Designing Oral-based Rubrics for Oral Language Testing with Undergraduate Spanish Students in ESP Contexts[1]

Carolina Girón-García[2] & Claudia Llopis-Moreno[3]

Abstract

During several decades, considerable attention has been paid to speaking assessment procedures. Probably, the reason for this relevance lies in the difficulties in dealing with oral assessment and in determining the adequate type of assessment (Campbell et al., 2001; Schwartz & Arena, 2013; Stoynoff, 2013). The main aim of this article is to analyze to what extent students' oral competences (such as their English competence and fluency) affect their peers' English oral production. More precisely, we intend to study effective procedures to assess oral production in ESP contexts. In order to assess our students, we have designed a speaking-based rubric as the main instrument (called 'Speaking Diagnostic Test'), based on previous research (Spandel, 2006; Wilson, 2006). The results obtained from both questionnaires will help us identify to what extent students' preferences in working with other classmates influence their English oral production and therefore to what extent these results could lead to the reformulation and modification of the assessment methods.

Key words: ESP contexts, rubric, 'Speaking Diagnostic Test', oral production.

---

[2] Universitat Jaume I, Spain., Email: gironc@uji.es
[3] Universitat Jaume I, Spain., Email: cllopis@uji.es

1. Introduction

In recent years, teachers of English as a second language (ESL) have paid particular attention to the relevance of communicative competence as an essential aspect in their English language courses. Canale and Swain (1980) and Canale (1983) defined the term of *communicative competence* as a combination of both, the idea of knowledge and skill needed for communication. The former was referred to as knowledge on how to use the language in different social contexts and, the latter implied the use of such knowledge for actual communication. As a result, ESL teachers have generally agreed on the idea of teaching students how to be communicatively competent in different social contexts and situations as the main goal of their English courses. Moreover, it is important to highlight students' perception on an existing gap between the relevance given to oral performance and the limited time devoted to develop it in the university context (Kavanagh & Drennan, 2008).

Higher Education (HE) English teachers are following the same line as primary and secondary ones, that is, fostering communication skills in their courses as an essential ability that undergraduate students need to acquire for their future professional careers. At university level, good communication skills (receptive and productive skills) have been emphasized as essential tools for students seeking future employment. Special interest has been laid on how to assess oral communication (Panadero & Jonsson 2013), and what the most suitable oral assessment methods are (Campbell et al., 2001; Schwartz & Arena, 2013; Stoynoff, 2013).

In this study, we will be paying special attention to the use of rubrics as an effective oral assessment tool in ESP contexts since they have been commonly used during the last years in Higher Education English courses. Particularly, we will be analyzing to what extent students' oral competences affect their peers' English oral production. In that respect, in our research we intend to study effective procedures so as to assess oral production in ESP contexts.

This paper is organized in six sections. In section 2, we will provide a definition of what a rubric is as well as its uses for oral assessment. In section 3, we will review what the different procedures for oral production assessment in ESP contexts are as well as comment on what students' preferences and attitudes are when being assessed in their oral skills. Section 4 describes the way this study has been carried out including a description of the number of participants, instruments and procedures that have been

used in it. Section 5 reports on the results obtained in the study as well as an in-depth discussion about such results followed by the general conclusions included in section 6 and some bibliographical references.

2. Towards the Design of Oral-Based Rubrics

Several definitions of the concept 'rubric' have been provided for several decades (Spandel, 2006; Wilson, 2006) and these 'rubrics' have become very popular in the field of language education, more precisely they have caused a great impact among teachers and students as a new tool to evaluate oral production in English.

Despite the vast amount of definitions for the word 'rubric' found, for the purpose of this research we use the term 'rubric' in the same way that researchers Allen and Tanner (2006) already pointed out to refer to "a type of matrix that provides scaled levels of achievement or understanding for a set of criteria or dimensions of quality for a given type of performance…" (p. 197). Along this line and concerning our work, our intention is to assess undergraduate students' oral performances with the design of a speaking-based rubric.

In order to use the adequate type of rubric, we need to focus on a specific kind that establishes a link between a particular content and the objectives that account for a given subject matter. At this respect, Allen and Tanner's (2006) description of analytical and holistic rubrics may lead us to a quite open view of how our rubric might be categorized. Thus, according to these authors, analytical rubrics "use discrete criteria to set forth more than one measure of the levels of an accomplishment for a particular task", whereas holistic rubrics are defined as those that "provide more general, uncategorized […] descriptions of overall dimensions of quality for different levels of mastery" (p.198).

As teachers, when we use a holistic rubric, we may observe that some students do not entirely fit into a given category. This idea is in relation to the fact that when evaluating with a holistic rubric, we might only use up to six levels of performance so as to not complicate things during the assessment process. However, a combination of holistic and analytic rubrics would be optimal for assessing students in particular cases. According to Taufiqulloh (2009) "analytic rating scales which are complicated and time consuming to use, are the most effective ways for diagnosing the communicating information, such as students' strength and needs" (p. 187). In accordance with these words, Underhill (1987) also suggests the use of

analytic rubrics, but in fact he puts particular emphasis to follow a balanced approach to use holistic and analytic rating scales. For this reason, both types of rubrics are not necessarily exclusive, but they could complement each other. Regarding this distinction, we find pertinent to highlight the fact that the rubric that we have designed for our research purposes is categorized concerning the second description (i.e. "holistic" rubric).

Many educators have mentioned and suggested relevant communication features so as to assess speaking in general speaking courses (Graham & Mignerey, 1990; Jones, 1994;). Moreover, students' learning assessment has traditionally and basically been focused on written exams; in contrast to non-traditional assessment procedures, such as portfolios and oral assessments. In fact, many oral assessment procedures require the use of a rubric; it is probably for that reason that non-traditional assessment is on the rise of many educators concerns at all educational levels, especially at universities.

Evaluators spend most of their time listening to students' speeches, and then discussing their oral assessment regarding each of the competencies reflected in a speaking rubric. At this respect, evaluators share the same points of view, but on some occasions they do not agree on the same students' score. For this reason, the use of rubrics could become a very influential tool for assessment procedures and results with regards to maintaining consistency among teachers (Dunbar et al., 2006).

When designing this rubric, we may ask ourselves the following question: "How should educators use a rubric in ESP contexts with undergraduate students?" Along this line, we could start designing our *'Speaking Diagnostic Test'* in ESP contexts (Appendix 1).

The Appendix I shows the speaking-based rubric designed for the purpose of our research. In this sense, students under study were asked to perform both a monologue and two dialogues. In the monologue students performed they had to develop the topic *"Why do you think studying English is important?"* on the other hand, they performed two dialogues (role-plays) with two different classmates (regarding their preferences in working with others, as they stated in the 'Student Questionnaire') on the topic of *"The language of socializing: A night at the opera"*. At this respect, the rubric was implemented and modified so as to include specific grammar and vocabulary in relation to these topics.

Finally, it is important to emphasize that the same 'Speaking Diagnostic Test' was employed in the three tasks (i.e. monologue and two dialogues).

3. Procedures for Oral Production Assessment in ESP Contexts

Teachers usually conduct assessment whose aim is to evaluate the progress of their students' English oral competence when s/he wants to achieve a certain language level requiring oral proficiency. Apart from this main purpose, we could also mention the existence of other aims for speaking assessment (O'Malley & Pierce, 1996, p. 63):
1. Initial identification and placements of students. In our study this stage was introduced by the use of a 'Speaking Diagnostic Test',
2. Movement from one level to another in an English program, and
3. Placement out of an ESL/bilingual program in to a grade-level classroom.

All these objectives help teachers to diagnose the students' progress in oral proficiency since speaking is considered as one of the most difficult skills to assess. The reason for this lies in the fact to the difficulty to determine which the optimal criteria to choose in assessing oral production in English are. In addition, it is of paramount importance to bear in mind that there are several components taking part in evaluating oral communication and that educators should take into consideration when it comes to his/her students' oral production evaluation. Components such as fluency, vocabulary, grammar, pronunciation, coherence, and communicative ability, contribute to attributing a more defined and clear-cut score to students' speeches in English. Bearing all these components in mind, educators should not forget what the setting standards are after establishing rubric and scoring parameters. In other words, the main goal of designing a rubric in ESP contexts in particular, should be being communicatively competent in a given situation such as "Socializing" with colleagues, workmates, etc.

The next section will be devoted to 'The Study' carried out for the purpose of our research. This includes on the one hand, the context and participants who took part in the experiment; and on the other hand, procedure and data collection, and instruments employed for the present study.

4. The study

4.1. Context and Participants

The aim of this study consists of the analysis of students' features (such as their English competence and fluency) in order to find out to what extent those features affect their peers' English oral production.

The study was carried out in four stages:

1. Elaboration and design of a 'Student Questionnaire' designed to find out a classmate they would like to take an oral exam with and another classmate they would not like to.

2. Elaboration and design of a 'Speaking Diagnostic Test'

3. Realization of a monologue on the topic *"Why do you think studying English is important?"*

4. Realization of two dialogues (role-plays) on the topic *"The language of socializing: A night at the opera".*

    4.1.Students perform the role-play with a classmate who they would like to take an oral exam with ('Ideal Partner', shortened as 'IDP').

    4.2. Students perform the role-play with a classmate who they would not like to take an oral exam with ('Unwanted Partner', shortened as 'UNP').

A total of 10 participants out of 30 were selected from the 1st year course at 'Universitat Jaume I' (Spain) according to the amount of relevant information they provided in the 'Student Questionnaire'. The 'Universitat Jaume I' (UJI) is a public university settled in the north of the Valencian Community, a region on the European Mediterranean coast located between the cities of Valencia and Barcelona. Among other things, we can highlight its convenient size, with about 15,000 students, and its integrated, modern, functional and sustainable campus.

Our selected students were recruited from different Engineering Degrees, including Electrical, Mechanical, Chemical, Agri-Food and Rural, and Computational Mathematics Engineering. In addition, students were enrolled in the Scientific English Subject (Modern Language) (code 1005). This subject was divided into three modules: (1) Theory, in which they practiced grammar and vocabulary related to their knowledge field (i.e. Engineering); (2) Problems, in which they practiced formal letters writing (enquiry, apology, and complaint letters); and (3) Laboratory, in which they practiced speaking (role-plays and individual oral presentations).

Although all the students gave their permission to use their data, an individual identification code was provided in order to safeguard their privacy.

4.2. Procedure and Data Collection

The study took place during the students' regular class time in the first semester of the 2013-2014 academic course. The module is compulsory

for all the students; and their participation in this study, though, not mandatory, was presented as one of the complementary activities to be given partial credit in addition to the final mark at the end of the semester.

In order for results to be available, we have carried out the following analysis procedures: First, we analyzed the students' questionnaires and we have selected those students who provided more detailed information about their preferences. Second, we observed the results obtained in the monologue in order to see their initial English level. Third, after this analysis, we distributed our selected students in pairs, according to (1) their preferences expressed in the 'Student Questionnaire'; and (2) the results obtained in the 'Speaking Diagnostic Test'. Finally, we compared the results obtained in both role-plays ('Ideal Partners' and 'Unwanted Partners').

To conclude, students' oral production in the different tasks of our analysis was recorded by means of a 'voice recorder', in which all students' monologues and dialogues performances were kept for the purpose of data collection in our study. In addition, this tool helped us to review students' performance in terms of vocabulary, grammar, communicative ability, pronunciation, and so forth. Therefore, this instrument has been useful in order to check and compare the rubric's scores obtained by each student, so as to support their assessment result.

4.3. Instruments

The main instrument for our study has been (a) a speaking-based rubric, also called 'Speaking Diagnostic Test'. Nevertheless, we have used other instruments endorsed for data collection: (b) 'Student Questionnaire', and (c) 'Voice Recorder'.

4.3.1. 'Student Questionnaire'

In order to determine the students' profile and information in their attitude towards learning English, a questionnaire was used, namely, 'Student Questionnaire' (Figure 1). This questionnaire elicits students' name and surname, age, gender, and two simple questions addressed to the students in order to get information about their preferences towards taking an oral exam with their classmates giving proper justifications.

Question 1 (Q1): Asks students to give a name and a reason for choosing a classmate they would like to take an oral exam with ('Ideal Partner').

Question 2 (Q2): Asks students to give a name and a reason for choosing a classmate they would not like to take an oral exam with ('Unwanted Partner').

These questions will help us classify students in pairs regarding their preferences towards taking an oral exam with their peers.

PERSONAL INFORMATION
Student's name and surname:
Age: **Gender: M / F**
MORE INFORMATION
Read the following questions carefully and answer them with honesty.
You can write as many names as you consider necessary.
1- What classmate(s) would you personally like to take an oral exam with? Justify your answer.

Name and Surname	Reason

2- What classmate(s) would you not like to take an oral exam with? Justify your answer.

Name and Surname	Reason

Figure 1. Student Questionnaire

4.3.2. 'Speaking Diagnostic Test'

For the purpose of this study, students' oral production was measured by means of a *'Speaking Diagnostic Test'* (see Section 2) that was previously designed considering the following categories: Fluency (Fl.), Vocabulary (Voc.), Grammar (Gr.), Pronunciation (Pr.), Coherence (Co.), and Communicative ability (Comm. ability). The score for each category could vary from 1 to 3, being 3 the highest score that could be obtained in each category. Thus, students could have a maximum of 18 points in this rubric, in case that their oral production was perfectly performed in the task.All the variables included in this test are considered as relevant (as in conventional rubrics) in order to assign a certain score in each of them to every student.

The 'Speaking Diagnostic Test' was used in two different stages of our study. Firstly, the rubric was used in order to assess a speaking monologue task in which students were asked the following question: *'Why do you think studying English is important?'* Secondly, it was used in a different task in which students had to perform a role-play that consisted of inviting a peer to go to the opera. Hence, the rubric was used in two different tasks with the aim to analyse to what extend students' oral production may be positively or negatively affected by their oral interaction with other peers. In order to analyse such a possible existing influence, we compared students' scores obtained in the rubric in their individual task (monologue) and in the interaction tasks (role-plays). The results and interpretation of the data obtained in this analysis will be revealed and discussed in the following section.

5. Results and Discussion

In this section, we present the different results obtained and which derive from our experiment with undergraduate students and a speaking-based rubric ('Speaking Diagnostic Test'). The results reveal the different elements or variables included in that rubric taken into account in both the monologue and dialogues performances on behalf of the students. Furthermore, we will consider the final results obtained in both tasks (monologue and dialogues). And finally, we will establish a comparison between the two role-plays students performed regarding the results they obtained. This comparison will consist in discussing the results obtained of one student performing a role-play with a classmate s/he preferred taking an oral exam with (according to their preferences stated in the 'Student Questionnaire'), and on the other hand, the same role-play performance with a student s/he did not want to take an oral exam with (with regards to the 'Student Questionnaire').

We believe that the results obtained from that comparison might reveal remarkable data concerning the influence in students' oral production in English. However, these data could lead to the reformulation and modification of the assessment methods.

The 'Speaking Diagnostic Test' shows the different variables (see section 2) that we considered in order assigning a certain score from 1 to 3 in each of them. The total score regarding all these variables is 18 points. Therefore, this test is our main instrument in order to get relevant data concerning students' results in the two tasks they performed: (a) Monologue

employed as a diagnostic placement test; and (b) Two dialogues (Role-plays) with classmates who wanted to work together ('Ideal Partners') and classmates who did not want to (Unwanted Partners'). These results will be discussed in the next subsection.

5.1. Students' Scores in the Monologue

As for the students' scores obtained in the monologue, the table below (Table 1) reflects the total amount of students who participated in the experiment, as well as the scores obtained in each of the variables and final scores.

Table 1.
Students' Scores in the Monologue

STUDENTS	MONOLOGUE						
	VARIABLES & SCORES						
	Fl.	Voc.	Gr.	Pr.	Co.	Comm. ability	FINAL SCORE
A	1	1	1	1	1	1	6/18
B	1	1	1	1	1	3	8/18
C	1	1	1	1	1	1	6/18
D	2	1	2	1	1	2	9/18
E	1	1	1	1	1	1	6/18
F	2	1	1	1	1	2	8/18
G	1	1	1	1	1	1	6/18
H	2	1	1	1	1	2	8/18
I	1	1	1	1	2	3	9/18
J	1	1	1	1	1	1	6/18

Due to research purposes, we are not going to analyse in full detail the results concerning the students' scores in the monologue, but we will proceed to comment on them in general terms since these results are the product of an initial proficiency level task ('Speaking Diagnostic Test').

This table (Table 1) provides indicative results, because of 10 students that are indicated in the table; only 2 students achieved an average result (9/18). These results seem to confirm that the students' initial English proficiency level did not reach our desirable expectations.

5.2. Students' Scores in the Dialogues

Concerning the two dialogues (role-plays), it is worth highlighting that in order to obtain relevant results, students indicated their preferences towards taking an oral exam with their peers in a 'Student Questionnaire' that was designed for the purpose of this research. Thus, according to their preferences ('IDP'/'UNP'), the following pairs were selected:

- 'Ideal Partner' ('IDP'): 'Student A' and 'Student E', 'Student B' and 'Student C', 'Student C' and 'Student F', 'Student D' and 'Student I', 'Student E' and 'Student D', 'Student F' and 'Student C', Student G' and 'Student B', 'Student H' and 'Student D', 'Student I' and 'Student E', and 'Student J' and 'Student D'.

- 'Unwanted Partner' ('UNP'): 'Student A' and 'Student I', 'Student B' and 'Student A', 'Student C' and 'Student A', 'Student D' and 'Student A', 'Student E' and 'Student F', 'Student F' and 'Student E', 'Student G' and 'Student H', 'Student H' and 'Student G', 'Student I' and 'Student D', and 'Student J' and 'Student A'.

Regarding Table 2 (Students' scores in the role-plays – 'Ideal Partners'), the 10 pairs who wanted to take an oral exam together ('IDP') are shown, providing their score in each of the variables as well as their final score.

Table 2.
Students' Scores in the Role-plays – 'Ideal Partners'

PAIRS (Students)	VARIABLES & SCORES						FINAL SCORE
	Fl.	Voc.	Gr.	Pr.	Co.	Comm. ability	
A / E	1/2	1/2	1/2	1/2	1/2	1/2	A (6/18) E (12/18)
B / C	1/2	1/2	1/2	2/2	1/2	1/2	B (7/18) C (12/18)
C / F	2/1	2/1	2/1	2/1	2/1	3/3	C (13/18) F (8/18)
D / I	2/2	2/1	2/2	2/1	2/2	3/2	D (13/18) I (10/18)
E / I	2/2	2/1	2/1	2/1	2/1	2/2	E (12/18) I (8/18)
F / C	1/2	1/2	1/2	1/2	1/2	3/3	F (8/18) C (13/18)

G / B	2/2	1/2	1/2	1/1	1/1	3/3	G (9/18) B (11/18)
H / D	2/2	2/2	2/2	2/2	2/2	3/3	H (13/18) D (13/18)
I / E	2/2	1/2	1/2	1/2	1/2	2/2	I (8/18) E (12/18)
J / D	1/2	1/2	1/2	1/2	2/2	1/3	J (7/18) D (13/18)

From these results we may say that only 5 individuals did not reach the average score (9/18) with regards to the final score (18 points). Thus, some examples such as students 'A', 'B', 'F', 'I', and 'J' do not reach the average class level. Accordingly, we will now comment these students in full detail:

'Student A' does not show fluent speech, which is reflected in long pauses and breaks. Vocabulary expressions and grammar use is rather poor. Furthermore, he does not employ specific vocabulary demanded in the subject. He pronounces words incorrectly, but he tries to be coherent. However, he does not adapt to his partner's level, which shows a great communicative ability in English. Finally, this student obtains 6/18 points in his final score.

'Student B' does not speak fluently. He uses long pauses, hesitations and long breaks. Furthermore, his vocabulary and grammar is very poor and does not use the specific topic vocabulary. The message he conveys in incoherent and difficult to understand, because he does not use linkers and connectors. Finally, his final score (7/18) is justified.

'Student F' does not speak fluently either and long pauses and hesitations are present during his speech. The vocabulary and expressions used are very poor, as well as poor grammatical structures with significant mistakes. Regarding pronunciation, most of the words he uses are incorrectly pronounced. Furthermore, the message he tries to transmit is difficult to understand, because of his lack in the use of connectors. Taking into account all these comments, we could certify that he obtained 8/18 in the final score.

'Student I' speaks mostly fluently, although he uses poor vocabulary and grammar. He does not try to employ the specific technical vocabulary and expressions required in this subject. In addition, he pronounces most words incorrectly with non-existing intonation or word stress. Finally, the message he conveys is completely incoherent due to the lack of connectors; however, he tries to adapt to his partner's level through the use of turn

taking and interrupting techniques. These data confirm his final score obtained (8/18).

'Student J' does not show fluent speech, which is reflected in long pauses, hesitations, and long breaks. Furthermore, vocabulary expressions and grammar use is poor. He does not employ the specific vocabulary demanded in the subject. Although he pronounces words incorrectly, he makes an effort in transmitting a coherent discourse and be understandable. However, he does not adapt to his partner's level, which shows a great communicative ability in English. Finally, for these reasons this student obtains 7/18 points in his final score.

On the other hand, students 'C' (13/18), 'D' (13/18), 'E' (12/18), 'G' (9/18), and 'H' (13/18) were perceived as to have obtained the highest scores in this role-play task, in which they could express their preferences when choosing their pairs. The five of them showed better oral skills than the rest of the subjects taking part in the same task. They spoke mostly fluently with a good command of vocabulary (using appropriate topic vocabulary); they used basic grammatical structures, although they still made some key mistakes. Furthermore, they were able to pronounce mostly all words correctly with correct intonation and word stress, so as their messages were mostly coherent but with few linking words. Finally, their communication ability was excellent, as these four students used turn taking and interrupting techniques very effectively.

In contrast to our previous data and taking Table 3 (Students' scores in the role-plays – 'Unwanted Partners') as a reference, the 10 pairs who did not want to take an oral exam together ('UNP') are shown, providing not only their score in each of the variables, but also their final score (up to 18 points).

Table 3.
Students' Scores in the Role-plays – 'Unwanted Partners'

PAIRS (Students)	Fl.	Voc.	Gr.	Pr.	Co.	Comm. ability	FINAL SCORE
DIALOGUES (ROLE-PLAYS)							
VARIABLES & SCORES							
A / I	1/2	1/1	1/2	1/1	1/2	1/2	A (6/18) I (10/18)
B / A	1/1	1/1	1/1	1/1	1/1	1/1	B (6/18) A (6/18)
C / A	1/1	1/1	1/1	2/1	1/1	2/1	C (8/18)

D / A	2/1	2/1	2/1	2/1	2/1	1/1	A (6/18) D (11/18)
E / F	3/1	3/1	3/1	2/2	2/1	2/2	A (6/18) E (15/18)
F / E	1/3	1/3	1/3	2/2	1/2	2/2	F (8/18) E (15/18)
G / H	1/2	1/2	1/1	1/1	1/2	1/3	F (8/18) G (6/18) H (11/18)
H / G	2/1	2/1	1/1	1/1	2/1	3/1	H (11/18) G (6/18)
I / D	2/2	1/2	2/2	1/2	2/2	2/3	I (10/18) D (13/18)
J / A	1/1	1/1	1/1	1/1	1/1	1/1	J (6/18) A (6/18)

As table 3 illustrates, the students who obtained lower scores in this case were 'A' (6/18), 'B' (6/18), 'C' (8/18), 'F' (8/18), 'G' (6/18), and 'J' (6/18). The six of them were below the average (9/18) when performing the task, as they were not able to speak fluently, because they needed time to plan what to say using long pauses and hesitations. The vocabulary and grammatical structures they employed were very poor and contained many significant mistakes. In addition, they made many pronunciation mistakes with non-existing intonation and word stress in their production. In general, their messages were incomplete and incoherent in most cases, as they did not use connectors. And finally, their communicative ability was very poor too, since they were not able to use turn taking or interrupting techniques. For all these reasons, their final scores in this task are justified.

In the light of all the results shown at the beginning of this section, further discussion is presented taking the following Research Question as starting point: *(RQ) "To what extent students' preferences in working with other classmates influence their English oral production?"*

The analysis of the data and the information obtained in tables 2 and 3 (see section 5.2) allows us to compare the scores obtained in both role-plays ('IDP' and 'UNP'). In addition, these data could lead us to suggest the following concepts: 'Higher Performance', 'Invariable Performance', and 'Lower Performance', since these terms could present a certain degree of ambiguity if we take into consideration the variability of assessment criteria. Let us recall that the rubric employed for the assessment of both role-plays included a total score of 18 points. Thus, we refer to 'Higher Performance' in the case of students who improved their performance results when

interacting with a classmate they chose in the 'Student Questionnaire'. We understand as 'Invariable Performance' in the case of students whose oral performance did not imply any change in their final score. And 'Lower Performance', in the case of students whose final scores dropped.

Figure 2 illustrates the final percentages obtained with regards to the comparison between students who wanted to take an oral exam together ('IDP') and those who did not want to interact with certain classmates ('UNP'), to what their preferences in the 'Student Questionnaire' are concerned. Therefore, these percentages have been calculated comparing the students' final scores in the rubric (up to 18 points) in tables 2 and 3 (see section 5.2). Thus, our initial hypothesis was that students who preferred working together would get higher results than those who did not want to. For this reason, we have only focused on the evolution of students comparing their results obtained in the 'IDP' role-play with regards to their results in the 'UNP' role-play.

Thus, taking into account the previous classification ('Higher', 'Invariable', and 'Lower Performance') and the percentages shown in Figure 2, we will now comment the following:

- 'Higher Performance': Students 'E' and 'I' obtained higher scores in the 'UNP' role-play than their scores obtained in the 'IDP' role-play. Student 'E' got 12/18 points ('IDP' role-play) and 15/18 points ('UNP' role-play). Student 'I' got 8/18 ('IDP' role-play) and 10/18 ('UNP' role-play).

- 'Invariable Performance': Students 'A' and 'F' remained the same in both role-plays results. Student 'A' got 6/18 and student 'F' got 8/18.

- 'Lower Performance': Students 'B', 'C', 'D', 'G', 'H', and 'J', obtained lower scores. Student 'B' got 7/18 ('IDP' role-play) and 6/18 ('UNP' role-play). Student 'C' got 13/18 ('IDP' role-play) and 8/18 ('UNP' role-play). Student 'D' got 13/18 ('IDP' role-play) and 11/18 ('UNP' role-play). Student 'G' got 9/18 ('IDP' role-play) and 6/18 ('UNP' role-play). Student 'H' got 13/18 ('IDP' role-play) and 11/18 ('UNP' role-play). Finally, student 'J' got 7/18 ('IDP' role-play) and 6/18 ('UNP' role-play).

All these results considered are illustrated in the following table (Table 4):

Table 4.
Students' Scores in the Role-plays – 'Ideal Partners' and 'Unwanted Partners'

ROLE-PLAYS		
STUDENTS	'IDEAL PARTNERS'	'UNWANTED PARTNERS'
A	6/18	6/18
B	7/18	6/18
C	13/18	8/18
D	13/18	11/18
E	12/18	15/18
F	8/18	8/18
G	9/18	6/18
H	13/18	11/18
I	8/18	10/18
J	7/18	6/18

To conclude, we find relevant to present the percentages of the 10 students who participated in this study, in order to observe to what extent their performances improved, remained the same, or decreased when interacting with someone they did not want to interact with.
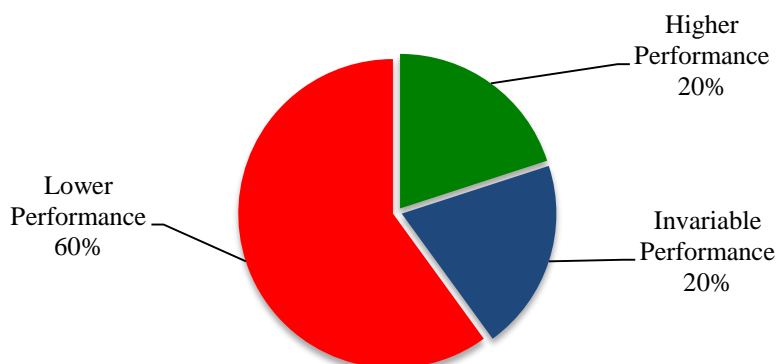
Figure 2. Percentage of results with 'Unwanted Partners'

5.3. Conclusions

The present chapter has been focused at presenting the results obtained from our study. These results have revealed the students' scores in the monologue and the students' scores in the dialogues.

As we have explained at the beginning of this chapter, we have established a comparison between the results obtained in both types of tasks (monologue and dialogue).

Bearing all the data obtained in mind, we may draw the following conclusions: Firstly, it is remarkable that our study goes beyond the objectives for speaking assessment suggested by O'Malley and Pierce (1996, p. 63) "initial identification" and "placement of students". The reason for this lies in the fact that our study adopts the initial identification objective represented by the 'Student Questionnaire' in an attempt to establish the students' profile and their personal information concerning their attitude towards learning English. This initial stage implies the success in the development of the second objective "movement from one level to another" as a means to stimulate the learning of English as a Second Language (ESL) in ESP contexts. Secondly, the elaboration and design of the 'Speaking Diagnostic Test' supports the idea proposed by Taufiqulloh (2009), in which he suggests the use of analytic rating scales in order to diagnose communicative information referred to students' needs and strengths. Therefore, we consider this test as a starting point for the teacher to establish and design further tasks in which s/he will be able to observe the students'

progress in their oral production. Nevertheless, we should not forget that the assessment process not only concerns the teacher's point of view, but also, this evaluation process should welcome students' active participation in it. Finally, regarding both the 'Speaking Diagnostic Test' and the 'Student Questionnaire', we draw to the conclusion that these are useful tools for the teacher. On the one hand, they allow the teacher to determine relevant features that could be considered in the distribution of students for the second task (role-play). On the other hand, our final results reveal that the students' scores in the 'Ideal Partner' role-play were significantly higher (60%) in relation to 'Unwanted Partners' (20%). Thus, to summarize these ideas, the teacher could employ all these data to foster students' oral production and mitigate their weaknesses when they have to interact with someone they do not feel comfortable with.

6. General Conclusions and Suggested Further Research

The present work focuses on analyzing to what extent students' oral competences (pronunciation, fluency, grammar, etc.) affect their peers' English oral production. Particularly, we intended to study effective procedures to assess oral production in ESP contexts.

We have seen that our study was based on the design of a rubric ('Speaking Diagnostic Test') that was employed so as to test students' initial performances in English. This rubric has been useful to assess students' performances in the two role-plays. Comparing the scores obtained in both role-plays we could conclude that our initial hypothesis has been supported by such scores. Then, when students interacted with classmates they had positively chosen in the 'Student Questionnaire', they performed significantly better than when they interacted with those ones they did not want to.

Our data have shown that there is a tendency towards a 'Lower Performance' (60%) when students perform a role-play with 'Unwanted Partners'. In contrast, we have observed that our initial hypothesis about a 'Higher Performance' (20%) is reflected in students who wanted to work together. Nevertheless, some students' oral performances remained the same ('Invariable Performance') final score without being affected by their partners (20%).

While the present study has attempted to examine to what extent students' preferences towards taking an oral exam with their peers influence their oral production, its results and further conclusions must be taken into

consideration in order to reformulate and modify the assessment methods. However, this study has some limitations, and the recognition of these should help refine future research efforts:

Firstly, it is important to consider the modification and/or implementation of the different variables used in the rubric in order to focus more in depth on students' weaknesses. This fact may lead to reconsider students' needs in English ESP courses.

Secondly, regarding the 'Student Questionnaire', we could include more specific questions and statements so as to obtain more precise information about the students' preferences. We could include a list of reasons why students would (not) choose a classmate, as we have found that students felt reluctant to provide such information.

Thirdly, we concede that the number of subjects has been a limitation in our study. Bigger groups of students were not available at the time, and with more students we could have found more nuances.

Finally, we think it is necessary to investigate not only on the use of rubrics on the part of the teacher ('Teacher assessment'), but also their use on behalf of students ('Peer assessment'). This last idea could contribute to the students' awareness on a series of limitations and aspects to consider when taking part in an oral exam in English for ESP contexts.

This study has only been a first step towards the design of more complex rubrics with bigger quantities of students as subjects. In the light of the results of our study, we view a quite open field to further explore the design and use of rubrics in ESP contexts.

References

Allen, D. & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE—Life Sciences Education, 5*, 197–203.

Campbell, et al. (2001). Peer versus self-assessment of oral business presentation performance. *Business Communication Quarterly, 64*(3), 23-42.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In Richards, J. C., & Schmidt, R. W. (Eds.), *Language and Communication*, 2-27. London: Longman.

Canale, M. & Swain, M. (1980). Theoretical Bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1-47.

Dunbar et al. (2006). Oral Communication Skills in Higher Education: Using a Performance-Based Evaluation Rubric to Assess Communication Skills. *Innovative Higher Education, 31*(2), 115-128. Springer.

Graham, E., & Mignerey, J. (1990). A longitudinal study of college students' communication competence. *Communication Education, 39*, 1-14.

Jones, E. (1994). *Essential skills in writing, speech and listening, and critical thinking for college graduates: Perspectives of faculty, employers, and policymakers.* University Park, PA: National Center for post-secondary Teaching, Learning, and Assessment.

Kavanagh, M., & Drennan, L. (2008). What skills and attributes does an accounting graduate need? Evidence from student perceptions and employer expectations. *Accounting and Finance, 48*(2), 279-300.

O'Malley, J, & Pierce, L. V. (1996). *Authentic assessment for English language learners*. Addison Wesley Publishing Comp, Inc.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*, 129-144.

Schwartz, D., & Arena, D. (2013). *Measuring What Matters Most. Choice-Based Assessments for the Digital Age.* Cambridge, Massachusetts. London, The MIT Press.

Spandel, V. (2006). In defense of rubrics. *English Journal, 96*(1), 19-22.

Stoynoff, S. (2013). Looking backward and forward at classroom-based language assessment. *ELT Journal, 66*(4), 523-532. Oxford University Press.

Taufiqulloh, S. (2009). Designing Speaking Test. Eksplanasi Volume 4 Nomor 8.

Underhill, N. (1987). *Testing spoken language.* Cambridge: Cambridge University Press.

Wilson, M. (2006). *Rethinking rubrics in writing assessment*. Porstmouth, NH: Heinemann.

Appendix 1.
Speaking Diagnostic Text

SPEAKING DIAGNOSTIC TEST – <u>EX 1005</u> (ROLE-PLAY INTERACTION) NAME GROUP DATE SCORE /18				
	1	2	3	COMMENTS
FLUENCY	Does not speak fluently. Uses long pauses, hesitations and long breaks	Speaks mostly fluently. Some pauses and hesitations.	Speaks very fluently. Few or non-existing hesitations and pauses.	
	1	2	3	COMMENTS
VOCAB.	Uses poor vocabulary and expressions. Does not use or has problems using specific topic vocabulary.	Uses basic vocabulary and expressions. Mostly uses some appropriate topic vocabulary	Uses an appropriate wide variety of specific vocabulary and expressions for the topic of the conversation.	
	1	2	3	COMMENTS
GRAMMAR	Uses poor grammatical structures. Many significant mistakes.	Uses basic grammatical structures. Several key mistakes.	Uses accurate and appropriate grammatical structures. Very limited mistakes	
	1	2	3	COMMENTS
PRON.	Pronounces most words incorrectly. Incorrect use or non-existing intonation	Pronounces almost all words correctly. Mostly uses correct	Pronounces all words correctly. Uses correct intonation and word stress.	

	and word stress.	intonation and word stress.		
COHERENCE	1	2	3	COMMENTS
	The message is incoherent and difficult to understand. Does not use connectors and linkers.	The message is mostly coherent. Uses few linkers and connectors.	The message is coherent. Uses suitable linkers and connectors. Correct content organization.	
COMM. ABILITY	1	2	3	COMMENTS
	Does not adapt to other speakers' levels. Does not use turn taking and interrupting techniques. Sometimes, he /she remains silent until the other speaker finishes.	Mostly tries to adapt to other speakers' levels. Few attempts and mistakes when using turn taking and interrupting techniques.	Flexibility to speakers of different levels. Uses turn taking and interrupting techniques effectively. Helps other students when they are stuck in the conversation.	