Index:

Resumen

Evaluar una prueba oral de inglés no es una tarea fácil. ¿Qué criterios tienen los examinadores en cuenta al valorar un examen oral de inglés? ¿Siguen los examinadores de distintos centros los mismos criterios? Este trabajo analiza qué aspectos consideran los expertos de inglés hoy en día al evaluar una prueba oral de inglés. Para ello, les pedimos a cuatro examinadores (2 especializados en exámenes de Cambridge y 2 de la Escuela Oficial de Idiomas) que valorasen as doce candidatos a los que previamente les habíamos hecho un examen oral. Posteriormente, los cuatro examinadores rellenaron un cuestionario sobre los criterios que habían seguido al evaluar a dichos candidatos. Los resultados mostraron una gran discrepancia tanto en la importancia que los examinadores dieron a los criterios de cómo en la evaluación de los candidatos. Así pues, como conclusión, consideramos que los descriptores del Marco Común Europeo de Referencia para las lenguas que los examinadores siguieron en los test orales no produjeron resultados estandarizados ni uniformes. Por ello, estas pautas deberían ser revisadas e implementadas con directrices más claras que los centros pudieran seguir de forma homogénea.

Palabras clave: criterios, examinadores, Cambridge, Escuela Oficial de Idiomas, Marco Común Europeo de Referencia para las lenguas, candidatos.

Abstract

Grading an oral English test is not an easy task. Which criteria do examiners take into consideration when assessing an oral test in English? Do examiners from different centres follow the same criteria? This work analyses what aspects English professional raters consider when assessing an oral test in English. Therefore, we asked four examiners (2 specialised in Cambridge examinations and 2 from the Official Language School) to rate twelve videotaped candidates who had been previously undergone an oral test. Afterwards, the four examiners filled in a questionnaire about the criteria that they had followed when evaluating the candidates. The results showed a big discrepancy both in the importance that examiners gave to the criteria and on the ratings of the candidates. To conclude, this study argues that the Common European Framework of Reference descriptors that examiners follow in oral tests do not seem to provide standardised and uniform results and so they should be revised-and implemented with clearer guidelines for accreditation centres to follow.

Keywords: criteria, examiners, Cambridge, Common European Framework of Reference descriptors, candidates.

1.Introduction

1.1. Motivation

What do we have to consider when grading an oral test in English? This issue is still far from clear for many experts on English language testing. According to the Common European Framework of Reference (CEFR)– which the standard scale that language centres theoretically follow-, examiners should consider five aspects or criteria[1]: a) range, which is related to the extension of the vocabulary; b) accuracy, which refers to the control and appropriate use of grammatical structures; c) fluency, which regards how spontaneously and smoothly the speaker can express himself/herself; d) interaction, which refers to the ability to interweave your contribution with suitable turn taking and cues; and e) coherence, which makes a referral to how well-structured the speech is. But do examiners really follow it to the letter?

From my own experience, I can say that it caused me some difficulties. In October 2016, I had to assess some students' oral proficiency in English who were doing a B.S. in Medicine and it was not an easy task, chiefly because I did not have any model or guideline to follow. A total of 22 candidates[2] were all applying for a scholarship to spend a couple of months abroad and they needed their spoken English to be graded. Therefore, the mark I gave them was decisive. The person who hired me asked me to do the oral test in pairs, but nothing else. Hence, I prepared two different tasks to be fulfilled in pairs and four different individual tasks. Then they had to choose randomly one of the in-pair tasks and two of the individual tasks, one per person. I used as a reference a PET Cambridge examination book. Another problem I had was that the candidates' English proficiency was very varied. There were some candidates whose English level was not better than a B1 (at least from my point of view), and some others whose English was a C2; in fact, there was a bilingual guy. The maximum mark that somebody could get was 100%, which was equivalent to a C2 (from 85% to 100%). Thus, sometimes it was very difficult to grade candidates whose English proficiency was similar, since the differences between one level and another (from A1 to A2 or B1 to B2, etc) were just of 15%-20%.

[1]All the definitions that will be given definitions on belong to the CEFR. *Common European Framework of Reference for Languages: Learning, teaching, assessment. Structured overview of all CEFR scales. 1.3. Qualitative aspects of spoken language use, 7.*

[2]Other terms such as test-takers, examinees or interviewees can be also used to refer to candidates taking a test, but I will always use the same terminology in this work, i.e. *candidate*, to avoid ambiguity.

As I did not have to follow any general rule when grading and by that time I was not aware of the qualitative aspects of the CEFR descriptors either, I made up my own template which would be used to take notes while listening to the candidates. The aspects that I took into consideration for the final mark were basically four. Hence, the template I used was divided into 4 sections: 1) Pronunciation, which I considered related to phonetics. Here, I also paid attention to the intelligibility; 2) The use of grammar in general together with lexical accuracy,- which would correspond to the accuracy and range aspects in the CEFR-, here the use of idioms or technical terms used accurately would give the candidate some extra mark in the evaluation of this section; 3) Fluency, here hesitations and corrections would penalise the candidate in this section; and 4) Discourse cohesion and coherence, i.e. how well-structured the speech was and how properly the candidate used connectors (which would correspond to coherence in the CEFR). Consequently, the correct use of connectors and conjunctions would be favourably viewed. However, I overlooked some aspects such as interaction that, as stated in the CEFR, is one of the 5 criteria to consider. I assumed that interaction had more to do with somebody's personality or even with the culture, as the conception of respect, turn-taking and silences vary from nation to nation. On top of that, what if the candidate made an exam on his/her own? Would we be able to assess him or her fully, even if there was no interaction? This is another issue I thought about and that made me research it deeper. According to the CEFR, this element is related to the ease and skill to make contact and cooperate with other speakers. Besides, this aspect seems to give importance to appropriate turn-taking.

It is likely that I gave too much importance to the pronunciation and to the accent of the candidate. But from my point of view, pronunciation played a key role in the oral exam. When I listened to someone whose English accent is very native-like, I assumed that his/her English in general was fine as well. I thought that having a good accent and a good pronunciation would give you a big advantage in an oral exam. In contrast, and it drew my attention, the CEFR does not place any importance on pronunciation, or at least they do not consider it as one of the five main criteria.

1. 2. The state of the question

The guidelines of every European centre should be based on the descriptors proposed by the CEFR. After having looked at the qualitative aspects of spoken language in the CEFR, the criteria were a bit clearer, but they still caused me some confusion since the terminology used

is not exactly the same as the terminology from the Cambridge Institutions other accreditation centres or even my own conception. For example, what I called *variety of vocabulary* in the template I used for the students of Medicine is called *range* in the CEFR. But both terms refer to the same thing, which is the extension/width of the speaker's lexis. Likewise, what I named use of grammar is what the CEFR descriptors call accuracy. Hence, I noticed that this wide variety of terms may sometimes lead to confusion to non-expert examiners, mainly with the aspects of *range* and *accuracy*, as it happened to me when rating candidates from the B.S. of Medicine. There are plenty of aspects to bear in mind when rating a student's spoken level of English. The definitions vary slightly from some guidelines to others and there is not much research focusing on the criteria, apart from some papers on fluency, which will be presented in the following paragraphs. As stated by the CEFR descriptors, fluency is:

> He or she can express him/herself at length with a natural, effortless, unhesitating flow. Pause sonly to reflect on precisely the right words to express his/her thoughts or to find an appropriate example or explanation. (Council of Europe, (n.d). Common European Framework of Reference for Languages: Learning, teaching, assessment. Structured overview of all CEFR scales)

Another definition of the concept of fluency was provided by Lennon (2000: 26), who defines it as "the rapid, smooth, accurate, lucid and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing". Segalowitz (2000) claimed that there are three types of fluency: cognitive fluency[3], utterance fluency[4], and perceived fluency[5]. At the same time, utterance fluency contains three different aspects of fluency as well (Skehan 2003; Tavakoli and Skehan 2005): breakdown fluency[6], speed fluency[7] and repair fluency[8]. Moreover, De Jong et al. (2012) argued in favour of the three kinds of fluencies mentioned before, the cognitive fluency, the utterance one and the perceived one.

[3] Cognitive fluency is defined by Segalowitz (2010: 225) as "the ability of the speaker to smoothly translate thoughts into speech"

[4] Utterance fluency is used by researches to measure speech-planning difficulties that surface in utterance by counting the number of filled pauses, corrections, and repairs, and by measuring the duration of pauses. (De Jong: 2012: 225).

[5]Perceived fluency pertains to the inference (raters) made on the basis of the utterance about the speakers' ability (i.e. about the speakers' cognitive fluency) (De Jong 2012: 225).

[6]"For breakdown fluency, the number and length of silent pauses were measured as well as the number of non-lexical filled pauses (such as "uh", "uhm", "er", "mm")" (De Jong 2012: 226).

[7]S*peed fluency (why in italics?) measured* the mean duration of syllables" (De Jong 2012: 226).

[8]"R*epair fluency deals with* the number of repetitions and the number of corrections measured" (De Jong 2012: 227).

De Jong et al. (2012) carried out an experiment about the relation between the cognition of a language and its speaking fluency. The paper deals with the question of whether second language (L2) measures of oral fluency, such as number of filled pauses, which should be adjusted to first language (L1) fluency behaviour to reflect L2-specific processing. As we will be dealing with non-native English speakers, it is necessary to know whether the L1 may influence on the L2. De Jong et al.'s (2012) experiment can give us an idea about whether L1 influences on L2. It is important to highlight that they focused on L2 utterance fluency- tested two groups of people whose L1s were very different from each other: English and Turkish.
The L2 of both groups of people was Dutch. The researchers considered that English is typologically close to Dutch, while Turkish is distant. With this study, they wanted to analyse the possible differences between L1 and L2 fluency behaviour. To find that out, the candidates performed tasks in their L1 and much the same in their L2, Dutch, but they avoided literal repetitions. After analysing the results, an important language effect was found. Overall, both Turkish and English participants were less fluent in their L2 than in their L1, but both were equally fluent in Dutch. However, there were differences in their L1 fluency. For example, English syllable duration in L1 was, on average, longer than Turkish syllables in L1. Meanwhile, in the L2 they did not show any difference in the syllable duration. The same happens with silent pauses per second, where the Turkish did fewer pauses in their L1, but the same in the L2 as English native speakers. Turkish speakers also produced fewer repetitions per second in their L1 than English native speakers, but, once again, no differences were shown in their L2.

The fact that English speakers produced more pauses than Turkish speakers was due to the length of Turkish and English words (De Jong et al. 2012). Turkish words are generally longer than English ones; therefore, for them, Turkish speakers find fewer moments to rest. Similarly, they came to the conclusion that English speakers probably produce more pauses because, having shorter words, they have bigger chances to repeat a word, since "stopping mid-word is non-preferred" (Levelt 1983: 239).

Considering what this research claims, we could assume that L1 behaviour does not influence sufficiently on L2 behaviour. This paper has helped us to know better one of the 5 main criteria to assess in an oral test, according to the CEFR. However, there are other 4 main criteria we did not find articles or studies about. Thus, in this research we will ask examiners

from different centres about their definition or perception of the other criteria.

Even though fluency is considered one of the main five criteria in the CEFR, not every accreditation centre embraces its descriptors in the same way. For example, the International English language Testing System (IETLS) does not follow the CEFR to the letter. The IETLS considers the four following criteria: a) Fluency and coherence, b) lexical resources[9], c) grammatical range and accuracy, and d) pronunciation[10]IETLS SPEAKING: Band descriptors (public version). (British Council, 2015). They place pronunciation as one of the main four criteria - CEFR does not -, overlooking an important aspect for the CEFR as is interaction. Additionally, IETLS puts together aspects that the CEFR proposes to mark separately; this is the case of fluency and coherence. What is more, IETLS considers as one aspect grammatical range and accuracy, and another one lexical resources. IETLS uses different terms from the CEFR, for example the latter does not split range into two criteria as IETLS does, which speaks instead about lexical resources and grammatical range. Although in the end they both consider somehow both aspects, they use different terms to refer to them. Therefore, as the experts refer to the same aspects with different terms and definitions, especially to range and accuracy, experts on the subject will be asked about it in this study.

The variation of terminology and aspects taken into consideration do not happen only in the IETLS system, but also in the two centres we will focus this research on, which are Cambridge ESOL examinations and Official Language School (EOI from now onwards)[12]. It is obvious that languages are dynamic and change throughout history, but what was not that obvious and what is actually happening is that the criteria examiners consider change too. In 2009, experts in Cambridge ESOL examinations were supposed to follow to the letter the five aspects that the CEFR proposed regardless of the level they were rating. Examples of speaking performance at CEFR levels A2 to C2 (University of Cambridge ESOL Examinations, 2009).

Nonetheless, in 2011 new guidelines were formulated to be followed by Cambridge experts and the criteria varied slightly depending on the level. We will focus on the B1, B2, C1, and

[9]Lexical resources refer basically to range and vocabulary accuracy
IETLS introduces within the pronunciation criterion the aspect of intelligibility. Accordingly, if the candidate's speech is often unintelligible his or her pronunciation will be rated very low (2 points in the 9point band).
[12] In Spanish, Escuela Oficial de Idiomas (EOI).

C2 levels[13]. For the B1 examination, 4 criteria need to be considered: a) grammar and vocabulary, equivalent to range and accuracy in the CEFR; b) discourse management, which considers fluency, coherence, and cohesion; c) pronunciation, including intelligibility and intonation; and d) interactive communication, equivalent to interaction in the CEFR. Assessing Speaking Performance – Level B1 (Cambridge English Language assessment, 2011). In the B2, they consider the same aspects as in the B1. In the C1 and the C2, the criteria they consider are the following: a) grammatical resources b) lexical resources, c) discourse management, d) pronunciation, and e) interactive communication. We can see how in the two highest levels they split grammar and vocabulary into 2 different aspects: grammatical resources and lexical resources**,** giving now more importance to them. Assessing Speaking Performance – Level C1 (Cambridge English Language assessment, 2011).
Assessing Speaking Performance – Level C2 (Cambridge English Language assessment, 2011).

It draws our attention that Cambridge ESOL gives more importance to grammar in the C1 and C2 than in lower levels. However, Maes (2011) claimed that different authors agree on the fact that grammar or grammar rules should be taken into account only in the four lowest levels, i.e. A1, A2, B1, and B2, and that in the C1 and C2 levels "grammar knowledge does not play an important role anymore. In other words, at B2 level, language learners already have to possess all the grammatical knowledge in order to be able to express themselves at the higher C1 and C2 level" (2011: 2). This is another issue to be considered; since depending on the level examiners are rating, they should give more importance to one aspect than to another. However, should grammar not play a significant role in the C1 and C2 levels? We will ask examiners from different centres.

At present, the criteria for the speaking tests from Cambridge ESOL are slightly different from the ones in 2011. For example, in the First Certificate (B2 level):

> The assessor gives 0-5 marks for each of the following criteria: Grammar and vocabulary; Discourse management; Pronunciation; and Interactive Communication. Marks for each of these criteria are doubled. The interlocutor gives a mark of 0-5 for Global Achievement. This

[13]Cambridge uses a different terminology to the CEFR when referring to the different levels. In this paper, we will be using the one from the CEFR to avoid ambiguity. The Preliminary test in the Cambridge Scale refers to the B1; the First Certificate in English to the B2; Cambridge English: Advanced to the C1; Cambridge English: Proficiency to the C2. The Cambridge English Scale explained. A guide to converting practice test scores to Cambridge English Scale scores (n.d.).

> mark is multiplied by four. Examiners may award half marks. Marks for all criteria are then combined, meaning there are 60 marks available in the Speaking test. The Cambridge English Scale explained (Cambridge English Language Assessment, n.d.).

Nowadays, they pay attention to the five criteria. And they have added global achievement, which is given even more importance than the rest, as you can get 20 points in this aspect and 10 in the other 4. This criterion basically refers to how the candidate handles communication on the topics, how fluent s/he is and coherent the discourse is, and how accurate the linguistic resources are. Therefore, the aspects of accuracy, coherence/cohesion, and fluency are important within this criterion. In C1 and C2 levels the Cambridge English Scale asserts the following:

> The assessor gives 0-5 marks for each of the following criteria: Grammatical Resource; Lexical Resource; Discourse Management; Pronunciation; and Interactive Communication. Marks for each of these criteria are doubled. The interlocutor gives a mark of 0-5 for Global Achievement. This mark is then multiplied by five. Examiners may award half marks. Marks for all criteria are then combined, meaning there are 75 marks available in the Speaking test.
> The Cambridge English Scale explained (Cambridge English Language Assessment, n.d).

So in these levels Cambridge examiners still split grammar and vocabulary as they did in the guidelines from 2011. Furthermore, they give more importance to global achievement, a criterion, together with pronunciation, which is not proposed by the CEFR. Although Cambridge examiners give different names to the criteria, they follow generally the CEFR as a reference.

Next, we will analyse the criteria followed by the other centres we will focus upon: the EOI. Experts in the EOI examinations take theoretically as a reference the CEFR, but they do not follow it to the letter either. They contemplate five criteria set out in the following way:

1) The grade of achievement[14]. This aspect refers to the fulfilment of the task caring about the candidate using the right register and form too.
2) Coherence/ cohesion, strategies of communication, and fluency[15]. Here they look at how well-structured the speech is, the appropriate use of connectors, the turn-takings when interacting, and the cohesion of the phrases.

[14]What they call in Spanish "adecuación".
[15]They refer to this in Spanish as "coherencia/ cohesion", "estrategias comunicativas" and" fluidez".

3) Pronunciation and intonation rate the pronunciation, how intelligible speakers are whether they use the right intonation, for example using the questioning intonation when asking a question.
4) Grammatical range and accuracy.
5) Finally, they consider the lexical range and accuracy.

As we can see, the EOI includes the five aspects proposed by the CEFR, but they approach them differently. In addition, they include other aspects not proposed by the CEFR such as pronunciation, intonation, and the grade of achievement. Another fact worth mentioning is that they do not give the same importance to every aspect. For the task achievement criterion, there are three possible marks: 10, 5 or 2, depending on how consistent the candidate is; where the higher the number, the higher the mark. For coherence/cohesion, strategies of communication, and fluency, the candidate can be rated with 15, 7 or 2 points. In the pronunciation and intonation criterion, s/he can be rated with 15, 8 or 2 points. For the last two criteria, they suggest 5 possible scores: For the grammatical range and accuracy, the rates can be 30, 23, 15, 8 and 3, being 30 excellent; whereas for the lexical range and accuracy,
they can be 30, 22, 15, 7, 3, with 30 being the highest. Thus, the EOI rates are noticeably different from the CEFR.

On the other hand, Cambridge ESOL examinations in a way try to adapt their criteria to the level being rated. Nonetheless, the EOI uses the same 5 criteria for every single level.
Obviously, the requirements vary from one level to another. We could also mention that the EOI in Cádiz follows the same pattern for all the levels and for all the languages taught in their centres (German, French, Italian and English). Pruebas terminales específicas de certificación. Guía del profesorado 2016/2017 (Junta de Andalucía, 2017).

1.3. Hypotheses

As we have seen, the criteria to take into consideration vary depending on the centre; even if both Cambridge and the EOI take the CEFR as a reference, there are noticeable differences between them. Therefore, we will put EOI and Cambridge examiners[16] to the test. Our

[16]Although the correct name would be examiners from Cambridge ESOL examination centres, we will name "Cambridge examiners" as a shortening for the full name.

questions are the following: Will the examiners from different accreditation centres follow their official guidelines when grading the exams? Or will they be carried along by impressions? And even if they follow different patterns, will they give the same mark to the candidates in the end? The best way to find out what experts consider when rating an oral test is by basing the research on the examiners themselves, and that is what we will deal with in this study. Expert examiners from Cambridge examinations and from the EOI will be asked to rate some students, according to the criteria from their centres. Our hypotheses in this study are the following:

1) If they place different importance on the criteria we propose in the questionnaire, the same candidate will receive different marks. The bigger the difference on the importance given to the criteria, the bigger the difference will be on the rates given to the candidates.

2) If examiners belong to the same centre, they will place similar importance to the criteria, and as a consequence the rates of the candidates will be similar too; bigger differences will arise among examiners from different centres.

3) Cambridge examiners will give more importance to some aspects than others depending on the level they are grading as their guidelines reflects, while for the EOI we expect them to give the same importance to the same five criteria for every level as the EOI guidelines to oral assessment state.

4) Cambridge examiners will give more importance to the aspects of accuracy, coherence/cohesion, and fluency than to pronunciation or interaction since the former determine the mark of the most important criterion, i.e. global achievement, as stated in the Cambridge guidelines.

5) For Cambridge examiners task fulfilment will not be as important as the other criteria as it is not included in their guidelines. Even though task fulfilment is not included among the 5 scoring criteria, Cambridge examiners will give some importance to it because when I was working with Cambridge examiners they stressed the importance of the fact that the candidates had to answer to the task provided

6) EOI examiners will consider more important range and accuracy than the other criteria

because in their official guidelines these two criteria can be rated up to 30 points, whilst the rest are assigned lower marks: task fulfilment with 10; pronunciation and intonation with 15; and coherence/cohesion, strategies of communication and fluency with 15.

7) As stated in the state-of-the-question section, experts use different terms to refer to the same idea. We believe that, though similar, they may not be on full agreement with the definitions of some criteria they will be given.

8) We assume that no other aspect apart from the eight proposed above -i.e. range, accuracy, interaction, coherence, fluency, pronunciation, intelligibility, and task fulfilment - will be taken into consideration by examiners from both centres.

In the next section, the methodology applied will be presented as well as the materials used in this study. Section three presents and discusses the results. Section four deals with the conclusions. Finally, the bibliography is provided in Section five followed by the appendices.

2. Materials and Methods

2.1. Participants

The participants in this study were 12 volunteer students from the University of Cadiz and 4 volunteer experts on English oral examinations. 9 out of the 12 students were Erasmus from four different nationalities who were studying for a semester or a year at the University of Cadiz: 3 female Polish students, 3 German students (2 females and 1 male), 3 Ukrainian students (2 females and 1 male), and 3 male Spanish students. All the participants were aged between 20 and 25 years old with an upper intermediate (B2) or advanced level of English (C1). In fact, all of them had the B2 level certificate, or at least they had completed the course Inglés Instrumental IV, equivalent to B2 (although it does not certify the level). In order to find non-native Spanish students, we asked for the 2016-17 list of Erasmus students at the University of Cádiz. We contacted some of them via e-mail and some others face to face. A few of them rejected to participate, which is why we could not gather the same amount of female and male students.

The examiners belonged to different official accreditation centres : 2 of them to the Cambridge examination centres (2 female examiners) and the other 2 to the EOI examination centres (1 female and 1 male examiner)[17]. They were all Spanish aged between 26 and 50 years old. It is also worth mentioning that one of the examiner from Cambridge is specialised in C1 and C2 examinations and the other one from this centre in B2 and C1. EOI examiners are experts of A1 and B1 one of them, and of B1 and B2 the other one. As I have recently worked with Cambridge examiners, it was not difficult to get them on this project. Contacting the EOI examiners was not difficult either, as I had studied in the EOI from 2009 to 2014, therefore I knew some teachers who would be willing to cooperate and accept to do the assessment.

2.2. Materials

Considering that for this research two experiments were performed -one with candidates and another with examiners-, this section will be divided in two parts as well: candidates' and examiners'.

[17]Even though the gender variable has not been taken into account, we have included it as a part of the data.

To carry out the candidate's test, they were first asked some icebreaker questions such as "What did you do yesterday?", "What are your plans for the following week?", "Why do you study English", "When did you start studying English?", "What do you like doing in your free time?"[18]. Then, an exam from the Cambridge book of B2 certificate was chosen, although just the interaction (part 3) was selected for the 12 students/candidates to do it (See Appendix 1). This part 3 was basically a prompt with words of different items which most people consider important nowadays (car, credit-car, hairdryer, mobile phone) and they had to discuss it. We chose this part from a Cambridge exam because I was more familiar with them than with EOI examinations, and for me Cambridge examinations were more accessible as I was working with them at that time.. An informed consent was elaborated for the candidates and administered to them before videotaping them (see Appendix 2).

For the examiners, we prepared an assessment sheet (See Appendix 3). On the assessment sheet examiners could rate the candidates with the following marks: B1, B1+, B2, B2+, C1, C1+, and C2. The piece of paper was divided into two parts one per candidate[19],, with a reference to their genre for examiners to be able to rate them, for example: Candidate 1 (girl) and Candidate 2 (boy). In the cases where the genre of both candidates was the same, a little description of them would be written referring to the colour of the hair and their outfit, for example: Candidate 1 (blonde girl in black jacket) & Candidate 2 (brunette girl in blue shirt). A wide area was left for the examiners to add comments about the candidate's mistakes or skills, just in case they wanted.

Moreover, a Likert questionnaire was also created for the examiners (see Appendix 4). The questionnaire contained 5 questions about the videotaped exams that they had previously rated and two questions about their professional experience. In the first question, the examiners had to rate in a 7-point Likert scale the importance they gave to different aspects when they rated the videotaped exams[20]; the aspects considered were 8: 1) range, 2) Pronunciation, 3·) Accuracy, 4) Fulfilment of the task, 5) Intelligibility, 6) Interaction, 7) Fluency, and 8) Coherence, which could be rated from 1 to 7 points for each of the criteria. Examiners were asked in each item to provide reasons why they gave such importance to the

[18] The icebreaker questions were not take from anywhere particularly, I just used those which I remember from previous oral speaking tests -both from Cambridge and EOI-.
[19]The candidates did the exam in couples, following the format of Cambridge examinations.
[20] 7 was equivalent to extremely important, 6 to very important, 5 to moderately important, 4 to neutral, 3 to slightly important, 2 to low importance, and 1 to not at all important. Likert-Type Scale Response Anchors (Vagias, 2006).

criteria. Then, in the second question they were asked whether they agreed with the definitions given about two of the aspects proposed, namely range and accuracy– the definitions were taken from the CEFR official scale. In the third item, they were asked whether they would follow the same criterion for every level or whether they would consider some more important than others depending on the level. In the fourth question, they could add possible criteria they could not find in the eight provided, and tell us how important they were for them. The last question was for them to comment on other aspect of the exams they had rated. In the last part of the questionnaire, they had to answer two questions about their professional experience, namely which proficiency level (From A1 to C2) and which examination (Cambridge or EOI) they were specialised in.

The professional questions were placed at the end of the questionnaire so as not to influence their answers to the previous items. If they had answered the academic questions before, they could have responded afterwards according to what they knew and not to what they actually thought. Once the examiners filled out the assessment sheets and the questionnaire, they were collected and set to be analysed.

2.3. Methodology

The methodology followed was a mixed methods research including both quantitative and qualitative materials. We decided to include some open-ended questions together with the Likert questionnaire because qualitative data allowed us to add an explanation to the quantitative data by means of the comments and thoughts from examiners, which were essential in this work.

First, we found 12 candidates. We tried to cluster the same amount of male and female candidates, but it turned out to be impossible, so in the end there were 7 women and 5 men[21]. Once they all had accepted, we arranged the couples and the time of the interviews. We managed to have some couples whose native language was the same and some others whose native language was different. Some candidates had met his/her partner for the exam before

[21]Nevertheless, we did not consider the gender variable as an important aspect for this study.

and some others had not. All the exams, except for one[22], took part in the same place: Sala de Juntas IV from the Faculty of Filosofia y Letras in Cádiz.

The 12 candidates did the English Oral test from the B2 Cambridge book. The exam consisted in an interaction to be done in pairs and one or two questions, which lasted 5 minutes in total. The candidates were first asked a couple of icebreaker questions, which lasted just one minute or less per each candidate. Subsequently, the interaction part was in connection modern life styles (See Appendix 1). They had to discuss it in pairs for 2 minutes approximately about the importance of some items[23] nowadays, and then in 1 minute they had to agree on the most important. The 6 tests the 12 candidates did were videotaped, and the recordings were provided to the examiners.

The four examiners were handed over the assessment sheets, the questionnaires, and the videos. In the case of Cambridge examiners, they watched the videotaped tests, rated the candidates, and filled up the questionnaire while I was with them. They did it after finishing their workday. They watched the videotaped exams just once. While watching them, they wrote down notes about the candidates– as in an official exam-. The 6 assessment sheets had the number of the video in the title (video 1, video 2, etc). The pauses between one video and another were just of a couple of minutes. Once they finished rating the candidates, I handed the examiners in the questionnaires[24]. It took them 30 minutes roughly to fill in the questionnaires. All this process took place in the academy where both examiners work, located in Cádiz.

In the case of the EOI examiners, the process was a bit different. I personally gave them the assessment sheets, the questionnaires, and the videos on a memory stick. However, they did not carry out the tasks in that moment, but during the following days and at their home. In any case, it took them the same amount of time to fill in the questionnaires -30 minutes roughly-. They also watched and rated the candidates in a row, but one of the EOI examiners (Examiner 3) told us that s/he had watched some videos twice just to corroborate the marks given,

[22]The last exam was planned to be done in the same room as the other 5. But the room was already taken for the date we had arranged, so we had to do it in another room.
[23]The items were already mentioned in the Materials section.
[24]The questionnaire had been previously administered via e-mail to 4 experts in English language – 3 of them native speakers of English and one Spaniard– for them to fill in as a pre-test. These experts suggested some corrections which were incorporated to the questionnaire handed in to the examiners.

though she had not changed the rates previously given or any other aspects.

3. Results and discussion section

First of all, we will present the marks that the candidates received from each examiner, and they will be discussed throughout this section. The exams were done in pairs; thus candidates 1 and 2 did it together, 3 and 4, and so on.

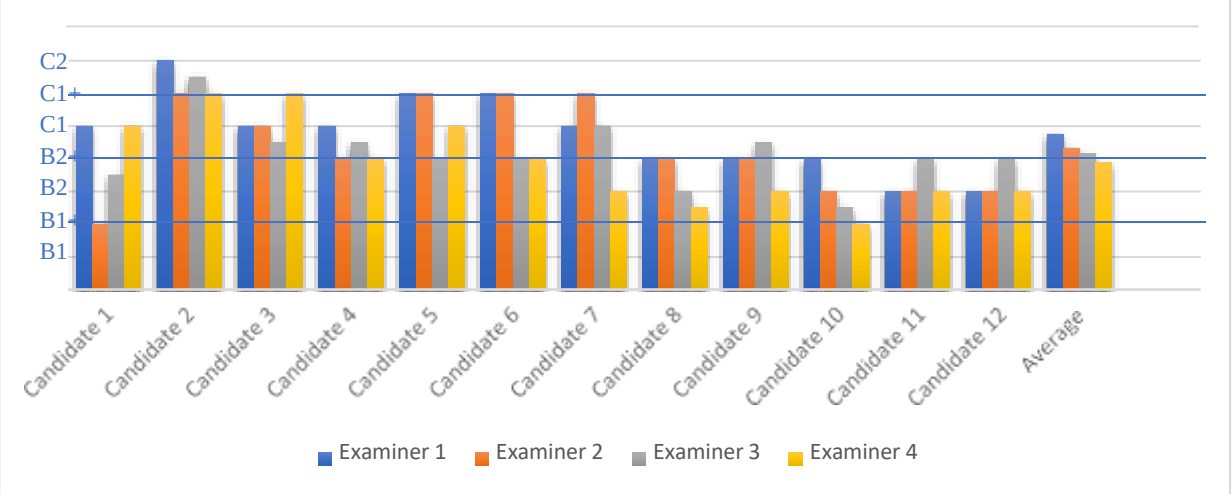


Figure 1. Rates given to the candidates by examiners.

We can see in the last pile of columns the average mark given to all candidates by each examiner. It is worth highlighting the difference between the average mark from Examiner 1, who belongs to a Cambridge Centre, with the average mark from the Examiner 4, who belongs to an EOI centre. The average mark given from Examiner 1 is very close to C1, whereas the average mark given Examiner 4 is B2, being almost a point difference between both ratings.

Curiously enough, Examiner 1, who rated the candidates with the highest average, is specialised in C1 and C2 examinations; whereas Examiner 4[25], with the lowest marks given, is specialised in the lowest levels: A1 and B1. Given that the exam was a B2 test taken from a Cambridge examination book, it is likely that Cambridge examiners, and especially Examiner 2, were more familiar with this kind of exams and candidates' level. Hence, the ratings given by Cambridge examiners may be generally closer to the real level of the candidates than the

[25]Examiner 2, from a Cambridge centre, was specialised in B2 and C1 exams; whereas examiner 3, from an EOI centre, in A1 and B1.

rates given by EOI examiners, as Cambridge examiners are more used to assessing these exams.

Now we will have a look at the criteria the examiners considered when grading. First, we will focus on Cambridge Examiners.

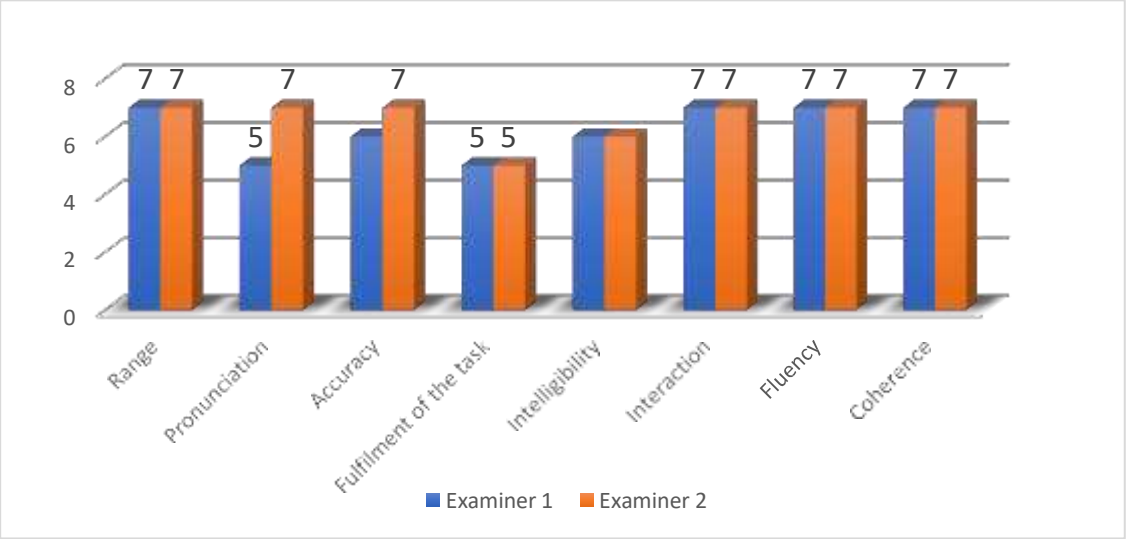


Figure 2. Importance given to each criterion by Cambridge examiners

Fig. 2 shows that both examiners agree with the importance of the criteria, with 7 being extremely important and 1 not important at all (See Appendix 3). The differences are found mainly in pronunciation, but also in accuracy. Did they also grade in such a similar way?

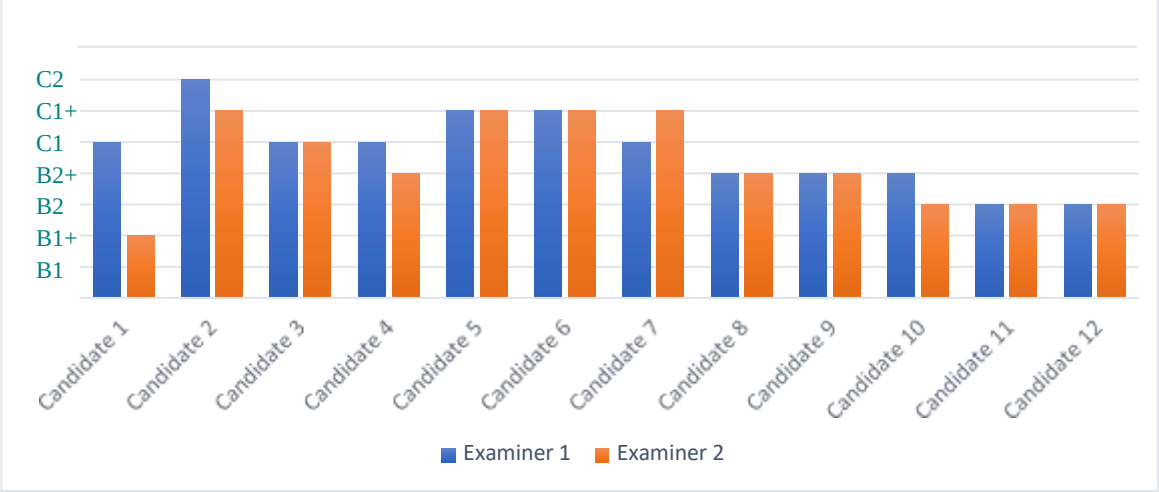


Figure 3. Grades given to the candidates by Cambridge examiners.

Fig. 3 shows small and few differences in the marks as well. We just found significant

differences in Candidate 1, and not very significant ones in Candidates 2, 4, 7, and 10. In order to find out whether these differences are related to the differences in the criteria, we looked at comments made by the examiners.

According to Examiner 2, Candidate 1 was very difficult to understand due to his intonation and pronunciation, which is why his mark was the lowest among all the candidates – pronunciation was a key role for Examiner 2. However, he was rated with a C1by Examiner 1.We can also see that Candidate 4 was rated differently depending on the examiners. If we look at the comments made by Examiner 2, they say the following: "Both show an appropriate vocabulary, but I think candidate one's (Candidate 3) pronunciation is much more native-like". Pronunciation played again a key role for Examiner 2, as played in the case of Candidate 7 and Candidate 10. Examiner 2 highlights the pronunciation of Candidate 7, as she pointed out that Candidate 1's accent was better than his/her partner's, whereas this examiner penalised Candidate 10 for the same reason: pronunciation. Examiner 1 rated Candidate 10 with a B2+, whereas Examiner 2 rated him with a B2 because "he has a Spanish accent and he does not pronounce the endings well".

The results from Cambridge examiners support our first hypothesis which stated that if the criteria varied, the grades would vary too. We notice that Examiner 2 placed different importance on pronunciation from Examiner 1, and so the ratings changed.

Can we confirm this hypothesis with EOI examiners too? *Fig. 4* indicates the importance given to every criterion by the EOI examiners.

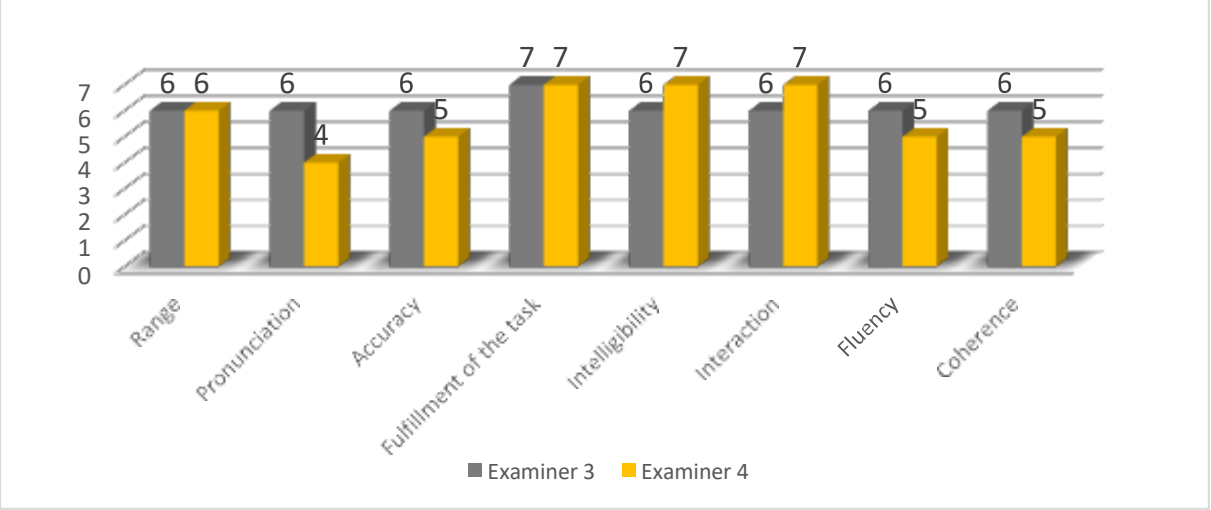


Figure 4. Importance given to each criterion by EOI examiners

Fig. 4 shows the disagreement between the EOI examiners in the importance attached to each criterion -which should be the same-, coinciding only on the range and task fulfilment criteria. *Fig. 4*, the criterion with which they disagreed the most is pronunciation. For Examiner 3, "pronunciation is an essential element in any oral performance" whereas for Examiner 4 pronunciation is just neutral[27]: "as long as they are intelligible, foreign accent is not important". This disagreement in the rating criteria correlates with the different grades given to the candidates, as shown in *Fig. 5*.

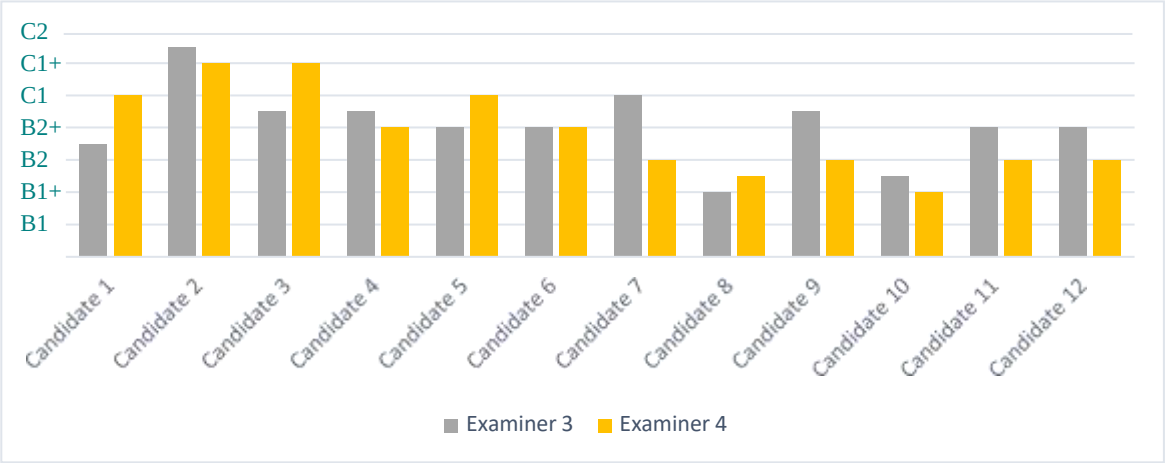


Figure 5. Grades given to the candidates by EOI examiners.

Analysing the candidates' rates in *Fig. 5*, the different mark given to Candidate 7 by both examiners is very significant. For Examiner 3, Candidate 7's English was equivalent to a C1 because "he is very fluent, speaks quite natural, and has a very good pronunciation and intonation". Thus, for Examiner 3 the main aspects considered were pronunciation and fluency. If we look at the importance they gave to every aspect, we notice that Examiner 3 gave more importance not only to pronunciation but also to fluency than Examiner 4 did.

Therefore, the grades are consistent with the answers given by examiners on the criteria. Candidate 1's pronunciation was again crucial for examiners. Examiner 3 commented that Candidate 1's pronunciation was not very clear because he mumbled a bit, which made it difficult to understand him at times. Thus, again Examiner 3 considered pronunciation (and intelligibility) a significant criterion and penalised the candidate's pronunciation deficiencies. The same happened with Candidate 9, to whom Examiner 3 awarded with a better mark than

[27]Neutral is equal to 4 points in the Likert-scale. (See appendix 3).

Examiner 4 arguing the following: "She didn't seem to participate as actively as her partner but she had a good level. Fluency, accuracy and very good pronunciation". Clearly pronunciation was crucial once more. Hence, our first hypothesis has been proved by the EOI examiners too, since the relevance attached to the criteria differed and so did the rates. It is true that examiners can identify where candidates falter and where they stand out. The problem is that examiners do not know which strengths and weaknesses should be highlighted in oral tests.

Regarding the issue of whether the centres have followed the official guidelines, *Fig. 6* shows the average importance given to the criteria by Cambridge examiners:

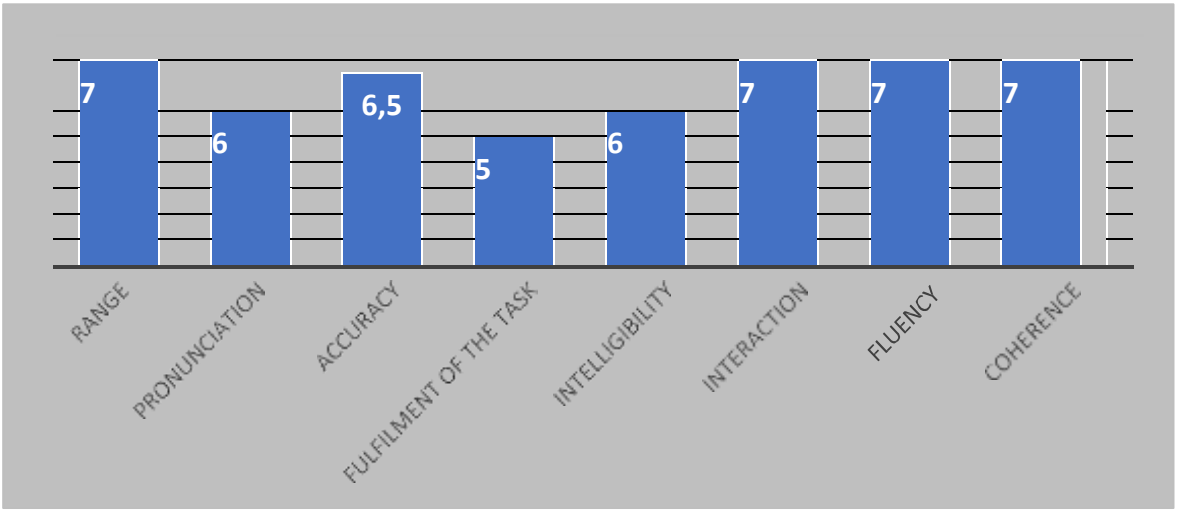


Figure 6. Average importance given to each criterion by Cambridge examiners.

Our fourth hypothesis claimed that for Cambridge examiners the criteria of accuracy, coherence/cohesion, and fluency would be the most important since these are within the main criteria according to their guidelines, i.e. global achievement. Nevertheless, we see in *Fig. 6* that fluency and coherence are among the most important, but not accuracy. Furthermore, range and interaction are as important as fluency and coherence. Therefore, we could say that the fourth hypothesis has not been completely fulfilled.

With this graph, we can also prove hypothesis number 5, which asserted that Cambridge examiners would consider task fulfilment the least deciding criterion when rating candidates, but not irrelevant. As we can see in Fig. 6, they considered task fulfilment moderately

important[29]. Hence, the fifth hypothesis has been verified.

If we focus on the criteria followed by EOI examiners. *Fig. 7* presents the average importance given to each criterion by EOI examiners.

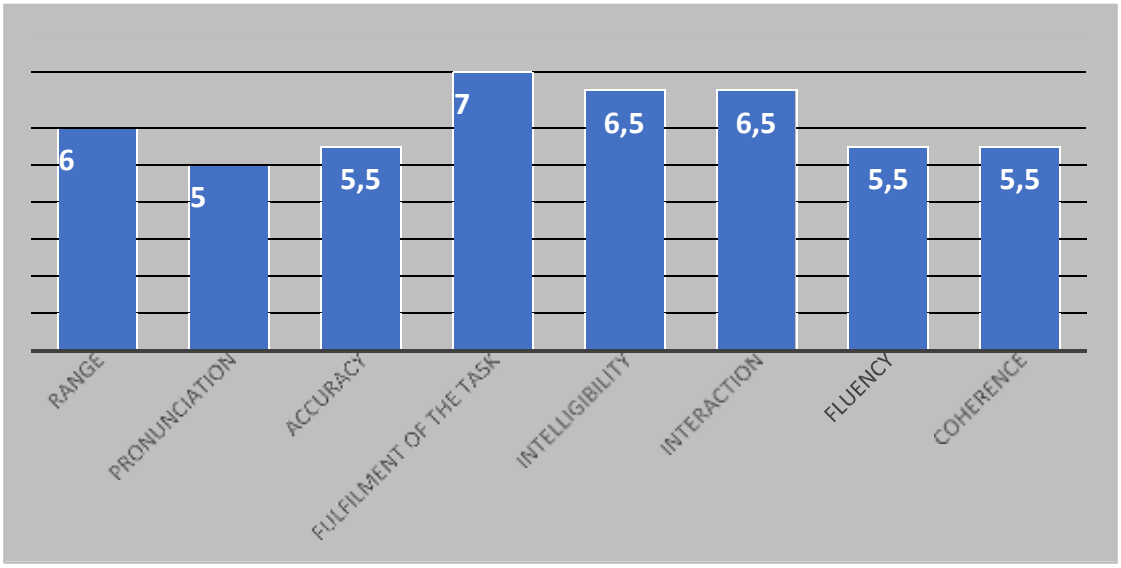


Figure 7. Average importance given to each criterion by Cambridge examiners.

Hypothesis number 6 claimed that EOI examiners would consider range and accuracy the most relevant criteria, as stated in their guidelines. However, we can see that they considered other aspects more important than range and accuracy, such as task fulfilment, interaction, and intelligibility. Moreover, according to the EOI official guidelines, task[30] fulfilment is the least important criterion whilst for the EOI examiners in this research it is the most relevant one. We find several contradictions between the official instructions and this sample. Hypothesis number 6 is clearly disproved then.

In order to see whether hypothesis number 2 is proven– i.e. whether examiners from the same centre would show more agreement with each other than examiners from different centres , we compared two graphs. Thus, *Fig. 8* illustrates the relevance of the criteria and *Fig. 9* the rates given by each examiner.

[29] Moderately important equals 5 points in the Likert-scale used in the questionnaire (See appendix 3).
[30]What they call in their guidelines "grade of achievement".

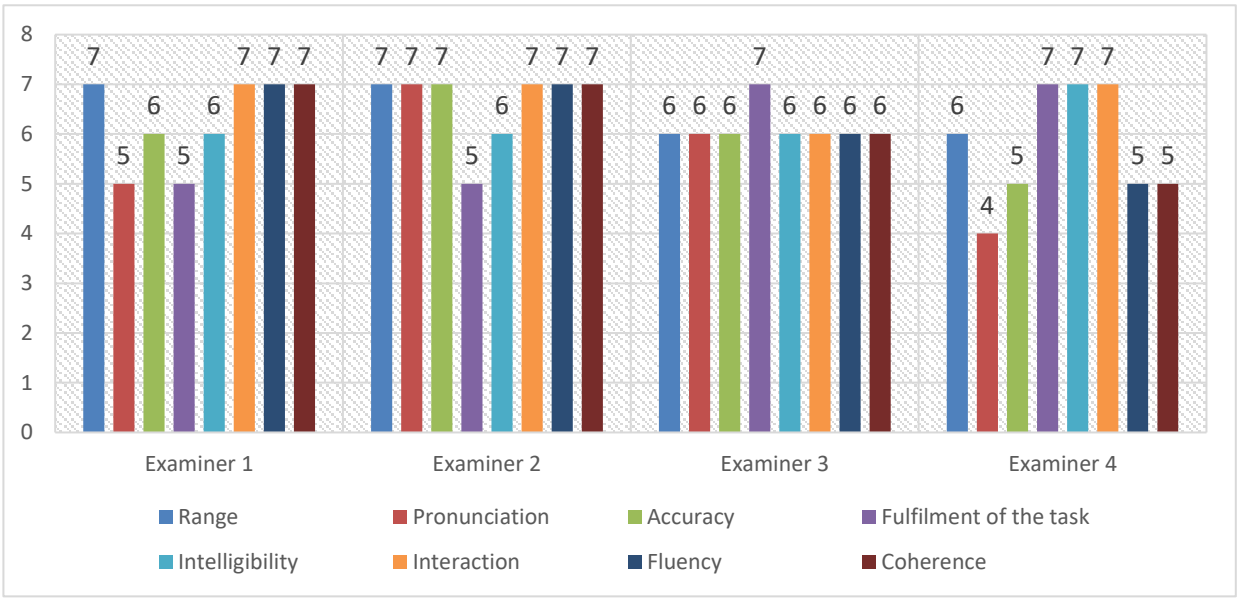


Figure 8. Importance given to each criterion by each examiner.

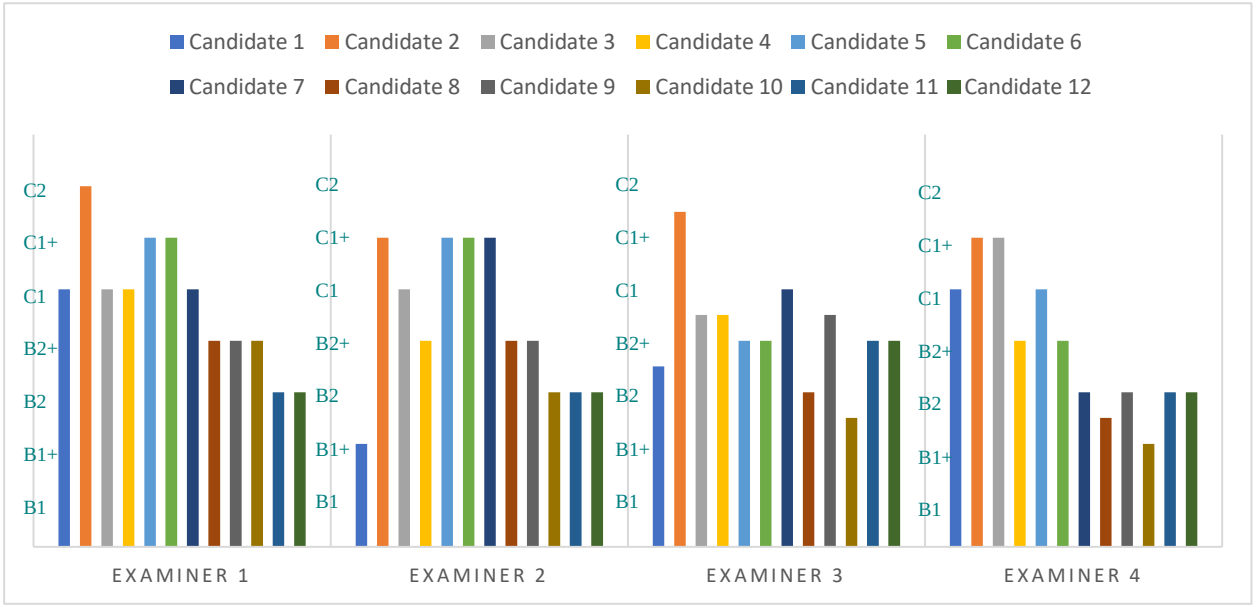


Figure 9. Grades given by each examiner.

In *Fig. 8* we see that Examiners 1 and 2, both from Cambridge, have the same opinion about 6 out of the 8 criteria. In addition, they also show agreement in the marks of 7 out of 12 candidates. In contrast, little agreement is shown by Cambridge examiners and EOI examiners. For example, Examiner 1 agrees with Examiner 4 in the relevance of only one criterion, and Examiner 2 has not rated any candidate with the same mark as Examiner 3. Therefore, in the case of Cambridge examiners, hypothesis number 2 is proven. Nonetheless, we do not see the same in the EOI examiners, as they show insufficient agreement between

them, and with Cambridge examiners (i.e. Examiners 1 and 2). Therefore, in the case of EOI examiner hypothesis number 2 is not proven. The differences between examiners from the same centre indicate that the importance given by examiners to the different criteria play an essential role in the evaluation of the oral test and should not be overlooked. In the case of EOI examiners if they do not have a template to follow, the importance of the criteria changes from one examiner to another and so do the results. Therefore, a template with the criteria to be rated seems to be essential for EOI examiners.

Three hypotheses still need to be considered; and they can be checked entirely with the answers from the questionnaires. For example, in hypothesis number 3 we asserted that Cambridge examiners would give different importance to the criteria depending on the level they were rating, and EOI examiners would give the same importance to every criterion no matter which level they would be grading, as their official guidelines state. We asked them about it in the questionnaire and they all said that the criteria to be followed were determined by the level to be graded; therefore, hypothesis number 3 is proven by Cambridge examiners, but not by EOI examiners.

Examiner 2, from Cambridge, claimed that in higher levels such as C1 and C2 the candidate had to show a wider range of vocabulary and more accuracy in terms of structures and grammar. Examiner 1, from Cambridge too, commented that for lower levels s/he would leave fluency a little bit behind, while giving more importance to vocabulary and grammatical structures; and for C1 and C2 all criteria were equally important. We can see that the opinion is not exactly the same. Theoretically at higher levels (C1 and C2), Cambridge examiners should place more importance in grammar and vocabulary than in lower levels, as examiner 2 commented.

On the other hand, Examiner 3(from the EOI) indicated that at lower levels one concentrates more on communication *per se*, whereas at higher levels the candidate needs to show a wider linguistic range & accuracy at all levels (grammar, vocabulary, pronunciation, etc). We see how this examiner contradicts the EOI guidelines, which recommend giving the same importance to the criteria in every level. Examiner 4 commented that each level should have different criteria, even though the basic criteria may stay the same throughout all levels. This opinion is a bit ambiguous as it is not clear which basic criteria s/he refers to. In any case, Examiner 4 claimed that the criteria should differ.

Another hypothesis which can be checked with the answers from the questionnaires is hypothesis number 7. We believed that, though similar, examiners would not fully agree with the definitions of some criteria they were be given. Contrary to our hypothesis, we found agreement from all the examiners on the definitions provided both for range[31] and for accuracy[32] in question 2 of the questionnaire, although Examiner 1 from Cambridge added that accuracy is not only what the definition said, but also that what the candidate says is accordance to what is required in the task. Moreover, we added some definitions in the footnotes about other three aspects (See appendix 4)[33], and they had the choice to comment on other elements if they did not agree with the definitions provided; but it seems that examiners from both Cambridge and EOI side with them. Therefore, our hypothesis number 7 has been in a way refuted, as we expected them to somewhat disagree with the definitions provided.

In hypothesis 8, we claimed that examiners would not take into consideration any other aspect apart from the eight proposed in the questionnaire (see Appendix 4), as we included the criteria recommended by CEFR, EOI and Cambridge official scales. However, three of the four examiners also considered another criterion. Examiner 2 (from Cambridge) mentioned "Self-confidence" as extremely important, i.e. 7 in the Likert-scale; Examiner 3 (from EOI) mentioned "a natural exchange" as very important (6 points); and Examiner 4 (from EOI) commented "Initiative" as very important (6 points) when assessing the candidates. This fact goes against hypothesis 8.

[31] The definition provided for range was "the extension of the vocabulary and the use of complex sentence forms".
[32] The definition provided for accuracy referred "to the control and appropriate use of grammatical structures". [33]All the definitions provided in the questionnaire – both in the footnotes and in question 2 – were taken from the CEFR

4. Conclusion

With this research, some of the hypotheses have been proved and some others have not. Hypothesis 1, which claimed that the variation in the criteria would also cause different ratings of the candidates, has been proven. Hypothesis 2, about the agreement between examiners from the same centre, has been proved by Cambridge examiners, though disproved by EOI examiners. Hypothesis 3, which claimed that EOI examiners would give the same importance to every criterion but not Cambridge examiners, has been verified by Cambridge examiners, but not by EOI examiners; so it has been partially proven. Hypothesis 4, on the importance given to each criterion by Cambridge examiners, has not been totally confirmed. Hypothesis 5, which argued that Cambridge examiners would place task fulfilment as the least deciding criterion when rating candidates, has been also proved. Hypothesis 6 stated that EOI examiners would consider range and accuracy the most relevant rating criteria; however, it was refuted as they gave more importance to other criteria. Hypothesis 7, about the definitions of the criteria, has been refuted as well. Finally, Hypothesis 8, which stated that examiners would not consider other aspects apart from the ones presented in the questionnaire, has also been disproved.

As the results demonstrate, it is evident that there is limited agreement when rating between the two accreditation centres examined and, in the case of the EOI, – there is even no agreement between the examiners from the same centre. Both centres use the CEFR descriptors as a reference, but they are not followed to the letter as every centre interprets them differently. Consequently, very dissimilar results arise, causing candidates to be rated differently depending on which criteria a given examiner considers more important.

Should we design a template for all the centres? Should we work on new descriptors or implementation of the actual descriptors for the Common European Framework, since some centres do not seem to pay much attention to them at present? Or should we just leave every centre to decide on their own guidelines? In that case, would every language certificate be an equally valid indicator of language proficiency? This research has proved that there are no standardised criteria for oral tests and that some language policy should be implemented in the CEFR descriptors to get a global agreement when rating oral tests. We personally suggest that the CEFR scales should be either revised and /or implemented, and that examiners from every centre should be urged to follow common scales to avoid such substantial differences.

Bibliography:

Assessing speaking performance at B1 level. (2011) Retrieved from: <http://www.cambridgeenglish.org/images/168618-assessing-speaking-performance-at-level-b1.pdf> [01-06-2017]

Assessing speaking performance at C1 level. (2011). Retrieved from: <http://www.cambridgeenglish.org/images/168620-assessing-speaking-performance-at-level-c1.pdf. > [03-06-2017]

Assessing speaking performance at C2 level. (2011) Retrieved from: <http://www.cambridgeenglish.org/images/182109-assessing-speaking-performance-at-level-c2.pdf,> [26-05-2017]

Bridgeman, B., Powers, D., Stone, E., & Mollaun, P., (2011) "Language Testing. TOEFL iBT speaking test scores as indicators of oral communicative language proficiency". SAGE journals, USA.

Burgess, T. (2001) *Guide to the Design of Questionnaires. A general introduction to the design of questionnaires for survey research.* University of Leeds, Leeds.

Cambridge English: First (FCE) Assessment commentary and marks: Florine. (2014). Retrieved from <http://www.cambridgeenglish.org/images/167891-cambridge-english-first-fce-from-2015-speaking-test-video.pdf> [21-05-2017]

Common European Framework of Reference for Languages: Learning, teaching, assessment. Structured overview of all CEFR scales. 1.3. (n.d.). Retrieved from: <http://ebcl.eu.com/wp-content/uploads/2011/11/CEFR-all-scales-and-all-skills.pdf > [26-05-2017]

De John, N., Groenhout, R., Schoonen, R. & Hulstijn, J. (2012). "Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behaviour". Cambridge University Press.

Dornyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford, New York.

Examples of speaking performance at CEFR levels to A2 to C2 (2009). Retrieved from: <http://www.cambridgeenglish.org/images/22649-rv-examples-of-speaking-performance.pdf,> [20-05-2017]

IETLS SPEAKING: Band descriptors (public version) (2015). Retrieved from: <https://takeielts.britishcouncil.org/sites/default/files/IELTS_Speaking_band_descriptors.pdf > [01-06-2017]

Junta de Andalucía. (2017) Pruebas terminales específicas de certificación. Guía del profesorado. Sevilla.

Maes, K. (2011) "CEFR Grammar: Which Rules at Which level?". *Language testing in Europe: Time for a New Framework?,* Antwerp (Belgium).

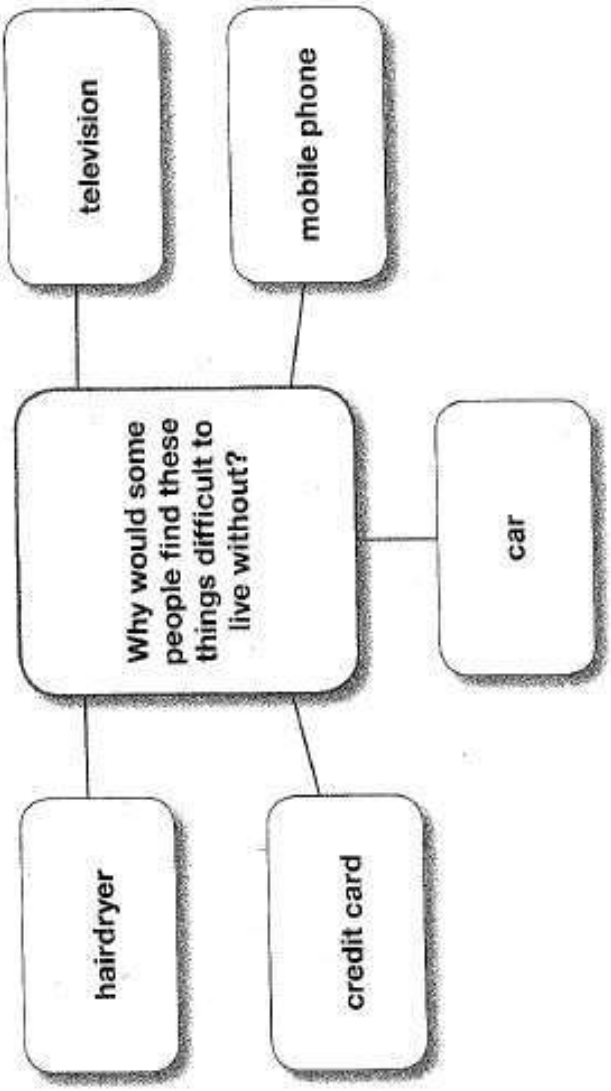
The Cambridge English Scale explained. A guide to converting practice test scores to Cambridge English Scale scores. (n.d) Retrieved from: http://www.cambridgeenglish.org/images/210434-converting-practice-test-scores-to-cambridge-english-scale-scores.pdf. [26-05-2017]

University of Cambridge ESOL examinations. Examples of speaking performance at CEFR levels A2 to C2 (2009). Retrieved from: http://www.cambridgeenglish.org/images/22649-rv-examples-of-speaking-performance.pdf [29-05-2017]

Vagias, W. (2006) Likert-Type Scale Response Anchors. Retrieved from: <https://www.uc.edu/content/dam/uc/sas/docs/Assessment/likert-type%20response%20anchors.pdf> [05-05-2017]

Appendixes:

Appendix 1:



Appendix 2: INFORMED CONSENT FOR VIDEOTAPING

I understand that I will be videotaped by the researcher in order to help him with his thesis about the Oral English Test. These tapes will be kept by the researcher in a locked filing cabinet. I understand that only the researcher will have access to these tapes and that they will be destroyed by 9/30/2017.

Video recording of study activities

Interviews may be recorded using video devices to assist with the accuracy of your responses. You have the right to refuse the video recording. Please select one of the following options:

I consent to video recording: Yes _____ No_____

Full name:

Signature:

Appendix 3: Assessment sheet:

Video 1

Candidate 1 (Boy)	Candidate 2 (Girl)
English level: *Circle the level that you consider this candidate has*. B1 B1+ B2 B2+ C1 C1+ C2	English level: *Circle the level that you consider this candidate has*. B1 B1+ B2 B2+ C1 C1+ C2
Comments:	Comments:

Video 2

Candidate 1 (Blonde girl in denim shirt)	Candidate 2 (Brunette girl in grey pullover)
English level: *Circle the level that you consider this candidate has*. B1 B1+ B2 B2+ C1 C1+ C2	English level: *Circle the level that you consider this candidate has*. B1 B1+ B2 B2+ C1 C1+ C2
Comments:	Comments:

Video 3

Candidate 1 (Boy)	Candidate 2 (Girl)
English level: *Circle the level that you consider this candidate has*. B1 B1+ B2 B2+ C1 C1+ C2	English level: *Circle the level that you consider this candidate has*. B1 B1+ B2 B2+ C1 C1+ C2
Comments:	Comments:

Video 4

Candidate 1 (Boy in turquoise jacket)	Candidate 2 (Boy in grey hoodie)
English level: *Circle the level that you consider this candidate has*. B1 B1+ B2 B2+ C1 C1+ C2	English level: *Circle the level that you consider this candidate has*. B1 B1+ B2 B2+ C1 C1+ C2
Comments:	Comments:

Video 5

Candidate 1 (Girl)	Candidate 2 (Boy)
English level: *Circle the level that you consider this candidate has*. B1 B1+ B2 B2+ C1 C1+ C2	English level: *Circle the level that you consider this candidate has*. B1 B1+ B2 B2+ C1 C1+ C2
Comments:	Comments:

Video 6

Candidate 1 (Girl wearing glasses)	Candidate 2 (Girl in black t-shirt)
English level: *Circle the level that you consider this candidate has*. B1 B1+ B2 B2+ C1 C1+ C2	English level: *Circle the level that you consider this candidate has*. B1 B1+ B2 B2+ C1 C1+ C2
Comments:	Comments:

Appendix 4: Questionnaire

First, we would like to thank you for accepting to participate in this questionnaire. The aim of this questionnaire is to know your opinion about the English of the videotaped candidates that you have previously seen and evaluated. Hence, there is neither wrong nor right answers. Also, you will find several footnotes to avoid confusion with the understanding of some aspects.

1. What importance did you give to the following aspects when grading the oral English videotaped test?

Circle the number from 1 to 7, where the correlations are the following:

***1**: Not important at all.* ***2**: Not very important.* ***3**: slightly important.* ***4**: Neutral.*

***5**: Moderately important.* ***6**: Very important* ***7:** Extremely important.*

Please also give your reasons for considering the aspect that are important/insignificant.

Range	1	2	3	4	5	6	7

Please tell us why you gave that importance to this aspect:

Pronunciation	1	2	3	4	5	6	7

Please tell us why you gave that importance to this aspect:

Accuracy	1	2	3	4	5	6	7

Please tell us why you gave that importance to this aspect:

Fulfilment of the task	1	2	3	4	5	6	7

Please tell us why you gave that importance to this aspect:

Intelligibility	1	2	3	4	5	6	7

Please tell us why you gave that importance to this aspect:

Interaction[34]	1	2	3	4	5	6	7

Please tell us why you gave that importance to this aspect:

Fluency[35]	1	2	3	4	5	6	7

Please tell us why you gave that importance to this aspect:

Coherence [36]	1	2	3	4	5	6	7

Please tell us why you gave that importance to this aspect:

[34] Interaction refers to the ability to interweave his/her contribution with suitable turn taking and cues. He/she can also help the discussion along.

[35] Fluency refers to how smoothly and spontaneously the speaker can express him/herself.

[36] Coherence refers to how well-structured the speech is.

2. Do you agree with the definitions of the following aspects? If not, please provide us with your definition.

-Range is the extension of the vocabulary and the use of complex sentence forms.

-Accuracy refers to the control and appropriate use of grammatical structures.

3. Would you follow the same criterion for every level? Or would you consider ones more important than others depending on the level grading? Please, give your reasons:

4. Did you also consider other aspects that were not mentioned in question 1 when grading? If any, please mention them and tell us how important you considered these.

Aspects:	**How important I considered it:**						
1.	1	2	3	4	5	6	7
2.	1	2	3	4	5	6	7
3.	1	2	3	4	5	6	7

5. Other elements (if any) you would like to mention and comment on:

And to conclude, please, tick next to the option that applies to you. .

I am specialised in:

-Cambridge examinations

-Escuela Oficial de Idiomas examinations

The level I am specialised in is:
Here, circle more than one, if necessary.

A1 A2 B1 B2 C1 C2

Thank you for your time!