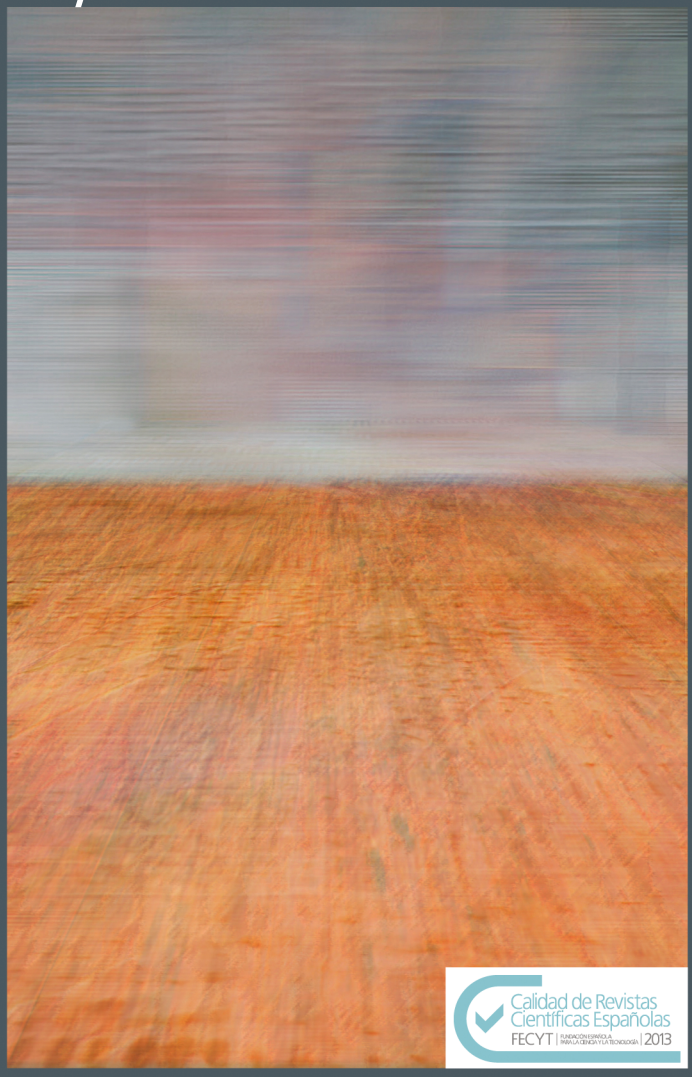
revista de EDUCACIÓN
Nº 369 JULIO-SEPTIEMBRE 2015

Rendimiento de los estudiantes de primer curso de universidad en las tareas orales en la simulación de un examen de acceso a la Universidad

Performance of first year university students in the speaking tasks of a simulated University Entrance Examination

Jesús García Laborda
Gloria Luque Agulló
Ana Isabel Muñoz
Margarita Bakieva



GOBIERNO DE ESPAÑA
MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE

Calidad de Revistas Científicas Españolas
FECYT | FUNDACIÓN ESPAÑOLA PARA LA CIENCIA Y LA TECNOLOGÍA | 2013

Performance of first year university students in the speaking tasks of a simulated University Entrance Examination

Rendimiento de los estudiantes de primer curso de universidad en las tareas orales en la simulación de un examen de acceso a la Universidad

DOI: 10.4438/1988-592X-RE-2015-368-292

Jesús García Laborda
Universidad de Alcala
Gloria Luque Agulló
Universidad de Jaén
Ana Isabel Muñoz
Universidad Católica de Ávila
Margarita Bakieva
Universidad de Valencia

Abstract

The debate over the Foreign Language Test (English) in the University Entrance Examination (PAU) has become a critical issue in the Spanish Educational system. Despite the Ministry of Education's interest in changing a test that has its strong emphasis on reading, writing and grammar but a general negligence towards listening and speaking, limited changes have been done by the regional administrations. This absence of evaluation of oral aspects in the exam may lead to a disregard for those aspects in the last levels of the educational process and, subsequently, low competence levels in oral language.

To test the oral competence of high school graduates, a set of speaking tasks were designed and delivered to 169 first semester students from three different Spanish universities who had recently taken the *PAU* to observe their speaking. Results showed that the large majority of the students in the research have a B1

Received: 17-11-2014 Accepted: 13-03-2015

with an enormous percentage of students in the A1-A2 band. These results do not significantly differ from a previous study from the Ministry of Education and show the need to include speaking tasks in the University Entrance Examination or the future High School graduation diploma. Qualitative studies also suggested that the way to deliver the tasks and the testing approach could possibly have an effect on the students' performance.

Keywords: Testing; oral competence; Common European Framework for modern languages (CEFR); impact; comparative studies.

Resumen

El debate sobre la Prueba de idiomas (inglés) en el Examen de Acceso a la Universidad (PAU) se ha convertido en un tema crítico en el sistema educativo español. A pesar del interés del Ministerio de Educación en el cambio de una prueba que tiene su fuerte énfasis en la lectura, la escritura y la gramática, sino una negligencia general hacia la comprensión y expresión oral, las administraciones regionales han hecho pocos cambios.

El hecho de no evaluar los aspectos orales en la prueba puede disminuir la importancia que estos aspectos reciben en los últimos niveles del ámbito educativo, y por ende, en el nivel oral de los alumnos. Para comprobar la competencia oral de los egresados de la escuela secundaria, un conjunto de tareas de habla fueron diseñadas y administradas a 169 estudiantes de primer semestre de tres universidades españolas diferentes que habían realizado recientemente el PAU para observar su competencia oral. Los resultados mostraron que la gran mayoría de los estudiantes en la investigación tienen un nivel B1 con un enorme porcentaje de estudiantes en el nivel A1- A2. Estos resultados no difieren significativamente de un estudio previo del Ministerio de Educación y muestran la necesidad de incluir las tareas que hablan en la Prueba de Acceso a la Universidad o el Examen Final de Bachillerato. Los estudios cualitativos también sugirieren que la forma de realizar las tareas y el enfoque de las pruebas podría tener un efecto sobre el rendimiento de los estudiantes.

Palabra clave: Exámenes; competencia oral; Marco Común Europeo para las lenguas modernas (MCER); impacto; estudios comparativos.

Introduction

The University Entrance Examination is the most important high-stakes general exam in Spain. The inclusion of a Foreign Language section dates back more than 25 years but despite the changes in language teaching

over time, this section has seen very limited changes in its construct. Obviously, the students' and social needs have changed in that period of time and, as a consequence, the test has become obsolete and provides only a limited quantity of information. Moreover, this information is used to obtain inferences about the students' linguistic potential and also to set the level to enter a university degree. In contrast with the exam outdated construct, oral skills have been one of the key issues in the recently passed Organic Law that Modifies the Quality of Education (LOMCE) (2014) in Spain. In fact, a great emphasis has been placed in the fact that Spanish students should graduate from high school with an excellent oral command or, at least, one foreign language. In Spain and many other European countries this has been an issue of major concern. However, what makes the case of Spain even more critical are the poor results found in international and national evaluations such as the European Survey of Language Competence (http://ec.europa.eu/languages/policy/strategic-framework/documents/language-survey-final-report_en.pdf). Three skills, reading, writing and listening were measured, and Spain was found to show one of the lowest results, particularly in the case of the oral skill, but also in the written ones. Besides, it is generally acknowledged in the labor market as well as by the higher educational institutions that Spanish students lack the necessary skills to pursue university studies in which English is either a communication language or, at least, a supporting tool for learning.

Recent studies done by the Institute of Educational Evaluation which depends on the Spanish Ministry of Education, Culture and Sports (MECD henceforth) state that just about 45% of the students would be able to achieve a B1 after graduating from high school. García Laborda, Amengual Pizarro & Litzler (2013) hesitate whether this language competence is acceptable to face the labor market or university studies. By using the Common European Framework for Languages (CEFR) in English for Specific Purposes, these authors consider that the number of situations in which the high school graduates would only be able to use very short dialogues (adscribing to a B1 level) is high, and thus they foresee serious communicative problems beyond social interaction when the grammar structures, but especially vocabulary and discourse, become the most important part of professional communication. In the same conference paper they also mention that the results of the European survey might be unrealistic, as interviewers had very limited training in

Received: 17-11-2014 Accepted: 13-03-2015

oral testing and they could have brought some pre-conceptions about the results and about the candidates they would be interviewing.

The University Entrance Examination in Spain

The English section of the University Entrance Examination in Spain has received very little attention in research as compared to other standardized foreign language tests. Only in the last decade a number of researchers have addressed such issue. Overall, there seems to be three main aspects that have centered most of the papers on the topic: washback (Amengual Pizarro, 2009), results analysis and proposals to change its current construct, which has not been modified for over 20 years (see, Fernández Álvarez & Sanz Sainz, 2005). Although there might seem to be clear distinction between these three aspects, in fact, they are clearly interwoven..

What seems to be clear is that with little variations in Galicia and Catalonia autonomous communities, the changes have been limited to format (such as the number of words in the composition section or the number of items in language use) and the addition of listening comprehension sections in the two Spanish communities mentioned above, which also have an additional official language. In 2005, a volume edited by Herrera Soler & García Laborda tried to indicate this lack of studies, particularly dealing with validation, the most relevant aspect.. However, up to that moment, and still today, the validation studies have been limited in content and scope. García Laborda (2006) pointed that the validation studies until then had only been done occasionally in four universities: *Granada, Complutense de Madrid, Baleares* and *Politécnica de Valencia*. After the 2005 volume there were some more papers published on this issue, of which it is worth mentioning the monograph of *Revista de Educación* in 2011, including aspects ranging from intercultural considerations to delivery through computers.

Given this lack of studies, the Spanish Ministry of Education, Culture & Sports (MECD) is aware of a number of provisions to be considered in order to overcome these problems. Firstly, there is a need to revise the educational paradigm in Spain, particularly in connection to language policies. When these policies are improved, assessment, evaluation and testing must fulfill a more relevant role than in previous educational

models. This increase in evaluation policies should be first, adequately shaped, as it may influence teaching outcomes, second, have a periodical role, as it will surely provide important benefits for schools, teaching and learning, and third, have a moderate impact and account for the socio-economic inequalities of the Spanish society, ensuring that not all resources provided to school depend on the results of these assessments.

Bearing in mind the MECD provisions, the OPENPAU project proposed different alternatives and followed two main lines to address the analysis of current limitations of Spanish students. On the one hand, the coordinator of the project established lines of cooperation between the research project and the MECD. The general idea was that the experience of the OPENPAU project[1] served to provide ideas to improve the current situation and also to revise an internal report on the high school leavers' English proficiency. As a counterpart, the MECD offered to provide information on the current research through the online delivery of the research database.

This paper looks at the students' speaking performance in paired interviews. The results are examined according to four criteria (accuracy, fluency, interaction and coherence) in three universities and comprise students from four autonomous regions from Spain. According to what has been said above, the study is justified by the lack of serious studies that can analyze the current speaking situation of high-school graduates and observe whether the results of the MECD can be contrasted by non-institutional research. While there is a general claim that there is a remarkable need to incorporate the speaking tasks in the University Entrance Exam, previous studies are limited, as far as it is currently known, to the one by the MECD and this one hereby presented. This is currently based on two premises: 1) test tasks have a significant impact in what is being taught; and 2) delivery could facilitate the implementation of speaking tasks.

The following study addresses first some of the current issues associated to pair and group speaking tests delivery, then it shows the research questions and then proceeds to the experimental analysis of speaking interviews by 169 university students. Finally, it addresses the results of this experimental research with the one done by the MECD and finishes with some conclusions that could provide with potential ideas for the current exam and the future High-School graduation diploma.

((1) Orientación, Propuestas y Enseñanza para la Sección de Inglés en la Prueba de Acceso a la Universidad.

Received: 17-11-2014 Accepted: 13-03-2015

Literature review

Although over the last decades there have been debates on the need of designing and implementing new speaking tasks in high-stakes tests, since the beginning of the 21st century, there have been two clear lines, one related to speaking tasks of online delivered speaking tasks (Bernhardt, Rivera & Kamil, 2004; Chapelle & Douglas, 2006; Vitiene & Miciuliene, 2008; Sawaki, Stricker, & Oranje, 2009; García Laborda, 2010b and others) and, the other, in-person delivered speaking tasks (Nakatsuhara, 2013). Among the latter, one of the most current trends focuses in pair and group assessments. Although integrated tasks have attracted both computer based and in-person test administrators (Sawaki, Stricker, & Oranje, 2009), it is worth considering what the advantages of group interviews have over computers in specific contexts. García Laborda & Royo (2007) mentions a number of difficulties that make computer based testing profitable on the long run, as high investments are necessary both in software and hardware. This may be the most important reason why many educational administrations have not implemented computer based assessments in high-stakes tests. However, face-to-face interviews or even telephone based interviews like, for instance, the Simulated Oral Proficiency Interview (SOPI) are time-consuming and still imply a high cost in human and resources costs (Heilmann, 2012). Thus, for many institutions pair/group in-person language tests are a feasible response. Given the current context in Spain and the very same compulsory nature of the PAU exam, issues related to individual features, grouping or personality could have a potential effect on test takers. However, it is necessary to promote forms of communication in tests that elicit interaction as a significant part of the communication construct (Brooks, 2009) and provide better inferences at lower costs (Dunbar, Brooks & Kubicka-Miller, 2006). Since the introduction of these tasks should aim at a more "real" communication, it would be expected that the use of the speaking tests should have a positive effect on the learners and thus provide positive wash back (Muñoz & Alvarez, 2010).

Paired speaking test tasks have become very common in many international tests, especially in the Cambridge board of examinations (Shaw & Weir, 2007), and Nakatsuhara (2013) mentions a large number of examples in many other parts of the word. Paired tests have a number

of benefits, including their reduced cost and time efficiency and their focus on co-constructed dialogue/speaking interaction (Galaczi, 2008; Nakatsuhra, 2006; Gan, 2008; Gan, 2010), particularly in countries like Spain, in which cooperation and sharing in education are highly regarded. Additionally, they may provide better opportunities for weaker learners (Elder, Iwashita & McNamara, 2002), facilitate fluency (Gan, Davison & Hamp-Lyons, 2009), provide a special role to body language and facilitate the observation of high level speaking functions

Although these factors have been observed internationally, up to now no formal test administered in Spain uses group or even pair assessment for oral discourse. In that sense, this paper seems to be a first approach especially in the context of high-stakes tests.

Design and implementation of research

Research questions

Given the current situation in Spain and the literature review, the following questions needed to be researched for the purpose of this study:

RQ1: *Are there any significant differences between the study on the 2nd Baccalaureate students done by the Instituto Nacional de Evaluación Educativa (INEE) and our current research?*

RQ2: *Is the approach proposed in this research valid to deliver the University Entrance Examination?*

It is worth considering whether group delivered speaking tests are more adequate than the one-to-one face-to-face interviews in the Spanish context.

Research method and study participants

Participants

The research team recorded a total of 85 paired interviews in four different autonomous communities in Spain, namely Castilla-La Mancha, Castilla León, Andalucía and Madrid (Table 2). These regions do not share

the same language entrance exam, but the one aspect they have in common is that they do not include an oral section (for further details on exams, see Bueno & Luque, 2012). A University was chosen from each autonomous community to make the sample representative and also maximize economic and personnel resources: Universidad de Alcalá (Madrid and Castilla La Mancha), Universidad Católica de Ávila (Castilla-León) and Universidad de Jaén (Andalucía). Students were volunteers in their first year at University. Table II shows the amount and frequency of participants and their region/community of origin. Table III indicates the number and frequency of participants who volunteered for the study depending on the faculty they had ascribed to. Because of the diverse communities of origin, universities and faculties students derive from, it was considered that the range and variety of participants evaluated in the study would show a comprehensive view of the level of oral competence in beginning University students in Spain.

TABLE II. Number of participants selected for study, percentages and Universities of origin

University	Frequency	Percentage
Andalucía (Universidad de Jaén)	39	23,1
CAM (****)	46	27,2
CLM (****)	21	12,4
Castilla- León (****)	55	32,5
Other (****)	8	4,8
Total	169	100,0

Received: 17-11-2014 Accepted: 13-03-2015

TABLE III. Faculties selected for study and percentage of participants.

Faculties	Frecuency	%	University
Psychology	38	22,5	Jaén-Andalucía
Pre-school teaching	16	9,5	Alcalá-Madrid/Castilla La Mancha
Primary Teaching	52	30,8	Alcalá-Madrid-Castilla La Mancha
Nursery	57	33,7	Universidad Católica de Ávila-Castilla León
Law	4	2,4	Universidad Católica de Ávila-Castilla León
Total	169	100,0	

Research tasks and data collection

Measure of proficiency levels.

In order to measure the students' proficiency, the researchers used the CEFR rating levels, ranging from A1 to B2, to assess the participants' overall oral competence. According to this, four rating criteria were used to measure the students' performance: accuracy, fluency, interaction and coherence. Scores from 0 to 3 were assigned according to their performance being 3= excellent, 2=average and 1=poor, 0 was assigned in very few cases to students who did not respond at all or their performance was unusually poor. The students' responses were video-recorded, the interviews numbered and finally partially transcribed.

Raters

Raters were trained teachers with long experience either in language testing for standardized tests or for the University Entrance Examination[2]. Interviews were then assessed and graded by six of these raters, who agreed on each student's global competence and then conveyed the

(2) An expert judgment was carried using the teachers/raters, who had ample experience in evaluation tasks for standardized tests or in the University Entrance Examination (a minimum of six years as evaluators). All had taken part in a training seminar. Once an agreement on the aspects to be evaluated was achieved, it was approved by all raters and used in all interviews.

Received: 17-11-2014 Accepted: 13-03-2015

grades assigned for each individual criterion. Agreement among raters was reached on how speaking tasks would be delivered before the process commenced. One interviewer/rater was in charge of asking and delivering the adequate questions. It was somehow active in task 1 (see below) and actively observing and supporting (in different degrees) in tasks 2 and 3. On these tasks the interviewer only participated if there was a clear breakthrough in the line of conversation.

Speaking tasks

To organize the design and delivery of the speaking tasks, a questionnaire was distributed to 16 PAU coordinators from all the Spanish communities to find out their opinions about the kind of tasks that they considered best in the hypothetical case that speaking tasks were implemented in the PAU from 2012. Their responses showed that they preferred three types of tasks: social-warm up personal questions, a picture description and a role-play. Also, according to their responses, the benchmark in the last year of high school should be a B1 in the CEFR and three tasks were considered and were delivered in the following order: 1) Informational dialogue; 2) Picture description+ question-response dialogue; 3) Prompt-based role play. Question one consisted in the interviewer asking individual questions on social topics such as sports, hobbies, family members, academic interests and so. In question two the student was assigned a picture randomly from a ten picture-set and a description for two minutes was to be provided. Then the classmate asked two opinion questions such as "why do you think they are here?" or "what do you think it will happen after?" Finally, for question three the students were assigned a card stating a case to discuss such as "organize a party at your place with the help of your partner" or "organize a study session for your next exam". Table 4 shows the questions and construct and objectives of each of these three tasks.

TABLE IV. Speaking tasks. Type of task, number of participants, information requested, goals and average time to complete each task.

Task type	Number of participants	Information requested	Grade of engagement	Goal	Task average time
Informational dialogue	2 (interviewer-test taker)	Personal questions of origin, family, hobbies and so	Rather limited to 1-1 interaction	Warm-up	3 minutes (per student)
Picture description + 2 question-response dialogue	2 (candidate-candidate)	Descriptive plus guessing or justification	Semi-passive (monologue) + semi-active (response to 2 questions)	Competence assessment Initial interaction	2 minutes for description + about 2 minutes per student
Prompt-based role play	2 (candidate-candidate)	Adequate to the prompt (case)	Active	Interactional participation Free speech Cooperative – Interactive task	4 to 5 minutes total

Data collection

The interviews were implemented between December 2012 and March 2013. Students volunteered and the interviews were given on class-days usually before or after the lessons. All of the students were enrolled in first semester English classes either for general or specific purposes. Students were constrained by their own class schedule so the research team decided to group them randomly without concerns for the proficiency level.

Results

The first thing to address in this study, given the number of issues that arose in the study of the Ministry was to observe what was the proficiency level that first year university students brought. Instead of choosing a criterion referenced assessment method, the research focused in the

Received: 17-11-2014 Accepted: 13-03-2015

proficiency levels on a norm reference assessment. That means that rather than observing whether students could achieve a B1 competence level, we observed how they could be placed according to their competence. Table V shows the global results obtained according to the CEFR.

TABLE V. Level of participants according to the CEFR.

Level	Number of students	Percentage
A1	32	18,9
A2	62	36,7
B1	57	33,7
B2	14	8,3
C1	4	2,4
Total	169	100,0

In our study about 55.6 did not obtain the minimum requirement of competence required in Foreign Languages by the Ministry of Education to graduate from high school. Moreover, the percentage of students who were above the required level was just 10.7. We considered this figure important because it was similar to that obtained in the European Survey of Language Competence.

Since we were observing the performance in a given test, the second aspect to be assessed consisted of observing how students were performing in the test within their own competence level across the test criteria. Table VI indicates the results of the students in the accuracy criteria. Accuracy was understood as "grammaticalness" or attachment to prescriptive standard grammar.

TABLE VI. Results of participants' performance for accuracy criteria. Measures range from zero to three. Each line indicates the total amount and percentage interviewers in different degrees gave to the participants' performance.

			Accuracy				Total
			0	1	2	3	
Degree	Psychology	Number of responses	2	20	12	4	38
		% within degree	5.3%	52.6%	31.6%	10.5%	100.0%
	Pre-school teaching	Number of responses	0	10	6	0	16
		% within degree	.0%	62.5%	37.5%	.0%	100.0%
	Primary teaching	Number of responses	0	25	26	1	52
		% within degree	.0%	48.1%	50.0%	1.9%	100.0%
	Nursery	Number of responses	0	35	15	7	57
		% within degree	.0%	61.4%	26.3%	12.3%	100.0%
	Law	Number of responses	0	1	3	0	4
		% within degree	.0%	25.0%	75.0%	.0%	100.0%
	Electrical Engineering	Number of responses	0	1	0	0	1
		% within degree	.0%	100.0%	.0%	.0%	100.0%
	Civil Engineering	Number of responses	0	0	1	0	1
		% within degree	.0%	.0%	100.0%	.0%	100.0%
Total		Number of responses	2	92	63	12	169
		%	1.2%	54.4%	37.3%	7.1%	100.0%

After observing the different frequencies in levels of accuracy, a further step was taken in the analysis to discover if different groups of students or different universities showed significant differences, that is to say, if the participants' origin or choice of degree would have a connection with their level of accuracy. The data (chi-square: 23.254) indicate that results did not show significant inter-group results and that within the groups most students tended to score either low or medium. This may well indicate a rater tendency to value students low or that actually students tend to underscore on specific oral tests. It was also observed that there was a degree dependency in the scores upon the degree of study. For instance, that was the case of the University of Alcalá school of education

Received: 17-11-2014 Accepted: 13-03-2015

where a significant difference between pre-service primary teachers tended to do better than their counterparts in pre-school. Finally, the Kendal Tau-b rank correlation coefficient (0.029) and a small asymptotic standard error (0.72) indicated the absence of association among the different samples. As a consequence, the results indicate that there are limited differences among the groups. However, there was a tendency to have lower scores in the pre-service teachers and the nurses.

The same analysis was applied to the other criteria referred to above, fluency, interaction and coherence (Tables VII to IX).

TABLE VII. Results of participants' performance for fluency criteria. Measures range from zero to three. Each line indicates the total amount and percentage interviewers in different degrees gave to the participants' performance.

			Fluency				Total
			0	1	2	3	
Degree	Psychology	Number of responses	1	9	17	11	38
		% dentro de Titulación	2.6%	23.7%	44.7%	28.9%	100.0%
	Pre-school teaching	Number of responses	0	5	7	4	16
		% within degree	.0%	31.3%	43.8%	25.0%	100.0%
	Primary teaching	Number of responses	0	12	31	7	50
		% within degree	.0%	24.0%	62.0%	14.0%	100.0%
	Nursery	Number of responses	0	21	25	11	57
		% within degree	.0%	36.8%	43.9%	19.3%	100.0%
	Law	Number of responses	0	2	2	0	4
		% within degree	.0%	50.0%	50.0%	.0%	100.0%
	Electrical Engineering	Number of responses	0	0	1	0	1
		% within degree	.0%	.0%	100.0%	.0%	100.0%
	Civil Engineering	Number of responses	0	0	1	0	1
		% within degree	.0%	.0%	100.0%	.0%	100.0%
Total		Number of responses	1	49	84	33	167
		%	.6%	29.3%	50.3%	19.8%	100.0%

Received: 17-11-2014 Accepted: 13-03-2015

TABLE VIII. Results of participants' performance for interaction criteria. Measures range from cero to 3. Each line indicates the total amount and percentage interviewers in different degrees gave to the participants' performance.

			Interaction				Total
			0	1	2	3	
Degree	Psychology	Number of responses	1	12	17	8	38
		% within degree	2.6%	31.6%	44.7%	21.1%	100.0%
	Pre-school teaching	Number of responses	0	6	6	4	16
		% within degree	.0%	37.5%	37.5%	25.0%	100.0%
	Primary teaching	Number of responses	0	14	32	6	52
		% within degree	.0%	26.9%	61.5%	11.5%	100.0%
	Nursery	Number of responses	1	24	21	11	57
		% within degree	1.8%	42.1%	36.8%	19.3%	100.0%
	Law	Number of responses	0	3	1	0	4
		% within degree	.0%	75.0%	25.0%	.0%	100.0%
	Electrical Engineering	Number of responses	0	1	0	0	1
		% within degree	.0%	100.0%	.0%	.0%	100.0%
	Civil Engineering	Number of responses	0	0	1	0	1
		% within degree	.0%	.0%	100.0%	.0%	100.0%
Total		Number of responses	2	60	78	29	169
		%	1.2%	35.5%	46.2%	17.2%	100.0%

Again, after observing the frequencies, an analysis was carried out to check whether significant differences could be found depending on group of students or university/faculty of origin. In this case the x^2 was 13,637 (not-significant) with an asymptotic significance of 0,752. This clearly indicates that the curve had a right tendency and that there was no significance in the chi-square inter-group results.

These statistic results indicate that the students performed better in this criterion. In fact, the tests show a clear tendency to the average score with a higher degree towards Excellency in all the groups. This excellence is not so extremely high in either group but it averages a total of 19.8%. Psychology in Jaén obtains better scores followed by the Primary teachers

Received: 17-11-2014 Accepted: 13-03-2015

of the school of education in Madrid. The Tau-b de Kendall (-0.101, Standard error 0.70) indicates a higher degree of association than in the previous criterion but it is still rather limited.

TABLE IX. Results of participants' performance for coherence. Measures range from cero to 3. Each line indicates the total amount and percentage interviewers in different degrees gave to the participants' performance.

			Interaction				Total
			0	1	2	3	
Degree	Psychology	Number of responses	1	12	17	8	38
		% within degree	2.6%	31.6%	44.7%	21.1%	100.0%
	Pre-school teaching	Number of responses	0	6	6	4	16
		% within degree	.0%	37.5%	37.5%	25.0%	100.0%
	Primary teaching	Number of responses	0	14	32	6	52
		% within degree	.0%	26.9%	61.5%	11.5%	100.0%
	Nursery	Number of responses	1	24	21	11	57
		% within degree	1.8%	42.1%	36.8%	19.3%	100.0%
	Law	Number of responses	0	3	1	0	4
		% within degree	.0%	75.0%	25.0%	.0%	100.0%
	Electrical Engineering	Number of responses	0	1	0	0	1
		% within degree	.0%	100.0%	.0%	.0%	100.0%
	Civil Engineering	Number of responses	0	0	1	0	1
		% within degree	.0%	.0%	100.0%	.0%	100.0%
Total		Number of responses	2	60	78	29	169
		%	1.2%	35.5%	46.2%	17.2%	100.0%

Interaction also shows greater performance scores which resemble those obtained in Fluency. The x^2 for Interaction was 15.49 with an asymptotic significance of 0.628 with a clear tendency towards the 2-3 values, and again there was no significance in the chi-square inter-group results. Also medium-high scores were observed for this criterion. High scores (3) were observed in three groups but in this case, low grades (1)

were even less than in the previous criterion. Especially significant is the case of primary pre-service teachers who showed a smaller percentage of low scores (1) but did hardly increase high scores (3). This was also supported by the Tau-b de Kendall value (-.092, Standard error 0.71).

Coherence also shows greater performance scores, which resemble those obtained in the two previous criteria. The x^2 for Interaction was 26.412 (not significant) with an asymptotic significance of 0.091 which also shows a tendency towards medium and high values, although in this case it is more centralized than in the previous two criteria. In reference to the results, Psychology students scored lower than in the second and third criteria but this change was not significant. Nor was it in the rest of the groups whose scores decreased but not in the same percentage. The Tau-b de Kendall value -.092, Standard error 0.71 shows that no significant differences were observed in relation to the curves of the other groups.

Global performance was considered important because it provided information on the overall grades of all the participants in the study (García Laborda, Amengual Pizarro, & Litzler, 2013), and could also be contrasted with the data obtained by the Ministry of Education. Table 10 shows the data obtained by the Ministry in their pilot study done in 2012 (http://www.mecd.gob.es/dctm/inee/documentos-de-trabajo/informe-pau-ingles.pdf?documentId=0901e72b8170cdc9), while the results of this research can be observed in the following two tables (Tables XI and XII).

TABLE X. Results of participants' global performance obtained by the Ministry of Education (http://www.mecd.gob.es/inee/Documentos-de-trabajo.html).

	Criterion					Total
	Scope	**Grammatical correction**	**Fluency**	**Interaction**	**Coherence**	
Part 1	65.08	60.39	64.66	68.17	68.51	63.65
Part 2	61.06	55.53	62.81	64.49	65.66	60.80
Total	60.22	54.77	61.89	63.82	64.57	61.39
P-value 1st and 2nd	0.04183	0.01614	0.34711	0.05709	0.13836	0.15091

Received: 17-11-2014 Accepted: 13-03-2015

TABLE XI. Results of participants' global performance in this study. The number of participants appears in the first line. The mean and standard deviation are shown in the third and fourth lines respectively.

	Accuracy	Fluency	Interaction	Coherence
Valid	169	167	169	168
Lost	0	2	0	1
Mean	1,50	1,89	1,79	1,83
Standard deviation	,647	,712	,731	,709

TABLE XII. Frequencies for each criterion evaluated and their corresponding values within the four total possible scores (from 0 to 3).

Score	Accuracy		Fluency		Interaction		Coherence	
	frec.	%	frec.	%	frec.	%	frec.	%
0	2	1,2	1	,6	2	1,2	1	,6
1	92	54,4	49	29,0	60	35,5	56	33,1
2	63	37,3	84	49,7	78	46,2	82	48,5
3	12	7,1	33	19,5	29	17,2	29	17,2
Total Pass	75	44.3%	117	59.2%	107	63.4%	109	65.7%
X^2 P ¿Significant MECD-OPENPAU?	6.5448 0.0105 Yes (p<.05)		4.35 0.037 Yes (p<.05)		0.0165 0.8978 No (p<.01)		0.0854 0,7701 No (p<.01)	

The Ministry's results and those obtained in our research were compared and analysed statistically and show that significant differences are limited to accuracy and fluency, and also that there was no significance in the chi-square inter-group results for the other criteria, that is to say, interaction and coherence. High scores (3) were observed in three groups in our research but in this case, low grades (1) were even less than in the previous criterion. Especially significant is the case of primary pre-service teachers who showed a smaller percentage of low scores (1) but did hardly increase high scores (3). This was also supported by the Tau-b de Kendall value (-.092, Standard error 0.71).

Although a number of interpretations are possible for this lack of significant differences between the study run by the MECD and our study regarding interaction and coherence, one of the most significant could be the way of delivery. Face-to-face interviews usually lead to higher anxiety (Woodrow, 2006; Hewitt & Stephenson, 2011) and this could have a special effect on how teachers approach the test. While the MECD used a highly cognitive methodology with limited interaction between the test taker and the interviewer, the facilitating attitude of the interviewers in our case lead to significant changes, which can be observed in the table below (TABLE XIII) .

Our interactive approach had an effect on students' fluency and accuracy. In this sense, an interactive approach with an active interviewer would benefit the learners. Nevertheless, these results are not conclusive and thus further research would be necessary.

Received: 17-11-2014 Accepted: 13-03-2015

TABLE XIII. Comparison of tasks' methodology. The first column reflects the aspects compared in the two studies. The second column shows the approach carried out by the Ministry of Education. The third column shows the approach developed in this study.

Observations	MECD (cognitivist approach)	OPENPAU (Interactive approach)
Competitiveness	Active	More passive, atmosphere is usually more relaxed
Cooperation – Speaking tasks	Clarification, questions, intere st in delivering	Completing, clarifying, productive questions, interest in meaning
Co-construction discourse	Tends to be two co -constructed monologues, little interest in the "other"	Dialogue tends to be engaging even with candidates with diverse proficiency levels
Attitudes	Cooperation is limited, instead students engage in limited realistic strategies/discourse	Cooperation is fundamental and realistic. Real daily life is constructed through language, body and context
Individual Factors affecting		Extroversion, cooperation, will to support
Grouping Dynamics	Pairs, trios but tend to attach to1 -1 turn taking. Even competence grouping	Interact more freely, turn taking tends to vary, may have different competences
Role of tester/interviewer	"listener", visually passive*, not active participant, does not correct	Mediator; active; facilitator, very active, moderates to produce adequate forms
Scoring	According to objectives	According to productivity

Discussion

The first interesting observation is that in our study most students (55.6%) were in the A1-2 competence band despite their many years of English learning. In fact, in most cases their starting age of learning English was 12. These results may indicate major flaws either in teaching processes, in choice of contents, in teaching materials, in the choice of methodologies, in rate of students per class or in a combination of all the factors contemplated. Another issue that needs to be considered is the foreign language competence requirements for graduation in high school (mostly B1 except in Andalucia – A2+- and Catalunia –B2-). Although researchers like Gómez Rodríguez (2010) have insisted that textbooks do not lead to the development of communicative competence, the results in the study show that students perform better in interactional criteria than in accuracy. This may indicate that pair interview support leads to higher scores, which is in accordance with current studies of pair/group (Nakasuhara, 2013) in which students with lower competence are matched with students with a higher one. The fact that there are no significant differences among groups in any of the criteria suggests that the tasks are valid for these groups.

As mentioned in the previous section, it was especially interesting to observe that there were no significant differences when comparing this study with the study run by the MECD in interaction and coherence. Whether further research would be desirable to justify these minimal differences, it may well mean that some aspects may be better developed in the classroom, that students tend to focus on specific aspects of communication or that raters' perceptions towards these criteria may be similar. These high correlations evidence the robusntness of the results hereby obtained.

Conclusions

This paper addressed three main issues:1) First year university students' competence following the CEFR, 2) Differences in rating criteria across university degrees and 3) Comparison with the previous report by the MECD. The results indicate that there is a relation between the results of this research and the MCD's. One of the outstanding features of the research is that there is a slight difference between the competence level

achieved by the end of 10th and 12th grade and that most students do not achieve levels beyond the B1. Another feature is that if communication is a final goal in the LOMCE, more attention should be given to tests especially aimed at obtaining inferences on the degree of oral interaction (García Laborda, 2010a).

There is a great gap between the competence students bring into university and the competence required by most universities to graduate (B2). This will obviously make universities increase the number of English (or foreign language) courses if this requirement is to be fulfilled. Testing has proved to be a valuable way to change education (Washback effect, Muñoz & Alvarez, 2010) and thus new items and forms of assessment need to be implemented in the Spanish educational system. Maybe, teachers, administrators and testers also need to reshape their concept of language testing and move towards more interactive approaches. Pair/group delivered speaking tests do not only favor the performance, which may be based on the fact that maybe by using interactive interviews, some communicative function that weaker students cannot show in face-to-face interaction due to stress and anxiety can be triggered by the presence of an equal, the other candidate (Horwitz, 2000; McCarthy & O'Keeffe, 2004).

There is still a way to walk before the new High School Final Exam is implemented in 2017 but maybe this humble paper will attract the attention of the educational politicians towards further research, which should focus first, on the possibilities of grouping, second, on how maximizing resources would increase competence levels and third, whether social interaction may have an effect in improving performance on tests. It should also address the major weakness of this paper which is the limitation of the sample, and the need to approach the results using corpus or pragmatic methodologies to achieve a sound construct that can be effectively validated both internal and externally.

Acknowledgments

The authors would like to express their gratitude to the Ministry of Science and Innovation for providing the funds for this research through the OPENPAU project (Orientacion, Propuestas y Enseñanza para la Seccion de Inglés en la Prueba de Acceso a la Universidad; FFI2011-22442) thanks to the co-financing between the Ministry of Economy and Innovation (MICINN) and the ERDF funds from the European Union.

References

Amengual Pizarro, Marian (2009). "Does the English Test in the Spanish University Entrance Examination influence the teaching of English?" *English Studies* 90 (5): 582-598

Bernhardt, E. B., Rivera, R. J., & Kamil, M. L. 2004. "The practicality and efficiency of web-based placement testing for college-level language programs". *Foreign Language Annals* 37(3): 356-366.

Brooks, L. 2009. "Interacting in pairs in a test of oral proficiency: Co-constructing a better performance". *Language Testing* 26(3): 341-366.

Bueno, C & Luque, G. 2012. "Competencias en lengua extranjera exigiblesen la Prueba de Acceso a la Universidad: una propuesta para la evaluación de los aspectos orales". *Revista de Educación* 357: 81-104.

Chapelle, C. A. & Douglas, D. 2006. *Assessing language through computer technology*. Cambridge, UK: Cambridge University Press.

Dunbar, N. E., Brooks, C. F., & Kubicka-Miller, T. 2006. "Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills". *Innovative Higher Education* 31(2): 115-128.

Fernández Álvarez, M., & Sanz Sainz, I. 2005. *Metodología para el diseño de una prueba de inglés en selectividad. Estudios y criterios para una selectividad de calidad en el examen de inglés.* Valencia: Universidad Politécnica de Valencia.

Galaczi, E. D. 2008. "Peer-peer interaction in a speaking test: The case of the «first certificate in English examination". *Language Assessment Quarterly* 5(2): 89-119.

Gan, Z. 2008. "Examining negotiation in peer group oral assessment: What are the implications?". *Australian Review of Applied Linguistics* 31(1): 3.

Gan, Z. 2010. "Interaction in group oral assessment: A case study of higher- and lower-scoring students". *Language Testing* 27(4): 585-602.

Gan, Z., Davison, C., & Hamp-Lyons, L. (2009). "Topic negotiation in peer group oral assessment situations: A conversation analytic approach". *Applied Linguistics* 30(3): 315-334.

García Laborda, J. 2006. *Analizando críticamente la Selectividad. Todos los estudiantes tiene las mismas posibilidades. Tesol Spain Newsletter*, 30(3): 9-13.

García Laborda, J. G., & Royo, T. M. 2007. Diseño y validación de la plataforma PLEVALEX como respuesta a los retos de diseño de exámenes de lenguas para fines específicos. Ibérica: *Revista de la Asociación Europea de Lenguas para Fines Específicos* (AELFE), (14): 79-98.

García Laborda, J. 2010a. "¿Necesitan las universidades españolas una prueba de acceso informatizada? El caso de la definición del constructo y la previsión del efecto en la enseñanza para idiomas extranjeros". *Revista de orientación y Psicopadagogía 21*(1): 71-80

García Laborda, J. 2010b. "Contextual clues in semi-direct interviews for computer assisted language testing". *Procedia: Social and Behavioral Sciences* 2: 3591-3595.

García Laborda, J., Litzler, M.F., & Amengual Pizarro, M. 2013. "Can Spanish high school students speak English?" In Colpaert, J., Mimons, M., Aerts, A., & Oberhofer, M. (Eds.) *International Conference Language Testing in Europe: Time for a New Framework* (pp.124-128), Antwerp 27-29 May 2013.

Gómez Rodríguez, L. F. 2010. "English Textbooks for Teaching and Learning English as a Foreign Language: Do They Really Help to Develop Communicative Competence?". *Educación y Educadores* 13(3): 327-346.

Heilmann, S. G. 2012. "Can you hear me now? Assessing students' classroom communication preferences via a telephone conference activity". *Journal of Educators Online* 9(1): 15.

Herrera Soler, H. & García Laborda, J. (Eds.) 2005. *Estudios y criterios para una selectividad de calidad en el examen de Inglés*. Valencia: Universidad Politécnica de Valencia.

Hewitt, E., & Stephenson, J. 2011. "Foreign Language Anxiety and Oral Exam Performance: A Replication of Phillips's MLJ Study". *The Modern Language Journal* 96(2): 170-189.

Horwitz, E. K. 2000. "It Ain't Over 'til It's Over: On Foreign Language Anxiety, First Language Deficits, and the Confounding of Variables". *The Modern Language Journal* 84: 256–259.

McCarthy, M. &A. O'Keeffe. 2004. Research in the teaching of speaking. *Annual Review of Applied Linguistics* 24: 26-43.

Elder, C., Iwashita, N., & McNamara, T. 2002. Estimating the difficulty of oral proficiency tasks: what does the test-taker have to offer?. *Language Testing*, 19(4): 347-368.

Muñoz, A. P., & Alvarez, M. E. 2010. "Washback of an oral assessment system in the EFL classroom". *Language Testing* 27(1): 33-49.

Murray, J. R. 2001. *Steps and recommendations for more placement test creation. ERIC reports*. Retrieved from http://files.eric.ed.gov/fulltext/ED453656.pdf

Nakatsuhara, F. 2006. The impact of proficiency-level on conversational styles in paired speaking tests. *Research notes*, 25, 15-19.Retrieved on March 12, 2015 from http://www.cambridgeenglish.org/images/23144-research-notes-25.pdf

Nakatsuhara, F. 2013. *The co-construction of conversation in group oral tests. Language Testing and Evaluation Series*, Frankfurt am Main, Germany: Peter Lang.

Sawaki, Y., Stricker, L. J., & Oranje, A. H. 2009. "Factor structure of the TOEFL internet-based test". *Language Testing* 26(1): 5-30.

Shaw, S. D., & Weir, C. J. 2007. *Examining writing: Research and practice in assessing second language writing* (Vol. 26). Cambridge University Press.

Vincent-Durroux, L., Poussard, C., Lavaur, J., & Aparicio, X. 2011. "Using CALL in a formal learning context to develop oral language awareness in ESL: An assessment". *ReCALL* 23(2): 86-97.

Vitiene, N., & Miciuliene, R. 2008. "Application of criteria-referenced assessment and qualitative feedback to develop foreign language speaking skills in the context of E-Teaching/Learning". *Quality of Higher Education,* 5*:* 152-168.

Yang, Q., Miller, M., & Bai, K. 2011. Testing the improvement of English as a foreign language instruction among Chinese college students through computerized graphic visuals. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Language* 45(1):170-182.

Woodrow, L. 2006. "Anxiety and speaking English as a second language". RELC journal, 37(3), 308-328. http://rel.sagepub.com/content/37/3/308.short (accessed 11/4/2014).

Contact address: Jesús García Laborda. Universidad de Alcalá. Dpto. de Filología Moderna. Despacho Virginia Woolf. C/ Trinidad, 3. 28801 Alcalá de Henares, Madrid. E-mail: jesus.garcialaborda@uah.es

Received: 17-11-2014 Accepted: 13-03-2015

La *Revista de Educación* es una publicación científica del Ministerio de Educación, Cultura y Deporte español. Fundada en 1940, y manteniendo el título de *Revista de Educación* desde 1952, es un testigo privilegiado de la evolución de la educación en las últimas décadas, así como un reconocido medio de difusión de los avances en la investigación y la innovación en este campo, tanto desde una perspectiva nacional como internacional. La revista es editada por la Subdirección General de Documentación y Publicaciones, y actualmente está adscrita al Instituto Nacional de Evaluación Educativa de la Dirección General de Evaluación y Cooperación Territorial.

GOBIERNO DE ESPAÑA
MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE

NIPO línea: 030-15-016-X
NIPO ibd: 030-15-017-5
ISSN línea: 1988-592X 0034-8082
ISSN papel: 0034-8082

www.mecd.gob.es/revista-de-educacion