

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

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

Jesús García Laborda

Universidad de Alcalá

Gloria Luque Agulló

Universidad de Jaén

Ana Isabel Muñoz

Universidad Católica de Ávila

Margarita Bakieva

Universidad de Valencia

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 sugirieron 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.

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 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.

Introducción

El Examen de Acceso a la Universidad es el examen general selectivo más importante en España. La inclusión de una sección de lengua extranjera

se remonta a hace más 25 años, pero a pesar de los cambios en la enseñanza de idiomas a lo largo del tiempo, esta sección ha visto cambios muy limitados en su estructura. Obviamente, las necesidades de los estudiantes y de la sociedad han cambiado en ese período de tiempo y, como consecuencia, el test ha quedado obsoleto y sólo ofrece una cantidad de información limitada. Más aún, esta información se usa para obtener inferencias sobre la capacidad lingüística de los estudiantes y también para fijar el nivel de acceso a un grado universitario. En contraste con la estructura obsoleta del examen, las habilidades orales han sido uno de los temas clave en la recién aprobada Ley Orgánica que Modifica la Calidad de Educación (LOMCE) (2014) en España. De hecho, se ha puesto un gran énfasis en el hecho de los estudiantes españoles deberían acabar la Educación Secundaria con un excelente dominio oral de, al menos, una lengua extranjera. En España y muchos otros países europeos este ha sido un tema de gran preocupación. Sin embargo, lo que hace el caso de España incluso más crítico son los malos resultados obtenidos en las evaluaciones internacionales y nacionales tales como la Encuesta Europea de Competencia Lingüística (http://ec.europa.eu/languages/policy/strategic-framework/documents/language-survey-final-report_en.pdf). Se midieron tres habilidades: lectura, escritura y comprensión oral, y España fue uno de los países con resultados más bajos, en particular en la habilidad de expresión oral así como en la habilidad escrita. Además se sabe en general en el mercado laboral, aparte de en las instituciones de educación superior, que los estudiantes españoles carecen de las habilidades necesarias para seguir estudios universitarios en los que el inglés es el idioma de comunicación o, al menos, una herramienta para el aprendizaje.

Recientes estudios realizados por el Instituto de Evaluación Educativa, dependiente del Ministerio español de Educación, Cultura y Deporte (MECD), sostienen que un 45% de estudiantes podrían alcanzar un nivel B1 al terminar la Enseñanza Secundaria y Bachillerato. García Laborda, Amengual Pizarro y Litzler (2013) dudan de si esta competencia lingüística es aceptable para enfrentarse al mercado laboral o a los estudios universitarios. Usando el Marco Común Europeo para las Lenguas (CEFR) en el inglés para Fines Específicos, estos autores consideran que el número de situaciones en las que los graduados de Bachillerato podrían solo usar diálogos muy breves (correspondientes a un nivel B1) es elevado, y por lo tanto prevén graves problemas

comunicativos más allá de la interacción social cuando las estructuras gramaticales, pero especialmente el vocabulario y el discurso, constituyan la parte más importante de la comunicación profesional. En el mismo artículo del congreso también mencionan que los resultados de la encuesta europea podrían ser poco realistas por cuanto los entrevistadores tenían una preparación muy limitada para realizar exámenes orales y podrían haber tenido algunas ideas preconcebidas sobre los resultados y sobre los candidatos que iban a examinar.

El Examen de Acceso a la Universidad en España

La parte de inglés del Examen de Acceso a la Universidad en España ha recibido muy poca atención en la investigación si se compara con otros exámenes estandarizados de lengua extranjera. Sólo en la pasada década un número de investigadores han afrontado este problema. En general, parece haber tres aspectos principales en los que se centra la mayoría de los artículos sobre el tema: washback (Amengual Pizarro, 2009), análisis de resultados y propuestas para cambiar su estructura actual, que no ha sido modificada durante más de 20 años (Fernández Álvarez y Sanz Sainz, 2005). Aunque podría parecer que existiera una clara distinción entre estos tres aspectos, en realidad están claramente interrelacionados.

Lo que parece estar claro es que con pocas variaciones en las Comunidades Autónomas de Galicia y Cataluña, los cambios se han reducido al formato (tales como el número de palabras en la parte de composición escrita o el número de palabras en la parte de uso de la lengua) y la incorporación de secciones de comprensión oral (*listening*) en las dos Comunidades arriba mencionadas, que tienen también una lengua oficial adicional. En 2005, un volumen editado por Herrera Soler y García Laborda intentó indicar esta carencia de estudios, en particular dedicados a la validación, el aspecto más relevante. Sin embargo, hasta ese momento, y aún hoy, los estudios de validación han sido limitados en contenido y amplitud. García Laborda (2006) señalaba que los estudios de validación hasta entonces solo se habían hecho ocasionalmente en cuatro universidades: Granada, Complutense de Madrid, Baleares y Politécnica de Valencia. Después del volumen de 2005 hubo algunos artículos más publicados en torno a este tema, de los cuales es digno hacer mención del monográfico de Revista de Educación en 2011, que

incluía aspectos que iban desde las consideraciones interculturales a la aplicación mediante ordenadores.

Dada esta ausencia de estudios, el Ministerio de Educación, Cultura y Deportes (MECD) es consciente de la necesidad de considerar un número de acciones para superar estos problemas. En primer lugar, se necesita revisar el paradigma educativo en España, particularmente en relación a las políticas lingüísticas. Cuando se mejoren estas políticas, la evaluación y exámenes deben desempeñar un papel más relevante que en los anteriores modelos educativos. Este aumento de las políticas de evaluación debería, en primer lugar, ser adecuadamente diseñado, ya que puede influir en los resultados de la enseñanza; en segundo lugar, tener un papel periódico, ya que con toda seguridad reportará importantes beneficios a los colegios, la enseñanza y el aprendizaje y, tercero, tener un impacto moderado y asumir a las desigualdades de la sociedad española, asegurando que no todos los recursos que se proporcionan a la escuela dependen de los resultados de estas evaluaciones.

Teniendo en cuenta las acciones del MECD, el proyecto OPENPAU proponía diferentes alternativas y seguía dos líneas principales para afrontar el análisis de las limitaciones actuales de los estudiantes españoles. Por un lado, el Coordinador del proyecto estableció líneas de cooperación entre el proyecto de investigación y el MECD. La idea general era que la experiencia del proyecto OPENPAU¹ servía para aportar ideas para mejorar la situación actual y también para revisar un informe interno sobre el dominio del inglés de los alumnos que terminan Bachillerato. Como contrapartida, el MECD se ofreció para proveer información sobre la investigación actual a través de la entrega online de bases de datos de la investigación.

Este artículo se centra en la realización de la prueba de expresión oral (*speaking*) en entrevistas por parejas. Los resultados se examinan atendiendo a cuatro criterios (precisión, fluidez, interacción y coherencia) en tres universidades y abarca estudiantes de cuatro comunidades autónomas de España. Según lo expuesto anteriormente, el estudio se justifica por la ausencia de estudios serios que puedan analizar la actual situación de expresión oral de los graduados de centros de enseñanza Secundaria y Bachillerato y observar si los resultados del MECD pueden contrastarse con investigación no institucional. Mientras que la opinión

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

generalizada es que hay una necesidad destacada de incorporar tareas de expresión oral en el Examen de Acceso a la Universidad, son limitados los estudios previos, hasta lo que se conoce actualmente, al realizado por el MECD y este que aquí presentamos. Este se basa en dos premisas: 1) las tareas del examen tienen un impacto significativo en lo que se enseña; y 2) la aplicación del examen podría facilitar la implementación de tareas de expresión oral (*speaking*).

El estudio que se presenta a continuación trata primero algunos de los problemas actuales asociados a la aplicación de exámenes orales en agrupamiento por parejas, muestra después las preguntas de investigación para proceder seguidamente al análisis experimental de las entrevistas orales de 169 estudiantes universitarios. Finalmente, contrasta los resultados de esta investigación experimental con los de la realizada por el MECD para terminar con algunas conclusiones que podrían aportar ideas para el examen actual y el futuro diploma de graduación en Bachillerato.

Revisión de la literatura

Aunque durante las últimas décadas ha habido debates sobre la necesidad de diseñar e implementar nuevas tareas de expresión oral en los exámenes selectivos, desde comienzos del siglo 21 ha habido dos claras líneas: una relacionada con las tareas de expresión oral de exámenes orales aplicados online (Bernhardt, Rivera & Kamil, 2004; Chapelle & Douglas, 2006; Vitienes & Miciuliene, 2008; Sawaki, Stricker, & Oranje, 2009; García Laborda, 2010b y otros) y, la otra, tareas de expresión oral aplicadas presencialmente (Nakatsuhara, 2013). Entre las últimas, una de las tendencias más comunes se centra en los exámenes en parejas y en grupo. Aunque las tareas integradas han atraído tanto a los examinadores online como presenciales (Sawaki, Stricker, & Oranje, 2009), merece la pena considerar cuáles son las ventajas que las entrevistas de grupo tienen sobre los ordenadores en contextos específicos. García Laborda & Royo (2007) mencionan un número de dificultades que hacen de los exámenes por ordenador rentables a largo plazo, ya que se requieren grandes inversiones tanto en software como en hardware. Esta puede ser la razón más importante por la que muchas administraciones educativas no han implementado exámenes por ordenador en los exámenes

selectivos. Sin embargo, las entrevistas cara a cara o incluso por teléfono como, por ejemplo, la *Simulated Oral Proficiency Interview (SOPI)* requieren bastante tiempo y aún implican un alto coste en personal y recursos (Heilmann, 2012). Por lo tanto, para muchas instituciones los exámenes de idiomas presenciales en parejas o en grupo son una respuesta asequible. Dado el contexto actual en España y la propia naturaleza obligatoria del examen PAU, cuestiones relacionadas con las características individuales, el agrupamiento o la personalidad podrían tener un efecto potencial en los examinandos. No obstante, es necesario promover formas de comunicación en los exámenes que requieren la interacción como una parte significativa del constructo comunicativo (Brooks, 2009) y proporciona mejores inferencias a unos costes más bajos (Dunbar, Brooks & Kubicka-Miller, 2006). Ya que la introducción de estas tareas debería tener como objetivo una comunicación más real, se esperaría que el uso de exámenes orales tuviera un efecto positivo en los estudiantes y, por tanto, proporcionara un wash back positivo (Muñoz & Alvarez, 2010).

Las tareas de exámenes de expresión oral en parejas se han hecho muy comunes en muchos exámenes internacionales, especialmente en el consejo examinador de Cambridge (Shaw & Weir, 2007), y Nakatsuhara (2013) menciona un gran número de ejemplos en muchas otras partes del mundo. Los exámenes en parejas tienen un número de beneficios, incluyendo su reducido coste y rendimiento del tiempo, y su enfoque en la interacción en diálogos o discurso construidos en común (Galaczi, 2008; Nakatsuhara, 2006; Gan, 2008; Gan, 2010), particularmente en países como España, en el cual la cooperación y el compartir en la educación están altamente considerados. Adicionalmente, pueden proporcionar mejores oportunidades para estudiantes más flojos (Elder, Iwashita & McNamara, 2002), facilitar la fluidez (Gan, Davison & Hamp-Lyons, 2009), dar un papel especial al lenguaje corporal y facilitar la observación de funciones de habla de alto nivel.

Aunque estos factores se han observado internacionalmente, hasta ahora ningún examen formal aplicado en España usa la evaluación en grupo o ni siquiera en parejas para el discurso oral. En ese sentido, este artículo parece ser un primer intento en el contexto de exámenes selectivos.

Diseño e implementación de la investigación

Preguntas de la investigación

Dada la situación actual en España y la revisión bibliográfica, las siguientes cuestiones necesitaban ser investigadas para el propósito de este estudio:

RQ1: *¿Hay diferencias significativas entre el estudio sobre estudiantes de 2º de Bachillerato realizado por el Instituto Nacional de Evaluación Educativa (INEE) y nuestra presente investigación?*

RQ2: *¿Es el enfoque propuesto en esta investigación válido para aplicar el Examen de Acceso a la Universidad?*

Merece la pena considerar si los exámenes de expresión oral aplicados en grupo son más adecuados que las entrevistas individuales cara a cara en el contexto español.

Método de investigación y participantes en el estudio

Participantes

El equipo de investigación grabó un total de 85 entrevistas en parejas en cuatro comunidades autónomas diferentes en España, a saber, Castilla-La Mancha, Castilla-León, Andalucía y Madrid (Tabla 2). Estas regiones no comparten el mismo examen de ingreso, pero el único aspecto que tienen en común es el de no incluir una sección oral (para más detalles sobre los exámenes, ver Bueno y Luque, 2012). Se eligió una universidad de cada comunidad autónoma para hacer representativa la muestra y también maximizar los recursos económicos y personales: Universidad de Alcalá (Madrid y Castilla-La Mancha), Universidad Católica de Ávila (Castilla-León) and Universidad de Jaén (Andalucía). Los estudiantes eran voluntarios en su primer año de universidad. La Tabla II muestra la cantidad y frecuencia de participantes y su región/comunidad de origen. La Tabla III indica el número y frecuencia de participantes que se ofrecieron voluntariamente para el estudio dependiendo de la Facultad a la que estaban adscritos. Debido a las diversas comunidades de origen, universidades y facultades de las que provenían los estudiantes, se consideró que el rango y variedad de participantes evaluado en el estudio mostraría una panorámica amplia del nivel de competencia oral de los estudiantes universitarios de nuevo ingreso en España.

TABLA II. Número de participantes seleccionados para elestudio, porcentajes y Universidades de origen

| Universidad | Frecuencia | Porcentaje |
|---------------------------------|------------|------------|
| Andalucía (Universidad de Jaén) | 39 | 23,1 |
| CAM (***) | 46 | 27,2 |
| CLM (***) | 21 | 12,4 |
| Castilla- León (***) | 55 | 32,5 |
| Otras (***) | 8 | 4,8 |
| Total | 169 | 100,0 |

TABLA III. Facultades seleccionadas para el estudio y porcentaje de participantes.

| Facultades | Frecuencia | % | Universidad |
|--------------------|------------|-------|---|
| Psicología | 38 | 22,5 | Jaén-Andalucía |
| Educación Infantil | 16 | 9,5 | Alcalá-Madrid/Castilla La Mancha |
| Educación Primaria | 52 | 30,8 | Alcalá-Madrid-Castilla La Mancha |
| Enfermería | 57 | 33,7 | Universidad Católica de Ávila-Castilla León |
| Derecho | 4 | 2,4 | Universidad Católica de Ávila-Castilla León |
| Total | 169 | 100,0 | |

Tareas de investigación y recogida de datos

Medida de los niveles de competencia lingüística.

Para medir la competencia lingüística de los estudiantes, los investigadores utilizaron los niveles de puntaje del CEFR, desde A1 a B2, para evaluar la competencia oral general de los participantes. Según esto, se usaron cuatro criterios de puntaje para medir la actuación de los

estudiantes: precisión, fluidez, interacción y coherencia. Se asignaron calificaciones de 0 a 3 según su actuación, siendo 3= excelente, 2=normal, 1=mala, 0 fue asignada en muy pocos casos a estudiantes que no respondieron en absoluto o cuya actuación fue excepcionalmente mala. Las respuestas de los estudiantes fueron grabadas en vídeo, las entrevistas numeradas y finalmente transcritas.

Examinadores.

Los examinadores eran profesores capacitados con una larga experiencia en examinar idiomas tanto para exámenes estandarizados como para los Exámenes de Acceso a la Universidad.² Las entrevistas fueron así examinadas y calificadas por seis de estos examinadores, los cuales se pusieron de acuerdo sobre cada competencia global y establecieron las calificaciones para cada criterio individual. Se llegó a un acuerdo en cómo se aplicarían las tareas de expresión oral (*speaking*) antes de comenzar el proceso. Uno de los entrevistadores se encargó de hacer las preguntas adecuadas. Era en cierto grado activo en la tarea 1 (ver más abajo) y estaba observando activamente y apoyando en distinto grado en las tareas 2 y 3. En estas tareas el entrevistador sólo participaba si había una clara oportunidad en la línea conversacional.

Tareas de expresión oral

Con el fin de organizar el diseño y la aplicación de las tareas de expresión oral, se distribuyó un cuestionario a 16 coordinadores de la prueba PAU de todas las Comunidades españolas para averiguar sus opiniones sobre el tipo de tareas de *speaking* que ellos consideraban eran las mejores en el caso hipotético de que se incorporaran en la PAU desde 2012. Sus respuestas evidenciaron que preferían tres tipos de tareas: preguntas personales de socialización, descripción de una foto y dramatización por medio de role-play. También, de acuerdo con sus respuestas, el nivel que debería alcanzarse en el último año de instituto debería ser un B1 según el CEFR y se consideraron y aplicaron tres tareas en el siguiente orden: 1) Diálogo informal; 2) Descripción de una foto + diálogo de pregunta-respuesta; Role-play sobre una situación planteada. La pregunta uno consistía en preguntas individuales planteadas por el

⁽²⁾ Se llevó a cabo un juicio de experto con la ayuda de los profesores/examinadores, que tenían una amplia experiencia en la evaluación de tareas en exámenes estandarizados o en el Examen de Acceso a la Universidad (un mínimo de seis años como evaluadores). Todos habían tomado parte en un seminario de capacitación. Una vez que se llegó a un acuerdo sobre los aspectos que debían ser evaluados, fue aprobado por todos los examinadores y usado en todas las entrevistas.

entrevistador sobre temas de relaciones sociales tales como los deportes, hobbies, miembros de la familia, intereses académicos y otros de este tipo. En la pregunta dos al estudiante se le asignaba una foto elegida al azar de entre un conjunto de diez y debía describirla durante dos minutos. Entonces el compañero hacía dos preguntas de opinión tales como “¿por qué crees que están aquí?” o “¿qué crees que ocurrirá después?”. Finalmente, para la pregunta tres a los estudiantes se les daba una tarjeta planteando un caso para discutir tales como “organiza una fiesta en tu casa con la ayuda de tu compañero” u “organiza una sesión de estudio para tu próximo examen”. La Tabla 4 muestra las preguntas así como la estructura y objetivos de cada una de estas tres tareas.

TABLA IV. Tareas de expresión oral. Tipo de tarea, número de participantes, información requerida, objetivos y tiempo medio para completar cada tarea.

| Tipo de tarea | Número de participantes | Información requerida | Grado de implicación | Objetivo | Tiempo medio de la tarea |
|---|------------------------------|---|--|--|---|
| Diálogo informal | 2 (entrevistador-examinando) | Cuestiones personales sobre el origen, familia, hobbies y otras | Bastante limitada a la interacción I a I | Romper el hielo | 3 minutos (porestudiante) |
| Descripción de una foto + Diálogo de 2 preguntas-respuestas | 2 (candidato-candidato) | Descriptiva y de averiguación o justificación | Semi-pasiva (monólogo) + semi-activa (respuesta a 2 preguntas) | Evaluación de competencia Interacción inicial | 2 minutos por descripción + unos 2 minutos por estudiante |
| Role play basado en una situación dada | 2 (candidato-candidato) | Adecuada a la situación (caso) | Activa | Participación interaccional Discurso libre Tarea cooperativa – Interactiva | 4 a 5 minutos total |

Recogida de datos

Las entrevistas se llevaron a cabo entre diciembre de 2012 y marzo de 2013. Los estudiantes se ofrecieron voluntarios y las entrevistas tuvieron lugar en días de clase, normalmente antes o después de las sesiones de clase. Todos los estudiantes estaban matriculados en el primer semestre de clases

de inglés, ya fuera general o para fines específicos. Los alumnos tenían que atenerse a su propio horario de clases, así que el equipo examinador decidió agruparlos al azar y sin tener en cuenta su nivel de competencia.

Resultados

El primer asunto que abordar en este estudio, dado el número de temas que surgieron en el estudio del Ministerio, fue observar cuál era el nivel de competencia que traían los estudiantes universitarios de primero. En lugar de elegir un método de evaluación referido a un criterio, la investigación se centró en los niveles de competencia en un examen con referencia a una norma. Esto significa que, más que observar si los alumnos podían alcanzar el nivel de competencia B1, nosotros observamos cómo podían ser clasificados según su competencia. La Tabla V muestra los resultados globales obtenidos según el CEFR.

TABLA V. Nivel de los participantes según el CEFR.

| Nivel | Número de estudiantes | Porcentaje |
|-------|-----------------------|------------|
| A1 | 32 | 18,9 |
| A2 | 62 | 36,7 |
| B1 | 57 | 33,7 |
| B2 | 14 | 8,3 |
| C1 | 4 | 2,4 |
| Total | 169 | 100,0 |

En nuestro estudio, cerca de 55,6 % no alcanzaron el requisito mínimo de competencia exigido en idiomas extranjeros por el Ministerio de Educación para graduarse en Bachillerato. Más aún, el porcentaje de estudiantes que estaban sobre el nivel requerido era solo de 10,7. Consideramos importante esta cifra porque era similar a la obtenida en la Encuesta Europea de Competencia Lingüística.

Dado que estábamos observando la actuación en un examen dado, el segundo aspecto a evaluar consistía en observar cómo los estudiantes se

comportaban en el examen dentro de su propio nivel de competencia a lo largo de los criterios del examen. La Tabla VI indica los resultados de los estudiantes en los criterios de precisión. La precisión se entendió como “gramaticalidad” o adecuación a la gramática estándar prescriptiva.

TABLA VI. Resultados de la actuación de los participantes para los criterios de precisión. Los valores varían de cero a tres. Cada línea indica la cantidad total y el porcentaje que los entrevistadores dieron a la actuación de los participantes en diferentes carreras de Grado.

| | | | Precisión | | | | Total |
|-------|---------------------|------------------------|-----------|--------|--------|-------|--------|
| | | | 0 | 1 | 2 | 3 | |
| Grado | Psicología | Número de respuestas | 2 | 20 | 12 | 4 | 38 |
| | | % dentro de Titulación | 5.3% | 52.6% | 31.6% | 10.5% | 100.0% |
| | Educación Infantil | Número de respuestas | 0 | 10 | 6 | 0 | 16 |
| | | % dentro de Titulación | .0% | 62.5% | 37.5% | .0% | 100.0% |
| | Educación Primaria | Número de respuestas | 0 | 25 | 26 | 1 | 52 |
| | | % dentro de Titulación | .0% | 48.1% | 50.0% | 1.9% | 100.0% |
| | Enfermería | Número de respuestas | 0 | 35 | 15 | 7 | 57 |
| | | % dentro de Titulación | .0% | 61.4% | 26.3% | 12.3% | 100.0% |
| | Derecho | Número de respuestas | 0 | 1 | 3 | 0 | 4 |
| | | % dentro de Titulación | .0% | 25.0% | 75.0% | .0% | 100.0% |
| | IngenieríaEléctrica | Número de respuestas | 0 | 1 | 0 | 0 | 1 |
| | | % dentro de Titulación | .0% | 100.0% | .0% | .0% | 100.0% |
| | Ingeniería Civil | Número de respuestas | 0 | 0 | 1 | 0 | 1 |
| | | % dentro de Titulación | .0% | .0% | 100.0% | .0% | 100.0% |
| | Total | Número de respuestas | 2 | 92 | 63 | 12 | 169 |
| | | % | 1.2% | 54.4% | 37.3% | 7.1% | 100.0% |

Después de observar las diferentes frecuencias en los niveles de precisión, se dio un paso más en el análisis para descubrir si diferentes

grupos de estudiantes o diferentes universidades mostraban diferencias significativas, es decir, si el origen de los participantes o su elección del grado tendrían una relación con su nivel de precisión. Los datos (chi-cuadrado: 23,254) indican que los resultados no mostraban relaciones intergrupales significativas y que dentro de los grupos la mayoría de los estudiantes tendían a obtener un puntaje bajo o medio. Esto podría bien indicar la tendencia de los examinadores a valorar a los estudiantes bajo o que realmente los estudiantes tendían a obtener bajos puntajes en exámenes orales específicos. También se observó que había una dependencia del grado en los puntajes sobre el grado de la Titulación. Por ejemplo, era el caso de la Escuela de Educación de la Universidad de Alcalá, donde una diferencia significativa entre las estudiantes de Educación Primaria tendía a hacerlo mejor que sus compañeras de Educación Infantil. Finalmente, el coeficiente de correlación de rango Kendal Tau-b (0,029) y un error standard asintótico pequeño (0,72) indicaron la ausencia de asociación entre las diferentes muestras. Como consecuencia, los resultados indican que hay diferencias limitadas entre los grupos. No obstante, hubo una tendencia a obtener puntuaciones más bajas en las futuras maestras y enfermeras.

Se aplicó el mismo análisis a los otros criterios referidos arriba: fluidez, interacción y coherencia (Tablas VII a IX).

De nuevo, tras observar las frecuencias, se llevó a cabo un análisis para comprobar si se podían encontrar diferencias significativas dependiendo del grupo de estudiantes o de la universidad/facultad de procedencia. En este caso, el χ^2 era 13,637 (no significativo) con una significancia asintótica de 0,752. Esto claramente indica que la curva tenía una tendencia derecha y que no había significancia en los resultados chi-cuadrados inter-grupales.

Estos resultados estadísticos indican que los estudiantes tuvieron una actuación mejor en este criterio. De hecho, los exámenes muestran una clara tendencia a un puntaje promedio con un mayor grado hacia la excelencia en todos los grupos. Esta excelencia no es tan extremadamente alta en ningún grupo pero tiene un promedio de 19,8%. Psicología en Jaén obtiene mejores puntajes seguida de los maestros de Educación Primaria de la Escuela de Educación de Madrid. La Tau-b de Kendall (-0,101, error standard 0,70) indica un grado más alto de asociación que en el criterio previo, pero es aún bastante limitado.

TABLA VII. Resultados de la actuación de los participantes para los criterios de fluidez. Las medidas varían de cero a tres. Cada línea indica la cantidad total y el porcentaje que los examinadores dieron a las actuaciones de los participantes en los diferentes grados.

| | | | Fluidez | | | | Total |
|-------|---------------------|------------------------|---------|-------|--------|-------|--------|
| | | | 0 | 1 | 2 | 3 | |
| Grado | Psicología | Número de respuestas | 1 | 9 | 17 | 11 | 38 |
| | | % dentro de Titulación | 2.6% | 23.7% | 44.7% | 28.9% | 100.0% |
| | Educación Infantil | Número de respuestas | 0 | 5 | 7 | 4 | 16 |
| | | % dentro de Titulación | .0% | 31.3% | 43.8% | 25.0% | 100.0% |
| | Educación Primaria | Número de respuestas | 0 | 12 | 31 | 7 | 50 |
| | | % dentro de Titulación | .0% | 24.0% | 62.0% | 14.0% | 100.0% |
| | Enfermería | Número de respuestas | 0 | 21 | 25 | 11 | 57 |
| | | % dentro de Titulación | .0% | 36.8% | 43.9% | 19.3% | 100.0% |
| | Derecho | Número de respuestas | 0 | 2 | 2 | 0 | 4 |
| | | % dentro de Titulación | .0% | 50.0% | 50.0% | .0% | 100.0% |
| | IngenieríaEléctrica | Número de respuestas | 0 | 0 | 1 | 0 | 1 |
| | | % dentro de Titulación | .0% | .0% | 100.0% | .0% | 100.0% |
| | Ingeniería Civil | Número de respuestas | 0 | 0 | 1 | 0 | 1 |
| | | % dentro de Titulación | .0% | .0% | 100.0% | .0% | 100.0% |
| | Total | Número de respuestas | 1 | 49 | 84 | 33 | 167 |
| | | % | .6% | 29.3% | 50.3% | 19.8% | 100.0% |

La interacción también muestra unos puntajes de actuación más altos que se asemejan a los obtenidos en el criterio de fluidez. La χ^2 para la interacción fue 15,49 con una significancia asintótica de 0,628 con una clara tendencia hacia los valores 2-3, y de nuevo no hubo significancia en los resultados inter-grupales chi-cuadrado. También se observaron puntajes medio-altos para este criterio. Se observaron altos puntajes (3)

en tres grupos pero en este caso, las notas bajas (1) fueron incluso menores que en el criterio anterior. Especialmente significativo es el caso de los futuros maestros de Educación Primaria, que mostraron un porcentaje menor de puntajes bajos (1) pero que apenas aumentaron los puntajes altos (3). Esto fue también refrendado por el valor Tau-b de Kendall (-.092, Error standard 0,71).

TABLA VIII. Resultados de la actuación de los participantes para el criterio de la interacción. Las medidas varían de cero a tres. Cada línea indica la cantidad total y el porcentaje que los entrevistadores en los diferentes grados dieron a las actuaciones de los participantes.

| | | | Interacción | | | | Total |
|-------|----------------------|------------------------|-------------|--------|--------|-------|--------|
| | | | 0 | 1 | 2 | 3 | |
| Grado | Psicología | Número de respuestas | 1 | 12 | 17 | 8 | 38 |
| | | % dentro de Titulación | 2.6% | 31.6% | 44.7% | 21.1% | 100.0% |
| | Educación Infantil | Número de respuestas | 0 | 6 | 6 | 4 | 16 |
| | | % dentro de Titulación | .0% | 37.5% | 37.5% | 25.0% | 100.0% |
| | Educación Primaria | Número de respuestas | 0 | 14 | 32 | 6 | 52 |
| | | % dentro de Titulación | .0% | 26.9% | 61.5% | 11.5% | 100.0% |
| | Enfermería | Número de respuestas | 1 | 24 | 21 | 11 | 57 |
| | | % dentro de Titulación | 1.8% | 42.1% | 36.8% | 19.3% | 100.0% |
| | Derecho | Número de respuestas | 0 | 3 | 1 | 0 | 4 |
| | | % dentro de Titulación | .0% | 75.0% | 25.0% | .0% | 100.0% |
| | Ingeniería Eléctrica | Número de respuestas | 0 | 1 | 0 | 0 | 1 |
| | | % dentro de Titulación | .0% | 100.0% | .0% | .0% | 100.0% |
| | Ingeniería Civil | Número de respuestas | 0 | 0 | 1 | 0 | 1 |
| | | % dentro de Titulación | .0% | .0% | 100.0% | .0% | 100.0% |
| Total | | Número de respuestas | 2 | 60 | 78 | 29 | 169 |
| | | % | 1.2% | 35.5% | 46.2% | 17.2% | 100.0% |

TABLA IX. Resultados de la actuación de los participantes para la coherencia. Las medidas varían de cero a tres. Cada línea indica la cantidad total y el porcentaje que los entrevistadores en los diferentes grados dieron a la actuación de los participantes.

| | | | Interacción | | | | Total |
|--------------|----------------------|------------------------|-------------|--------|--------|-------|--------|
| | | | 0 | 1 | 2 | 3 | |
| Grado | Psicología | Número de respuestas | 1 | 12 | 17 | 8 | 38 |
| | | % dentro de Titulación | 2.6% | 31.6% | 44.7% | 21.1% | 100.0% |
| | Enseñanza Infantil | Número de respuestas | 0 | 6 | 6 | 4 | 16 |
| | | % dentro de Titulación | .0% | 37.5% | 37.5% | 25.0% | 100.0% |
| | Enseñanza Primaria | Número de respuestas | 0 | 14 | 32 | 6 | 52 |
| | | % dentro de Titulación | .0% | 26.9% | 61.5% | 11.5% | 100.0% |
| | Enfermería | Número de respuestas | 1 | 24 | 21 | 11 | 57 |
| | | % dentro de Titulación | 1.8% | 42.1% | 36.8% | 19.3% | 100.0% |
| | Derecho | Número de respuestas | 0 | 3 | 1 | 0 | 4 |
| | | % dentro de Titulación | .0% | 75.0% | 25.0% | .0% | 100.0% |
| | Ingeniería Eléctrica | Número de respuestas | 0 | 1 | 0 | 0 | 1 |
| | | % dentro de Titulación | .0% | 100.0% | .0% | .0% | 100.0% |
| | Ingeniería Civil | Número de respuestas | 0 | 0 | 1 | 0 | 1 |
| | | % dentro de Titulación | .0% | .0% | 100.0% | .0% | 100.0% |
| | Total | Número de respuestas | 2 | 60 | 78 | 29 | 169 |
| | | % | 1.2% | 35.5% | 46.2% | 17.2% | 100.0% |

La coherencia también muestra mayores puntajes de actuación, que se asemejan a los obtenidos en los dos criterios anteriores. La χ^2 para la coherencia fue 26,412 (no significativo) con una significancia asintótica de 0,091 que también muestra una tendencia hacia los valores medios y altos, aunque en este caso está más centralizada que en los dos criterios anteriores. En referencia a los resultados, los estudiantes de Psicología

obtuvieron puntajes más bajos que en el segundo y tercer criterio pero este cambio no fue significativo. Tampoco lo fue en el resto de los grupos, cuyos puntajes disminuyeron pero no en el mismo porcentaje. El valor Tau-b de Kendall $-.092$, Error standard 0.71 muestra que no se observaron diferencias significantes en relación a las curvas de los otros grupos.

La actuación global se consideró importante porque proporcionó información sobre las calificaciones generales de todos los participantes en el estudio (García Laborda, Amengual Pizarro y Litzler, 2013) y pudo también ser contrastada con los datos obtenidos por el Ministerio de Educación. La Tabla X muestra los datos obtenidos por el Ministerio en su estudio piloto realizado en 2012 (<http://www.mecd.gob.es/dctm/inee/documentos-de-trabajo/informe-pau-ingles.pdf?documentId=0901e72b8170cdc9>), mientras que los resultados de esta investigación se pueden observar en las dos tablas siguientes (Tablas XI y XII).

TABLA X. Resultados de la actuación global de los participantes obtenida por el Ministerio de Educación (<http://www.mecd.gob.es/inee/Documentos-de-trabajo.html>).

| | Criterio | | | | |
|----------------|----------|-----------------------|---------|-------------|------------|
| | Alcance | Corrección gramatical | Fluidez | Interacción | coherencia |
| Parte 1 | 65.08 | 60.39 | 64.66 | 68.17 | 68.51 |
| Parte 2 | 61.06 | 55.53 | 62.81 | 64.49 | 65.66 |
| Total | 60.22 | 54.77 | 61.89 | 63.82 | 64.57 |
| Pvalor 1ª y 2º | 0.04183 | 0.01614 | 0.34711 | 0.05709 | 0.13836 |

Los resultados del Ministerio y los obtenidos en nuestra investigación fueron comparados y analizados estadísticamente y mostraron que las diferencias significativas se limitan a la precisión y fluidez, y también que no hubo significancia en los resultados chi-cuadrado inter-grupo para los otros criterios, es decir, interacción y coherencia. Se observaron altos puntajes (3) en los tres grupos en nuestra investigación, pero en este caso las notas bajas fueron incluso menores que en el criterio anterior. Es

especialmente significativo el caso de los estudiantes de Educación, los cuales mostraron un menor porcentaje de notas bajas (1) pero que apenas aumentaron las notas altas (3). Esto también fue apoyado por el valor Tau-b de Kendall (-.092, error standard 0.71).

TABLA XI. Resultados de la actuación global de los participantes en este estudio. El número de participantes aparece en la primera línea. La media y la desviación típica se muestran en la tercera y cuarta línea respectivamente.

| | Precisión | Fluidez | Interacción | Coherencia |
|--------------------------|------------------|----------------|--------------------|-------------------|
| Válidos | 169 | 167 | 169 | 168 |
| Perdidos | 0 | 2 | 0 | 1 |
| Media | 1,50 | 1,89 | 1,79 | 1,83 |
| Desviación típica | ,647 | ,712 | ,731 | ,709 |

Aunque es posible dar varias interpretaciones a esta falta de diferencias significativas entre el estudio llevado a cabo por el MECD y nuestro estudio en lo que se refiere a la interacción y a la coherencia, una de las más contundentes podría ser la forma de aplicación del test. Las entrevistas unipersonales cara a cara producen una mayor ansiedad (Woodrow, 2006; Hewitt&Stephenson, 2011) y esto podría tener un efecto especial en cómo los profesores enfocan el examen. Mientras el MECD usó una metodología altamente cognitiva con una limitada interacción entre el examinando y el entrevistador, la actitud facilitadora de los entrevistadores en nuestro caso llevan a cambios significativos, lo que puede ser observado en la tabla más abajo (TABLA XIII).

TABLA XII. Frecuencias para cada criterio evaluado y sus correspondientes valores dentro de los cuatro puntajes posibles (de 0 a 3).

| Puntaje | Precisión | | Fluidez | | Interacción | | Coherencia | |
|--|------------------------------------|-------|---------------------------------|-------|-----------------------------------|-------|-----------------------------------|-------|
| | 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 Aprobados | 75 | 44.3% | 117 | 59.2% | 107 | 63.4% | 109 | 65.7% |
| X² P ¿Significativo 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) | |

TABLA XIII. Comparación de la metodología de las tareas. La primera columna refleja los aspectos comparados en los dos estudios. La segunda columna muestra la metodología utilizada por el Ministerio de Educación. La tercera columna muestra la desarrollada en el presente estudio.

| Aspectos observados | MECD (enfoque cognitivista) | OPENPAU (enfoque interactivo) |
|--|--|---|
| Competitividad | Activa | Más pasiva, la atmósfera es normalmente más relajada |
| Cooperación – Tareas de expresión oral (<i>Speaking</i>) | Clarificación, preguntas, interés en la aplicación | Completar, aclarar, preguntas productivas, interés en el significado |
| Discurso co-construido | Tiende a ser co-construido por dos monólogos, poco interés en el “otro” | El diálogo tiende a ser implicativo incluso con candidatos de distintos niveles de competencia lingüística |
| Actitudes | La cooperación es limitada. En su lugar, los estudiantes se implican en un discurso/estrategias limitadamente realista/s | La cooperación es fundamental y realista. Se construye la vida diaria real a través del lenguaje, el cuerpo y el contexto |
| Factores individuales que influyen | | Extroversión, cooperación, voluntad de colaborar |
| Dinámica de grupo | Pares, tríos pero tienden a atenerse a las intervenciones la I. Incluso agrupación por nivel de competencia | Interactúan más libremente, el turno de palabra tiende a variar; pueden tener diferentes niveles de competencia |
| Papel del examinador/entrevistador | “Oyente”, prácticamente pasivo*, participante no activo, no corrige | Mediador; activo; facilitador, muy activo, modera para producir formas adecuadas |
| Puntaje | Según los objetivos | Según la productividad |

Nuestra metodología interactiva tuvo un efecto en la fluidez y precisión de los estudiantes. En este sentido, un enfoque interactivo con un entrevistador activo beneficia a los estudiantes. No obstante, estos resultados no son conclusivos y sería necesaria una investigación ulterior.

Discusión

La primera observación interesante es que en nuestro estudio la mayoría de los estudiantes (55,6%) estaban en la banda de competencia A1-2 a pesar de sus muchos años de aprendizaje del inglés. De hecho, en la mayoría de los casos su edad de comienzo del aprendizaje del inglés fue a los 12 años. Estos resultados pueden indicar serias deficiencias, ya sea en los procesos de enseñanza-aprendizaje, en la elección de contenidos, en los materiales de enseñanza, en la elección de metodología, en la ratio de estudiantes por clase o en una combinación de todos estos factores. Otro tema que necesita considerarse es el de los requisitos de nivel de competencia en un idioma extranjero para graduarse en el Bachillerato (mayoritariamente B1 excepto en Andalucía (A2+) y Cataluña (B2)). Aunque investigadores como Gómez Rodríguez (2010) han insistido que los libros de texto no conducen al desarrollo de la competencia comunicativa, los resultados en el estudio muestran que los estudiantes actúan mejor en los criterios interaccionales que en la precisión. Esto puede indicar que las entrevistas en parejas conlleva puntajes más altos, lo que concuerda con los estudios actuales de agrupación en parejas (Nakasuhara, 2013) en los cuales los estudiantes con competencia más baja son emparejados con estudiantes de competencia más alta. El hecho de que no haya diferencias significativas entre los grupos en ninguno de los criterios sugiere que las tareas son válidas para estos grupos.

Como se mencionó en el apartado anterior, fue especialmente interesante observar que no hubo diferencias significativas cuando se comparó este estudio con el desarrollado por el MECD en interacción y coherencia. Si bien sería deseable una ulterior investigación para justificar estas diferencias mínimas, puede bien significar que algunos aspectos pueden ser mejor desarrollados en el aula, que los estudiantes tienden a concentrarse en aspectos específicos de la comunicación o que las percepciones de los examinadores hacia estos criterios pueden ser similares. Estas altas correlaciones evidencian la contundencia de los resultados obtenidos en el presente estudio.

Conclusiones

Este artículo afrontó tres temas principales: 1) La competencia de los estudiantes universitarios de primer curso según el CEFR; 2) Las

diferencias en los criterios de evaluación en diferentes grados universitarios y 3) La comparación con el informe previamente realizado por el MECD. Los resultados indican que hay una relación entre los resultados de esta investigación y los del MECD. Uno de los aspectos más destacados de la investigación es que hay una ligera diferencia entre el nivel de competencia alcanzado al final de 4º de la ESO y 2º de Bachillerato y que la mayoría de los estudiantes no alcanza niveles más allá de un B1. Otra conclusión es que si la comunicación es un objetivo final en la LOMCE, debería darse más importancia a los exámenes encaminados especialmente a obtener inferencias sobre el grado de interacción oral (García Laborda, 2010a).

Hay un gran desnivel entre la competencia con que los estudiantes llegan a la universidad y la competencia exigida por la mayoría de universidades para graduarse (B2). Esto hará, obviamente, que las universidades aumenten el número de cursos de inglés (o idioma extranjero) si este requisito debe ser cumplido. Se ha demostrado que la forma de evaluar es una manera valiosa de cambiar la educación (Efecto Washback, Muñoz y Álvarez, 2010) y por lo tanto se necesitan implementar nuevos exámenes y formas de evaluación en el sistema educativo español. Quizás los profesores, administrativos y examinadores necesiten también reconfigurar su concepto de evaluación de idiomas y evolucionar hacia enfoques más interactivos. Los exámenes de expresión oral (*speaking*) realizados en parejas no sólo favorecen la actuación, la cual puede estar basada en el hecho de que quizás por el uso de entrevistas interactivas, alguna función comunicativa que los estudiantes más débiles no pueden mostrar en una interacción cara a cara debido al estrés y ansiedad, se active por la presencia de un igual, el otro candidato (Horwitz, 2000; McCarthy & O’Keeffe, 2004).

Aún queda camino por recorrer antes de que el Examen Final de Secundaria se implemente en 2017 pero quizás este humilde estudio atraiga la atención de políticos educativos hacia otras investigaciones posteriores, que deberían enfocarse, primero, en las posibilidades de agrupamiento; segundo, en cómo maximizando recursos aumentarían los niveles de competencia y tercero, comprobar si la interacción social puede tener un efecto en la mejora de actuación en los exámenes orales. Debería igualmente afrontar la mayor limitación de este artículo, que es el tamaño de la muestra, y la necesidad de comprobar los resultados usando un corpus o metodologías pragmáticas para conseguir un constructo sólido que pueda ser validado eficazmente tanto interna como externamente.

Reconocimientos

Los autores desean expresar su gratitud al Ministerio de Ciencia e Innovación por haber provisto los fondos para esta investigación a través del proyecto OPENPAU (Orientación, Propuestas y Enseñanza para la Sección de Inglés en la Prueba de Acceso a la Universidad; FFI2011-22442) gracias a la co-financiación del Ministerio de Economía e Innovación (MICINN) y los fondos ERDF de la Unión Europea.

Referencias

- 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 exigibles en 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 Psicopedagogía* 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: PeterLang.
- 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).

Dirección de contacto: 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