



Educación XX1

ISSN: 1139-613X

ISSN: 2174-5374

educacionxx1@edu.uned.es

Universidad Nacional de Educación a Distancia
España

Ruiz-Lázaro, Judit; González Barbera, Coral; Gaviria Soto, José-Luis
LAS PRUEBAS DE INGLÉS PARA ACCEDER A LA UNIVERSIDAD.
UNA COMPARACIÓN ENTRE COMUNIDADES AUTÓNOMAS 1
Educación XX1, vol. 24, núm. 1, 2021, -Junio, pp. 233-270
Universidad Nacional de Educación a Distancia
España

DOI: <https://doi.org/10.5944/educXX1.26746>

Disponible en: <https://www.redalyc.org/articulo.oa?id=70666127010>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UNED 

Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

10

LAS PRUEBAS DE INGLÉS PARA ACCEDER A LA UNIVERSIDAD. UNA COMPARACIÓN ENTRE COMUNIDADES AUTÓNOMAS ¹

(THE COLLEGE ENTRANCE EXAMS. A COMPARISON BETWEEN AUTONOMOUS COMMUNITIES)

Judit Ruiz-Lázaro
Coral González Barbera
José-Luis Gaviria Soto
Universidad Complutense de Madrid

DOI: 10.5944/educXX1.26746

Cómo referenciar este artículo/How to reference this article:

Ruiz-Lázaro, J., González Barbera, C. y Gaviria Soto, J.L. (2021). Las pruebas de inglés para acceder a la Universidad. Una comparación entre Comunidades Autónomas. *Educación XX1*, 24(1), 233-270, <http://doi.org/10.5944/educXX1.26746>

Ruiz-Lázaro, J., González Barbera, C. & Gaviria Soto, J.L. (2021). The college entrance exams. A comparison between Autonomous Communities. *Educación XX1*, 24(1), 233-270, <http://doi.org/10.5944/educXX1.26746>

RESUMEN

El sistema de acceso a la universidad en España es un tema social de interés. Este estudio pretende comparar las pruebas de Lengua Extranjera-Inglés de las 17 comunidades autónomas que dan acceso a la universidad en España. La finalidad es reflexionar sobre la posible influencia de las diferencias en el diseño de las pruebas sobre el rendimiento promedio obtenido por los examinados en función de la comunidad autónoma a la que pertenecen. Para ello, se plantea un diseño no experimental y transversal de carácter descriptivo y comparado. Por un lado, se estudia la estructura de 68 exámenes, correspondientes a los modelos A y B de las convocatorias ordinaria y extraordinaria de 2017, mediante un análisis descriptivo-analítico basado en una metodología comparada adaptando el enfoque

planteado por García-Garrido (1991). Por otro, se recoge información de las calificaciones promedio obtenidas por los aspirantes en las pruebas y se analiza su posible relación con la estructura de las pruebas por comunidades autónomas, gracias a los datos publicados anualmente por el Sistema Integrado de Información Universitaria. Los resultados muestran diferencias en la evaluación de las destrezas lingüísticas. Por ejemplo, el *Speaking* no se evalúa en ninguna comunidad autónoma y el *Listening*, únicamente, en Galicia y Cataluña. Asimismo, hay disparidades en el peso de cada bloque de contenidos dentro de la prueba, en el formato de los ítems y en la nota promedio obtenida por los aspirantes en cada una de las comunidades autónomas. Se concluye que no puede descartarse que el actual sistema de evaluación de la competencia lingüística en Lengua Extranjera - Inglés esté produciendo importantes desequilibrios e inequidades asociadas a la comunidad autónoma de procedencia.

PALABRAS CLAVE

Rendimiento, pruebas externas, acceso a la universidad, segunda lengua, lengua inglesa

ABSTRACT

The system university access in Spain is a social issue of interest. This study aims to compare the Foreign Language-English tests of the 17 Autonomous Communities that give access to university in Spain. The purpose is to reflect on the possible influence of the differences in the design of the tests on the average performance obtained by those examined, depending on the Autonomous Community to which they belong. To this end, a non-experimental and transversal design of a descriptive and comparative nature is proposed. On the one hand, the structure of 68 examinations, corresponding to models A and B of the 2017 initial and secondary rounds, is studied by means of a descriptive-analytical analysis based on a comparative methodology adapting the approach proposed by García-Garrido (1991). On the other hand, information is collected on the average scores obtained by the candidates in the tests, and their possible relationship with the structure of the tests by Autonomous Community is analysed, thanks to the data published annually by the Integrated University Information System. The results show differences in the assessment of language skills. For example, Speaking is not assessed in any autonomous community and Listening is only assessed in Galicia and Catalonia. Likewise, there are disparities in the weight of each content block within the test, in the format of the items and in the average mark obtained by the candidates in each of the Autonomous

Communities. It is concluded that it cannot be ruled out that the current system of assessing linguistic competence in Foreign Languages-English is producing important imbalances and inequalities associated with the Autonomous Community of origin.

KEYWORDS

Achievement, external testing, access to university, second language, English language

INTRODUCCIÓN

La Evaluación para el Acceso a la Universidad (EvAU) se implantó en el curso académico 2016-2017. Inicialmente se diseñó con la finalidad de ser una evaluación de fin de etapa de Bachillerato (Real Decreto 310/2016). Sin embargo, hasta la entrada en vigor de la normativa resultante del Pacto de Estado Social y Político por la Educación, estas pruebas no tienen ningún efecto para la obtención del título de Bachillerato. Por lo tanto, sus características, por el momento, son similares a las de la anterior Prueba de Acceso a la Universidad (PAU) (Real Decreto-ley 5/2016).

Aunque en España existe un marco normativo común para la elaboración de las pruebas para el acceso a la universidad, y aunque en principio el diseño de las pruebas es competencia del Estado, la realidad es que la misma legislación concede autonomía a cada comunidad para su elaboración (Real Decreto 310/2016; Real Decreto-ley 5/2016). Por lo tanto, aunque las comunidades autónomas tienen competencia para desarrollar sus propias pruebas, parecería razonable que dichas pruebas abordasen un mismo contenido, planteasen los mismos tipos de preguntas, utilizarasen el mismo sistema de corrección y el mismo baremo. La realidad es que, actualmente, existen 17 pruebas con diferencias en la estructura de algunos de sus componentes y en el tipo de ítems. Esto permite sospechar que, a pesar de que el objetivo perseguido es ‘homogeneizar y estandarizar’ las calificaciones de los estudiantes, en la práctica no pueda garantizarse que se estén evaluando las mismas habilidades (Ruiz-Lázaro y González, 2017).

La lengua extranjera es una de las tres materias comunes que todos los alumnos deben realizar dentro del conjunto de pruebas para acceder a la universidad en España. Según los datos publicados anualmente por el Sistema Integrado de Información Universitaria, en los últimos cinco años más del 97% de los aspirantes elijen inglés como lengua extranjera, tanto en las convocatorias ordinarias como en las extraordinarias (SIIU, 2017).

Antecedentes

La evaluación de la lengua extranjera empieza a ser considerada de importancia en los años 60, en los que se desarrollan modelos tanto teóricos como empíricos. Un primer modelo, que sirvió de base para evaluaciones a gran escala en Estados Unidos, fue definido por Carroll (1961) y Lado (1961). En este modelo se diferenciaban, por un lado, las habilidades o destrezas como la comprensión auditiva, la lectura, la escrita y la producción oral y, por otro, los elementos como la pronunciación, estructura gramatical, vocabulario y aspectos culturales. Desde entonces, han surgido otros modelos: los que apoyan la competencia lingüística extranjera como una habilidad unitaria, a partir de 1970; los defensores de pruebas directas o evaluación del desempeño, a partir de 1980; el rendimiento basado en tareas, o el rendimiento basado en competencias (a partir del 2000 y hasta la actualidad). En este contexto, se han establecido pautas o estándares de evaluación internacionales en los que sustentar este tipo de evaluaciones educativas de una segunda lengua o L2. Entre ellas, destacan las elaboradas por los miembros de ALTE (*Association of Language Testers in Europe* en 1994), las diseñadas por la JLTA (*Japan Language Testing Association* en 2002) y las propuestas por EALTA (*European Association for Language Testing and Assessment* en 2006).

Existen numerosas investigaciones que se centran tanto en la importancia de evaluar el rendimiento en el uso de una lengua extranjera, como en los métodos más adecuados para hacerlo. Por ejemplo, algunos estudios internacionales como los de Lee y Greene (2015), Harrington y Roche (2014) y Carlsen (2018) han mostrado evidencia de la correlación positiva y significativa entre las puntuaciones en la prueba de inglés para acceder a la universidad y el dominio y éxito académico futuro en diferentes países, lo que convierte el análisis y posible mejora de las pruebas de inglés para acceder a la universidad en una cuestión relevante.

En cuanto a la metodología de evaluación, a raíz de los cambios producidos en el modo de acceso a la educación superior en España, el Ministerio de Educación, Cultura y Deporte publicó en 2012 un monográfico sobre “Nuevas perspectivas en la sección de idiomas de la Prueba de Acceso a la Universidad” donde se propone, entre otras cosas, una metodología para evaluar los aspectos orales en la prueba de inglés que da acceso a la universidad, en ocasiones a través de un instrumento informatizado (Amengual-Pizarro y Méndez, 2012; Bueno y Luque, 2012; Díez-Bedmar, 2012; García-Laborda, 2012).

Asimismo, el tema en el que se centra el presente artículo ha sido abordado en ocasiones en la literatura científica nacional. Por ejemplo, entre

las investigaciones que comparan las pruebas diseñadas por comunidades autónomas en materia de lengua extranjera inglesa, destaca la realizada, hace más de dos décadas, por Muñoz-Repiso et al. (1997). En ella se estudiaban las pruebas de acceso a la universidad en España, entre ellas las de inglés. Posteriormente, Martín-Martín (2012) analizó el peso que tenía el conocimiento léxico en la prueba de inglés de cada comunidad autónoma y se mostraba la tendencia a dejar de evaluar esta sección al implantar actividades orales en los exámenes.

Contexto

En este artículo se aborda esta misma cuestión asumiendo los principales resultados obtenidos en los estudios anteriores. Por una parte, damos por supuesta la importancia de la competencia lingüística en lengua extranjera, fundamentalmente en inglés, no solo por su valor como vehículo de información y modo de acceso a los resultados más relevantes de investigación en la universidad, sino por su demostrada capacidad como predictor de éxito universitario. Por otra parte, nos preocupa la posibilidad de que en este ámbito la evaluación en la EvAU no alcance los estándares que por normativa se establecieron y que no cumpla las funciones para las que se diseñó.

El contexto en el que se evalúa esta competencia lingüística para acceder a la universidad en España empezó a modificarse con la introducción de la PAU en el curso académico 2008-2009. Los cambios más notorios fueron dos: por un lado, su adaptación al Marco Común Europeo de Referencia para las Lenguas, cuyas directrices se basan en las publicadas por la ALTE (MCER, Consejo de Europa, 2001) y, por otro, su adecuación al currículo que regula el contenido en la etapa de bachillerato.

Con respecto al primer punto, es en el MCER donde se establecen las competencias que debe cubrir el aprendizaje de esta área. En él se definen 3 niveles y 6 subniveles: usuario básico (A1, A2), usuario independiente (B1, B2) y usuario competente (C1, C2) (Consejo de Europa, 2001). En España, de acuerdo con García-Laborda (2012), el nivel que mide la prueba de inglés que da acceso a la universidad varía entre un B1-1 y un B1-2 en función de la comunidad autónoma en la que se diseñe. El MCER establece que un hablante de este nivel tiene que ser capaz de:

Utilizar el idioma con cierta seguridad y flexibilidad, receptiva y productivamente, tanto en forma hablada como escrita, así como para mediar entre hablantes de distintas lenguas, en situaciones cotidianas y menos corrientes que requieran comprender y producir textos de cierta extensión y

complejidad lingüística, en una variedad de lengua estándar, con estructuras habituales y un repertorio léxico común no muy idiomático, y que versen sobre temas generales, cotidianos o en los que se tiene un interés personal. (Consejo de Europa, 2001, p. 26).

Para la determinación del nivel, el MCER establece cuatro destrezas lingüísticas primarias: comprensión lectora (*Reading*), comprensión auditiva (*Listening*), expresión e interacción oral (*Speaking*) y expresión escrita (*Writing*). Esto se refleja en la normativa que regula las pruebas de acceso a la universidad (Real Decreto 1892/2008, Real Decreto 1105/2014) en la que se incorporan ejercicios de comprensión y expresión oral.

Por su parte, el currículo que regula los contenidos en la etapa de Bachillerato, el otro referente de las pruebas, indica que en él se “incorpora el enfoque orientado a la acción recogido en el Marco Común Europeo de Referencia para las Lenguas” (Real Decreto 1105/2014, 422) y, por tanto, los organiza en cuatro bloques: *Reading*, *Listening*, *Speaking* y *Writing*. Estos bloques deberían tener un peso propio en las pruebas de Lengua Extranjera-Inglés para acceder a la universidad, tal y como se establece en la matriz de especificaciones regulada por la Orden ECD/1941/2016 (Tabla 1).

Tabla 1

Contenidos evaluables de Lengua Extranjera-Inglés de 2º de Bachillerato en la EvAU

Módulos	Bloques de contenidos establecidos en el currículo	Agrupación y peso (%) en la EvAU*
Módulo I. Comprensión de textos orales y escritos	Bloque 1. Comprensión de textos escritos (<i>Reading</i>)	Agrupación de contenidos I 60%
	Bloque 2. Comprensión de textos orales (<i>Listening</i>)	
Módulo II. Producción de textos orales y escritos: expresión e interacción	Bloque 3. Producción de textos orales (<i>Speaking</i>)	Agrupación de contenidos II 40%
	Bloque 4. Producción de textos escritos (<i>Writing</i>)	

*Este peso viene determinado en la matriz de especificaciones de acuerdo con la Orden ECD/1941/2016 y el Real Decreto 310/2016.

Fuente: elaboración propia a partir del Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato

Además de la referencia al MCER, el currículo menciona los contenidos sintáctico-discursivos que deben enseñarse en Bachillerato y que, por lo

tanto, podrán ser objeto de evaluación en la prueba de Lengua Extranjera-Inglés de la EvAU.

Una primera inconsistencia entre la EvAU y la normativa que la regula puede observarse en la tabla 1. En ella los contenidos sintáctico-discursivos no tienen peso asignado en la matriz de especificaciones ni se reflejan entre los estándares de aprendizaje evaluables para el diseño de la prueba que da acceso a la universidad.

Por otra parte, las calificaciones promedio obtenidas en las pruebas para acceder a la universidad, publicadas anualmente por el SIIU, nos permiten ver diferencias sustanciales en función de la comunidad autónoma. Los promedios de las pruebas de Inglés de 2017 varían más de un punto y medio entre el País Vasco, Cantabria o La Rioja con 7.44, 7.4 y 7.3, respectivamente, y Cataluña, Asturias y Baleares con una puntuación promedio de 5.92, 5.86 y 5.85, respectivamente. Por lo tanto, cabe preguntarse si el diseño de las pruebas podría ser un factor explicativo de estas diferencias.

Estas inconsistencias permiten preguntarse si los otros cuatro bloques de contenidos, o destrezas lingüísticas, son evaluados en todas las comunidades autónomas y de qué manera. Aunque, a grandes rasgos, todas las administraciones implicadas hacen referencia a estos componentes, varios autores (Martín-Martín, 2012; Suárez-Álvarez et al., 2014) señalan la necesidad de analizar las pruebas de acceso a la universidad para determinar si responden a las directrices MCER.

Las inconsistencias mencionadas, y alguna otra que señalaremos más adelante, podrían estar generando un impacto potencial en el futuro académico de los estudiantes examinados debido a dos circunstancias. Se trata, por una parte, de la tasa de ocupación² en las universidades, cercana al 89.8% (Sistema Integrado de Información Universitaria, 2019). Por otra, de la posibilidad de que los aspirantes puedan elegir la comunidad autónoma y, por consiguiente, la universidad en la que cursar sus estudios. En consecuencia, estas dos circunstancias hacen que el acceso a la universidad, especialmente a las carreras profesionales y académicas más demandadas, sea un recurso público y escaso. Por ello, debería ser una exigencia que la estructura de las pruebas de las distintas comunidades autónomas fuese similar.

Aunque en la Constitución Española no se mencionan expresamente los principios que deben regular el acceso a la universidad, se podrían aplicar, por analogía, los principios reguladores del acceso a la función pública. Tal y como ocurre en este último caso, las plazas universitarias son un bien público y escaso en el que hay más aspirantes a disfrutarlo que posibilidades de satisfacción. En efecto, en el artículo 103.3 de la

Constitución Española, contemplado en el título IV sobre el Gobierno y la Administración, se establece que:

La ley regulará el estatuto de los funcionarios públicos, el acceso a la función pública de acuerdo con los principios de mérito y capacidad, las peculiaridades del ejercicio de su derecho a sindicación, el sistema de incompatibilidades y las garantías para la imparcialidad en el ejercicio de sus funciones. (art. 103, CE 1978)

Desde el punto de vista de la legalidad vigente, estos principios deberían aplicarse por analogía al acceso a la universidad. Todos los estudiantes deberían tener los mismos derechos y las mismas oportunidades de acceso, respetándose los principios de igualdad, mérito y capacidad. Estos principios excluyen la posibilidad de que dos estudiantes con el mismo nivel de conocimientos, méritos y capacidades tengan diferente probabilidad de acceder a la universidad por presentarse en comunidades autónomas diferentes.

En consecuencia, este estudio considera el principio de igualdad como aquel que reconoce que todos los estudiantes deben ser tratados de la misma forma por la ley y por la aplicación práctica de la misma, considerando sus méritos y sus capacidades personales, independientemente de cuáles sean sus circunstancias o su procedencia geográfica. Dicho en otras palabras, en igualdad de condiciones todos deben tener la misma probabilidad de acceso a la universidad.

Por todo ello, el propósito general de este trabajo es determinar si puede establecerse una relación entre el rendimiento promedio obtenido por los examinados de las diferentes comunidades autónomas y el diseño de las pruebas de Lengua Extranjera-Inglés con las que son evaluados. El rendimiento se operacionaliza a través de la medida del resultado promedio obtenido por los estudiantes de una misma comunidad autónoma en la prueba de Lengua Extranjera-Inglés que da acceso a la universidad. En cuanto al diseño de la prueba se operacionaliza a través de la descripción pormenorizada de su estructura, enumeraciones de las destrezas lingüísticas que pretende medir, número y tipo de preguntas que utiliza para evaluarlas y puntuación que le otorga a cada una de ellas.

Se destacan 3 objetivos específicos en este estudio:

- Determinar si las pruebas de Lengua Extranjera-Inglés que se diseñan para la EvAU en las distintas comunidades autónomas están en consonancia con las directrices que establece el currículo de Bachillerato y con el MCER.

- Determinar si las pruebas de Lengua Extranjera-Inglés que se diseñan para la EvAU en las distintas comunidades autónomas tienen un contenido, una forma de corrección y puntuación similares y comparables.
- Describir y ordenar los resultados de rendimiento promedio obtenido por los aspirantes en las pruebas y relacionar estos rendimientos con las diferencias de diseño observadas.

MÉTODO

Se trata de un diseño no experimental y transversal de carácter descriptivo y comparado, que incluye un análisis de datos secundarios o *secondary research*.

Para abordar el primer objetivo específico, se estudió la estructura de las pruebas de 17 comunidades autónomas mediante el análisis descriptivo-analítico de 68 exámenes, correspondientes a las convocatorias ordinaria y extraordinaria, opción A y B, de 2017. Se utilizó una metodología comparada en la que se realiza una adaptación del enfoque planteado por García-Garrido (1991). Las dos etapas principales establecidas por ese autor son, la fase analítica (o descriptiva) y la fase sintética (o comparativa).

- En la fase descriptiva o analítica se recopilaban los datos necesarios de fuentes primarias según cuatro criterios de selección rigurosos: ámbito del objeto de estudio, contexto, tiempo y espacio.

Tabla 2
Fuentes de información y documentación analizada

Fuentes primarias		Información analizada
Ministerio de Educación	Legislación	Reales Decretos Decretos Órdenes Resoluciones
Organismos nacionales	Universidades	Pruebas de Lengua Extranjera-Inglés para el acceso a la universidad
Organismos internacionales	Marco Común Europeo de Referencias para las Lenguas (MCER)	Association of Language Testers in Europe (ALTE). Modelos de pruebas de: <ul style="list-style-type: none"> • Inglés General • ACLES • Trinity
	Pruebas estandarizadas	<ul style="list-style-type: none"> • IELTS • TOEFL

Fuente: elaboración propia

Se realizó una revisión del marco legal vigente que regula el acceso a la universidad en España, tanto a nivel nacional como autonómico. Seguidamente se compilaron las pruebas de Inglés que se aplicaron en el año 2017, en las convocatorias ordinaria y extraordinaria, y se examinaron, clasificaron y evaluaron los datos referidos al contenido.

En la fase comparativa, o sintética, se establecieron tres niveles de concreción (dimensión, parámetro e indicador) que permitieron delimitar el objeto de estudio (Tabla 3).

Tabla 3
Sistema de unidades de comparación

Dimensión	Parámetro	Indicador
ESTRUCTURA GENERAL	Bloque: Comprensión de textos escritos. <i>Reading</i>	Nº de palabras del texto Optatividad Nº de ítems Tipo de tareas a realizar Puntuación otorgada a cada tarea sobre el total (sobre 10 puntos) Puntuación esperada por azar
	Bloque: Comprensión de textos orales. <i>Listening</i>	Nº de ítems Tipo de tareas a realizar Puntuación otorgada a cada tarea sobre el total (sobre 10 puntos) Existencia de penalización en los errores Puntuación esperada por azar
	Bloque: Producción de textos escritos. <i>Writing</i>	Nº de ítems Tipo de tareas a realizar Tipo de evaluación Nº de palabras requeridas Optatividad Puntuación otorgada a cada tarea sobre el total (sobre 10 puntos)
	Bloque: Producción de textos orales. <i>Speaking</i>	-
	Bloque: Contenidos sintáctico-discursivos	Nº de ítems Tipo de tareas a realizar Optatividad Puntuación otorgada a cada tarea sobre el total (sobre 10 puntos)

Dimensión	Parámetro	Indicador
EVALUACIÓN	Ponderación (%) de cada bloque de contenidos sobre la puntuación total	% Reading % Listening % Speaking % Writing % Contenidos sintáctico-discursivos

Fuente: elaboración propia

Posteriormente, se diseñó una pauta de registro *ad hoc* para organizar la información. Con el programa ATLAS.ti 8.0 se codificó y categorizó la información y se comparó con el peso que tiene cada uno de los bloques de contenidos en cada una de las pruebas. En cada prueba se analizaron los ítems y se clasificaron en función de su formato (número, tipología, carácter obligatorio u optativo y puntuación).

Como documentación de referencia para valorar los ítems en cada versión de la EvAU se hizo uso de varias fuentes. Por una parte, la guía del usuario para examinadores (ALTE, 2005) y, por otro lado, el análisis de pruebas internacionales estandarizadas (Inglés General, ACLES, Trinity, IELTS y TOEFL).

Para el segundo objetivo específico se hizo uso de los datos publicados anualmente por el Sistema Integrado de Información Universitaria. Se recogió información de las calificaciones promedio obtenidas por los aspirantes en la prueba en función de la comunidad autónoma en la que la realizaron.

Finalmente, se analizó la posible relación entre el rendimiento promedio y la estructura de las pruebas comparando por comunidades autónomas.

RESULTADOS

Los resultados se presentan en dos secciones: por un lado, los referentes a la estructura de las pruebas y su adecuación al MCER y, por otro, los que aluden al rendimiento promedio obtenido por los aspirantes en las mismas y su posible relación con la estructura de las pruebas.

Resultados referentes a la estructura de las pruebas

En un primer punto los resultados se presentan agrupados en cinco categorías: comprensión de textos escritos, comprensión de textos orales, producción de textos escritos, producción de textos orales y contenidos sintáctico-discursivos.

Seguidamente, se presenta la ponderación de cada una de las agrupaciones de contenidos sobre la puntuación total de la prueba en cada una de las comunidades.

Por último, se valora si las pruebas se adecúan a las directrices establecidas por el MCER.

Bloque: Comprensión de textos escritos (*Reading*)

Este bloque está orientado a evaluar la comprensión global de un texto dado. El usuario con un nivel B1 debe “leer textos sencillos sobre hechos concretos que tratan sobre temas relacionados con su especialidad con un nivel de comprensión satisfactorio” (Consejo de Europa, 2001, p. 72).

Según la normativa el texto que evalúa la comprensión lectora puede ser de índole informativa, argumentativa, descriptiva o narrativa, e ir acompañado por algún tipo de apoyo visual, a través de multi-textos, sobre noticias reales, textos inferenciales, estímulos icónicos, etc. Sin embargo, la realidad es que en todas las comunidades las pruebas evalúan un solo tipo de texto y sin apoyo visual. Según la ALTE, los textos deben ser auténticos, escritos por nativos y proceder de periódicos o revistas. Se trata de evaluar el uso de la lengua en la vida real, y no el conocimiento de la lengua como estructura. Estos textos podrán adaptarse para que los aspirantes comprendan los elementos sociales y culturales que están implícitos. Las pruebas de Baleares, Canarias, Cantabria, Castilla-La Mancha, Cataluña, La Rioja y la Comunidad Valenciana especifican en la misma prueba de dónde ha sido extraído y adaptado el texto. En todos los casos se trata de textos procedentes de medios de comunicación.

La **longitud del texto** cambia de unas a otras comunidades. En doce comunidades el texto tiene entre 200 y 300 palabras mientras que, en Canarias, Baleares, Navarra y Cantabria el texto se sitúa en una horquilla entre 300 y 400. En el caso de Cataluña, se presenta un texto de una mayor longitud con alrededor de 500 palabras. A pesar de que el MCER no establece un número determinado de palabras para cada nivel, los exámenes de Inglés General y ACLES para el nivel B1 se aproximan a las 200 palabras, mientras

que el Trinity College ISE I presenta en torno a las 400 y IELTS y TOEFL-ibt unas 700. Hay que indicar que estos últimos están diseñados para medir diferentes niveles a partir del mismo texto.

En cuanto a la **optatividad**, las preguntas del *Reading* son obligatorias en todas las comunidades excepto en Canarias, que incluye ítems optativos.

La **puntuación** asignada a este bloque por comunidades autónomas varía entre 3 y 7 puntos, tal y como se puede observar en la columna “puntuación total” de la tabla 4.

En las dos últimas columnas de la misma tabla se ha incluido el cálculo de la puntuación esperada al contestar por azar, por cada tarea y en el total de la prueba. Ese cálculo se corresponde con lo siguiente:

Si hacemos N = número de ítems, n = número de opciones de respuesta, u_{ij} es la puntuación obtenida por la respuesta i (correcta o incorrecta) en el ítem j , cuando no se penalizan los errores, $u_{0i} = 0$ con lo que, si n fuese 4, tendríamos:

$$\varepsilon(U) = \sum_{j=1}^N \sum_{i=0}^1 u_{ij} P(u_{ij}) = \sum_{j=1}^N \left(u_{0i} * \frac{n-1}{n} + u_{1i} * \frac{1}{n} \right) = u_{11} * \frac{1}{n} + u_{12} * \frac{1}{n} + \dots + u_{1N} * \frac{1}{n} = N/n = N/4$$

En Asturias y en Galicia, debido al tipo de preguntas que se plantean en sus pruebas, la probabilidad de contestar bien al azar es inapreciable. En Cataluña, única comunidad donde se penalizan los errores, la probabilidad de responder bien al azar queda neutralizada por el algoritmo de calificación. En el resto de las comunidades autónomas, sobre un total de 10 puntos, la puntuación esperada al responder por azar varía desde .375 en Murcia hasta 1.916 en Cantabria. Desde cualquier punto de vista se trata de una diferencia excesivamente grande.

En la Tabla 4 se recoge la información más relevante, referida a los tipos de ítems empleados y los tipos de tareas que se piden.

Tabla 4

Análisis del bloque: "Comprensión de textos escritos. Reading". Año 2017

Comunidad autónoma	Nº de palabras del texto	Optatividad	Nº de ítems	Tipo de tarea	Puntuación		Puntuación esperada por azar	
					Por tarea	Total	Por tarea	Total
Andalucía	250-280	No	10	2	1		.25	
				4	2	4	1	1,25
				4	1		-	
Aragón	200-250	No	8	2	2		1	
				2	2	5	-	1
				4	1		-	
Asturias	300	No	4	4	PA	4	4	- 0
Balears	350-400	No	7	2	VF	1	.5	
				1	PA	1	-	.5
				4	LV	1	-	
Canarias	320-345	Sí	6	3	VF	1.5	.75	
				3	LV	1.5	-	.75
Cantabria	350-400	No	14	5	VF	2.5	1.25	
				5	LV	1.5	-	1.916
				4	EM	2	.666	
Castilla-La Mancha	250-280	No	8	4	VF	2	1	
				4	LV	1	-	1
Castilla y León	200	No	9	1	PA	2	-	
				3	VF	1.5	.75	
				1	EM	0.5	.1665	.9165
				5	LV	2	-	
Cataluña	500	No	8	8	EM	4	4	0 0
Comunidad Valenciana	200-250	No	12	2	PA	2	-	
				3	VF	1.5	.75	
				4	LV	1	-	1.25
				3	EM	1.5	.5	
Extremadura	250-280	No	5	2	VF	2	1	
				3	LV	1.5	-	1
Galicia	300	No	7	1	RS	1	-	
				4	LV	1	-	0
				2	PA	2	-	
Madrid	300	No	8	2	VF	2	1	
				2	PA	2	-	1
				4	LV	1	-	

Comunidad autónoma	Nº de palabras del texto	Optatividad	Nº de ítems	Tipo de tarea	Puntuación		Puntuación esperada por azar	
					Por tarea	Total	Por tarea	Total
Murcia	250-280	No	5	LV	1		-	
			8	2 EM	1.5	4.5	.375	.375
			1	PA	2		-	
Navarra	350-400	No	2	PA	2		-	
			11	4 VF	2	5	1	1
			5	LV	1		-	
País Vasco	300	No	4	PA	4		-	
			10	2 VF	2	7	1	1
			4	LV	1		-	
Rioja (La)	300	No	6	EM	3		1	
			4	LV	1		-	
			4	RH	1	6	-	1
			4	PA	1		-	

EM: Elección múltiple

PA: Pregunta abierta

VF: Verdadero o falso

LV: Localización de vocabulario

RS: Resumen

RH: Relleno de huecos

Fuente: elaboración propia

Por tipo de ítems y tareas se observa lo siguiente:

- Elección múltiple. Andalucía, Cataluña y Murcia diseñan ítems con 4 opciones de respuesta y solo una correcta, mientras que Cantabria, Castilla y León, la Comunidad Valenciana y La Rioja los diseñan con 3 alternativas. Este tipo de ítems se evalúan con una puntuación de entre .5 (Castilla y León) y 4 puntos (Cataluña). En todos los casos los ítems no son interdependientes, los distractores son lo suficientemente plausibles y no están redactados de forma negativa.

La prueba de Lengua Extranjera-Inglés debe medir el grado de conocimientos o destrezas lingüísticas que tienen los aspirantes, por lo tanto, se espera que si un alumno no sabe ninguna respuesta debe obtener un cero. Sin embargo, las pruebas que se diseñan en algunas comunidades autónomas no penalizan los errores en los ítems de opción múltiple de manera que pueden responder bien por azar.

Cataluña, donde se utilizan 8 ítems, es la única que penaliza cada respuesta errónea. El cálculo de la penalización responde al siguiente algoritmo: como este bloque tiene un valor máximo de 4 puntos y se evalúa con 8 ítems, a cada ítem respondido correctamente le corresponde .5 puntos. En un ítem de n alternativas y una sola correcta, la esperanza matemática para una respuesta por azar debería ser cero. Luego

$$\varepsilon(U) = \sum_{i=1}^N u_i P(u_i) = u_0 * \frac{n-1}{n} + u_1 * \frac{1}{n} = u_0 * \frac{n-1}{n} + 0,5 * \frac{1}{n} = 0$$

$$\begin{aligned} u_0 &= -\frac{n}{n-1} * \frac{u_1}{n} = \\ &= -\frac{0,5}{n-1} \end{aligned}$$

u_0 = penalización por error

u_1 = 0,5 = puntuación por respuesta correcta

Para el caso de $n=4$, y $u_1 = 0,5$, $u_0 = -0,166$.

Esta es una diferencia importante respecto de otras comunidades autónomas que no penalizan los errores, por lo que la esperanza al responder por azar es distinta de cero.

- Verdadero o falso. Doce comunidades autónomas utilizan este tipo de ítems. En 7 de ellas se debe indicar la frase literal que aparece en el texto como respuesta a la pregunta, mientras que en el País Vasco hay que justificarlo con palabras propias. En Andalucía y Baleares se puede elegir entre ambas opciones y en Castilla y León y en la Comunidad Valenciana basta con indicar la línea donde se encuentra la evidencia. La puntuación otorgada a las preguntas de verdadero o falso varía entre 1.5 y 2 puntos a excepción de Baleares en donde se otorga únicamente 1 punto. Como en el caso de las preguntas de opción múltiple también deberían penalizarse los errores. En las dos últimas columnas de la tabla pueden verse las importantes diferencias que existen entre las comunidades en este apartado.
- Localización de vocabulario. Es la tarea que aparece con más frecuencia en el conjunto de las pruebas de todas las comunidades

autónomas. Todas las comunidades autónomas, a excepción de Asturias y Cataluña, incluyen ítems de localización de sinónimos o antónimos o palabras que tengan el mismo significado que una expresión dada. Cabe destacar que la dificultad no es la misma en todas las comunidades autónomas dado que en algunas se indica el número del párrafo en el que se encuentra la palabra que hay que encontrar como ocurre en Cantabria, Madrid y Navarra lo que facilita la tarea. En el caso de Canarias se indica la línea exacta y se demanda escribir una frase para evaluar la producción escrita. Su valor, en la mayoría de las pruebas, es de 1 o 1.5 puntos, excepto en Castilla y León donde se valoran con 2 puntos a estos ítems.

- Relleno de huecos. La Rioja evalúa la riqueza léxica, la capacidad de encontrar palabras literales en el texto, deducir su significado y adaptarlas a un nuevo contexto a través de ítems de este tipo con un valor de la pregunta de 1 punto.
- Pregunta abierta. Diez comunidades autónomas utilizan este tipo de ítems para medir el grado de comprensión lectora. El valor de esta pregunta varía desde 1 (Baleares y La Rioja) hasta 4 puntos (Asturias y País Vasco). Asturias solo utiliza 4 preguntas de este tipo para evaluar este bloque de contenidos referido a la comprensión lectora.
- Resumen. Galicia es la única comunidad autónoma en la que se solicita un resumen de 50 palabras. Su valor es de 1 punto.

Las pruebas estandarizadas que tomamos como referencia utilizan preguntas cerradas y objetivas para medir la comprensión lectora: preguntas de elección múltiple, de emparejamiento, etc. Por el contrario, se observa que diez comunidades autónomas utilizan en sus pruebas preguntas abiertas donde los aspirantes utilizan su propio lenguaje y entra en juego la ortografía, la exactitud de la respuesta y, por tanto, el juicio humano del evaluador. En este sentido, la ventaja de utilizar ítems de opción múltiple es que se aísla el componente de comprensión lectora y se evita medir el componente de expresión escrita, es decir, son independientes de la capacidad de escritura, tal y como establece el MCER.

Bloque: Comprensión de textos orales (*Listening*)

En el MCER se establece “En las actividades de comprensión auditiva, el usuario de la lengua como oyente recibe y procesa una información de

entrada (*input*) en forma de enunciado emitida por uno o más hablantes” (Consejo de Europa, 2001, p. 68). El usuario con nivel B1 debe:

Comprender información concreta relativa a temas cotidianos o al trabajo e identifica tanto el mensaje general como los detalles específicos siempre que el discurso esté articulado con claridad y con un acento normal.

Comprender las principales ideas de un discurso claro y en lengua estándar que trate temas cotidianos relativos al trabajo, la escuela, el tiempo de ocio, incluyendo breves narraciones. (Consejo de Europa, 2001, p. 69)

Al analizar este bloque de contenidos, se observa que solo dos comunidades autónomas evalúan la comprensión auditiva: Cataluña y Galicia. Ambas lo hacen a través de ítems de opción múltiple; sin embargo, existen diferencias sustanciales entre ambas.

Un aspecto destacable es que las pruebas internacionales que tomamos como referencia utilizan monólogos o entrevistas (se escuchan varias veces) y utilizan preguntas cerradas para medir el grado de comprensión lectora a través de preguntas de elección múltiple con 3 o 4 respuestas posibles, relleno de huecos, verdadero o falso (Inglés General) y emparejamiento de frases (ACLES). En la prueba de Trinity es el examinador el que realiza preguntas de comprensión auditiva vinculadas al audio. Ambas se ajustan al nivel B1 ya que el *Listening* consiste en una entrevista donde los personajes son de sexo opuesto, lo que facilita el nivel de comprensión auditiva.

En Cataluña se propone una actividad en la que se emite una conversación o un programa de radio. En las instrucciones se indica que, en primer lugar, se realiza una lectura previa de las preguntas y, a continuación, se procede a escuchar la grabación una única vez. Se han de contestar 8 ítems de opción múltiple con 4 alternativas de respuesta. El valor total de este bloque es de 2 puntos. Como en Cataluña cada respuesta acertada tiene un valor de 0.25, aplicando la expresión anterior para u_0

$$u_0 = -\frac{n}{n-1} * \frac{u_1}{n} = -\frac{0,25}{n-1} = -\frac{0,25}{3} = -.083$$

Y esta es la penalización que allí se aplica por respuesta errónea. Por otro lado, en Galicia se propone una actividad en la que se emite una conversación entre varias personas. En primer lugar, se realiza una lectura previa de las preguntas (2 minutos) y, a continuación, se procede a escuchar la grabación hasta 3 veces. Se han de contestar 10 preguntas de opción múltiple con 3 posibles respuestas. Los errores no penalizan y la puntuación otorgada a este bloque es de 1 punto. La puntuación esperada

que pueden obtener los aspirantes al responder por azar es de .333 puntos. Otra vez una diferencia importante. Incluso en estas comunidades, que son las únicas que incluyen este bloque, encontramos diferencias importantes. Para que los resultados fuesen formalmente equiparables, en Galicia deberían penalizar los errores, igual que en Cataluña, pues de otro modo los resultados están parcialmente sobrevalorados. A eso hemos de añadir la dificultad adicional en las pruebas de Cataluña en donde la audición se produce una sola vez.

Bloque: Producción de textos escritos (*Writing*)

En las actividades de expresión escrita, un usuario con un nivel B1 debería “escribir textos sencillos y cohesionados sobre una serie de temas cotidianos dentro de su campo de interés enlazando una serie de distintos elementos breves en una secuencia lineal” (Consejo de Europa, 2001, p. 64).

La actividad principal en este bloque es la redacción de un texto. La **longitud** varía en extensión en función de la comunidad autónoma en la que se realice la prueba. En este sentido, tal y como muestra la Tabla 5, en nueve comunidades autónomas se pide una redacción de entre 80 y 125 palabras, mientras que, en otras siete (Baleares, Cantabria, País Vasco, Valencia, Madrid, La Rioja y Navarra), ha de tener una longitud de entre 120 y 150. En Murcia se solicitan dos redacciones de entre 25 y 50 palabras cada una. Extremadura, Cataluña, Canarias, País Vasco y Navarra ofrecen varias alternativas a elegir una.

Como referencia, los exámenes de Inglés General, ACLES y Trinity College ISE I para el nivel B1 demandan textos de entre 100 y 140 palabras. Estas pruebas también incluyen tareas de contenidos sintáctico-discursivos. A pesar de que el MCER no establece un número determinado de palabras para cada nivel, las pruebas de las comunidades autónomas se adaptan al marco común a excepción de la que se desarrolla en Murcia. De hecho, no parece posible demostrar un adecuado nivel de expresión escrita produciendo un texto tan breve.

Tabla 5
Análisis del bloque: “Producción de textos escritos. Writing”. Año 2017

Comunidad autónoma	Nº de ítems		Tipo de tarea	Evaluación	Nº de palabras requeridas	Optatividad	Nº de opciones a elegir/ totales	Puntuación	
								Por tarea	Total
Andalucía	1	1	RTX	Holística	120	No	-	3	3

Comunidad autónoma	Nº de ítems		Tipo de tarea	Evaluación	Nº de palabras requeridas	Optatividad	Nº de opciones a elegir/ totales	Puntuación	
								Por tarea	Total
Aragón	1	1	RTX	Holística	80-120	No	-	3	3
Asturias	1	1	RTX	Holística	100-120	No	-	4	4
Baleares	1	1	RTX	Holística	120-150	No	-	4	4
Canarias	5	4 1	CS RTX	Analítica	6-20/situación 100-120	Sí Sí	4/6 1/2	2 3	5
Cantabria	1	1	RTX	Analítica	120-150	No	-	4	4
Castilla-La Mancha	1	1	RTX	Holística	125	No	-	3	3
Castilla y León	1	1	RTX	Holística	100-120	No	-	3	3
Cataluña	1	1	RTX	Holística	100	Sí	1/2	4	4
Comunidad Valenciana	1	1	RTX	Analítica	130-150	No	-	4	4
Extremadura	2	1 1	RTX CS	Analítica	80 20	Sí No	1/2	3 1	4
Galicia	1	1	RTX	Holística	120	No	-	3	3
Madrid	1	1	RTX	Holística	100-150	No	-	3	3
Murcia	1	1	RTX	Analítica	25-50	No	-	2	2
Navarra	1	1	RTX	Holística	150	Sí	1/2	3	3
País Vasco	1	1	RTX	Analítica	130	Sí	-	3	3
Rioja (La)	2	1 1	RTX RTX	Holística	140 30-50	No	-	3 1	4

RTX: Redacción de un texto

CS: Comentario de una situación dada

Fuente: elaboración propia

Otra tarea del examen es el comentario de una o varias situaciones (Canarias) que en ocasiones es la continuación de un diálogo (Extremadura) o la redacción de un texto complementario con menor extensión (La Rioja). Sin embargo, la tarea de Canarias tiene carácter optativo, es decir, se deben elegir 4 situaciones de 6 dadas.

La puntuación estipulada para este bloque de contenidos en las diferentes pruebas varía entre 3 y 4 puntos, por lo que no se observan diferencias sustanciales entre comunidades autónomas a excepción de Canarias donde se otorgan hasta 5 puntos sobre 10.

No es posible analizar los ítems de producción de forma independiente a sus criterios de evaluación, es decir, los ítems de producción necesitan su regla o norma de corrección. Todas las comunidades autónomas utilizan criterios de corrección como la riqueza léxica, organización, coherencia, vocabulario, etc. tal y como establece la ALTE. Sin embargo, no todas las pruebas establecen los mismos instrumentos para evaluar este tipo de ítems. La mayoría de las comunidades autónomas los evalúan a través de evaluaciones holísticas, es decir, globales, con menor nivel de especificidad y, por tanto, más subjetivas. Por el contrario, en Canarias, Cantabria, Valencia, Extremadura, Murcia y País Vasco se evalúan estos ítems a través de un proceso analítico. Para ello se dispone de una rúbrica en la que se desglosan los criterios observables para cada nivel de ejecución aportando un mayor detalle y objetividad a las respuestas.

Puede verse en definitiva que en este bloque hay también diferencias importantes entre las comunidades, tanto en la valoración total del bloque, en las tareas requeridas en el mismo, como en las formas de evaluación. No puede decirse tampoco que haya equivalencia formal entre las comunidades.

Bloque: Producción de textos orales (*Speaking*)

La producción de textos orales, *Speaking*, no se evalúa en ninguna comunidad autónoma, por lo tanto, no cubren las destrezas orales que establece el MCER para la adquisición de una lengua extranjera. En ningún documento oficial de los analizados se justifica que no se deba evaluar esta destreza.

Por otro lado, la evaluación de la expresión oral en las pruebas internacionales como en la de Inglés General, ACLES o Trinity se realiza a través de varias tareas como entrevistas personales, diálogos con interlocutores y compañeros y descripción de fotografías, dando respuesta a las directrices del MCER.

Ciertamente hay razones de dificultades logísticas para el desarrollo de un bloque como este. Sin embargo, no se aprecia ningún esfuerzo por suplir estas carencias con alguna alternativa más factible. En contra de lo que se establece en la normativa este bloque queda vacío de contenido en la EvAU.

Bloque: Contenidos sintáctico-discursivos

Este bloque incluye la sección de gramática (morfología, sintaxis y fonología) en la que se evalúa el conocimiento sobre la lengua.

La mayoría de las comunidades autónomas evalúa este bloque a excepción de Cantabria, Cataluña, Valencia, País Vasco y La Rioja (Tabla 6).

Tabla 6

Análisis de los “Contenidos sintáctico-discursivos”. Año 2017

Comunidad autónoma	Nº de ítems		Tipo de tarea	Optatividad	Puntuación	
					Por ítem	Total del bloque
Andalucía	8	4 4	RH RO	No	1 2	3
Aragón	4	4	RO	No	2	2
Asturias	8	4 4	RH RO	No	1 1	2
Baleares	10	4 2 4	RH RO FN	No	1 1 1	3
Canarias	8	8	RH	No	2	2
Castilla-La Mancha	10	6 4	RO FN	No	3 1	4
Castilla y León	2	2	RO	No	1	1
Extremadura	2	2	RO	No	1.5 y 1	2.5
Galicia	2	2	RO	No	2	2
Madrid	7	6 1	RH RO	No	1.5 0.5	2
Murcia	10	4 6	RO RH	No	2 1.5	3.5
Navarra	8	8	RH	No	2	2

RH: Relleno de huecos

RO: Reformulación de oraciones

FN: Fonología

Fuente: elaboración propia

Los ítems que se utilizan para medir el conocimiento sobre la gramática son de tres tipos:

- Rellenar huecos. Siete comunidades autónomas utilizan este tipo de ítems donde tienen que completar espacios con diferentes formas verbales, preposiciones o una serie de palabras dadas.

En el caso de Canarias, se deben completar espacios que forman parte de un diálogo, simulando una conversación.

En el caso de Asturias, el número de sustantivos que podrían llenar el espacio vacío es considerable (*things, reasons, motifs, incentives, etc.*).

“Why become an au-pair? There are many (A)_____ but perhaps the first and (B)_____ important is the opportunity to travel...”

En los casos de Andalucía, Baleares, Madrid o Navarra se proporciona al aspirante una serie de palabras clave, como en las pruebas de:

“She is very good_____ (preposition) singing” o “Don’t go to school if you... (don’t/couldn’t/will/may not) feel well”.

- Reformulación o transformación gramatical de oraciones. Más de la mitad de las comunidades autónomas demanda transformar oraciones utilizando un pronombre relativo, el estilo indirecto o la pasiva, entre otras.
- Fonología. Se espera que los aspirantes reconozcan e interpreten los fonemas o sonidos. Únicamente dos comunidades autónomas presentan ítems relacionados con la fonología: Baleares y Castilla-La Mancha.

Los ítems que evalúan este bloque de contenidos sintáctico-discursivos tienen carácter obligatorio en todas las comunidades autónomas. La puntuación otorgada a este bloque varía entre 1 (Castilla y León) y 4 puntos (Castilla-La Mancha).

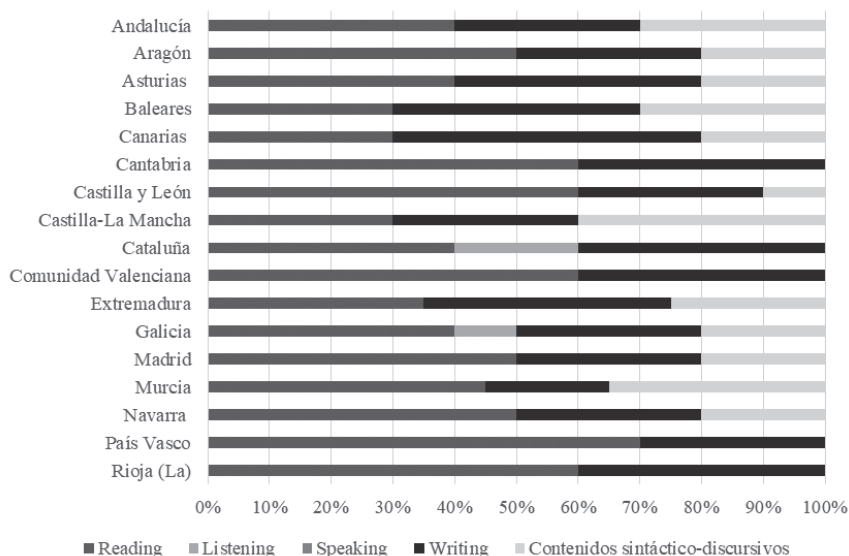
De nuevo son relevantes las diferencias tanto en puntuación como en las tareas requeridas.

Ponderación de cada bloque de contenidos

En la Figura 1 se muestra la ponderación, en términos porcentuales, de cada bloque sobre la puntuación total. El bloque de contenidos que más peso

tiene en la mayoría de las comunidades autónomas es el de comprensión de textos escritos, *Reading*.

Figura 1
Ponderación de cada bloque de contenidos



En las pruebas internacionales, cada bloque de contenidos representa un 25% del total de la prueba, mientras que los contenidos sintáctico-discursivos se evalúan de forma implícita dentro de cada uno de los anteriores. En contraste con esas pruebas, y tal y como se indicó en la tabla I, el peso del *Reading* más el *Listening* en la prueba de Lengua Extranjera-Inglés debería ser del 60% mientras que el *Speaking* y el *Writing* debería ser del 40%. Sin embargo, se puede observar que ninguna comunidad autónoma diseña sus pruebas en consonancia con las directrices establecidas en la normativa (Orden Ministerial ECD 1941/2016) dado que el peso varía significativamente.

Valoración de la adecuación de las pruebas a las directrices establecidas por el MCER y por las pruebas internacionales estandarizadas

En resumen, podemos decir que ninguna de las 17 pruebas se adapta, en su totalidad, a las directrices establecidas por el MCER ni a las pruebas internacionales estandarizadas que hemos tomado como referencia y que son generalmente aceptadas como un estándar de hecho (Figura 2).

Figura 2
Síntesis del grado de adecuación de las pruebas a las directrices establecidas por el MCER

Comunidad autónoma	Reading	Listening	Speaking	Writing	Contenidos sintáctico-discursivos
Andalucía	●			■	
Aragón	● ○			■	▲
Asturias	○*			■	▲
Baleares	● ○			■	
Canarias	● □			□	
Cantabria	●				
Castilla-La Mancha	●			■	
Castilla y León	● ○			■	
Cataluña				■	
Comunidad Valenciana	● ○				
Extremadura	●				
Galicia	○*			■	
Madrid	● ○			■	
Murcia	● ○			△	
Navarra	● ○			■	
País Vasco	● ○				
Rioja (La)	● ○			■	

Criterios:

● No se penalizan los errores de elección múltiple

○ Preguntas abiertas sin rúbricas de corrección

□ Optatividad en el n° de ítems a realizar

■ Evaluación general y holística

△ Número de palabras requeridas insuficiente

▲ Redacción inadecuada de ítems

* En los casos de Asturias y Galicia, un solo criterio da lugar a un grado de adaptación incompleto en *Reading* dado que la primera evalúa este bloque únicamente a través de preguntas abiertas mientras que la segunda lo hace, en mayor proporción, a través de preguntas abiertas y un resumen.

	Completo
	Parcial: no se adapta en un criterio
	Incompleto: no se adapta en ≥2 criterios
	No se evalúa

En función de los análisis realizados en este estudio y los resultados obtenidos, se puede afirmar que ninguna de las 17 pruebas, en su totalidad, cumple con las características propias de un examen que tiene la finalidad de determinar si los estudiantes son aptos para cursar la enseñanza superior universitaria, en primer lugar, y establecer un orden de prioridad en la elección de centros y estudios. Sin embargo, los componentes *Reading* y *Listening* de la prueba de Cataluña, así como los componentes *Writing* y los contenidos sintáctico-discursivos de la prueba de Extremadura son los que muestran un mayor grado de adaptación al MCER. Por el contrario, las pruebas que menos se adaptan al MCER son las diseñadas en la Comunidad Valenciana, Murcia y La Rioja.

Paralelamente, tras este análisis se pueden clasificar las pruebas en función de su facilidad, de manera que las más sencillas son aquellas que solo evalúan *Reading* y *Writing*, que se caracterizan por tener ítems optativos y que la puntuación esperada por azar es mayor: País Vasco, Cantabria, La Rioja, Comunidad Valenciana, Navarra, Aragón, Canarias y Murcia. Por el contrario, las más complejas son aquellas que evalúan más de dos destrezas lingüísticas, que utilizan únicamente preguntas abiertas para medir el grado de adquisición de alguna de las habilidades y que la puntuación esperada por azar es menor: Asturias, Baleares y Cataluña.

Resultados referentes a las posibles relaciones que puedan existir entre el rendimiento promedio y la estructura de las pruebas

Como ya hemos señalado, para cualquier alumno debería ser indiferente para sus probabilidades de éxito en la elección de centro y carrera, el lugar en donde realiza su prueba de acceso a la universidad. Sin embargo, hemos visto que las diferencias formales entre comunidades son muy importantes, puesto que no es lo mismo que se requiera la elaboración de una composición de 150 palabras (Valencia, Madrid o Navarra) o de 25, tal y como ocurre en Murcia. En este apartado tratamos de comprobar si las diferencias aparentes en dificultad son concomitantes con diferencias en rendimiento medio de los alumnos. En este contexto, entendemos el concepto de dificultad como la complejidad del contenido de las pruebas o la carga de trabajo que suponen las tareas asignadas al estudiante. Si este fuera el caso, no podríamos descartar que las diferencias en rendimiento se debiesen a diferencias en estructura y dificultad.

En la Tabla 7 presentamos los resultados promedio en cada comunidad autónoma junto con una indicación del nivel de facilidad aparente de las pruebas y de las puntuaciones esperadas por simple azar en cada caso.

Tabla 7

Relación entre la estructura de las pruebas de Lengua Extranjera-Inglés y el rendimiento promedio obtenido por los examinados en las mismas en el año 2017

Comunidad autónoma	Nota media obtenida	Nivel de facilidad de las pruebas	Puntuación esperada por azar	Nota media final esperada
País Vasco	7.44		1	6.44
Cantabria	7.4		1.916	5.484
Rioja (La)	7.3		1	6.3
Comunidad Valenciana	7.17		1.25	5.92
Navarra	7.11		1	6.11
Aragón	6.85		1	5.85
Canarias	6.83		.75	6.08
Andalucía	6.78		1.25	5.53
Extremadura	6.71		1	5.71
Galicia	6.65		.333	6.317
Madrid	6.63		1	5.63
Castilla y León	6.45		.9165	5.5335
Castilla-La Mancha	6.15		1	5.15
Murcia	6.14		.375	5.765
Cataluña	5.92		0	5.92
Asturias	5.86		0	5.86
Baleares	5.85		.5	5.35

Fuente: elaboración propia

	Nivel alto de facilidad
	Nivel medio de facilidad
	Nivel bajo de facilidad

Se observan tres categorías de casos distintos, junto a las posibles características de las pruebas que pueden estar afectando a las puntuaciones de los aspirantes:

- Las puntuaciones promedio más altas (>7). Estas se corresponden con las pruebas cuyo nivel de facilidad es mayor por disponer de las siguientes características:
 - Las cuatro comunidades autónomas con mayor promedio son las únicas que no evalúan el bloque de contenidos sintáctico-discursivos en sus pruebas.
 - En tres de ellas se establecen criterios de corrección analíticos para el *Writing*.
 - Lo más importante de todo, tienen una puntuación esperada por azar alta.

Destacan tres excepciones por tener otras características que influyen en que el nivel de facilidad de las pruebas sea mayor:

- Aragón: única prueba en la que el evaluador puede bonificar o penalizar con 1 punto al aspirante desde su punto de vista en función del buen uso de la lengua o el número de faltas en el *Writing*, lo que otorga mayor subjetividad al proceso de corrección.
 - Canarias: única prueba en la que se da opción de elegir entre varios ítems, en el *Reading*, o entre varias opciones, en el *Writing*.
 - Murcia: única prueba en la que se exige una menor longitud en el *Writing* (de 25 a 50 palabras).
- Las puntuaciones promedio (6-7) que se mantienen en la media.
 - Las puntuaciones promedio bajas (<6). Estas se corresponden con las pruebas cuyo nivel de facilidad es menor por tener las siguientes características:
 - Tienen una puntuación esperada por azar baja: la de Cataluña y Asturias es 0.
 - Asturias, junto a Galicia, son las únicas que utilizan preguntas abiertas para evaluar el *Reading*.
 - Cataluña, junto a Galicia, son las únicas que evalúan el *Listening*. Sin embargo, Galicia no penaliza los errores en los ítems de opción múltiple, lo que podría ser compatible con una

explicación de por qué los estudiantes gallegos obtengan un promedio mayor que los de Cataluña e incluso que los de otras comunidades autónomas.

DISCUSIÓN Y CONCLUSIONES

El presente estudio aporta evidencias acerca de los tres objetivos específicos que se plantean sobre la prueba de Lengua Extranjera-Inglés: si se cumplen las directrices marcadas en normativa, si existen diferencias significativas en su diseño y estructura entre las comunidades autónomas y si cabe la posibilidad de que las diferencias observadas en el rendimiento promedio de los estudiantes por comunidades autónomas estén producidas por las diferencias en el diseño de las pruebas.

Una primera conclusión es la relacionada con el primer objetivo específico que se plantea en este estudio: ninguna comunidad autónoma cumple con las directrices establecidas en el currículo (Real Decreto 1105/2014) ni en el MCER. En ambos documentos, que sirven como pilares fundamentales para el diseño de las pruebas de Lengua Extranjera-Inglés, se establece la medición de cuatro destrezas lingüísticas para la obtención de cualquiera de los niveles que establece el Consejo de Europa (2001). Sin embargo, ninguna comunidad autónoma evalúa la producción oral, o *Speaking*, mientras que solo Galicia y Cataluña incorporan ítems para medir la comprensión auditiva, o *Listening*. En esta misma línea, solo Castilla-La Mancha y Baleares contemplan ítems dirigidos a medir la competencia fonológica. Pero, incluso en estas dos comunidades, la medida de esta competencia se hace desde una perspectiva mecánica y con una presentación desligada de la producción oral. Esta inadecuación de la producción oral fue resaltada en los estudios de Palomeque (2015), Tamayo-Beritán y Fernández-Terry (2017) y Erreyes y García (2017). En cualquier caso, en estas comunidades en las que se consideran estas competencias ligadas a la comprensión y producción oral, aunque de manera parcial o insuficiente, la puntuación otorgada a estos ejercicios no supera el 20% del total de la prueba. En consecuencia, la comprensión y la producción oral no forman parte sustantiva de la estructura de la prueba. Esta carencia tiene consecuencias negativas a medio y largo plazo. Por una parte, cabe destacar que la enseñanza en la etapa de bachillerato se ciñe a la preparación de los contenidos evaluables en la EvAU por lo que los contenidos que no se evalúan en el acceso a la universidad terminan difuminándose y pierden relevancia. En efecto, este hecho ha sido señalado por Sevilla-Pavón et al., 2017 y los de Sevilla-Pavón et al., 2017. Por otra parte, y como consecuencia de lo anterior, el contexto universitario se ve afectado debido a que muchos de los estudiantes acceden con un nivel inferior al esperado en comprensión

y expresión oral del inglés, tan necesario en estudios superiores. La exclusión de dimensiones esenciales de la comunicación en una segunda lengua en la prueba EVAU, y como efecto perverso de la enseñanza para el examen, aumenta el riesgo de convertir la asignatura de lengua extranjera en la fase final del bachillerato en una preparación libresca y desligada de la función comunicativa de una lengua viva. En esta línea, autores como Lee y Greene (2015), Harrington y Roche (2014) y Carlsen (2018), aconsejan la inclusión, tanto en la etapa de formación preuniversitaria como en la misma universidad, de la evaluación del componente oral en las pruebas de Lengua Extranjera-Inglés.

Una segunda conclusión está relacionada con el segundo objetivo específico que se plantea en este estudio: las pruebas de Lengua Extranjera-Inglés que se diseñan para la EvAU en las distintas comunidades autónomas no son equiparables, esto es, aunque dicen medir el mismo constructo no lo hacen. Tras el análisis de las 68 pruebas recogidas, se comprueba que existen diferencias sustanciales en el contenido, en las formas de corrección y en las puntuaciones otorgadas a cada destreza lingüística. Es más, en la evaluación de la comprensión lectora, o *Reading*, se utilizan 6 tipos de ítems diferentes: elección múltiple, verdadero o falso, localización de vocabulario, relleno de huecos, preguntas abiertas y resumen. Esto implica que, con algunos tipos de ítems los aspirantes únicamente tienen que mostrar su comprensión lectora, mientras que con otros tienen que mostrar, además, su capacidad para expresarse por escrito. En algunas comunidades autónomas se evalúa el *Reading* combinando estos tipos de ítems por lo que, consecuentemente, la comprensión lectora no se evalúa de manera independiente de la expresión escrita. Este hecho dificulta que el peso otorgado a cada uno de los bloques de contenidos en cada una de las pruebas se ajuste al peso orientativo que establece la normativa (Orden ECD/1941/2016). Desde un punto de vista psicométrico, esta evaluación conjunta resulta inapropiada porque no se determina el nivel real de comprensión lectora al penalizar a un estudiante por su déficit en escritura. La falta de independencia de la medida de la capacidad de comprensión lectora y de la capacidad de expresión escrita ya fue puesta de manifiesto por Esteban (2017) y Molina et al. (2018). En este sentido, solo en Cataluña se evalúan las dos destrezas de forma independiente a través de ítems objetivos de elección múltiple. Sería deseable que todas las comunidades autónomas utilizaran el mismo tipo de ítems para medir las competencias y, además, que las competencias a medir estuviesen claramente diferenciadas evitando todo lo posible que interfiera el dominio de una de ellas en la ejecución de otra. El uso de este tipo de ítems de respuesta cerrada para medir la comprensión lectora y de preguntas abiertas para la expresión escrita podría ser una cuestión a considerar para conseguir esa independencia de las competencias.

Todo lo anterior nos lleva a concluir que no se está midiendo un mismo constructo en todas las comunidades autónomas. Desde el punto de vista de la estructura y composición de las pruebas analizadas, de ningún modo puede sostenerse que haya una unidad de composición teórica de los exámenes utilizados. Por lo tanto, no se puede asumir que un mismo alumno pudiese obtener calificaciones equiparables si hipotéticamente se sometiese a más de uno de estos exámenes. Consecuentemente, estas pruebas no cumplen con la función de igualación de oportunidades que, supuestamente, tienen. Si esta falta de equivalencia es un hecho, pierde todo su sentido y justificación la idea de que exista una prueba común para determinar el orden y modo de acceso de los aspirantes a las plazas universitarias. Para la mayoría de los estudiantes que tratan de acceder a universidades de su propia comunidad autónoma, la consecuencia de esta disparidad es que el nivel de inglés con el que se accede a la universidad puede variar sustancialmente de una a otra comunidad autónoma. Esto, unido a diferencias encontradas en otros estudios similares al que aquí se presenta (Ruiz-Lázaro y González, 2017), lleva a que a largo plazo el efecto de la disparidad sobre el rendimiento en la universidad pueda incrementarse, llegando, en último extremo, a conformar grupos de universidades de primera y de segunda.

Una tercera conclusión está relacionada con el tercer objetivo específico que se plantea en este estudio: las diferencias observadas en el rendimiento promedio de los estudiantes en la prueba de Lengua Extranjera-Inglés de la EvAU son concomitantes con las diferencias observadas en el diseño de la misma. Aunque se trata supuestamente del mismo constructo, las proporciones de los contenidos que evalúa la prueba de cada comunidad, así como el formato de los ítems impiden su adecuada comparación al no estar estandarizadas. Por lo tanto, con 17 conjuntos de pruebas diferentes y una gran variedad de ítems, existen diferencias importantes en el rendimiento promedio de los aspirantes que se asocian con las diferencias estructurales de las pruebas. Estos resultados coinciden con los que Ruiz-Lázaro y González (2017) destacaron en un estudio similar comparando las pruebas de Lengua Castellana y Literatura. Si, como las evidencias sugieren, se puede estar obteniendo un mayor o menor rendimiento promedio debido a la estructura de la prueba aplicada en cada comunidad autónoma, entonces, los aspirantes pueden verse beneficiados o perjudicados para ser aceptados en las universidades por motivos ajenos a su desempeño en el examen. Desde un punto de vista técnico, estas pruebas deberían cumplir con las características técnicas de fiabilidad y validez. En cuanto a la fiabilidad, se debería controlar la subjetividad del evaluador a la hora de corregir, estableciendo pruebas con el mayor grado de objetividad posible. Por otra parte, deberían utilizarse algoritmos de puntuación que contemplasen la corrección del efecto del azar. Como hemos visto, Cataluña es la única comunidad en la que se penalizan los errores. Pero, en

cualquier caso, en todas las comunidades autónomas debería utilizarse el mismo procedimiento, pues de otra forma se generan importantes agravios comparativos.

Un ejercicio que aporta mucha luz sobre esta última cuestión es el cálculo de la nota media cuando se compensan las respuestas por azar. En efecto, si hacemos,

$$X = \text{nota media obtenida} - \text{puntuación esperada por azar} = \text{nota media final esperada}$$

resulta que la nota promedio obtenida en todas las comunidades autónomas es más homogénea. Las diferencias se reducirían desde una amplitud de 1.59 puntos (7.44 correspondiente al País Vasco y 5.85 a Islas Baleares) a una diferencia máxima de 1.29 puntos (5.15 puntos en Castilla-La Mancha y 6.44 en el País Vasco). Solo con esta corrección se produciría una disminución relevante en la variabilidad de los resultados observados, por lo que la puntuación esperada en cada comunidad autónoma estaría más equilibrada. Esto parece indicar que, si se establecieran unos criterios comunes en cuanto al diseño, estructura construcción y calificación de las pruebas, se podría subsanar el agravio comparativo entre aspirantes de las diferentes comunidades autónomas. Este podría ser un paso para, de algún modo, garantizar la homogeneización y estandarización tanto de las pruebas como de los resultados promedio para cumplir con el objetivo de equidad para todos los aspirantes españoles. Por supuesto, el objetivo no tiene que ser la desaparición de las diferencias en rendimiento medio de los alumnos de las distintas comunidades autónomas, sino que las diferencias que se observen sean diferencias reales en preparación y conocimientos y no diferencias artificiales producidas por deficiencias en la coordinación entre las administraciones públicas implicadas.

En cuanto a las limitaciones de este estudio, todas ellas se derivan de las características específicas del mismo. Se trata de un estudio no experimental y, por tanto, las relaciones observadas podrían tener una naturaleza espuria debido a que no se ha manipulado la variable independiente. Así, mientras que en los estudios experimentales la seguridad de las relaciones causales se basa, fundamentalmente, en el supuesto de independencia de la asignación de los tratamientos y las variables de respuesta (según el paradigma Neyman-Rubin, (Rubin, 1974; Holland, 1986)), en los estudios no experimentales esa seguridad es menor. Esto quiere decir que, en estos últimos, la consistencia entre las observaciones y las explicaciones teóricas ofrecidas es la que aporta seguridad en las conclusiones obviando cualquier explicación causal alternativa. Asimismo, el paradigma de Suppes (1970) determina que un evento A es causa de B si, además de darse ciertas condiciones, no existe un evento C que sea explicación alternativa de

B. En esta línea y en el contexto de este trabajo, la seguridad de que las diferencias en el rendimiento medio son producidas por la heterogeneidad de los exámenes está asociada a nuestra seguridad de que no existe una explicación alternativa que lo justifique como, por ejemplo, que los alumnos de las comunidades que obtienen un rendimiento medio más alto lo hagan porque el sistema educativo de esas comunidades sea claramente superior o porque sus capacidades difieran significativamente. Un modo de solventar esta limitación sería aplicando la misma prueba a una muestra aleatoria de sujetos de las distintas comunidades autónomas. Algo así sería posible solo si una decisión de orden político apoyase tal medida.

En relación con la limitación anterior, y por tratarse de un estudio secundario de datos, el modo y procedimiento de recogida de la información no fue planificado por los autores del presente estudio. Por lo tanto, al ser otra la finalidad de la recogida de la información, no se han recogido los datos de forma aleatoria. Sin embargo, un atenuante para esta limitación es que se trata de un estudio censal y se cuenta con estadísticas publicadas de todos los alumnos evaluados en las convocatorias analizadas.

Otro aspecto importante es la limitación temporal del estudio. Los datos analizados proceden de las convocatorias, ordinaria y extraordinaria del año 2017. En la medida en la que se produzcan cambios en los modos de diseño y construcción de las pruebas EvAU estos resultados podrían quedar obsoletos. Sin embargo, la relativa constancia de los procedimientos de evaluación desde que comenzó la aplicación de pruebas para ordenar el acceso a la universidad proporciona cierta garantía de que los resultados aquí obtenidos no son una anécdota puntual sino una tendencia con una cierta constancia en el tiempo. Como prospectiva y para atenuar esta limitación, este estudio podría complementarse con otro de carácter causal en el que se considerasen varias cohortes de estudiantes o momentos de aplicación de las pruebas, de forma que se pudiera analizar la descomposición de la varianza en las puntuaciones promedio de las comunidades autónomas en diferentes momentos temporales y los motivos que la provocan.

Por último y en relación con las aportaciones de Freedman (2002), la naturaleza agregada de los datos evidencia la peligrosidad de la falacia ecológica, es decir, de afirmar que lo que es verdad para los grupos agregados (en este caso las comunidades autónomas) sea verdad para los subgrupos o los individuos. Ciertamente, cabe la posibilidad de que algunos subgrupos específicos de sujetos hubiesen obtenido resultados similares sin importar con qué prueba hubiesen sido evaluados, como por ejemplo aquellos con altas capacidades o aquellos con mayores necesidades de preparación. En ambos casos, no importaría la comunidad autónoma en la que hubiesen

sido evaluados ya que sus resultados hubiesen sido similares en cualquiera de ellas.

En definitiva, este trabajo muestra una heterogeneidad de las pruebas entre comunidades autónomas y una discrepancia, mayor o menor, pero nunca coincidencia, de cada una de las pruebas con el MCER. En consecuencia, el acuerdo entre las comunidades autónomas y el gobierno central es urgente con el fin de elaborar una prueba de acceso a la universidad que garantice el principio de igualdad. Esto quiere decir que la igualdad de oportunidades sea efectiva y garantice que las diferencias en el ingreso a cualquier universidad pública de España se deban, únicamente, a diferencias en los conocimientos y en el desempeño de los aspirantes en las pruebas y no a otros factores relacionados con características propias de los instrumentos utilizados para medirlos.

NOTAS

- 1 Este estudio forma parte de una tesis doctoral financiada por la Universidad Complutense de Madrid y el Banco Santander a través de una ayuda para contratos predoctorales de personal investigador en formación (Referencia CT27/16-CT28/16) dentro del Programa de Financiación UCM-Santander 2016.
- 2 Relación porcentual entre la matrícula de nuevo ingreso por preinscripción y el número de plazas ofertadas.

REFERENCIAS BIBLIOGRÁFICAS

- ALTE (2005). *Materials for the guidance of test item writers*. <https://bit.ly/2OzYMLL>
- Amengual-Pizarro, M. & Méndez, M. C. (2012). Implementing the Oral English Task in the Spanish University Admission Examination: an International Perspective of the Language. *Revista de Educación*, 357, 105-128.
- Bueno, M.C. y 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.
- Carroll, J.B. (1961). *Fundamental considerations in testing English proficiency of foreign students*. In *Testing the English Proficiency of Foreign Students*, (pp. 30-40). Center for Applied Linguistics.
- Carlsen, C.H. (2018). The Adequacy of the B2 Level as University Entrance Requirement. *Language Assessment Quarterly*. <http://dx.doi.org/10.1080/15434303.2017.1405962>
- Consejo de Europa (2001). *Marco Común Europeo de Referencia para las Lenguas: aprendizaje, enseñanza, evaluación*. Anaya.
- Constitución Española (1978). Boletín Oficial del Estado, 29 de diciembre de 1978, núm. 311, 29313 a 29424.
- Díez-Bedmar, M.B. (2012). The Use of the Common European Framework of Reference for Languages to Evaluate Compositions in the English Exam Section of the University Admission Examination. *Revista de Educación*, 357, 55-80.
- Erreyes, H.M.B., y García, T.M.B. (2017). Las competencias lingüísticas español-inglés en el sistema educativo ecuatoriano. *REFCaLE: Revista Electrónica Formación y Calidad Educativa*, 5(2), 33-42.
- Esteban, A. (2017). Estado actual de la comprensión lectora en Educación Primaria. *Revista Fuentes*, 19(1), 15-38. <http://dx.doi.org/10.12795/revistafuentes.2017.19.1.01>
- Freedman, D.A. (2002). *The Ecological Fallacy*. University of California.
- García-Garrido, J.L. (1991). *Fundamentos de Educación Comparada*. Dykinson.
- García-Laborda, J. (2012). Presentación. De la Selectividad a la Prueba de Acceso a la Universidad: pasado, presente y un futuro no muy lejano. *Revista de Educación*, 357, 17-27.
- Harrington, M. & Roche, T. (2014). Identifying academically at-risk students in an English-as-a-Lingua-Franca university setting. *Journal of English for Academic Purposes*, 15, 37-47. <http://dx.doi.org/10.1016/j.jeap.2014.05.003>
- Holland, P.W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396), 945-960. <http://dx.doi.org/10.1080/01621459.1986.10478354>
- Lado, R. (1961). *Language Testing*. McGraw-Hill.
- Lee, Y. & Greene, J. (2015). The predictive validity of an ESL placement test. *Journal of Mixed Methods Research*, 1(4), 366-389. <http://dx.doi.org/10.1177/1558689807306148>

- Martín-Martín, J.M. (2012). El conocimiento léxico en la sección de inglés de la Prueba de Acceso a la Universidad. *Revista de Educación*, 357, 129-142.
- Molina, N.E.C., Velásquez, A.A.R., y Abril, P.E.C. (2018). Rúbricas y su incidencia en el desarrollo de la destreza del habla en el idioma inglés. *Revista Magazine de las Ciencias*, 3(2), 83-94.
- Muñoz-Repiso, M., Murillo, F.J., Arrimadas, I., Navarro, R., Díaz-Caneja, P., Martín, A.I., y Fernández, E. (1997). *El sistema de acceso a la universidad en España: tres estudios para aclarar el debate*. CIDE.
- Orden ECD/1941/2016, de 22 de diciembre, por la que se determinan las características, el diseño y el contenido de la evaluación de Bachillerato para el acceso a la Universidad, las fechas máximas de realización y de resolución de los procedimientos de revisión de las calificaciones obtenidas, para el curso 2016/2017. Boletín Oficial del Estado, 309, de 23 de diciembre de 2016, pp.89890-89949.
- Palomeque, N.G. (2015). El desarrollo de la competencia comunicativa: una mirada hacia la formación de maestros. *Revista de la Facultad de Educación*, 22, 65-70.
- Real Decreto-ley 5/2016, de 9 de diciembre, de medidas urgentes para la ampliación del calendario de implantación de la Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa. Boletín Oficial del Estado, 298, de 10 de noviembre de 2016, pp.86168-86174.
- Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato. Boletín Oficial del Estado, 3, de 3 de enero de 2015, pp.169-546.
- Real Decreto 310/2016, de 29 de julio, por el que se regulan las evaluaciones finales de Educación Secundaria Obligatoria y de Bachillerato. Boletín Oficial del Estado, 183, de 30 de julio de 2016, pp.53049-53065.
- Real Decreto 1892/2008, de 12 de noviembre, por el que se regulan las condiciones para el acceso a las enseñanzas universitarias oficiales de grado y los procedimientos de admisión a las universidades públicas españolas. Boletín Oficial del Estado, 283, de 24 de noviembre de 2008, pp. 46932-46946.
- Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5), 688-701. <http://dx.doi.org/10.1037/h0037350>
- Ruiz-Lázaro, J. y González, C. (2017). Análisis de la Prueba de Lengua Castellana y Literatura que da acceso a la universidad: comparación entre las comunidades autónomas. *Bordón. Revista de Pedagogía*, 69(3), 175-195.
- Sevilla-Pavón, A., García-Laborda, J., y Gimeno-Sanz, A.M. (2017). Actitudes de los profesores de la Comunidad Valenciana ante la tipología de ejercicios para la evaluación asistida por ordenador de las destrezas orales en la futura Prueba de Acceso a la Universidad. *Educacao e Pesquisa*, 43(4), 1-18.
- Sevilla-Pavón, A., Gimeno-Sanz, A., y García-Laborda, J. (2017). Actitudes docentes hacia los ejercicios de la Prueba de Acceso a la Universidad informatizada. *Educação e Pesquisa*, 43(4), 1179-1200.

- Sistema Integrado de Información Universitaria (2017). *Estadísticas de las Pruebas de Acceso a la Universidad*. Ministerio de Educación, Cultura y Deporte.
- Sistema Integrado de Información Universitaria (2019). *Estadística de Universidades, Centros y Titulaciones. Curso 2018-2019*. Ministerio de Educación, Cultura y Deporte.
- Suárez-Álvarez, J., González-Prieto, C., Fernández-Alonso, R., Gil, G., y Muñiz, J. (2014). *Evaluación psicométrica de la expresión oral en inglés de las Pruebas de Acceso a la Universidad*. Ministerio de Educación, Cultura y Deporte.
- Suppes, P. (1970). *A Probabilistic Theory of Causality*. North-Holland Publishing.
- Tamayo-Beritán, AL. y Fernández-Terry, J.A. (2017). La importancia de diseñar instrumentos eficaces para evaluar la habilidad oral en el Idioma Inglés. *Educación y Sociedad*, 15(1), 121-135.

PERFIL ACADÉMICO Y PROFESIONAL DE LOS AUTORES

Judit Ruiz-Lázaro. ORCID: <http://orcid.org/0000-0003-2036-0428>

Personal Investigador en Formación en el Departamento de Investigación y Psicología en Educación (área de Métodos de Investigación y Diagnóstico en Educación) en la Universidad Complutense de Madrid. Actualmente realiza su tesis doctoral centrada en la evaluación del acceso a la universidad, en la misma universidad. Es miembro del grupo de investigación Medida de Evaluación y Sistemas Educativos (MESE). E-mail: judruiz@ucm.es

Coral González Barbera. ORCID: <http://orcid.org/0000-0002-1947-6303>

Doctora en Educación por la Universidad Complutense (2003). Profesora titular en el Departamento de Investigación y Psicología en Educación (área de Métodos de Investigación y Diagnóstico en Educación). Principal línea de interés: rendimiento académico y elementos con los que se relaciona. Miembro del grupo de investigación Medida de Evaluación y Sistemas Educativos (MESE) desde su consolidación. E-mail: corala@ucm.es

José-Luis Gaviria Soto. ORCID: <http://orcid.org/0000-0002-7398-9943>

Doctor en Filosofía y Ciencias de la Educación (UCM) desde 1988 en el Departamento de Investigación y Psicología en Educación (área de Métodos de Investigación y Diagnóstico en Educación) en la Universidad Complutense de Madrid. Especial interés en los problemas metodológicos de la evaluación educativa. Es miembro del grupo de investigación Medida de Evaluación y Sistemas Educativos (MESE). E-mail: josechogs@gmail.com

Fecha Recepción del Artículo: 17. Febrero. 2020

Fecha Modificación del Artículo: 04. Junio. 2020

Fecha Aceptación del Artículo: 08. Junio. 2020

Fecha Revisión para Publicación: 24. Junio. 2020