ONOMÁZEIN
Revista de lingüística, filología y traducción

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE LETRAS

Differences and Similarities between Oral and Written Competence in Spanish Pre-University Students: A Correlational Study

Jesús García Laborda
Universidad de Alcalá
España

María Belén Díez-Bedmar
Universidad de Jaén
España

Elena Martín Monje
Universidad Nacional a Distancia (UNED)
España

Miguel Fernández Álvarez
Universidad Politécnica de Madrid
España

ONOMÁZEIN 42 (diciembre de 2018): 172-196
DOI: 10.7764/onomazein.42.05
ISSN: 0718-5758

Jesús García Laborda: Modern Philology Department, Universidad de Alcalá, España. | Correo electrónico: jesus.garcialaborda@uah.es
María Belén Díez-Bedmar: Universidad de Jaén, España. | Correo electrónico: belendb@ujaen.es
Elena Martín Monje: Universidad Nacional a Distancia (UNED), España. | Correo electrónico: emartin@flog.uned.es
Miguel Fernández Álvarez: Universidad Politécnica de Madrid, España. | Correo electrónico: m.fernandezupm.es

Fecha de recepción: diciembre de 2016
Fecha de aceptación: septiembre de 2017

Abstract

The Foreign Language (FL) section of Spanish University Entrance Examination (EFL-PAU) has had few revisions over the last twenty years. The Spanish government has substituted the old EFL-PAU University Entrance Examination by a high stakes Baccalaureate Final Evaluation. However, further changes are expected in the coming years. Among the most important ones for the Foreign Language Section is a deep and necessary revision with the inclusion of new types of tasks. To try and inform the decisions made for the new Baccalaureate Final Evaluation, this article reports on the quantitative and qualitative analyses conducted thanks to a pilot oral test carried out with 772 recorded candidate performances. Three main goals were established for this research: 1) to find out the most important variables which characterize foreign language learning in secondary school in Spain; 2) to find out if there is any correlation between the oral competence of students at the end of their non-compulsory secondary education (as obtained from the pilot study) and the marks students obtain in the University Entrance Examination (which does not include the oral skill); and 3) to find out any aspects related to EFL teaching and learning which may potentially enhance oral performance. The results of this paper provide rich information on the students' foreign language learning context, the strong correlation between the written and oral competences and the need to pay attention to three variables which foster the development of oral foreign language in secondary school classrooms.

Keywords: University entrance examination; correlational study; foreign language.

1. Introduction

The teaching and learning of languages has not only academic implications in the classroom, but also a social impact that started being explored in the field of Applied Linguistics only a few decades ago. The role that foreign and second languages play inside the classroom is somewhat related to the use of the language that students make in the real world, which ultimately has an impact on pedagogical decisions and instructional practices.

Traditionally, language has been perceived as a set of separate skills that could be taught independently, depending on the context and needs of the students. A review of the history of second language teaching reveals different approaches focusing on discrete language skills being used for many years. Since the introduction of the concept of "communicative competence" (Hymes, 1972) language is considered from a more global and holistic perspective, as the main focus should be communication, both inside and outside the classroom. Considering this holistic approach to language, the present study focuses on the social and linguistic impact of foreign language teaching in Spain. An important focus on writing, reading and grammar renders less interest in oral skills. It is the aim of the study to investigate the differences and similarities between oral and written competence in pre-university students with the ultimate goal of determining what changes need to be made in order to improve the educational system.

The research presented in this study is motivated by the interest of the Spanish Ministry of Education to improve its current language assessments in order to introduce the adequate improvement measures in language testing and learning and replace the current University Entrance Examination (run by universities) by the Baccalaureate Final Evaluation (whose prospective form will be organized by the Ministry and delivered and rated by high school teachers). This study also intends to suggest an approach for other countries where similar changes may be currently under consideration. After a review of the literature, focused on social influences of language learning and sociocultural factors in language performance, we present the context of the study and describe the research. The discussion and conclusions at the end offer some guidelines for the improvement of both teaching and testing of languages in the high school context, if the oral skill is to be fostered

2. Social Influences of Language Learning

The study of foreign languages has increased its importance over time mostly due to the world professional and social mobility. Consequently, languages and especially English as a Lingua Franca have become of outstanding importance in general education planning and curriculum implementation (Smolin & Clayton, 2009; Holme, Richards, Jimerson & Cohen, 2010; Musoleno

& White, 2010; Kearns, 2011; Scott, 2012; Van Rijn, Beguin & Verstralen, 2012). What is done and taught in the classroom is influenced by the language and its use in real life. Spolsky (1989: 13-14) reminds us that "language learning is individual but occurs in society, and while the social factors are not necessarily direct in their influence, they have strong and traceable indirect effects," which has strong implications on the attitudes displayed by teachers and students.

According to Barkhuizen (2004: 553), "language learning takes place in a social context which consists of a number of influential social factors." Some of the factors identified include the setting where learning and assessment occurs (in most cases, the classroom) and the participants. Due to their active role in the learning process, learners can be seen as some of those primary participants in both processes. However, the definition of participant can be extended to include other stakeholders, such as teachers, administrators, parents, language testers, educational policy makers, and by extension politicians and society in general, as everyone plays a direct or indirect role in the process of language teaching, learning and assessing (Rea-Dickins, 1997).

Testing, as mentioned above, is one of those areas that has become a cornerstone in language education (Tsang, Katz, & Stack, 2008; Heilig, 2011; Gutiérrez Eugenio & Saville, 2017), considering both its benefits and drawbacks (Bracey, 2009; Fitchett & Heafner, 2010; Howie, 2012; Koretz, 2016). Language testing is currently used to take decisions in student mobility, student's acceptance in bilingual higher education programs and also to assess language programs (Mahon, 2006; Wright & Choi, 2006; Solano-Flores & Li, 2013; Hill & McNamara, 2015). However, although large scale national assessments tend to be more and more common, they also tend to ignore the relationship between the candidates' personal factors and their grades in high stakes exams (Qian, 2009; Gorges & Kandler, 2012; Bai, Hudson, Millwater, & Tones, 2013) mostly because a great disparity may evidence that practice is not adequate or that socioeconomic problems may be creating significant differences within a country (Fitchett & Heafner, 2010; Kim, 2011; Solano-Flores & Li, 2013).

The role of second and foreign language teaching is furthermore affected by factors such as accountability, which nowadays influences many of the decisions that teachers and administrators have to make (Kang & Chang, 2014; Ercikan, Roth, & Asil, 2015). With the goal of comparing and assuming international common policies (Morgan & Taylor Poppe, 2012; Wilby, 2012), accountability practices can be translated into higher demands on teachers and more pressure on students to pass standardized tests. Improvements at a national level also require the analysis of not only linguistic but other socioeconomic and educational aspects (Wall, 2005). Educational authorities need to analyze these aspects to introduce significant changes in their policies in order for the countries to evolve in national curriculum development and syllabus design, and for local schools or teachers to address their specific students' needs.

3. Sociocultural Factors in Language Performance

In order to define language performance, it is important to have a clear understanding of what the actual "performance" implies and what factors have an influence on it. Spolsky's (1989) conditions on second language learning are based on factors such as nature of L2 knowledge, language use, testing and measurement, individual learner factors, linguistic social context, natural learning, and formal learning.

However, there seems to be a need for instruction that ensures that learners (1) develop both a rich repertoire of formulaic expressions and a rule-based competence, (2) focus mainly on meaning, and (3) also focus on form (Ellis, 1999), making language competence and performance a complex task that requires much more than isolated skills, grammar knowledge or just comprehensible input (Krashen, 1981, 1985). Learners, in fact, also need high quality input (not just comprehensible), ample opportunities to practice, high quality feedback and individualized content (Zhao & Lai, 2007).

While research focuses on best practices in the teaching of second languages, as noted above, the reality in the classroom is much different. Although the teacher's use of the target language seems to be an old issue in language learning, most teachers would agree that the ideal situation in a foreign language classroom is that in which the student is surrounded by the use of the target language (TL) as much as possible. However, despite the beneficial use of the mother tongue in the foreign language classroom (Liu & Zeng, 2015), it is difficult to see to what extent students in classes which may be taught in the speaker's first language (L1) or where the interaction is mostly in their L1 can lead to student's language development (Moussu & Llurda, 2008).

High stakes testing policies have also been seriously criticized because teachers mostly teach to the test, there is an abuse of test strategies versus real knowledge (Xu & Wu, 2012), there are differences in socioeconomic aspects (Russell & Kavanaugh, 2011), because external and impact validation need to be considered in the test framework (Weir, 2013), and because of the need of equal and fair achievement opportunities (Wright & Li, 2008), especially among social groups (Escamilla, Chavez, & Vigil, 2005).

However, on the other hand, language tests have a direct influence on how classes are taught, syllabi and curricula designed, teachers trained and the students' evolution (Changing Language Teaching through Language Testing), an effect called washback (Alderson & Wall, 1993; Fan, 2017; Kwon, Lee, & Shin, 2017; Zou & Xu, 2017, among others). Many studies have tended to focus on the pros and cons of washback in Spain—especially the negative issues—(Amengual Pizarro, 2010; Fernández Álvarez, 2012; Bueno-Alastuey, García Laborda, Alcón, & Luque Agullo, 2014) because the very own existence of washback is very difficult to deny in Spain and many international contexts. Assessment-based educational systems are based on the potential ben-

eficial effects of test in learning while most negative issues are related with accountability in education. Many practitioners criticize them because results tend to benefit some and have negative effects on the rest; for instance, language tests can benefit students of certain classes who may have more opportunity to learn them in some special schools or language centers, or those who travel or their parents speak to them in a L2. However, good practices are usually based on the design, development, and delivery of valid tests. Likewise, the design of the tests requires of field studies that consider the linguistic initial situation, the target expected proficiency, and the factors that can account for differences among the prospective students.

4. The Study Context

Europe has been using joint educational and linguistic policies for the last twenty years. However, believing that the educational situations in all the countries of the European Union are similar is probably a deceptive illusion. It is certainly true that there is an increasing interest in standardizing education and increasing the number of languages that its citizens speak but that is extremely difficult because there are deep differences among and within the different OECD countries (see for example European Commission, 2012). There are clear differences in foreign languages competence but also in sciences, mathematics, and first language. Spanish students usually are among the worst countries in international assessments like PISA and others (OECD, 2012).

The teachers' common response to this issue is that other countries prepare specifically for these international assessments. Although this assumption may "feel" true, there is little empirical evidence to support it. In relation to foreign languages, Spain is among the European countries with the lowest competence in English. The situation of French is relatively better than English (European Commission, 2012), the world of business and culture is currently driven in the English language. Thus a strong competence in English is not only desirable for educational purposes but also to succeed in studying and working abroad, a current need in a country going through a deep economic crisis with an immense percentage of unemployment.

Given all of these premises and having in mind the current tendency towards accountability in education, the Spanish Ministry of Education, Culture and Sports is currently planning a strong educational reform in which National Assessments are the cornerstone. One of the aspects that the Ministry has emphasized is the development of adequate foreign languages tests after 6th, 8th, 10th, and 12th grades. Besides, the former University Entrance Examination will eventually be replaced by the Baccalaureate Final Evaluation at the end of 12th grade. The major change in the Foreign Languages section will be the inclusion of speaking tasks that have never been included in any high stakes test in Spain before. Before proceeding to a thorough revision and implementation of a new test, the educational authorities in Spain considered it necessary to pilot a sound study of the current situation of speaking

in EFL across the country. This pilot study was done by the Spanish Institute of Educational Evaluation (Instituto Nacional de Evaluación Educativa, INEE), and then the data offered to research groups for their use. The data were obtained from speaking tests and the recordings of scoring by trained raters in seven educational communities from Spain (Asturias, Aragón, Islas Baleares, Castilla-La Mancha, Comunidad Valenciana, La Rioja, and Madrid) totalling 772 11th and 12th grade (last year of high school) students.

To provide the authorities with results which would inform any future decisions made concerning the effective inclusion of an oral exam to the current EFL-PAU (which only focuses on written skills nowadays), the present study used the data resulting from that pilot study to answer the following research questions:

1) Which are the most important variables which determine the learning of a Foreign Language in Spain nowadays?
2) Is there a correlation between the oral competence of students at the end of their non-compulsory secondary education, as obtained from the pilot study, and their written competence as assessed in the University Entrance Examination nowadays?
3) What aspects related to EFL teaching and learning may potentially enhance oral performance?

4.1. Participants

A total of 772 students in their 1st and 2nd year of Bachillerato (equivalent to 11th and 12th grades) took part in the study, of which 54% were male and 46% female. In order to achieve a representative sample, the schools were randomly chosen in the 7 participating autonomous regions following the sample design shown in table 1 below.

Each of the regions involved in the project interviewed about 200 students of three types of cities (less than 10,000 inhabitants, cities with 10,000-100,000 inhabitants, and cities with more than 100,000 inhabitants) including, whenever possible, both private and public schools.

4.2. Instruments

Pilot oral test

The tasks included in the speaking test were constructed taking the Common European Framework of Reference (CEFR) B1 level as a benchmark, since it is the predominant level for

TABLE 1
Sample design for random choice of schools in participating autonomous regions

	TYPE OF SCHOOL		
Size of town	State	Private	Total
>100,000 inhabitants	2	2	4
<100,000 inhabitants	2 (1 private school wherever available)		2
>10,000			
< 10,000 inhabitants	2 (1 private school wherever available)		2
Total			8

the EFL-PAU (Díez-Bedmar, 2011). Students at this level can understand the main points of a text on familiar matters regularly encountered in work, school, leisure, etc., and can produce simple, connected texts on topics which are familiar or of personal interest. They can also describe experiences and events and briefly give reasons and explanations for opinions and plans (Council of Europe, 2001: 24).

The total duration of the pilot oral test was 8-10 minutes and consisted of two differentiated parts. The first part lasted 2-3 minutes and focused on spoken interaction: a conversation between the examiner and the student which dealt with personal information (name, family, school, free time, holidays, past experiences, future plans). The second part was longer (6-7 minutes) and involved both spoken production and spoken interaction: the student was given some photographs which he/she had to describe. After that, they were asked some questions related to the photographs which would encourage discussion and the students' expression of agreement or disagreement on topics arising from the visual aids.

Three raters were present at the time of the interview: one delivered the questions and rated the student's performance while the other two just acted as raters. Given the extensive number of students and the difficulty associated to obtaining images from under agers (18 years old) in Spain, the Ministry administrators decided not to record the sessions. All the interviews took place between March and April 2012.

Questionnaire

The purpose of the questionnaire was to collect data related to the students' background and profile, their EFL learning and the contexts in which they use this foreign language. With that aim, the questionnaire was divided into two differentiated parts: 1) data related to the stu-

dent and his/her personal background: age, sex, school record, socioeconomic background of their families, country of origin, autonomous region where they live, mother tongue; and 2) data related to EFL learning and actual use: size of classes, duration of EFL class, number of EFL classes, use of EFL in class, pedagogic materials and resources, EFL use outside the educational context (extra-curricular activities, school trips to English-speaking countries).

For the purposes of this paper only the data that have shown a significant correlation to the students' oral competence (which are mainly related to the second part of the questionnaire) will be analysed in detail.

4.3. Procedure

The speaking test was evaluated following a rubric which focused on five criteria: 1) range, 2) grammatical correctness, 3) fluency, 4) interaction, and 5) coherence, with a score from 1 to 10, 1 being the lowest. Three different examiners/raters were used for this study, with the purpose of obtaining a higher reliability in the evaluation, being their final mark the mean obtained from their respective evaluations.

Furthermore, additional data relevant to this study were collected: students' final grade in the school subject "Foreign Language: English", students' overall final grade for that academic year, and the results of their EFL-PAU. The latter were used to establish a possible correlation between the pilot oral test and the EFL-PAU, which only evaluates written competence, so as to find out whether it was worthwhile to implement an oral test in the EFL-PAU.

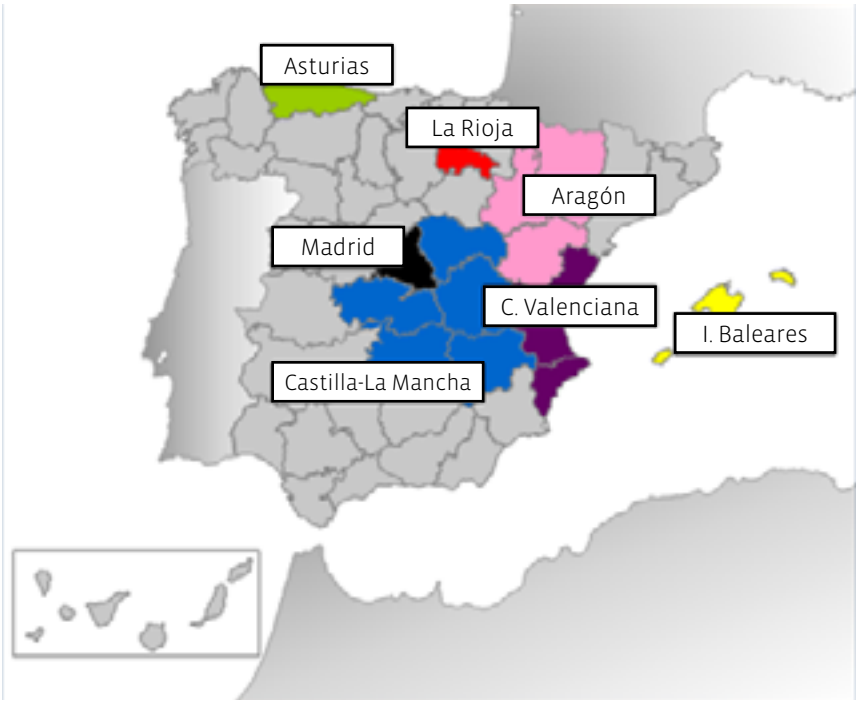
5. Data Analysis and Discussion
5.1. Students and Their Context: An Overview of the Variables in EFL Learning

The data corresponding to the students' variables regarding themselves, their families, and personal context were the following. The students' age ranged between 16 and 21, most of the students being 17 (47.9%), and most of them (54.7%) being women. During their studies most of these students did not need to retake any school year either at primary education (98.2%), compulsory secondary education (93.3%), or optional secondary education (93.3%).

The participants' origin (i.e., their *Comunidad Autónoma* or autonomous region) was also analysed in this study. The seven participating autonomous regions are shown in figure 1. Out of these, the ones from which most students are were found to be Aragón (14.5%), Castilla-La Mancha (13.7%), Baleares (13.4%), Madrid (13.1%), or Valencia (12.8%). Regarding emigration, 6.4% of the participants in the study are immigrant students. The language that is spoken at

the participants' home is Spanish in most of the cases (85.5%), then followed by Catalan (7.8%), Valencian (3.4%), or any other language (2.3%).

FIGURE 1
Autonomous regions participating in the study



5.2. Contextualization of Students' EFL Learning

5.2.1. Research Question 1

Variables related to the English classroom

Three variables were analysed regarding the English classroom, namely the ratio of students per class, the duration of the classes, and the number of classes per week. First, most students take their classes in a group of 21-25 students (33.2%), 26-30 students (30.3%), or 16-20 students (26.7%). Only 4.5% of the students benefit from classes with 10-15 students and 4.2% attend overcrowded classes with 31-35 students. Second, half of the participants in this questionnaire attended classes which lasted 50 minutes, and 31.7% of the students took classes which took 55 minutes. 60-minute classes were a reality for 12.2% of the students, and classes taking 40 or 45 minutes were only attended by 1% or 4% of the sample, respectively. As far as the number of classes per week is concerned, most of the students (76.2%) took three classes per week, 15.8% took four classes per week, and only 4.6% or 2.3% took five or six classes per week, respectively.

Variables related to the language and methodology used

Four variables were considered to describe the language and the methodology used in the English classroom: 1) the use of English in the class on the part of the teacher, 2) on the part of the student, 3) the presentations made in English, and 4) the use of books, music, or videos in English. First, the teachers' use of English in the class was the most frequent situation, with 42.1% of the teachers using English in all the classes and 34.6% doing so in most of the classes. However, 5.4% of the teachers were reported to never or almost never use the foreign language in their classes, which is highly surprising in foreign language teaching. Students, on the contrary, affirm that this is not the case for the students' participation in class. In fact, most of the students (44.2%) report to use English in some classes, 23.6% of the students do so in most of the classes, and, in a similar proportion, 23.3% of the students never or almost never use English in the classroom. Only 7.6% of the students claim that they use English in the English classroom on all occasions. Third, half of the students (49.5%) report to make presentations in the foreign language in some classes. However, 32.8% of the students also report to never or almost never make presentations in the foreign language. Only 4.2% of the students do so in all the classes. In the fourth place, the use of books, music, and videos in English is found to be 47.4% in some classes and 21.6% in most of the classes. Unfortunately, the percentage of students who have access to those resources in the English classroom in all the classes and the ones which do not use those resources in their classes is quite close, with 12.1% in the former and 17.6% in the latter. Finally, 33.5% of the students are motivated to use English in most of the classes, 27.8% are motivated to do so in all the classes, 25.1% in some of the classes, and 12.3% never or almost never.

The resources used in the language classroom were also considered in this study. Thus, the use of audiovisuals, printed materials, online materials, computer materials, language labs, textbooks, and reading books will be analysed in what follows. First, audiovisual materials were found to be used on some occasions along the academic year (34.9%), once per month (17.6%), on some occasions in a month (18.7%), and in almost all the classes (5.2%). However, they were also reported not to be used or little used (22.4%). In the case of printed materials, almost half of the students (43.6%) reported that they were never used or almost never used in the classroom, a tendency which was followed in the rest of the students since these materials were found to be employed only on some occasions along the academic year (34.3%), once a month (12%), on some occasions per month (7.6%), or in almost all the classes (1.3%). Therefore, a tendency not to use printed material can be seen in the data.

Similarly, online resources were not very much used either. In fact, half of the students reported that they were never or almost never used in the classroom (44.1%), they were used on some occasions in the academic year (25.2%), once a month (12.6%), on some occasions in a month (12.1%), and only 4.7% of the students used it in almost all the classes. A similar scenar-

io was found in the use of computer programmes, since 54.1% of the students never or almost never use them, 21% use them on some occasions in the academic year, 10.5% use them once a month, 7.3% use them on some occasions in a month, and only 4.9% of the students use them in all their classes. In the case of computer labs, their use was even more limited, as can be seen in the data. 81.7% of the students have never or almost never used them, and it is only 3.1% of the students that use them once a month, and 1.1% of the students use them in all the classes. The limited use of the materials analysed so far indicates that another resource in the English classroom, i.e., the textbook, continues on being the material more widely used. This is so to such an extent that 68.8% of the students use it in all the classes, and only 3.4% of the students never or almost never use it. Finally, in the case of the reading texts in the classroom, the situation found reveals that they are never used or almost never used in 33.1% of the classes, or they are used occasionally in the academic year (32.7%). On the contrary, 7.2% of the students use them in every class.

Time devoted to homework

Regarding the time devoted to homework, most of the students report that they devote little time to do so (60.7%), and 28.3% claim that they devote some time to do so. 2.5% of the participants in the study are the ones who report that they devote much time, as opposed to 7.2% of the students who confess they do not save time for homework at all.

Use of EFL outside the classroom

Other aspects which were of interest for the study were the ones related to the use of English outside the classroom. First, most of the students were found to use English in chats, emails, or letters on few occasions (67.3%). The same scenario was found in the use of English with a relative or friend (74%), with a neighbour (85.2%), with tourists (76.5%), or on the Internet (67.9%). The main difference found in their use of the foreign language in those contexts is that it is when they are online that participants reported to use English on several occasions per week in the highest percentage (10.1%).

Contact with EFL through media

Data were obtained regarding the students' use of songs, films without subtitles, TV shows/programmes without subtitles, TV shows/programmes with subtitles, computer programmes, books, comics, and web pages. The analysis of these data revealed that the highest percentage of use (78.7%) is that of songs, which are used on some occasions per week by the par-

ticipants. In order of importance, songs were followed by TV shows/programmes without subtitles (50.4%), comics (50%), TV shows/programmes with subtitles (42.1%), and computer programmes (30.4%). Finally, films without subtitles (33.2%), books (45.7%), and web pages (24.8%) are only used on some occasions along an academic year.

5.2.2. Research Question 2

Adding an oral part to an exam such as the EFL-PAU entails considerable effort on the part of the stakeholders and everyone involved in the process, including teachers and students. Two analyses were conducted to check if there was any correlation between the mark in the EFL-PAU and the marks (mean obtained from the three raters' marks) for each of the five criteria (*Range*, *Grammatical Correctness*, *Fluency*, *Interaction*, and *Coherence*) considered when evaluating the two oral tasks (providing personal background and describing pictures) in the pilot study of the oral test in the EFL-PAU.

In the following analysis the criteria applied for the first task of the pilot oral test have been labelled *Range1*, *GrammCorrect1*, *Fluency1*, *Interaction1*, and *Coherence1*, and the ones applied for the second task are *Range2*, *GrammCorrect2*, *Fluency2*, *Interaction2*, and *Coherence2*.

Since the data obtained from the EFL-PAU results proved non-normally distributed (see table 3 below), two Spearman's Rank Order correlations were run (one per task).

TABLE 3
Descriptive statistics for the EFL-PAU scores

	N	MEAN	STD. DEVIATION	SKEWNESS		KURTOSIS	
	STATISTIC	STATISTIC	STATISTIC	STATISTIC	STD. ERROR	STATISTIC	STD. ERROR
EFL-PAU score	772	6.3899	1.95101	-.254	.088	-.492	.176
Valid N (listwise)	772						

As a result of the Spearman's Rank Order correlation run with the data obtained from the first oral task (i.e., providing personal information), a statistically significant strong relationship (see Salkind, 2000; Pallant, 2005) was found between the mark obtained in the English Exam in the UAE and the mean of the mark given to each of the five criteria considered for the evaluation of the oral test, namely *Range1* (r= .628, $p\leq$ .001), *GrammCorrect1* (r= .628, $p\leq$ .001),

Fluency1 (*r*= .616, *p*≤ .001), *Interaction1* (*r*= .623, *p*≤ .001), and *Coherence1* (*r*= .629, *p*≤ .001). In the case of the analysis of the relationship between the five criteria, the correlations also proved to be very strong from the statistical point of view, as can be seen in the results obtained when comparing *Range1P* and the other four criteria, namely *GrammCorrect1* (*r*= .976, *p*≤ .001), *Fluency1* (*r*= .975, *p*≤ .001), *Interaction1* (*r*= .973, *p*≤ .001), and *Coherence1* (*r*= .973, *p*≤ .001). The full correlation coefficients are found in table 4 below:

TABLE 4

Correlations between EFL-PAU and first task in the pilot oral test

			EFL-PAU SCORE	RANGE 1	GRAMM-CORRECT1	FLUEN-CY1	INTER-ACTION1	COHE-RENCE1
Spearman's rho	EFL-PAU score	Correlation Coefficient	1.000	.634**	.628**	.616**	.623**	.629**
		Sig. (2-tailed)	.	.000	.000	.000	.000	.000
		N	772	772	772	772	772	772
	Range1	Correlation Coefficient	.634**	1.000	.976**	.975**	.973**	.973**
		Sig. (2-tailed)	.000	.	.000	.000	.000	.000
		N	772	1194	1194	1194	1194	1194
	Gramm Correct1	Correlation Coefficient	.628**	.976**	1.000	.967**	.963**	.968**
		Sig. (2-tailed)	.000	.000	.	.000	.000	.000
		N	772	1194	1194	1194	1194	1194
	Fluency1	Correlation Coefficient	.616**	.975**	.967**	1.000	.981**	.977**
		Sig. (2-tailed)	.000	.000	.000	.	.000	.000
		N	772	1194	1194	1194	1194	1194
	Inter-action1	Correlation Coefficient	.623**	.973**	.963**	.981**	1.000	.984**
		Sig. (2-tailed)	.000	.000	.000	.000	.	.000
		N	772	1194	1194	1194	1194	1194
	Cohe-rence1	Correlation Coefficient	.629**	.973**	.968**	.977**	.984**	1.000

		Sig. (2-tailed)	.000	.000	.000	.000	.000	.
		N	772	1194	1194	1194	1194	1194

**. Correlation is significant at the 0.01 level (2-tailed).

As can be seen in table 5, the results obtained with the data from the second oral task (i.e., describing pictures) reveal a similar scenario, regarding the strong relationship between the mark obtained in the EFL-PAU and the mean of the mark granted to each of the five criteria used for the evaluation of the oral task, as well as the very strong relationship found between the five criteria themselves (which were the same as the ones used for the first oral task). In fact, the strong correlation between the mark obtained in the EFL-PAU and the mean of the mark given to each of the five criteria considered showed that *Range2* (r= .641, p≤ .001), *GrammCorrect2* (r= .651, p≤ .001), *Fluency2* (r= .634, p≤ .001), *Interaction2* (r= .633, p≤ .001), and *Coherence2* (r= .641, p≤ .001). In the case of the correlation between the five criteria used to evaluate the oral task, the results obtained are the following when correlating *Range2* with *GrammCorrect2* (r= .977, p≤ .001), *Fluency2* (r= .977, p≤ .001), *Interaction2* (r= .973, p≤ .001), and *Coherence2* (r= .974, p≤ .001).

TABLE 5
Correlations between EFL-PAU and second task in the pilot oral test

			EFL-PAU SCORE	RANGE 2	GRAMM-CORRECT2	FLUEN-CY2	INTER-ACTION2	COHE-RENCE2
Spearman's rho	EFL-PAU score	Correlation Coefficient	1.000	.641**	.651**	.634**	.633**	.641**
		Sig. (2-tailed)	.	.000	.000	.000	.000	.000
		N	772	772	772	772	772	772
	Range2	Correlation Coefficient	.641**	1.000	.977**	.977**	.973**	.974**
		Sig. (2-tailed)	.000	.	.000	.000	.000	.000
		N	772	1194	1194	1194	1194	1194
	Gramm-Correct2	Correlation Coefficient	.651**	.977**	1.000	.968**	.964**	.970**
		Sig. (2-tailed)	.000	.000	.	.000	.000	.000
		N	772	1194	1194	1194	1194	1194
	Fluency2	Correlation Coefficient	.634**	.977**	.968**	1.000	.984**	.980**

		Sig. (2-tailed)	.000	.000	.000	.	.000	.000
		N	772	1194	1194	1194	1194	1194
	Interaction2	Correlation Coefficient	.633**	.973**	.964**	.984**	1.000	.986**
		Sig. (2-tailed)	.000	.000	.000	.000	.	.000
		N	772	1194	1194	1194	1194	1194
	Coherence2	Correlation Coefficient	.641**	.974**	.970**	.980**	.986**	1.000
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.
		N	772	1194	1194	1194	1194	1194

**. Correlation is significant at the 0.01 level (2-tailed).

As a result, it can be claimed that there are strong correlations between the mark obtained in the English Exam Section in the UAE (composed of written tasks) and the marks granted to each of the five criteria considered for the evaluation of the oral task (with similar results in the two tasks performed by the student). Similarly, very strong correlations are found between the marks given to each of the five criteria considered for the evaluation of the oral task (again with similar results in the two tasks).

5.2.3. Research Question 3

Finally, to answer Research Question 3, a more in-depth analysis has been carried out in order to establish whether some of the following variables analysed in the questionnaire may have an impact on the students' oral competence: a) frequency of teacher's input in English (*TeacherSpeaksEFL*); b) frequency of students' production and interaction in English (*StudentsSpeakEFL*); c) frequency of oral tasks done in class (*SpeakingTasks*); d) use of English multimedia resources (*EFLBooksMusicVideos*); and e) teacher's encouragement for students to speak in English (*MotivationEFLSpeaking*).

Table 6 below shows the correlations between these variables and the scores in the first task of the pilot oral test.

When all these variables were analysed by means of Pearson's correlation, three variables were found to correlate positively with all the other variables. These are *StudentsSpeakEFL*, *SpeakingTasks* and *MotivationEFLSpeaking*.

TABLE 6

Correlations between EFL learning variables from questionnaire and first task in pilot oral test

			RANGE1	GRAMM-CORRECT1	FLUENCY1	INTER-ACTION1	COHE-RENCE1
Spearman's rho	Teacher-SpeaksEFL	Correlation Coefficient	.041	.047	.053	.051	.056
		Sig. (2-tailed)	.158	.105	.071	.078	.054
		N	1179	1179	1179	1179	1179
	Students-SpeakEFL	Correlation Coefficient	.073*	.070*	.079**	.086**	.083**
		Sig. (2-tailed)	.012	.016	.007	.003	.005
		N	1179	1179	1179	1179	1179
	Speaking-Tasks	Correlation Coefficient	.084**	.082**	.094**	.102**	.103**
		Sig. (2-tailed)	.004	.005	.001	.000	.000
		N	1179	1179	1179	1179	1179
	EFLBooks-MusicVideos	Correlation Coefficient	.048	.041	.053	.056	.056
		Sig. (2-tailed)	.102	.160	.071	.053	.056
		N	1179	1179	1179	1179	1179
	Motivation-EFLSpeaking	Correlation Coefficient	.097**	.087**	.110**	.112**	.109**
		Sig. (2-tailed)	.001	.003	.000	.000	.000
		N	1179	1179	1179	1179	1179

**. Correlation is significant at the 0.01 level (2-tailed).

The data obtained are the following ones:

a) The variable *StudentsSpeakEFL* correlated with all the other variables, since *Range1* (r= .073, $p\leq$ .05), *GrammCorrect1* (r= .079, $p\leq$ .05), *Fluency1* (r= 0.79, $p\leq$ .05). The correlation is strongly significant with *Interaction1* (r=.86, $p\leq$ .005) and *Coherence1* (r= .083, $p\leq$ .005), as can be seen in the p values.

b) The variable *SpeakingTasks* also correlated with all the other variables, as can be seen in *Range1* (r= .084, $p\leq$ .005), *GrammCorrect1* (r= .082, $p\leq$ .005), *Fluency1* (r= .094, $p\leq$ .005). As it was the case with *StudentsSpeakEFL*, the correlation is strongly significant with *Interaction1* (r= .102, $p\leq$ .001) and *Coherence1* (r= .103, $p\leq$ .001).

c) The variable *MotivationEFLSpeaking* correlated with *GrammCorrect1* (r= .087, $p\leq$ .005), and in a more significant way with *Range1* (r= .097, $p\leq$ .001), *Fluency1* (r= .110, $p\leq$ .001), *Interaction1* (r= .112, $p\leq$ .001), and *Coherence1* (r= .109, $p\leq$ .001).

Other two variables, *TeacherSpeaksEFL* and *EFLBooksMusicVideos* only correlated with some other variables in a marginally statistically significant way. In the case of *TeacherSpeaksEFL*, the correlation was found with *Coherence1* (r= .056, $p\leq$ .054) and in the case of *EFLBooksMusicVideos*, that was the case with *Interaction1* (r= .056, $p\leq$ .053) and *Coherence1* (r= .056, $p\leq$ .056).

Table 7 below shows the correlations between these variables and the scores in the second task of the pilot oral test, the oral description of a picture followed by one or two questions about the visual prompt.

TABLE 7
Correlations between EFL learning variables from questionnaire and second task in pilot oral test

			RANGE2	GRAMM-CORRECT2	FLUENCY2	INTER-ACTION2	COHE-RENCE2
Spearman's rho	Teacher-SpeaksEF	Correlation Coefficient	.049	.060*	.069*	.074*	.077**
		Sig. (2-tailed)	.092	.041	.017	.011	.008
		N	1179	1179	1179	1179	1179
	Students-SpeakEFL	Correlation Coefficient	.078**	.075*	.090**	.102**	.094**
		Sig. (2-tailed)	.007	.010	.002	.000	.001
		N	1179	1179	1179	1179	1179
	Speaking-Tasks	Correlation Coefficient	.087**	.078**	.099**	.108**	.108**
		Sig. (2-tailed)	.003	.007	.001	.000	.000
		N	1179	1179	1179	1179	1179
	EFLBooks-MusicVideos	Correlation Coefficient	.051	.046	.065*	.066*	.059*

		Sig. (2-tailed)	.077	.114	.026	.024	.041
		N	1179	1179	1179	1179	1179
	Motivation-EFLSpeaking	Correlation Coefficient	.100**	.089**	.112**	.120**	.117**
		Sig. (2-tailed)	.001	.002	.000	.000	.000
		N	1179	1179	1179	1179	1179

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

When the data obtained from the oral test and question 19 in the questionnaire were analysed by means of Pearson's correlation, three variables were found to correlate positively with all the other variables. These are *StudentsSpeakEFL*, *SpeakingTasks* and *MotivationEFL-Speaking*. The data obtained are the following ones:

a) The variable *StudentsSpeakEFL* correlated with all the other variables, since *Range2* (r= .078, $p\leq$ .05), *GrammCorrect2* (r= .075, $p\leq$ .05). The correlation is strongly significant with *Fluency2* (r= 0.090, $p\leq$ .05), *Interaction2* (r=.102, $p\leq$ .005), and *Coherence2* (r= .094, $p\leq$ .005), as can be seen in the p values.

b) The variable *SpeakingTasks* also correlated with all the other variables, as can be seen in *Range2* (r= .087, $p\leq$ .005), *GrammCorrect2* (r= .078, $p\leq$ .005), and *Fluency2* (r= .099, $p\leq$ .005). As it was the case with *StudentsSpeakEFL*, the correlation is strongly significant with *Interaction2* (r= .108 $p\leq$ .005) and *Coherence2* (r= .108 $p\leq$ .005).

c) The variable *MotivationEFLSpeaking* correlated strongly with *Range2* (r= .100, $p\leq$ .001), *GrammCorrect2* (r= .089, $p\leq$ .005), *Fluency2* (r= .112, $p\leq$ .001), *Interaction2* (r= .120, $p\leq$ .001) and *Coherence2* (r= .117, $p\leq$ .001).

As with the first task, the other two variables, *TeacherSpeaksEFL* and *EFLBooksMusicVideos*, correlated with the other variables in a statistically less significant way. In the case of *TeacherSpeaksEFL*, the correlation was found with *GrammCorrect2* (r= .060, $p\leq$ .05), *Fluency2* (r= .069, $p\leq$ .05), *Interaction2* (r=.074, $p\leq$ .005), and *Coherence2* (r= .077, $p\leq$ .005). In the case of *EFLBooksMusicVideos*, it correlated with *Fluency2* (r= .065, $p\leq$ .05), *Interaction2* (r=.066, $p\leq$ .005), and *Coherence2* (r= .059, $p\leq$ .005).

As a summary, it is possible to see that three variables, namely *StudentsSpeakEFL*, *SpeakingTasks*, and *MotivationEFLSpeaking*, correlate with all the other variables. Similarly, *Interaction1* and *Coherence1* are two variables which correlate with all the variables in

a significant way (even though some are marginally significant), but with *Interaction1* and *TeacherSpeaksEFL*.

6. Conclusion and Limitations of the Study

The present study correlated the grades of the writing Spanish University Entrance Examination and 772 speaking interviews of Bachillerato (non-compulsory education) students in order to observe whether corresponding results would be obtained. It was ascertained that there is a strong correlation between oral and written competence and three factors were highlighted as the most influential ones in the students' performance: 1) the students' participation in English in their classes, 2) the setup of oral production tasks as part of the EFL syllabus and 3) the students' motivation for EFL learning. Thus, it would be advisable for teachers, curriculum developers and policy makers to take these considerations into account as enhancers of EFL oral performance in Spanish pre-university students: teachers should encourage to use EFL in the educational context, oral presentations should be a compulsory part of the EFL curriculum, and there should be assurances so that high school students are positively motivated to learn EFL.

In linguistic terms, the findings of this study also imply that maybe the differences between writing and speaking might not be as clear as one may expect from the time devoted in class to each skill. This could erroneously lead to thinking that no special effort may be necessary to improve the speaking in high school. However, although Oller (1979) warned of this affect, he also agreed that there is a considerable step to reach balanced language learning.

Another finding in this paper is that, after 12 years of foreign language classes, most learners are likely to have only acquired a CEFR B1 level of English. With the increase of bilingual (English-Spanish) programs in Spain, significant changes are expected to be found even on a short term (Pavón Vázquez, Ávila López, Gallego Segador, & Espejo Mohedano, 2015). There is no question that this limited performance needs to be improved, in both oral and written language, if current students are to study and work abroad. Thus, it is necessary not only to revise the learning process but also to design sound test that ensures a positive impact in the education (García Laborda & Fernández Álvarez, 2011; García Laborda & Litzler Jerman, 2015) especially if delivered online (García Laborda, Magal Royo, Litzler, & Giménez López, 2014). Although, in general, this has been associated to positive washback, the intended effect should go beyond the test and have a permanent and deeper effect. We strongly believe that tests could trigger such effect but it requires more than just revision of the EFL-PAU: it requires a strong planning that at this point still needs to be implemented.

7. Acknowledgements

The researchers would like to express their gratitude to the Ministry of Economy and Competitiveness (MICINN) for supporting the development and implementation the OPENPAU research project (FFI2011-22442) with Co-founding with ERDF funds under the 2008-2011 plan. Finally, Jesús García Laborda wants to express that this paper could not have been possible without the visit to Penn State University supervised by Dr. Joan Kelly Hall through the Research Mobility for Senior researchers grant (with funding from the Spanish Ministry of Education, Culture & Sports, PRX12/00376) who actually collaborated by revising part of the research design and data analysis.

8. Works Cited

Alderson, J. Charles, & Dianne Wall, 1993: "Does Washback Exist?", *Applied Linguistics* 14 (2), 115-129.

Amengual Pizarro, Marian, 2010: "Exploring the Washback Effects of a High-stakes English Test on the teaching of English in Spanish Upper Secondary Schools", *Revista Alicantina de Estudios Ingleses* 23, 149-170.

Bai, Li, Peter Hudson, Jan Millwater, & Megan Tones, 2013: "Development of a survey instrument to measure TEFL academics' perceptions about, individual and workplace characteristics for conducting research", *International Journal of Research & Method in Education* 36 (1), 52-66.

Barkhuizen, Gary, 2004: "Social Influences on Language Learning" in Alan Davies & Catherine Elder (eds.): *The Handbook of Applied Linguistics*, Malden, MA: Blackwell Publishing, 552-575.

Bracey, Gerald W., 2009: "Time to say goodbye to high school exit exams", *Principal Leadership* 10 (2), 64-66.

Bueno-Alastuey, Mª Camino, Jesús García Laborda, Ana Isabel Alcón, & Gloria Luque Agullo, 2014: "Setting the standards for the foreign language speaking tasks of the new baccalaureate general test", *Theory and Practice in Language Studies* 4 (9), 1763-1769.

Council of Europe, 2001: *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.

Díez-Bedmar, María Belén, 2011: "Spanish pre-university students' use of English: CEA results from the university entrance examination", *International Journal of English Studies* 11 (2), 141-158.

Ellis, Rod, 1999: "Item versus system learning: explaining free variation", *Applied Linguistics* 20, 460-80.

Ercikan, Kadriye, Wolff-Michael Roth, & Mustafa Asil, 2015: "Cautions about inferences from international assessments: The case of PISA 2009", *Teachers College Record* 117 (1), 28.

Escamilla, Kathy, Loretta Chavez, & Peter Vigil, 2005: "Rethinking the 'gap:' high-stakes testing and Spanish-speaking students in Colorado", *Journal of Teacher Education* 56 (2), 132-144.

European Commission, 2012: *First European survey on language competencies: final report*, Brussels: European Commission.

Fan, Jinsong, & Xun Yan, 2017: "From test performance to language use: Using self-assessment to validate a high-stakes English proficiency test", *The Asia - Pacific Education Researcher* 26 (1-2), 61-73.

Fernández Álvarez, Miguel, 2012: "Test Impact Inside and Outside the Classroom" in ACLES: *Multilingüismo en los Centros de Lenguas Universitarios: Evaluación, Acreditación, Calidad y Política Lingüística*, Granada: Centro de Lenguas Modernas, 49-69.

Fitchett, Paul G., & Tina L. Heafner, 2010: "A national perspective on the effects of high-stakes testing and standardization on elementary social studies marginalization", *Theory and Research in Social Education* 38 (1), 114-130.

García Laborda, Jesús, & Miguel Fernández Álvarez, 2011: "Teachers' opinions towards the integration of oral tasks in the Spanish university examination", *International Journal of Language Studies* 5 (3), 1-12.

García Laborda, Jesús, & Mary Frances Litzler Jerman, 2015: "External validation of the foreign speaking tasks of the high school leaving diploma", *International Journal of Language Studies* 9 (2), 97-112.

García Laborda, Jesús, Teresa Magal Royo, Mary Frances Litzler, & José Luis Giménez López, 2014: "Mobile phones for Spain's university entrance examination language test", *Educational Technology & Society* 17 (2), 17-30.

Gorges, Julia, & Christian Kandler, 2012: "Adults' learning motivation: Expectancy of success, value, and the role of affective memories", *Learning and Individual Differences* 22 (5), 610-617.

Gutiérrez Eugenio, Esther, & Nick Saville, 2017: "Policy review: The role of assessment in European language policy - a historical overview", *Assessment in European Language Policy, Languages, Society and Policy* [available at http://www.meits.org/files/policy_papers/uploads/Gutierrez_Eugenio_Saville_JLSP_2017.pdf on July 25, 2017].

Hill, Kathryn, & Tim McNamara, 2015: "Validity inferences under high-stakes conditions: A response from language testing", *Measurement: Interdisciplinary Research and Perspectives* 13 (1), 39-43.

Jellison Holme, Jennifer, Meredith P. Richards, Jo B. Jimerson, & Rebecca W. Cohen, 2010: "Assessing the effects of high school exit examinations", *Review of Educational Research* 80 (4), 476-526.

Howie, Sarah, 2012: "High-stakes testing in South Africa: Friend or foe?", *Assessment in Education: Principles, Policy & Practice* 19 (1), 81-98.

Hymes, Dell, 1972: "On communicative competence" in John B. Pride & Janet Holmes (eds.): *Sociolinguistics*, Harmondsworth: Penguin, 269-93.

Kang, Mun-koo, & Hyung-ji Chang, 2014: "Ensuring validity of practical English certification test of local office of education in Korea", *Journal of Pan-Pacific Association of Applied Linguistics* 18 (2), 111-122.

Kearns, Laura-Lee, 2011: "High-stakes standardized testing & marginalized youth: An examination of the impact on those who fail", *Canadian Journal of Education* 34 (2), 112-130.

Kim, Jinok, 2011: "Relationships among and between ELL status, demographic characteristics, enrolment history, and school persistence", *CRESST report 810 National Center for Research on Evaluation, Standards, and Student Testing (CRESST)* [http://www.cse.ucla.edu/products/reports/R810.pdf].

Koretz, Daniel, 2016: "Making the term 'validity' useful", *Assessment in Education: Principles, Policy & Practice* 23 (2), 290-292.

Krashen, Stephen, 1981: *Second language learning and second language acquisition*, Oxford: Pergamon Press.

Krashen, Stephen, 1985: *The Input Hypothesis: issues and implications*, New York: Longman.

Kwon, Suh Keong, Moonbok Lee, & Dongkwang Shin, 2017: "Educational assessment in the republic of korea: Lights and shadows of high-stake exam-based education system", *Assessment in Education* 24 (1), 60-77.

Liu, Yinggin, & Annie Ping Zeng, 2015: "Loss and gain: Revisiting the roles of the first language in novice adult second language learning classrooms", *Theory and Practice in Language Studies* 5 (12), 2433-2440.

Mahon, Elizabeth A., 2006: "High-stakes testing and English language learners: Questions of validity", *Bilingual Research Journal* 30 (2), 479-497.

Morgan, Stephen L., & Emily S. Taylor Poppe, 2012: "The consequences of international comparisons for public support of K-12 education: Evidence from a national survey experiment", *Educational Researcher* 41 (7), 262-268.

Moussu, Lucie, & Enric Llurda, 2008: "Non-native English-speaking English language teachers: History and research", *Language Teaching* 41 (3), 315-348.

Musoleno, Ronald R., & George P. White, 2010: "Influences of high-stakes testing on middle school mission and practice", *RMLE Online: Research in Middle Level Education* 34 (3), 1-10.

OECD, 2012: *PISA* 2009 *Technical Report*, Paris: OECD Publishing.

Oller, John, 1979: *Language tests at school*, London: Longman.

Pallant, Julie, 2005: *SPSS Survival Manual: a step by step guide to data analysis using SPSS for Windows*, Berkshire, UK: Open University Press.

Pavón Vázquez, Víctor, Javier Ávila López, Arturo Gallego Segador, & Roberto Espejo Mohedano, 2015: "Strategic and organisational considerations in planning content and language integrated learning: A study on the coordination between content and language teachers", *International Journal of Bilingual Education and Bilingualism* 18 (4), 409-425.

Qian, David, 2009: "Comparing direct and semi-direct modes for speaking assessment: Affective effects on test takers", *Language Assessment Quarterly* 6, 113-125.

Rea-Dickins, Pauline, 1997: "So, why do we need relationships with stakeholders in language testing? A view from the UK", *Language Testing* 14 (3), 304-314.

Russell, Michael, & Maureen Kavanaugh, 2011: *Assessing students in the margin: Challenges, strategies, and techniques*, Charlotte, NC: IAP - Information Age Publishing.

Salkind, Neil J., 2000: *Exploring research*, Upper Saddle River, NJ: Prentice Hall.

Scott, Evan Poon, 2012: "Short-term gain at long-term cost? how resit policy can affect student learning", *Assessment in Education: Principles, Policy & Practice* 19 (4), 431-449.

Smolin, Andrea, & Jennifer K. Clayton, 2009: "Standardized testing: How prepared are we?" *Journal of Cases in Educational Leadership* 12 (4), 29-36.

Solano-Flores, Guillermo, & Min Li, 2013: "Generalizability theory and the fair and valid assessment of linguistic minorities", *Educational Research and Evaluation* 19 (2), 19.

Spolsky, Bernard, 1989: *Conditions for second language learning*, Oxford: Oxford University Press.

Tsang, Sau-Lim, Anne Katz, & Jim Stack, 2008: "Achievement testing for English language learners, ready or not?", *Education Policy Analysis Archives* 16 (1), 29.

Van Rijn, Peter W., Anton A. Beguin, & Hubb H. Verstralen, 2012: "Educational measurement issues and implications of high stakes decision making in final examinations in secondary education in the Netherlands", *Assessment in Education: Principles, Policy & Practice* 19 (1), 117-136.

Vasquez Heilig, Julian, 2011: "Understanding the interaction between high-stakes graduation tests and English learners", *Teachers College Record* 113 (12), 2633-2669.

Wall, Dianne, 2005: *The impact of high-stakes examinations on classroom teaching: a case study using insights from testing and innovation theory*, Cambridge: Cambridge University Press.

Weir, Cyril, 2013: "Measured constructs: A history of Cambridge English language examinations 1913-2012", *Research Notes* 51, 2-10.

Wilby, Peter, 2012: "Can we believe the international league tables? FORUM: For Promoting 3-19", *Comprehensive Education* 54 (2), 341-344.

Wright, Wayne E., & Daniel Choi, 2006: "The impact of language and high-stakes testing policies on elementary school English language learners in Arizona", *Education Policy Analysis Archives* 14 (13), 1-75.

Wright, Wayne E., & Xiaoshi Li, 2008: "High-stakes math tests: How 'no child left behind' leaves newcomer English language learners behind", *Language Policy* 7 (3), 237-266.

Xu, Yun, & Zunmin Wu, 2012: "Test-taking strategies for a high-stakes writing test: An exploratory study of 12 Chinese EFL learners", *Assessing Writing* 17 (3), 174-190.

Zhao, Yong, & Chun Lai, 2007: "Technology and second language learning: Promises and problems" in L. A. Parker (ed.): *Technology-based learning environments for young English learners: In and out of school connections*, Mahwah, NJ: Erlbaum, 167-205.

Zou, Shen, & Quian Xu, 2017: "A washback study of the test for English majors for grade eight (TEM8) in china-from the perspective of university program administrators", *Language Assessment Quarterly* 14 (2), 140-159.