Associating speaker variables with English pronunciation ratings in Spanish tertiary education

Mateusz Pietraszek
Universidad Francisco de Vitoria

Received: 2023-01-10 / Accepted: 2023-09-15
DOI: *https://doi.org/10.30827/portalin.vi41.27083*
ISSN paper edition: 1697-7467, ISSN digital edition: 2695-8244

ABSTRACT: Although the use of English in the ever-changing context of bilingual tertiary education has been extensively studied to date, pronunciation research in the field is still scarce. This paper aims at providing empirical data regarding the hypothesised impact of bilingual education programmes on pronunciation performance operationalised through the notions of intelligibility, comprehensibility and foreign-accentedness. For this purpose, 60 advanced speakers of English from a Spanish university (enrolled both in bilingual and monolingual degree courses) were recorded and assessed by 330 international English speakers from 30 different L1s. Both the speakers and the listeners completed a basic socio-demographic questionnaire. The independent variables were related to the participants' (1) reported exposure to bilingual instruction at the previous and current levels of education, as well as (2) their overall level in English and (3) their exposure to English abroad. The findings showed that bilingual education was in no way associated with differences in pronunciation ratings. However, the speakers' level and their experience abroad were correlated with the comprehensibility and foreign-accentedness scores. The results are not only consequential for the understanding of the (lack of) impact of bilingual variables on pronunciation, but they also allow to examine the adequacy of the CEFR phonological descriptors.
Keywords: bilingual education, EMI, level, English pronunciation, intelligibility

La asociación de las variables del hablante con la pronunciación en inglés en la universidad española

RESUMEN: Aunque el uso del inglés en el contexto cambiante de la educación superior bilingüe ha sido extensamente estudiado, la investigación sobre la pronunciación sigue siendo escasa. El objetivo de este artículo es proporcionar datos empíricos sobre un hipotético impacto de los programas bilingües en la pronunciación concebida mediante las nociones de inteligibilidad, comprensibilidad y grado de acento extranjero. Con este fin, 60 universitarios españoles (matriculados tanto en grados bilingües en monolingües) fueron grabados y – posteriormente – evaluados por 330 hablantes de inglés con 30 lenguas nativas diferentes. También se recogieron datos sociodemográficos básicos sobre los participantes. Las variables independientes están basadas en la información proporcionada por los hablantes relativa a (1) su participación en programas bilingües antes de la universidad y durante su etapa universitaria, (2) su nivel de inglés y (3) su experiencia en el extranjero. Los análisis no han demostrado asociación alguna entre la educación bilingüe y el desempeño en pronunciación. Sin embargo, los resultados muestran una correlación entre el nivel de inglés y las estancias en el extranjero, por un lado, y la comprensibilidad y el grado de acento extranjero, por otro. Los resultados del estudio no solo son relevantes para entender mejor el impacto de la educación bilingüe en la pronunciación de los alumnos (o la falta del mismo), sino que también permiten examinar la idoneidad de los descriptores fonológicos del MCER.

Palabras clave: enseñanza bilingüe, EMI, nivel, pronunciación del inglés, inteligibilidad

1. Introduction

English pronunciation is often considered to be a difficult linguistic skill for Spanish learners of English – an opinion which has often made its way to the Spanish media (cf. El Confidencial, 2014; Galván, 2010; La Razón, 2019). It is also one of the first aspects of language based on which impressionistic judgments are made regarding the speakers' linguistic competence (Cutillas Espinosa, 2017; Lippi-Green, 1994). As the exact measurement of pronunciation skills may be a burdensome task and the way it is construed may prove somewhat elusive, it has been one of the least studied (and taught) aspects of second language acquisition or even – as Gilbert (2010, p. 3) put it – "the orphan of ESL/EFL".

This paper focuses on the relationship between some characteristics of Spanish university students as predictors of their pronunciation performance in an educational context where the presence of English has been increasing exponentially for the past two decades (Diogo et al., 2019; Fielden Burns et al., 2018; Foskett, 2010; Ramos García, 2013; Ramos García & Pavón Vázquez, 2018). Said variables comprise the different contexts in which the students have been exposed to English as well as their level. The data presented in this article are part of a broader research project investigating the variability of Spanish students' pronunciation scores considering characteristics related to the speakers, listeners and phonological features of the sampled speech.

The measurement of pronunciation has been operationalised by means of three different variables: *intelligibility* (INT), *comprehensibility* (COM) and *foreign-accentedness* (FA), each related to a different aspect of the participants' phonological performance. The first one, i.e., *intelligibility*, can be best defined as the objective measure of the segmental decoding of the speech of a person by a listener (Jenkins, 2000; Walker, 2010). On the other hand, the more subjective dimensions of c*omprehensibility* and *foreign-accentedness* – as defined in studies conducted by Derwing and Munro (2015) – reflect the ratings given by listeners on Semantic Differential Scales (Dörnyei, 2007; Poljak 2019). While the former gauges the perceived ease of understanding, the latter is concerned with the perceived degree of divergence from L1 speakers' pronunciation. The *intelligibility* variable in this paper also roughly follows the way it is construed by Derwing and Munro (2009, 2015) in the sense that it involves real *understanding*. However, said understanding is reduced to the decoding of phonological component by eliminating as much semantic information as possible using Semantically Unpredictable Sentences (SUS), as explained later in the methodology section (Benoît et al., 1996; Wang, 2007). This differs from intelligibility studies where the variable is measured as the understanding of semantically meaningful texts by means of transcriptions or other tasks focused on the meaning of the text presented in the speech samples (Derwing & Munro, 1997, 2015). In fact, *intelligibility* has been tested through a variety of tasks to date, including content questions (multiple choice, true/false, specific information), translation, transcriptions (of meaningful sentences), SUS and reaction times (Gooskens, 2013). Additionally, the term *intelligibility* has also been used to denote what is conceptualised as *comprehensibility* in this study, i.e., the listeners' opinions of how understandable they found a speaker rather than their real understanding of speech samples (Gooskens, 2013; Haugen, 1966).

Focusing on the phonological component aligns with the views held by English as a Lingua Franca researchers (Jenkins, 2000; Walker, 2010). Their concept of *intelligibility* can be traced back to that of Smith and Nelson (1985), for whom it is the first stage of language processing based solely on phonological information. Jenkins (2000, p. 82, 83, 90) noticed that even proficient ELF users use more bottom-up strategies when decoding meaning than native speakers, who rely on contextual cues to a much greater extent (top-down strategies). This, in turn, may lead to increased difficulties in communication based on phonological misunderstandings as opposed to morphosyntactic errors. These observations were corroborated by further ELF research (Deterding, 2013; Field, 2005; Meierkord, 2004) and informed the decision regarding the conceptualisation of *intelligibility* in this study as the targeted listener sample herein is comprised by ELF speakers from all around the globe. Indeed, many researchers (inside and outside the ELF paradigm) have investigated the impact of segmental features on *intelligibility* (Barrass et al., 2020; Jenkins, 2000; Jeong et al., 2017; Joto et al., 2007; Levis, 2018; Minematsu et al., 2011; O'Neal, 2015; Walker, 2010; Wang, 2007), and *comprehensibility* and *accentedness* (Barrass, 2017; Crowther et al., 2015; Saito et al., 2016; Trofimovich & Isaacs, 2012) outside the Spanish context.

As regards the variables related to the speakers' characteristics considered herein, a relatively high number of studies have been conducted on the effects of bilingual programmes at pre-university levels (e.g., Gallardo del Puerto & Martínez Adrián, 2013; Ruiz de Zarobe, 2011). While certain studies showed a positive effect of CLIL (Content and Language Integrated Learning) instruction on pronunciation (Lasagabaster, 2008; Pérez-Cañado, 2018; Rallo Fabra & Juan-Garau, 2010), others showed no such effect (Dalton-Puffer, 2008; Gallardo del Puerto et al., 2009; Gallardo del Puerto & Gómez Lacabex, 2017; Pérez-Cañado & Lancaster, 2017; Rallo Fabra & Jacob, 2015). The results of the research being inconclusive at best, Merino and Lasagabaster (2018, p. 21) still claimed that "pronunciation happens to be the least affected of such dimensions", thus reiterating a similar statement by Dalton-Puffer (2011).

At the tertiary level of education, there has been a dearth of studies dealing with language competence – let alone pronunciation. Regarding phonetic performance in EMI (English Medium Instruction), some evidence was found for its impact on foreign accent in the Austrian context (Richter, 2015, p. 221) when comparing EMI students with those enrolled on an English Language degree. The former outperformed the latter as their degree of foreign accent diminished to an extent significantly greater than in the case of the English language degree students. In Spain, a preliminary study (Gómez Lacabex & Gallardo del Puerto, 2021, p. 133) with 12 EMI students reported the participants' comprehensibility to be rated at a 4.9/9 and intelligibility - 0.43/1. However, the research does not explicitly compare EMI and non-EMI students and, therefore, one cannot ascertain whether the EMI variable is at work here all other factors being equal.

Apart from the different types of exposure to English in domestic contexts, another variable that is taken into account in this research is the speakers' experience living abroad. Flege (1988) suggested that accents were largely settled after a one-year stay abroad. Derwing and Munro (2015, p. 40-43) provide an overview of the studies dealing with the effect of the length of residence in native immersion contexts. Out of the 13 surveyed papers, 9 provide evidence in favour of such effects, including 3 out of the 4 studies dealing specifically with accent. At the same time, research into stays abroad from the Spanish perspective indicates

that even stays as short as 3 months can lead to a milder foreign accent in both children and adults (Muñoz & Llanes, 2014, p. 442). Other studies suggested there was no little to no effect of a short study abroad on certain phonological features, such as vowel length or Voice Onset Time (Avello & Lara, 2014). Additionally, studying abroad had no significant impact on the degree of accentedness rated by non-native listeners (Avello et al., 2012). Juan-Garau (2014) points out that while studying abroad may be advantageous for the development of oral accuracy, it is not always the case and speaker-related variables (e.g. pre-departure level or motivation) may often be associated with the end result. A recent literature review of the study abroad experience research (Moore et al. 2021, p.1) highlights that in spite of the number of published articles, "the evidence remains inconclusive".

Finally, it is only logical that proficiency level, one of the independent variables considered in this research, should bear on all aspects of pronunciation – including phonetic performance. Language proficiency has been proved to be a determining factor in predicting many differences in accentedness, comprehensibility and intelligibility ratings (Barrass, 2017; Derwing & Munro, 1997; Saito et al., 2016). Since the publication of the Common European Framework of Reference for Languages (Council of Europe, 2001), attempts have been made at a systematic description of all aspects of language skills and usage for teaching and assessment purposes. However, as shown in Table 1, the original description of phonological control was scarce. Not only did it contain highly impressionistic labels (*clear, natural, some effort*) but it also used the terms *native speaker* and *foreign accent* without specifying their exact meaning or suggesting ways of operationalising them in practice.

Table 1. *Phonological control descriptors: CEFR 2001*

C2	As C1
C1	Can vary intonation and place sentence stress correctly in order to express finer shades of meaning.
B2	Has acquired a clear, natural, pronunciation and intonation.
B1	Pronunciation is clearly **intelligible** even if a **foreign accent** is sometimes evident and occasional mispronunciations occur.
A2	Pronunciation is generally clear enough to be understood despite a **noticeable foreign accent**, but conversational partners will need to ask for repetition from time to time.
A1	Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by native speaker used to dealing with speakers of his/her language group.

Note: Council of Europe, 2001, p. 117.

In a later document (Council of Europe, 2018), it was acknowledged that the phonological descriptors were not specific enough in their 2001 form, which was even dubbed "unrealistic" (p. 134). Also, the *intelligibility* principle (cf. Levis, 2005) was applied as the phonological descriptors in the original CEFR were the only ones to make a reference to a native model, albeit indirectly through the expression *foreign accent*. In the CEFR companion (2018, p. 134), intelligibility is finally defined as "accessibility of meaning for listeners, covering also the listeners' perceived difficulty in understanding (normally referred to as *comprehensibility*)". Further in the document (p. 135), it is added that the intelligibility descriptors account for "the extent of influence from other languages", "control of sounds"

and "control of prosodic features" making it even harder to pinpoint. It can, however, be assumed that the *intelligibility* construct in the CEFR is actually – to a large extent and as stated in the document – *comprehensibility* the way it is understood in this research and previous studies (Derwing & Munro, 2015), especially because no objective measurement methods are suggested, and only impressionistic descriptors are offered. Quite interestingly, references to accent (although not *foreign* accent anymore) are present at all levels.

Table 2. *Phonological control descriptors: CEFR Companion 2018*

C2	[…] **Intelligibility** and effective conveyance of and enhancement of meaning are **not affected in any way by features of accent** that may be retained from other language(s).
C1	Can employ the full range of phonological features in the target language with sufficient control **to ensure intelligibility throughout**. […] some features of accent retained from other language(s) may be noticeable, but **they do not affect intelligibility**.
B2	[…] **accent** tends to be influenced by other language(s) he/she speaks, but has **little or no effect on intelligibility**.
B1	Pronunciation is **generally intelligible.**
A2	Pronunciation is generally clear enough to be understood. […] **A strong influence from other language(s)** he/she speaks on stress, rhythm and intonation may **affect intelligibility**, requiring collaboration from interlocutors […]
A1	**Pronunciation of a very limited repertoire** of learnt words and phrases can be **understood with some effort** by interlocutors.

Note: Council of Europe, 2018, p. 136.

The new descriptors (Table 2) show clear progression in *intelligibility* (as defined by the CEFR Companion 2018 and roughly equivalent to *comprehensibility* in this study (Derwing & Munro, 2015). Full proficiency – where there are no intelligibility difficulties caused by phonological control – is achieved around level C1. Indeed, there is little (if any) difference between the reference to "intelligibility" in the CEFR C1 and C2 descriptors. As discussed later on in this paper, the data herein shed new light on the applicability of these descriptors at least in the Spanish context of this study.

The research questions (RQs) addressed in this paper are related to the aforementioned speaker characteristics which have been hypothesised to bear a potential relationship to the learners' pronunciation performance.

RQ1: What association is there between the speakers' level of English and their COMFA-INT scores?

RQ2: What is the effect of the type of exposure to English (pre-university CLIL, EMI and stays abroad) on the participants' pronunciation scores?

2. Methodology

At the initial stage of this research, 60 Spanish speakers of English were recorded at a Spanish university (including 34 males and 26 females) after obtaining their informed con-

sent through a signed form. Sociodemographic data were also collected. The participation in the study was voluntary and was not used for the purposes of course assessment. 32 of the speakers were enrolled on English Medium Instruction (EMI) courses while the remaining 28 were monolingual degree students (non-EMI). 39 had attended bilingual schools prior to university while 21 had no such experience. They were all in their 2nd, 3rd or 4th year, and came from 5 different degrees. Their average age was 21.2. (*Min* = 19, *Max* = 26, *SD* = 1.44) and they all had at least a B2 level in English according to institutional requirements and placement tests. However, only those who explicitly reported holding a certified B2 (or higher) level of English were considered in the analyses regarding level (n = 50) later in this paper while the remaining 10 were excluded. In this part of the sample, there were 27 B2, 15 C1, and 8 C2 speakers. Regarding their experience living abroad ("Do you have experience living abroad"), 15 speakers reported having none, while 45 answered "yes". The amount of time spent abroad ("How much time did you spend abroad?") ranged from 0 to 6 years with an average situated at 1.16 years. 21 out of 45 students spent less than a year abroad, 7 were away for a year and the remaining 17, for more than a year.

Before gathering the data, the instruments were validated by two independent experts and piloted with a group of 28 students with satisfactory results. The research tools included: (1) a series of 40 Semantically Unpredictable Sentences (henceforth SUS, see APPENDIX) (Benoît et al., 1996; Wang, 2007), (2) an original Elicitation Paragraph (EP, see APPENDIX), (3) a sociodemographic (background) questionnaire, and (4) an attitudinal questionnaire (not addressed in this paper). The SUS were thought to be an appropriate tool as the limited semantic information they contain would force the listeners to interpret phonetic information without resorting to the semantic context (cf. Kang et al. 2018).

The recordings took place in empty university classrooms and lasted around 5 minutes per speaker. The participants had time to read the written prompts before the recording and they could repeat any output they were not satisfied with. After the recording and the collection of questionnaire data, the final online *intelligibility*, *comprehensibility* and *foreign-accentedness* tests were set up and administered by means of an online platform (Limesurvey Gmbh., 2019) to 330 listener informants from nearly 30 L1 backgrounds, including L1 Spanish and L1 English speakers. The listeners were recruited through an online self-selection process and volunteered to participate in the research. There were 206 (62.4%) females and 120 (36.4%) males in the listener sample whereas 4 (1.2%) people did not identify with either of the two genders. The average age of the listeners was 30 (*Min* = 14, *Max* = 70, *SD* = 11.1). Regarding their L1s, 33% were native speakers of English 25.2% were native speakers of Spanish and 20.3% were Polish[1]. Among the people whose mother tongue was not English, 94% reported to have a B2 or higher level in the language.

Each listener rated a random selection of five speakers assigned by the platform. The *intelligibility* part of the test consisted in the orthographic transcription of auditory input extracted from the speakers' SUS, while *comprehensibility* and *foreign-accentedness* were measured using 6-point semantic differential scales based on 25-30-word excerpts from the EP. Only the ends of the scales were labelled in the *comprehensibility* and *foreign-accentedness*

[1] All the L1s reported in the study were: Arabic, Basque, Cantonese, Catalan, Croat, Danish, Dutch, English, Finnish, French, Ilocano, Italian, Lithuanian, Mandarin, Nepali, Persian, Polish, Portuguese, Romanian, Russian, Serbian, Spanish, Swedish, Tagalog, Tamil, Turkish, Urdu, Vietnamese.

items of test. Each listener answered five questions evaluating the speakers' *comprehensibility* by completing the sentence "The speaker is ___ to understand" on a scale from (1) *very difficult* to (6) *very easy*. The five items assessing accentedness requested the listeners to "rate the speaker's degree of foreign accent" on a scale from (1) *native-like* to (6) *very foreign*. Each listener listened to and was asked to transcribe 20 different SUS, i.e., 4 SUS for each of the 5 speakers randomly selected for that listener by the platform. The same SUS (pronounced by different speakers) were never present in one listener's test. The order of the 30 items (20 SUS transcriptions and 10 scalar items) in the test were randomised by the platform. No time limit for the completion of the survey was set and the average time necessary to complete the task was 21 minutes.

The 6,600 SUS transcribed by the listeners were converted into numeric intelligibility values using a scoring protocol devised for the purposes of the study. The total number of points a speaker could obtain from each listener on the SUS test was 30. There was a total of 15 words in the 4 SUS coming from each speaker, each bearing an equal weight of 2 points. The full 2 points were allotted to those transcriptions which were fully correct, and one point was allotted if one mistake was present in a word, i.e., when a phoneme was not recognised correctly. Evident spelling mistakes or typos (e.g., **boughr* for *bought*) not resulting in other existing words and homophones (e.g., *weak* for *week* or *cell* for *sell*) were treated as correct transcriptions. Otherwise, points were deducted. The average number of listeners who rated an individual speaker was 27.5. Each individual speaker's COM and FA scores were obtained by calculating the means of the points assigned by the listeners who rated their speech. The speaker's INT rating was the average of their listeners' scores on the SUS transcription task. The different magnitudes of the scales used for COM/FA (*Max* = 6) and INT (*Max* = 30) were not problematic as no cross-variable statistics were calculated for this study except for correlations for which the range of the scale is irrelevant. If mean comparisons were necessary, the scores can be easily scaled by dividing the intelligibility ratings by 5.

The following independent variables were hypothesised to bear a relationship with the three dependent variables in the study (i.e., intelligibility, comprehensibility and foreign-accentedness).

a) English level (RQ 1)
b) pre-university exposure to CLIL (RQ 2)
c) university exposure to EMI (RQ 2)
d) staying abroad (RQ 2)

The data was based on the speaker characteristics extracted from the questionnaires accompanying the recordings. All statistics were conducted using IBM SPSS version 26.0 (IBM Corp., 2020).

3. Results

The findings of the statistical tests are presented in this section for every independent variable: English level, CLIL and EMI instruction, and staying abroad. The scores used in the statistical tests correspond to the means calculated by averaging all intelligibility, com-

prehensibility and foreign-accentedness scores obtained from all the listeners who evaluated each speaker. The descriptive statistics regarding the COM, FA and INT values in each group are presented in tables. The correlation and pairwise comparison effect sizes were interpreted *sensu* Cohen (1988), while the ANOVA effect sizes were calculated following Tomczak and Tomczak (2014).

3.1. English level

The level of English competence as self-reported by 50 speakers ranged from a CEFR B2 through C1 to C2[2]. A one-way ANOVA was conducted to check whether INT, COM and FA differed across levels. The test demonstrated there was a significant difference in the overall comprehensibility depending on the speakers' level with a large effect size[3], $F(2, 47) = 4.345$, $p = .019$, $\eta^2 = .16$. Post-hoc analyses using the HSD Tukey test ($p = .014$) showed that the scores were significantly higher for C2 level students than for B2 students. However, there was no significant difference ($p = .609$) across levels B2 and C1 or C1 and C2 ($p = .122$). Similarly, foreign-accentedness differed significantly amongst speakers at different levels, the effect size of level being large, $F(2, 47) = 6.950$, $p = .002$, $\eta^2 = .228$. Again, post-hoc HSD Tukey test results revealed a statistically significant difference ($p = .002$) only between levels B2 and C2 but not between C1 and C2 ($p = .078$) or B2 and C1 ($p = .260$). Unlike comprehensibility and foreign-accentedness, intelligibility did not differ across levels, $F(2, 47) = 1.284$, $p = .286$.

Table 3. *COM, FA and INT by level: descriptive statistics*

	COM	FA	INT
B2	$M = 3.87$, $SD = .74$,	$M = 4.4$, $SD = .80$	$M = 21.18$, $SD = 3.25$
C1	$M = 4.09$, $SD = .62$,	$M = 3.98$, $SD = .86$	$M = 21.34$, $SD = 2.8$
C2	$M = 4.71$, $SD = .73$	$M = 3.18$, $SD = .81$	$M = 23.07$, $SD = 2.13$

Apart from the aforementioned effect sizes, non-parametric correlations were also calculated between level and the three dependent variables. As can be seen in Table 4, level was found to be positively correlated with COM ($r_s = .348$, $p = .013$) and negatively with FA ($r_s = -.453$, $p = 001$), but not with INT ($r_s = .152$, $p = .293$). Just like in the case of effect sizes, the correlation is stronger between level and foreign-accentedness than between level and comprehensibility.

Table 4. *Non-parametric correlations: level and dependent variables*

	COM	FA	INT
Level	$r_s = .348^{*}$[4], $p = .013$	$r_s = -.454^{**}$, $p = .001$	$r_s = .152$, $p = .293$

[2] As mentioned earlier in the methodology section, only those speakers who reported to be holders of language certificates, i.e., in the tests including the level variable, the sample size was 50, not 60.

[3] $\eta^2 = .01$ SMALL, $\eta^2 = .06$ MEDIUM, $\eta^2 = .16$ LARGE (Tomczak & Tomczak 2014).

[4] * is used to denote *p*-values below .05, while ** is used for values below .01.

3.2. Bilingual education: CLIL and EMI programmes

Regarding the exposure of the speaker sample to CLIL at previous education stages, no statistically significant differences were found in overall dependent variable means between the groups of students previously exposed to CLIL (CLIL group) and those with no such exposure (non-CLIL group). The analyses did not indicate statistically significant differences in the COM of the students in both groups, $t(58) = -.84$, $p = .402$. Likewise, FA did not differ significantly between the two groups, $t(58) = 1.47$, $p = .146$, and neither did INT, $t(58) = -.17$, $p = .863$.

Table 5. *COM, FA and INT by medium of instruction: descriptive statistics*

	COM	FA	INT
CLIL	$M = 4.14$, $SD = .70$,	$M = 4.01$, $SD = .91$	$M = 21.73$, $SD = 3.16$
non-CLIL	$M = 3.97$, $SD = .57$,	$M = 4.35$, $SD = .74$	$M = 21.59$, $SD = 2.41$
EMI	$M = 4.01$, $SD = .75$	$M = 4.01$, $SD = .75$	$M = 21.86$, $SD = 3.01$
non-EMI	$M = 4.11$, $SD = .71$	$M = 4.12$, $SD = .92$	$M = 22.36$, $SD = 3.02$

Similarly, none of the differences in overall INT, COM and FA means between the EMI and non-EMI student groups were found to be statistically significant. An independent sample t-test did not show differences in the comprehensibility of the sampled EMI students and non-EMI students, $t(58) = .260$, $p = .796$. Moreover, no significant differences were detected in an independent sample t-test between the FA scores of in both groups, $t(58) = -.076$, $p = .940$. Finally, the INT of the EMI sample was not statistically different from that of the non-EMI sample, $t(58) = -.493$, $p = .624$.

3.3. Staying abroad

The students' experience abroad was also hypothesised to be potentially associated with their phonological performance. However, no statistically significant differences between groups were supported by the results of *t*-tests. COM proved not to differ significantly between those with experience abroad (group A) and those without it (group B), $t(58) = -.691$, $p = .493$. Neither did the level of FA vary significantly between groups, $t(58) = .351$, $p = .727$. Finally, INT levels were not significantly different, $t(58) = -.074$, $p = .941$.

Table 6. *COM, FA and INT by length of staying abroad: descriptive statistics*

	COM	FA	INT
group A (YES)	$M = 4.12$, $SD = .69$	$M = 4.11$, $SD = .81$	$M = 21.70$, $SD = 2.7$,
group B (NO)	$M = 3.97$, $SD = .86$	$M = 4.20$, $SD = 1.05$,	$M = 21.64$, $SD = 3.53$

However, despite the lack of group differences, certain correlations were detected as described below. The time spent abroad was positively correlated with COM ($r = .402$, $p = .001$) and negatively correlated with FA ($r = -.315$, $p = .001$). Nevertheless, the INT level

was not significantly correlated with the amount of time spent by the speakers abroad (r = .143, p = .275).

Table 7. *Parametric correlations: years abroad and dependent variables*

	COM	FA	INT
Tears abroad	r = .402**, p = .001	r = -.315**, p = .001	r = .143, p = .275

Guided by previous research, which suggests that the first year of residence is crucial for the establishment of accent (Flege, 1988), and in the light of statistically significant correlations in our study, another series of post-hoc t-tests was carried out to find out whether those students who had spent a year or more abroad differed significantly from those who had spent less or no time abroad.

Significant results were found for comprehensibility and foreign-accentedness. Comprehensibility differed significantly between those with at least one year of foreign experience (group A) and those without it (group B), F(58) = -2.812, p = .007, d = .741. The effect size was moderate, yet approaching large. Similarly, moderate foreign-accentedness differences were revealed between the groups, F(58) = 2.060, p = .044, d = .543. Again, intelligibility was not affected (F(58) = -.764, p = .448) as no statistical differences were detected.

Table 8. *COM, FA INT by staying abroad for 1 year or more: descriptive statistics*

	COM	FA	INT
group A (YES)	M = 4.4, SD = .69,	M = 3.86, SD = .83	M = 22.04, SD = 2.41
group B (NO)	M = 3.87, SD = .68	M = 4.32, SD = .91	M = 21.45, SD = 3.19

4. Discussion

Let us now critically examine the results presented in the previous section in light of previous research. It should not come as a surprise that the overall level of competence in English, which quite logically affects the four main skills (reading, writing, listening and speaking), also has an effect on pronunciation performance (COM-FA-INT). The statistical analyses carried out in the present research bear out this assumption for comprehensibility and foreign-accentedness as defined in this study, but not for intelligibility.

As already mentioned in the introduction, phonological skills are mentioned in the CEFR (Council of Europe, 2001, p. 91) alongside other primary linguistic skills related to syntax and the lexicon and obviously vary across levels just like vocabulary range or the fluency in the use of grammatical structures. The findings of the inferential statistics herein corroborate the hypothesis that there were indeed significant and large differences between levels on foreign-accentedness and comprehensibility tests. The differences were larger for foreign-accentedness than for comprehensibility as shown by both large effect size calculations and moderate non-parametric correlations. However, ANOVA post-hoc tests showed that those differences were not distributed equally. The surveyed listeners assigned different comprehensibility and foreign-accentedness scores to B2 as opposed to C2 students but no differences were found between B2 and C1 and C1 and C2. Obviously, it does not conclu-

sively deny the existence of any differences, which may not have been detected due to the limited C-level sample sizes (15 C1 and only 8 C2 speakers). On the contrary, it would be logical to suspect that they do exist, especially as the limits between competence levels are not clear-cut and those are best viewed as a continuum. Yet – at least at higher levels – the results hint that a change of more than one level up or down may be necessary for the differences in comprehensibility and foreign-accentedness to become apparent in statistical analyses. This might suggest the B2, C1 and C2 descriptors might need reconsidering and further research to draw clearer boundaries between them.

As explained earlier, the CEFR (Council of Europe, 2001, p. 117, Table 1) as well as the 2018 Companion (Council of Europe, 2018, p. 136, Table 2) suggest little to no difference in the phonological "control" between C1 and C2 levels. However, the findings in our study suggest that it might exist although further research is needed to corroborate this claim. It is true that the tests were not sensitive enough to detect B1-C1 differences, they did, however, find B2-C2 differences, which might be interpreted as evidence that there is an underlying and yet undescribed C2 phonological competence as recognised by listeners. It is noteworthy that a foreign accent is mentioned as "sometimes evident" by the CEFR only up to a B1 level and is not referred to at all in the higher-level descriptors from 2001. This study suggests that foreign-accentedness is at play and evolving at higher levels as shown by the post-hoc ANOVA tests. This might tentatively seem to align with the 2018 descriptors. Yet – upon closer inspection – it is hard to establish any clear-cut differences between the level of foreign accent in the CEFR descriptors: its presence is simply acknowledged throughout.

On the other hand, intelligibility tests rendered no statistically significant results depending on the speakers' levels. Considering that intelligibility within this research design was measured solely based on the number of segmental cues decoded by listeners, the results suggest that segmentals are not sufficient to tell apart different levels of competence. Quite interestingly, this could broadly confirm CEFR descriptors, where a "clearly intelligible" pronunciation is achieved by level B1. However, such interpretation would be misleading and only possible if the term *intelligibility* were vaguely interpreted as what is considered *comprehensibility* in the paradigm followed in this research. And even so, comprehensibility differences were detected here as shown in the previous paragraph. This might mean that – while a person's speech is actually intelligible – a stronger accent could be subjectively perceived as harder to understand.

Regarding the type of degree course, the speakers were enrolled on, no evidence was found for any differences in pronunciation performance (measured as COM-FA-INT) between EMI and NON-EMI groups. Similarly, no differences were found between those students who had been schooled in CLIL contexts before university and those schooled exclusively in Spanish. Although the environment of this research is the university and not primary or secondary CLIL, this study's findings related to the lack of differences in pronunciation performance between EMI and non-EMI groups certainly confirm observations by Dalton-Puffer (2008, 2011), according to whom pronunciation is hardly affected by CLIL. Those observations have been borne out in numerous quantitative studies in the Spanish context, most of which come from the Basque country (Ruiz de Zarobe, 2007; Gallardo del Puerto et al., 2009; Rallo Fabra & Juan-Garau, 2015; Gallardo del Puerto & Gómez Lacabex, 2017). However, contradictory research results have also been reported in CLIL contexts (e.g. Lasagabaster,

2008). In other studies (Pérez-Cañado 2018; Pérez-Cañado & Lancaster, 2017; Rallo Fabra & Juan-Garau, 2010), the time of exposure to CLIL was found to be positively correlated with pronunciation performance. Although the amount of EMI exposure was not controlled for in this research, it might be an interesting factor to consider in prospective studies.

On the other hand, the sometimes-divergent findings of CLIL research may be quite confidently attributed to the disparity in the methods and assessment criteria. Thus, more unified, and pronunciation-oriented research would be required at all levels of education for definitive conclusions to be drawn. For the time being, most studies seem to corroborate Dalton-Puffer's (2011) statement about the scarce effects of bilingual education on pronunciation and in those where statistically significant relationships are reported, they are usually small. Some tentative explanations offered by previous researchers suggest it might be due not only to the non-native status of CLIL teachers (e.g. Gallardo del Puerto et al., 2009, p. 65), but also to the fact of the intelligibility benefit (Gallardo del Puerto & Gómez Lacabex, 2017, p. 46) in contexts where both teachers and pupils share the same mother tongue.

Be that as it may, it should not go unnoticed that CLIL and EMI may be considered incomparable altogether as many CLIL programmes start with the pupils having very low levels of competence in English (A1-A2)[5] while the level of English required to enrol on an EMI programme is usually a B2. Therefore, as CLIL pupils' basic English skills improve, so does their pronunciation, leaving the non-CLIL groups behind regarding both level and pronunciation as its correlate. In this study, the level distribution between both groups was the same[6], so it was assumed it could be discarded as a confounding variable before running statistical analyses. At the time of writing, only a few studies regarding EMI pronunciation performance amongst students were found (cf. Gómez Lacabex & Gallardo del Puerto, 2021; Richter, 2015) apart from a handful of studies referring to lecturer accent perception (cf. *Hellekjær,* 2010; Volín et al., 2018). While Richter (2015) did find an improvement in the degree of foreign-accent among EMI students, the Spanish study (Gómez Lacabex & Gallardo del Puerto, 2021) was rather listener-centred and did not compare EMI students with non-EMI speakers, which implies that no conclusions can be drawn on the impact of EMI classes as a variable affecting pronunciation. Whereas more research is clearly wanting, the analyses conducted for this study show that the findings regarding the limited or non-existent effects of CLIL on pronunciation can – at least for the time being – be extended onto the tertiary education context and EMI students.

Finally, one independent variable related to the participants' language exposure was analysed, namely, their experience in English speaking contexts abroad. In a first round of tests, no group differences were found between those who had spent some time abroad (YES group) and those who had not (NO group). To shed light on a possible interpretation of this finding, it should be borne in mind that the image which emerges from the descriptive statistics is that of a very heterogeneous group of people in the YES group encompassing a wide array of student profiles ranging from those who may have spent a couple of sum-

[5] In the Autonomous Region of Madrid, the bilingual programme started 2004 at the primary level and 2010 at the secondary level. A minimum A2 level is required for the students to enter bilingual programmes at the secondary level (Otto & Estrada, 2019, p. 32).

[6] Results of the level distribution difference test between EMI and non-EMI students render non-significant results ($U = 272.5$, $p = .441$, $N = 50$).

mer months abroad to people who lived abroad for years. Hence, it is understandable that, while group differences were insignificant, the very amount of time was still moderately correlated with less foreign-accentedness and increased comprehensibility with a high level of statistical significance. It was logical to at least tentatively conclude based on the data that those students who spent more time abroad performed significantly better while short stays (weeks or months) had no such consequential effects. The correlations were higher for comprehensibility than for foreign-accentedness. As this apparent incoherence between the *t*-test and correlation results required further examination, another series of tests were conducted to check whether those who spent at least one year or more time abroad obtained better pronunciation scores in the study. Those pairwise analyses shed further light on our previous findings and corroborated there were indeed statistical differences between those whose time spent abroad was equal to or exceeded one year and those who spent less or no time abroad. Those with more time exposure quite logically performed better, i.e., were less accented and more comprehensible. Just like in the case of correlations, the statistical significance of the difference in comprehensibility was higher than in foreign-accentedness. Moreover, the effect sizes were high for comprehensibility and medium in the case of foreign-accentedness. All these findings suggest that stays abroad (1 year+) are a better predictor of comprehensibility than of a lesser degree of foreign accent, although both these proxies for pronunciation performance are associated[7]. They are also partially aligned with previous research, notably Flege's (1988) suggestion that accents are established after a one-year period abroad. This early research inspired the seemingly arbitrary breakdown (1-year limit) in our analysis set-up. Indeed, our results provide evidence for an association between longer stays and an improvement in pronunciation: differences in COM and FA were detected for those who resided and spoke English abroad for 1 year or more but not if the groups were split between those who have spent 0 time abroad and the rest (Muñoz & Llanes, 2013). The amount of time spent abroad was also correlated with higher COM and lower FA. Also, Derwing and Munro's (2015) overview of the effects of the length of residence on pronunciation seems to be at least partially corroborated by our results keeping in mind the difference in contexts (i.e., the speakers' current place of residence, which is not a native speaking environment like in most of the quoted studies). However, the results herein should be interpreted with caution due to the fact that causation cannot be established without measuring the speakers' performance before and after their stay/study abroad period.

5. Conclusions

The main aim of this paper was to reveal potential associations between certain speaker characteristics (English level, EMI and CLIL exposures, experience abroad) and pronunciation performance measured through three independent variables: comprehensibility (COM), foreign-accentedness (FA) and intelligibility (INT).

After presenting and discussing the data, let us now review the research questions (RQs) formulated at the beginning of this paper. Regarding RQ1, the speakers' level turned

[7] The distribution of levels in both groups (for those who reported level) in this test was the same, $U = 335.5$, $p = 435$, $N = 50$, indicating that level was not a confounding variable in this case.

out to be associated with comprehensibility and foreign-accentedness. The level positively correlated with the former and was negatively correlated with the latter. The participants' educational background (RQ2) had no bearing on their pronunciation scores. However, the amount of time spent abroad correlated positively with comprehensibility and negatively with foreign-accentedness. All the data suggest intelligibility may be the most elusive of the three dependent variables and that subjective measurements (COM, FA) correlate better with advanced (B2-C2) speakers' characteristics than a more objective measurement of their intelligibility. This might be due to the intelligibility measuring method in the present research or other cognitive processing factors, e.g., due to the fact that INT was measured using a SUS task, which eliminates all contextual information and forces the listener to focus on the phonological component of the input and employ more bottom-up processing strategies. Comprehensibility and foreign-accentedness, however, were measured using meaningful excerpts which could be interpreted using top-down strategies as well.

Perhaps the most noteworthy insight of this study is that it cannot be expected that bilingual programme graduates will be significantly more phonologically proficient than those who have been schooled in Spanish. The fact that level is a fairly effective predictor of accentedness and comprehensibility seems to demonstrate that language proficiency is associated with better pronunciation unlike exposure to class content in a foreign language (EMI) or being taught content in a foreign language alongside that language (CLIL). Moreover, from a purely practical point of view, the recently revised CEFR descriptors might still be insufficient for Spanish learners of English. The data herein – showcasing differences exist between B2 and C2 levels – warrant their further revisiting. However, more research is needed due to the small C1 and C2 samples in this study, which limits the power of the conclusions that can be drawn from the data.

Lastly, as mentioned in the introduction, the research presented in this article tackles only variables derived from the differences between the speakers. The data related to listener-related variables and possible interactions between speaker and listener differences remain to be addressed in follow-up publications.

6. References

Avello, P., & Lara, A. R. (2014). Phonological development in L2 speech production during study abroad programmes differing in length of stay. In Pérez-Vidal, Carmen (ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 137-166). John Benjamins.

Avello, P., Mora, J. C., & Pérez-Vidal, C. (2012). Perception of FA by non-native listeners in a study abroad context. *Research in Language, 10*(1), 63-78.

Barrass, J. (2017). *The intelligibility of Korean English pronunciation from a lingua franca perspective* [Unpublished master's thesis]. University of Oxford.

Barrass, J. P., Baffoe-Djan, J. B., Rose, H., & Boggs, J. A. (2020). Intelligibility and comprehensibility of Korean English speakers' phonological features in lingua franca listening contexts. *The Journal of AsiaTEFL, 17*(1), 124-142. doi:10.18823/asiatefl.2020.17.1.8.124

Benoît, C., Grice, M., & Hazan, V. (1996). The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using semantically unpredictable sentences. *Speech Communication, 18*(4), 381-392.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second ed.). New York Psychology Press.

Council of Europe. (2001). *Common European framework for languages: learning, teaching, assessment*. Cambridge University Press.

Council of Europe. (2018). *Common European framework of reference for languages: learning, teaching, assessment. Companion volume with new descriptors*. Council of Europe.

Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2015). Second language comprehensibility revisited: investigating the effects of learner background. *TESOL Quarterly, 49*(4), 814-837.

Cutillas Espinosa, J. A. (2017). "A relaxing cup of Lingua Franca Core": local attitudes towards locally-accented English. *Atlantis (Salamanca, Spain), 39*(1), 11-32.

Dalton-Puffer, C. (2008). Outcomes and processes in content and language integrated learning (CLIL): current research from Europe. In W. Delanoy, & L. Volkmann (Eds.), *Future perspectives for English language teaching* (pp. 139-157). Carl Winter.

Dalton-Puffer, C. (2011). Content-and-language integrated learning: from practice to principles? *Annual Review of Applied Linguistics, 31*, 182-204.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: evidence from four L1s. *Studies in Second Language Acquisition, 19*(1), 1-16.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*(2), 478-490.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals. Evidence-based perspectives for L2 teaching and research.* John Benjamins.

Deterding, D. (2013). *Misunderstandings in English as a Lingua Franca: an analysis of ELF interactions in South-East Asia*. De Gruyter.

Diogo, S., Queirós, A., Carvalho, T., Manatos, M., & Soares, D. (2019). 20 years of the Bologna Declaration – a literature review on the globalisation of higher education reforms. *EDULEARN19 Proceedings* (pp. 5196-5204). 10.21125/edulearn.2019.1283

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

El Confidencial. (2014). Nuestro acento en inglés nos delata: estos son los errores más comunes. El Confidencial (24 January 2014). https://www.elconfidencial.com/alma-corazon-vida/2014-01-24/nuestro-acento-en-ingles-nos-delata-estos-son-los-errores-mas-comunes_79791/

Field, J. (2005). Intelligibility and the listener: the role of lexical stress. *TESOL Quarterly*, *39*(3), 399-424.

Fielden Burns, L. V., Rico García, M., & Hernández Linares, R. (2018). The growth of EMI tertiary programs in Spain: novel approaches to developing linguistic competence in instructors, *LSP in multi-disciplinary contexts of teaching and research. Papers from the 16th International AELFE Conference, Vol. 3* (pp. 40-46). https://easychair.org/publications/paper/tRR3. https://doi.org/10.29007/qpdg

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America, 84*(1), 70-79.

Foskett, N. (2010). Global markets, national challenges, local strategies. In F. Maringe, & N. Foskett (Eds.), *Globalization and internationalization in higher education. Theoretical, strategic and management perspectives* (pp. 35-50). Continuum.

Gallardo del Puerto, F., & Gómez Lacabex, E. (2017). Oral production outcomes in CLIL: an attempt to manage amount of exposure. *European Journal of Applied Linguistics, 5*(1), 31-54.

Gallardo del Puerto, F., & Martínez Adrián, M. (2013). ¿Es más efectivo el aprendizaje de la lengua extranjera en un contexto AICLE? Resultados de la investigación en España. *Padres y Maestros, 263*, 25-28.

Gallardo del Puerto, F., Gómez Lacabex, E., & García Lecumberri, M. L. (2009). Testing the effectiveness of content and language integrated learning: the assessment of English pronunciation. In Y. Ruiz de Zarobe, & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning. Evidence from research from Europe* (pp. 63-80). Multilingual Matters.

Galván, F. (2010). Por qué a los españoles se nos da mal el inglés. El País (9 July 2010). https://elpais.com/diario/2010/07/19/opinion/1279490411_850215.html

Gilbert, J. B. (2010). Pronunciation as orphan: what can be done? *Speak Out! IATEFL Pronunciation Special Newsletter, 43*, 3-7.

Gómez Lacabex, E., & Gallardo del Puerto, F. (2021). Pronunciation in EMI. A prelimiary study of Spanish university students' intelligibility and comprehensibility. In D. Lasagabaster, & A. Doiz (Eds.), *Global perspectives on language aspects and teacher development in English medium Instruction* (pp. 126-144). Routledge, Taylor & Francis.

Gooskens, C. (2013). Experimental methods for measuring intelligibility of closely related language varieties. In R. Bayley, R. Cameron & C. Lucas (Eds.), *The Oxford Handbook of Sociolinguistics* (pp. 195-213). Oxford University Press.

Haugen, E. (1966). Semicommunication: The language gap in Scandinavia. *Sociological Inquiry 36*, 280–97.

Hellekjær, G. O. (2010). Language matters: assessing lecture comprehension in Norwegian English-medium higher education, *Language use and language learning in CLIL classrooms* (pp. 233-258). John Benjamins.

IBM Corp. (2020). IBM SPSS [computer software]. IBM.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.

Jeong, H. J., Thorén, B., & Othman, J. (2017). Mutual intelligibility of Malay- and Swedish-accented English: an experimental study. *Indonesian Journal of Applied Linguistics, 7*(1), 43-53.

Joto, A., Nagase, Y., & Funatsu, S. (2007). Effect of VOT on intelligibility of English voiceless stops produced by native speakers of Japanese [Conference Presentation]. PTLC2007: Phonetics Teaching and Learning Convention, London, United Kingdom. http://www.phon.ucl.ac.uk/ptlc/proceedings/ptlcpaper_31e.pdf

Juan-Garau, M. (2014). Oral accuracy growth after formal instruction and study abroad: onset level, contact factors and long-term effects. In Pérez-Vidal, Carmen (ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 87-109). John Benjamins.

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk, & H. Widdowson (Eds.), *English in the world: teaching and learning the language and the literature*. Cambridge University Press.

Kang, O., Thomson, R. I., & Moran, M. (2018). Empirical approaches to measuring the intelligibility of different varieties of English in predicting listener comprehension. *Language Learning, 68*(1), 115-146. doi:10.1111/lang.12270

La Razón. (2019). Los españoles son los europeos que peor hablan inglés. La Razón (13 July 2019). https://www.larazon.es/familia/los-espanoles-son-los-europeos-que-peor-hablan-ingles-PO24191608/

Lasagabaster, D. (2008). Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal, 1*(1), 30-41.

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press.

LimeSurvey Gmbh. (2019). *LimeSurvey: An Open Source Survey Tool.* https:// www.limesurvey.org.

Lippi-Green, R. (1994). Accent, standard language ideology, and discriminatory pretext in the courts. *Language in Society*, *23*(2), 163-198.

Meierkord, C. (2004). Syntactic variation in interactions across international Englishes. *English World-Wide, 25*(1), 109-132.

Merino, J. A., & Lasagabaster, D. (2018). The effect of content and language integrated learning programmes' intensity on English proficiency: A longitudinal study. *International Journal of Applied Linguistics, 28*(1), 18-30.

Minematsu, N., Okabe, K., Ogaki, K., & Hirose, K. (2011). Measurement of objective intelligibility of Japanese accented English using ERJ (English Read by Japanese) Database, *Interspeech Proceedings,* Florence, Italy. (pp. 1481-1484).

Moore, I., Torgerson, C., & Beckmann, N. (2021). Systematic review measuring the efficacy of study abroad in undergraduate language learners on linguistic proficiency gains. *Review of Education*, 9(3).

Muñoz, C., & Llanes, À. (2014). Study abroad and changes in degree of foreign accent in children and adults. *The Modern Language Journal, 98*(1), 432-449. https://doi:10.1111/j.1540-4781.2014.12059.x

O'Neal, G. (2015). Interactional intelligibility: the relationship between consonant modification and pronunciation intelligibility in English as a lingua franca in Japan. *Asian Englishes, 17*(3), 222–239.

Otto, A., & Estrada, J. L. (2019). Towards an understanding of CLIL assessment practices in a European context: main assessment tools and the role of language in content subjects. *CLIL Journal of Innovation and Research in Plurilingual and Pluricultural Education, 2*(1), 31-42. https:// doi:10.5565/rev/clil.11

Pérez Cañado, M. L., & Lancaster, N. K. (2017). The effects of CLIL on oral comprehension and production: a longitudinal case study. *Language, Culture, and Curriculum, 30*(3), 300-316.

Pérez-Cañado, M. L. (2018). CLIL and educational level: A longitudinal study on the impact of CLIL on language outcomes. *Porta Linguarum, 29*, 51-70.

Poljak, L. (2019). Trends in accentedness and comprehensibility research with respect of L2 speech ratings. *SFU Educational Review, 12*(2), 36-63.

Rallo Fabra, L., & Jacob, K. (2015). Does CLIL enhance oral skills? Fluency and pronunciation errors by Spanish-Catalan learners of English. In M. Juan-Garau, & M. Salazar Noguera (Eds.). *Content-based language learning in multilingual educational environments* (pp. 163-177). Springer.

Rallo Fabra, L., & Juan-Garau, M. (2010). Intelligibility and foreign accentedness in a context-and-language-integrated-learning (CLIL) setting [Conference Paper]. *6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 373-378).

Ramos García, A. M. (2013). Higher education bilingual programmes in Spain. *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras, 19*(19), 101-112. https:// doi:10.30827/Digibug.29632

Ramos García, A. M., & Pavón Vázquez, V. (2018). The linguistic internationalization of higher education: a study on the presence of language policies and bilingual studies in Spanish universities. *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras, Monográfico III,* 31-46.

Richter, K. (2015). *English-medium instruction and pronunciation a longitudinal case study of Austrian business students at the University of Applied Sciences Vienna.* [PhD Dissertation]. University of Vienna.

Ruiz de Zarobe, Y. (2007). CLIL in a bilingual community: similarities and differences. *Vienna English Working Papers, 16*(3), 47-52.

Ruiz de Zarobe, Y. (2011). Which language competencies benefit from CLIL? An insight into applied linguistics research. In Y. Ruiz de Zarobe, J. M. Sierra & F. Gallardo del Puerto (Eds.). *Content and foreign language integrated learning: contributions to multiliingualism in European contexts* (pp. 129-154). Peter Lang.

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics 37(2),* 217-240.

Smith, L. E., & Nelson, C. L. (1985). International intelligibility of English: directions and resources. *World Englishes, 4*(3), 333-342.

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences, 1*(21), 19-25.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism, 15*(4), 905-916.

Volín, J., Skarnitzl, R., & Henderson, A. (2018). Perceptual impact of foreign-accented speech. In J. Volín, & R. Skarnitzl (Eds.), *The pronunciation of English by speakers of other languages* (pp. 73-94). Cambridge Scholars Publishing.

Walker, R. (2010). *Teaching the pronunciation of English as a lingua franca*. Oxford University Press.

Wang, H. (2007). *English as a lingua franca: Mutual intelligibility of Chinese, Dutch and American speakers of English*. LOT.

7. Appendix

Semantically Unpredictable Sentences (SUS) (adapted from Wang, 2007)

1. The state sang by the long week.
2. The man lay through the wide war.
3. The day hung to the great night.
4. The year smiled through the young head.
5. The time breathes with the high side.
6. The way ran of the hot room.
7. The thing hung from the small line.
8. The grass lied on the blue night.
9. The school stayed for the new tube.
10. The hand fell of the high form.
11. The real field made the vote.
12. The white home got the art.
13. The clear friend brought the ground.
14. The black van held the zone.
15. The whole month bans the air.
16. The thin job got the road.
17. The poor vet hit the tax.
18. The short field said the youth.
19. The full home took the term.
20. The white sense ate the zoo.

21. Use the game or the hair.
22. Ask the trial and the tree.
23. Leave the sport and the thought.
24. Call the club and the growth.
25. Turn the love or the test.
26. Spend the sale or the nose.
27. Start the age or the price.
28. Show the plant or the sound.
29. Stop the stock and the list.
30. Live the sport and the fund.
31. When does the charge like the late plane?
32. Where does the band sell the low set?
33. Why does the cell like the deep length?
34. When does the gun like the deep bed?
35. Why does the range watch the fine rest?
36. When does the sign lead the red roof?
37. How does the voice plan the cold fear?
38. How does the chance send the deep roof?
39. Why does the juice bear the jump trade?
40. How does the cloud watch the low text?

Elicitation paragraph (originally designed for this research)

The sun was rising slowly and the birds were singing. The view from the top of the hill was amazing. Susan turned round and stepped heavily into the kitchen. She took a plastic cup and filled it with juice. Then, she sat down on a chair and started thinking what food she would cook for lunch. Suddenly, her dog, Zoe, jumped onto her, spilling her drink. She threw the empty cup away into a small waste bin. She hadn't been feeling good for years. Her job as a university nurse didn't give her pleasure and she hated starting in the early morning. At her age, it wasn't easy to deal with people. As she couldn't hear well, she was getting used to being on her own. She decided to skip work, stay in and enjoy the first day of spring. Then she remembered it was Saturday anyway.