States, and that's it...that is, to know more to be able to give therapy to bilingual people, or "gringos" and things like that.</i>	HCT 9	Aparte que tendríamos la oportunidad de trabajar por ejemplo en el extranjero, en Estados Unidos y ya... o sea saber más para poder dar terapias a personas bilingües o gringas y así
14.	Mmhmh, What do you mean that the purpose of having this test as a graduation requirement would be fulfilled?	Int	Mmhmh ¿A qué te refieres con que se cumpliría el objetivo de que se tenga este examen como requisito de egreso?
15.	<i>That, well, the purpose of finishing your study program knowing English is that you could face hum...situations like going abroad and getting a job already knowing English related to your study program.</i>	HCT 10	A que, bueno, pues el objetivo de salir de tu carrera sabiendo inglés es que puedas presentarte este... ante las situaciones de irte al extra



## Appendix X Statistics Q1

### Descriptive statistics

#### Mean and standard deviation comparison of results among disciplinary groups

##### Section 1

##### Study or work-related language, items 2 to 5

		N	Mean	Std. Deviation
TTQ2 The Exit Test included activities (in written form) that are similar to what I will need to do in my professional life	1 Unknown discipline	300	2,78	1,026
	2 Social sciences	380	2,82	1,070
	3 Phys & Nat Scs	30	2,83	1,020
	4 Health sciences	176	2,69	1,046
	Total	886	2,78	1,048
TTQ4 The Exit test I just took tested my ability to understand the main points of clear standard speech on familiar matters regularly encountered in workplaces	1 Unknown discipline	300	2,64	,963
	2 Social sciences	380	2,54	1,005
	3 Phys & Nat Scs	30	2,73	1,081
	4 Health sciences	176	2,51	1,080
	Total	886	2,57	1,009
TTQ12 The Exit Test I just took tested my ability to enter unprepared into conversation on topics that are familiar or of professional interest	1 Unknown discipline	300	2,80	0,950
	2 Social sciences	380	2,84	0,985
	3 Phys & Nat Scs	30	2,80	0,761
	4 Health sciences	176	2,81	0,972
	Total	886	2,82	0,963
TTQ3 The Exit test tested the language abilities that are required to study a master's degree in an English-speaking country	1 Unknown discipline	300	2,80	,989
	2 Social sciences	380	2,80	1,048
	3 Phys & Nat Scs	30	3,07	1,015

Appendix X

	4 Health sciences	176	2,78	1,008
	Total	886	2,81	1,019
TTQ5 The Exit Test I just took tested my ability to understand the main points of clear standard speech on familiar matters regularly encountered in school	1 Unknown discipline	300	2,69	,958
	2 Social sciences	379	2,63	,972
	3 Phys & Nat Scs	30	2,60	,968
	4 Health sciences	176	2,55	,996
	Total	885	2,63	,971

Appendix Table 2 Mean comparison among groups: items 2 to 5

**Receptive skills: items 6 to 9**

		N	Mean	Std. Deviation
<b>listening</b>				
TTQ6 The Exit Test I just took tested my ability to understand the main points of clear standard speech on familiar matters regularly encountered in leisure activities	1 Unknown discipline	300	2,73	1,023
	2 Social sciences	380	2,92	,971
	3 Phys & Nat Scs	30	2,93	1,202
	4 Health sciences	176	2,81	1,018
	Total	886	2,83	1,009
TTQ7 The Exit Test I just took tested my ability to understand the main point of many radio or TV programs on current affairs or topics of personal interest	1 Unknown discipline	300	2,76	1,039
	2 Social sciences	379	2,83	,985
	3 Phys & Nat Scs	30	2,87	1,106
	4 Health sciences	175	2,67	1,052
	Total	884	2,77	1,021
<b>reading</b>				
	1 Unknown discipline	300	2,71	,987

## Appendix X

TTQ8 The Exit Test I just took tested my ability to understand simple technical information, such as operating instructions for everyday equipment	2 Social sciences	379	2,72	1,050
	3 Phys & Nat Scs	30	2,80	1,031
	4 Health sciences	176	2,69	,948
	Total	885	2,71	1,007
	<hr/>			
TTQ9 The Exit Test I just took tested my ability to understand the description of events, feelings and wishes in personal letters	1 Unknown discipline	300	2,78	,981
	2 Social sciences	380	2,81	1,006
	3 Phys & Nat Scs	30	2,80	,961
	4 Health sciences	176	2,75	,977
	Total	886	2,79	,989

Appendix Table 3 Mean comparison among groups: items 6 to 9

**Productive skills: items 10 to 18**

		N	Mean	Std. Deviation
<b>speaking</b>				
TTQ10 The Exit Test I just took tested my ability to deal with common situations likely to arise whilst travelling in an area where English is spoken.	1 Unknown discipline	300	2,89	,919
	2 Social sciences	380	3,02	,907
	3 Phys & Nat Scs	30	3,03	,890
	4 Health sciences	176	2,88	,972
	Total	886	2,95	,925
<hr/>				
TTQ11 The Exit Test I just took tested my ability to enter unprepared into conversation on topics that are familiar or of personal interest.	1 Unknown discipline	300	2,94	,910
	2 Social sciences	380	3,11	,902
	3 Phys & Nat Scs	30	3,17	,834
	4 Health sciences	176	3,06	,867

	Total	886	3,04	,898
TTQ13 The Exit Test I just took tested my ability to enter unprepared into conversation on topics that are familiar and related to everyday life	1 Unknown discipline	300	2,96	,876
	2 Social sciences	380	3,09	,962
	3 Phys & Nat Scs	30	3,10	,759
	4 Health sciences	176	3,01	,910
	Total	886	3,03	,918
TTQ14 The Exit Test I just took tested my ability to describe experiences and events, my dreams, hopes or ambitious	1 Unknown discipline	300	2,92	,978
	2 Social sciences	380	3,07	,936
	3 Phys & Nat Scs	30	3,10	,960
	4 Health sciences	176	2,96	1,016
	Total	886	3,00	,968
TTQ15 The Exit Test I just took tested my ability to briefly give reasons and explanations for opinions and plans	1 Unknown discipline	300	2,89	,969
	2 Social sciences	380	3,03	1,014
	3 Phys & Nat Scs	30	3,20	,961
	4 Health sciences	176	2,93	,944
	Total	886	2,97	,985
TTQ16 The Exit Test I just took tested my ability to narrate a story orally or relate a plot of a book or film and describe my reactions	1 Unknown discipline	300	2,78	,932
	2 Social sciences	380	2,92	1,008
	3 Phys & Nat Scs	30	3,07	,828
	4 Health sciences	176	2,92	,988
	Total	886	2,88	,974
<b>writing</b>				
TTQ17 The Exit Test I just took tested my ability to write a simple connected text on topics that are familiar or of personal interest	1 Unknown discipline	300	2,88	,973
	2 Social sciences	380	3,06	,945
	3 Phys & Nat Scs	30	3,20	,925

	4 Health sciences	176	2,97	,976
	<b>Total</b>	<b>886</b>	<b>2,99</b>	<b>,963</b>
TTQ18 The Exit Test I just took tested my ability to write personal letters describing experiences and impressions	1 Unknown discipline	300	2,76	1,065
	2 Social sciences	380	2,97	1,004
	3 Phys & Nat Scs	30	3,03	,765
	4 Health sciences	176	2,88	,978
	<b>Total</b>	<b>886</b>	<b>2,88</b>	<b>1,016</b>

Appendix Table 4 Mean comparison among groups: items 10 to 18

## Section 2

What test takers are interested in and the importance of each skill in their future job

		N	Mean	Std. Deviation
TTQ27 The Exit Test I just took NEEDS to measure English related to the world of entertainment (music, movies, videogames)	1 Unknown discipline	300	3,05	,982
	2 Social sciences	379	3,11	,977
	3 Phys & Nat Scs	30	3,27	,944
	4 Health sciences	176	2,93	,992
	<b>Total</b>	<b>885</b>	<b>3,06</b>	<b>,982</b>
TTQ28 The Exit Test I just took NEEDS to measure English related to travelling for business	1 Unknown discipline	300	2,88	1,028
	2 Social sciences	380	3,01	1,016
	3 Phys & Nat Scs	30	3,07	,980
	4 Health sciences	176	2,84	1,003
	<b>Total</b>	<b>886</b>	<b>2,93</b>	<b>1,017</b>
TTQ29 The Exit Test I just took NEEDS to measure English	1 Unknown discipline	300	2,96	1,050
	2 Social sciences	380	3,08	1,019

Appendix X

related to doing a postgraduate in an English-speaking country	3 Phys & Nat Scs	30	3,07	1,015
	4 Health sciences	176	2,94	,939
	<b>Total</b>	<b>886</b>	<b>3,01</b>	<b>1,015</b>
	<hr/>			
TTQ30 The Exit Test I just took NEEDS to measure English related to talking about the things that I like to do, eat, etc.	1 Unknown discipline	300	3,22	,907
	2 Social sciences	380	3,25	,930
	3 Phys & Nat Scs	30	3,23	1,073
	4 Health sciences	176	3,15	,922
	<b>Total</b>	<b>886</b>	<b>3,22</b>	<b>,925</b>
<hr/>				
TTQ31 The Exit Test I just took NEEDS to measure English related to the world of work (useful for my future professional life)	1 Unknown discipline	300	3,10	,978
	2 Social sciences	380	3,20	1,001
	3 Phys & Nat Scs	30	3,23	,971
	4 Health sciences	176	3,09	,949
	<b>Total</b>	<b>886</b>	<b>3,15</b>	<b>,982</b>
<hr/>				
TTQ32 The Exit Test I just took NEEDS to measure English related to talking about my family	1 Unknown discipline	300	3,14	,976
	2 Social sciences	380	3,24	,928
	3 Phys & Nat Scs	30	3,30	,877
	4 Health sciences	175	3,08	,912
	<b>Total</b>	<b>885</b>	<b>3,18</b>	<b>,941</b>
<hr/>				
TTQ33 The Exit Test I just took NEEDS to measure English related to travelling for pleasure	1 Unknown discipline	300	3,17	,968
	2 Social sciences	380	3,27	,948
	3 Phys & Nat Scs	30	3,37	,809
	4 Health sciences	176	3,14	,893
	<b>Total</b>	<b>886</b>	<b>3,21</b>	<b>,940</b>

Appendix Table 5 Mean comparison among groups items 27 to 33

## The importance of each skill in their future job

		N	Mean	Std. Deviation
TTQ34a I think a job related to my disciplinary area requires me to speak in English	1 Unknown discipline	299	2,76	1,151
	2 Social sciences	378	2,58	1,179
	3 Phys & Nat Scs	30	2,53	1,167
	4 Health sciences	176	2,43	1,184
	Total	883	2,61	1,174
TTQ34b I think a job related to my disciplinary area requires me to understand what is said to me in English	1 Unknown discipline	299	3,04	1,006
	2 Social sciences	378	2,91	1,036
	3 Phys & Nat Scs	30	2,90	,995
	4 Health sciences	176	2,76	1,038
	Total	883	2,92	1,028
TTQ34c I think a job related to my disciplinary area requires me to write in English	1 Unknown discipline	299	2,81	1,153
	2 Social sciences	378	2,65	1,141
	3 Phys & Nat Scs	30	2,67	1,155
	4 Health sciences	176	2,47	1,175
	Total	883	2,67	1,157
TTQ34d I think a job related to my disciplinary area requires me to read in English.	1 Unknown discipline	299	3,03	1,080
	2 Social sciences	378	2,88	1,147
	3 Phys & Nat Scs	30	2,93	1,081
	4 Health sciences	176	2,72	1,126
	Total	883	2,90	1,122

Appendix Table 6 Mean comparison among groups items 34a to 34d

**Section 3** The relation between language classes and the test in use

		N	Mean	Std. Deviation
TTQ19 The English classes I took, in general, helped me prepare for the test	1 Unknown discipline	300	2,49	1,152
	2 Social sciences	380	2,59	1,162
	3 Phys & Nat Scs	30	2,77	,935
	4 Health sciences	176	2,41	1,153
	Total	886	2,53	1,151
TTQ20 The English classes I took provided opportunities to practice the type of activities included in the test	1 Unknown discipline	300	2,55	1,086
	2 Social sciences	380	2,55	1,144
	3 Phys & Nat Scs	30	2,63	,890
	4 Health sciences	176	2,43	1,159
	Total	886	2,53	1,119
TTQ21 The English classes I took included activities that are similar to the type of activities I may need to perform in English in my future job	1 Unknown discipline	300	2,52	1,152
	2 Social sciences	380	2,57	1,166
	3 Phys & Nat Scs	30	2,73	,868
	4 Health sciences	176	2,48	1,116
	Total	886	2,54	1,142

Appendix Table 7 Mean comparison among groups items 19 to 21

**Section 4** What test takers think of the test difficulty and testing format

		N	Mean	Std. Deviation
TTQ1 The Exit Test was easy to respond to.	1 Unknown discipline	300	2,57	,974
	2 Social sciences	380	2,71	1,051



Appendix X

	3 Phys & Nat Scs	30	2,73	1,015
	4 Health sciences	176	2,59	1,021
	<b>Total</b>	<b>886</b>	<b>2,64</b>	<b>1,019</b>
TTQ22 instructions in Spanish	1 Unknown discipline	300	2,28	1,225
	2 Social sciences	380	2,26	1,239
	3 Phys & Nat Scs	30	2,47	1,224
	4 Health sciences	175	2,39	1,159
	<b>Total</b>	<b>885</b>	<b>2,30</b>	<b>1,218</b>
TTQ23 more open-ended questions	1 Unknown discipline	300	2,27	1,206
	2 Social sciences	380	2,21	1,202
	3 Phys & Nat Scs	30	1,97	1,098
	4 Health sciences	176	2,15	1,149
	<b>Total</b>	<b>886</b>	<b>2,21</b>	<b>1,189</b>
TTQ24 to be multiple-choice only	1 Unknown discipline	300	2,92	1,087
	2 Social sciences	380	3,03	1,083
	3 Phys & Nat Scs	30	3,20	,961
	4 Health sciences	176	3,00	1,131
	<b>Total</b>	<b>886</b>	<b>3,00</b>	<b>1,090</b>
TTQ25 more time to be completed	1 Unknown discipline	300	3,17	1,043
	2 Social sciences	380	3,20	1,057
	3 Phys & Nat Scs	30	3,17	1,085
	4 Health sciences	176	3,14	1,097
	<b>Total</b>	<b>886</b>	<b>3,18</b>	<b>1,060</b>
TTQ26 to be longer	1 Unknown discipline	300	1,95	1,120
	2 Social sciences	380	1,97	1,113

Appendix X

3 Phys & Nat Scs	30	2,07	1,230
4 Health sciences	176	1,93	1,034
Total	886	1,96	1,103

Appendix Table 8 Mean comparison among groups items 1, 22 to 26

## Appendix XI Q2 Statistics

### Speaking and listening-related language competences

#### Group Statistics

	Mean		Mean		Mean	
	Social Sciences		Physical and Natural Sciences		Health Sciences	
	12 Sts SSc N=267	22 Pfs SSc N=332	13 Sts PNSc N = 159	23 Pfs PNSc N = 98	14 Sts HSc N= 280	24 Pfs HSc N = 153
<b>Speaking and listening</b>						
C1 offers advice to clients within their own job area on simple matters.	2.67	2.45	2.79	1.90	2.58	1.99
C2 requests routine services related to their own area of work.	2.81	2.11	2.96	1.84	2.65	1.86
C3 asks questions of a fact-finding nature.	2.87	2.44	2.88	1.94	2.68	2.07
C8 asks for clarification when something is not clear (e.g. "Can you repeat that please?", "Is it .. or ...?", "So, the....is/has/needs...")	2.82	2.30	2.88	1.80	2.71	2.05
C9 participates in job-related meetings and seminars (e.g. making suggestions or asking questions: "Why don't we..."; "Can we...", "Is it possible to...").	2.84	2.38	2.91	1.65	2.75	1.97
C10 expresses opinions in simple terms, for example, "I don't agree", "I think..." provided that the question/issue has been put clearly and simply.	2.71	2.36	2.84	1.77	2.63	2.04
C11 provides information or opinions, such as when planning a company event (e.g. "We could...", "There is...", "I think...").	2.62	2.26	2.76	1.67	2.61	1.92
C12 greets a visitor and engages in a limited conversation for a short time, for example, enquiring about a visitor's journey, hotel, etc. (e.g. "How was your ...", "Did you like...?", "Is/Was your .... ?").	2.94	2.34	2.97	1.92	2.86	2.20
C13 provides instructions and directions, such as giving clients clear directions to the business/company. (e.g. "Take a taxi to. ...", "Go straight...", "walk two blocks, then turn...").	2.59	2.20	2.69	1.78	2.61	1.88

C15 makes phone calls requesting information or services (e.g. "Do you have...", "I would like...", "How much/many...?")	2.66	2.09	2.76	1.68	2.71	1.84
C16 make job-related phone calls, leave a phone message, place a service or materials order (e.g. "I need...", "Can I speak to...", "I'm looking for...", "I")	2.66	2.05	2.72	1.59	2.54	1.81
<b>TOTAL MEAN</b>	<b>2.74</b>	<b>2.27</b>	<b>2.83</b>	<b>1.78</b>	<b>2.67</b>	<b>1.97</b>

Appendix Table 9 Speaking and listening. Mean comparison of students and professionals per disciplinary area (highlighted values are above the total mean value for the group of competences within the group of participants).

## Listening and speaking

### Group Statistics

Listening and speaking	Mean		Mean		Mean	
	Social Sciences		Physical and Natural Sciences		Health Sciences	
	12 Sts SSc N = 267	22 Pfs SSc N = 332	13 Sts PNSc N = 159	23 Pfs PNSc N = 98	14 Sts SSc N = 280	24 Pfs SSc N = 153
C4 takes a routine order (restricted to quantity, delivery date, etc.).	2.58	2.03	2.75	1.66	2.52	1.75
C5 receives phone calls (e.g. take messages/makes notes, etc.).	2.54	2.14	2.69	1.67	2.67	1.87
C14 deals with predictable requests from a visitor, for example, "Can you arrange a taxi to the airport?", "Can I use a projector in the presentation?", "Where can I...?"	2.75	2.14	2.74	1.73	2.61	1.95
<b>TOTAL MEAN</b>	<b>2.62</b>	<b>2.11</b>	<b>2.73</b>	<b>1.69</b>	<b>2.60</b>	<b>1.85</b>

Appendix Table 10 Listening and speaking. Mean comparison of students and professionals per disciplinary area.

## Listening

### Group Statistics

Listening	Mean		Mean		Mean	
	Social Sciences		Physical and Natural Sciences		Health Sciences	
	12 Sts SSc N = 267	22 Pfs SSc N = 332	13 Sts PNSc N = 159	23 Pfs PNSc N = 98	14 Sts HSc N = 280	24 Pfs HSc N = 153

C6 understands spoken directions, i.e. safety procedures at work.	2.85	2.40	3.05	1.79	2.70	2.03
C7 follows a simple presentation/ demonstration and understanding explanations concerning a product or topic within one's own area of expertise.	3.03	2.62	3.11	1.91	2.82	2.17
<b>TOTAL MEAN</b>	2.94	2.51	3.08	1.85	2.76	2.10

Appendix Table 11 Listening. Mean comparison of students and professionals per disciplinary area (highlighted values are above the total mean value for the group of competences within the group of participants).

## Reading

### Group Statistics

Reading	Mean		Mean		Mean	
	Social Sciences		Physical and Natural Sciences		Health Sciences	
	12 Sts SSc N = 267	22 Pfs SSc N = 332	13 Sts PNSc N = 159	23 Pfs PNSc N = 98	14 Sts HSc N = 280	24 Pfs HSc N = 153
C17 recognizes and understands at least partially the general meaning of a non-routine letter within its own work area.	2.88	2.28	2.89	1.92	2.76	2.14
C18 understands straightforward factual information on routine job-related faxes, memos, emails, etc.	2.88	2.40	2.96	1.89	2.67	2.12
C19 obtains basic relevant information (product specifications, professional or commercial leaflets, advertisements, internet pages, etc.....)	3.01	2.42	2.99	2.00	2.69	2.08
C20 recognizes and understands at least partially the general meaning of a routine letter within its own work area.	2.80	2.24	2.95	1.80	2.63	2.05
C21 understands general written instructions (for example, installation, functioning or maintenance manuals)	3.14	2.52	3.16	1.86	2.72	2.18

C22 uses several pieces of information, such as a work schedule to see who should be working and several reports on a process to identify the source of a problem.

2.57      2.10      2.64      1.56      2.49      1.83

**TOTAL MEAN**

2.88      2.33      2.93      1.84      2.66      2.07

Appendix Table 12 Reading. Mean comparison of students and professionals per disciplinary area (highlighted values are above the total mean value for all competences within the group of participants).

## Writing

### Group Statistics

Writing	Mean		Mean		Mean	
	Social Sciences		Physical and Natural Sciences		Health Sciences	
	12 Sts					
	SSc N = 267	22 Pfs SSc N = 332	13 Sts PNSc N = 159	23 Pfs PNSc N = 98	14 Sts HSc N = 280	24 Pfs HSc N = 153
C23 writes short reports, faxes, emails, and memos on basic everyday job-related matters (e.g. simple factual information related to a recent event, the time and place of an activity, people attending, and topics to be addressed).	2.57	2.06	2.65	1.49	2.46	1.88
C24 writes requests for goods, services, etc., on a range of routine matters but MAY need to get these checked.	2.68	2.00	2.77	1.59	2.49	1.75
C25 makes notes on routine matters, such as taking/placing orders.	2.51	2.00	2.53	1.54	2.30	1.75
C26, fill out a form, such as an incident/accident report form or a medical record.	2.54	1.96	2.56	1.44	2.35	1.69
C27 reports events or incidents involving products, machines or personnel (workers, clients), i.e. a workplace accident, a complaint, or a failure in a process.	2.42	1.95	2.45	1.44	2.31	1.61
C28 makes notes for personal use on non-routine aspects such as a client's requirements and basic characteristics of a new product or service.	2.43	2.04	2.47	1.54	2.36	1.78
<b>TOTAL MEAN</b>	2.53	2.00	2.57	1.51	2.38	1.74

Appendix Table 13 Writing. Mean comparison of students and professionals per disciplinary area (highlighted values are above the total mean value for all competences within the group of participants).

## Appendix XII Example of language use test task



Universidad Juárez  
del Estado de Durango



### INSTITUTIONAL LANGUAGE PROGRAMME

#### B1 ORAL CERTIFICATION TEST

PAIRS

PROCEDURE:

**1. Greet students (Ss) as they sit:**

I: Good morning....afternoon.... (as appropriate)

**2. Ask Ss for their names:**

I: Hello, my name is \_\_\_\_\_. This is my colleague; he/she will be listening to us. And your names are \_\_\_\_?

*(WRITE DOWN Ss' NAMES TO KEEP TRACK OF THEM DURING THE EXAM -- student A is the one on your left, and student B is the one on your right).*

THIS BOOKLET IS DESIGNED TO BE USED WITH PAIRS.

**REMEMBER TO ADDRESS STUDENTS BY THEIR NAMES AT ALL TIMES.**

I: (*Candidate A*), where are you from?

I: And you, (*Candidate B*), where are you from?

I: Thank you

I: (*Candidate B*), how do you spell your last name?

I: And what about you (*Candidate A*), how do you spell your last  
name?

I: (*Candidate A*), what is your phone number?

I: (*Candidate B*), where do you live/what is your address?

I: Thank you

I: In the next part, you are going to talk to each other.



PART 1 A

PAIRS

**REMEMBER TO ADDRESS STUDENTS BY THEIR NAMES**

C3. CAN ask questions of a fact-finding nature.

**The words in bold indicate stress. Stress out words in bold to help candidates understand the task and the expected outcome.**

I: In this part of the test, you are going to **talk to each other**. I am going to ask you to imagine a situation.

You both work for different companies. (*Candidate A*), you placed an order with the company where (*Candidate B*) works. (*Candidate A*), your order hasn't arrived, so you call the company to check on your order. Read the instructions and information on your card (hand out the card to *Candidate A*) to ask about your order. (*Candidate B*) answer (*Candidate A*)'s questions and ask for clarification if required. Use the information in your card (hand out the card to *Candidate B*).

I'll give you a moment to look at your card.

(*Candidate A*) you will start.

(Allow about 30 or 40 seconds)

Is that clear? Do you want me to repeat the instructions? (Repeat if required).

(*Candidate A*) please start.

Student A	Part 1a pairs
A client is asking about an order.	
This is the information you have. <b>If something is not correct, you may need to spell it for him/her. If there is a mistake, apologize and say you will correct the mistake.</b>	
<b><i>You work for OWL Inc. Greet the client and offer to help.</i></b>	
Date of order: May 6 <sup>th</sup> (today is May 17 <sup>th</sup> .) Expected arrival: May 11 <sup>th</sup>	
<b><i>Verify the name of the person who placed the order.</i></b>	
Order placed by: <b>Diana/Dante (use the name according to the gender of Student B) Ramírez</b>	
15 items size A (not in stock)	
10 blue items standard size	
12 boxes of small rectangular containers (10cms. height x 12 cms. long)	
Tell the client you will check and call back in an hour. Apologize and offer a 10% discount on the next order.	

Student B	Part 1a pairs
You are calling to check on an <b>order</b> you placed, as it has <b>not arrived</b> .	
<ul style="list-style-type: none"> <li>• Your name is <b>Denisse/Daniel (choose one) Rodriguez</b>. Say why you are calling: You want to <ul style="list-style-type: none"> <li>○ know why the order has not arrived (say when you placed it).</li> <li>○ check the number and type of items ordered</li> <li>○ know when the order will arrive</li> </ul> </li> </ul>	
Date of order: May 6 <sup>th</sup>	

**50** items size A

10 blue items standard size

**2** boxes of small rectangular containers (10cms. height x 12 cms. long)

Expected **arrival date: May 11<sup>th</sup>** (today is May 17<sup>th</sup>.)

When they finish, thank them and retrieve cards from candidates.

I: Thank you. May I have the cards, please?

I: Now, you are going to **change roles**. (*Candidate B*) you placed the order, and (*Candidate A*), you work for the company where the order was placed.

(*Candidate B*) you think there was a mistake in the order placed. You also think there is a mistake in the delivery address. Ask (*Candidate A*) what items were ordered and the shipping address. Ask for corrections to be made if necessary.

I'll give you a moment to look at your card.

(*Candidate B*) you will start.

(Allow about 30 or 40 seconds.)

Is that clear? Do you want me to repeat the instructions? (Repeat if required).

(*Candidate B*) please start.

Student B	Part 1b pairs
You are calling to check on an order placed because you think there was a mistake.	
<p><b>Identify yourself (name and name of company). Explain why you are calling.</b></p>	
<p>You work for RODD Industries. You are calling to ask about an order placed two days ago (<b>today is February 6<sup>th</sup></b>). You think there was a mistake in the amount and type of items ordered. You also think there was a mistake in the address.</p>	
<ul style="list-style-type: none"> <li>• Ask <i>Candidate A</i> if the information you have is the same he/she has. The order should be:           <ul style="list-style-type: none"> <li>○ <b>28 items D25F</b></li> <li>○ 150 bronze clips size <b>E</b></li> <li>○ 23 mts. <b>plain</b> chain in <b>silver</b> finishing</li> </ul> </li> </ul>	
<p><i>Shipping address:</i></p>	
<p>Paseo del <b>Canario # 2547</b></p>	
<p>Third floor</p>	
<p>Las Aves, <b>82430</b></p>	
<p>Durango, Dgo.</p>	

Student A	Part 1b pairs
A client is asking about an order.	
<p>This is the information you have. <b>If something is not correct, you may need to spell it for him/her. If there is a mistake, apologize and say you will correct the mistake.</b></p>	
<p>Date of order: February 4<sup>th</sup> (today is February 6<sup>th</sup>.) Expected shipping date: February 9<sup>th</sup></p> <ul style="list-style-type: none"> <li>○ <b>23 items code B25F</b></li> <li>○ 150 bronze clamps size <b>i</b></li> <li>○ 43 mts. <b>snake</b> chain in <b>gold</b> finishing</li> </ul>	
<p><i>Shipping address:</i></p>	

Paseo del **Calvario # 2617**

Third floor

Las Aves, **82130**

Durango, Dgo.