Pautas para elaborar pruebas de ELE válidas y confiables[1]

Ana Vine Jara
Universidad de Concepción

Anita Ferreira Cabrera
Universidad de Concepción

1. Introducción

En la evaluación del Español como Lengua Extranjera con Fines Específicos Académicos (EFA) se observa una serie de problemáticas que evidencian la falta de claridad respecto de lo que se debería evaluar en una prueba de EFA (Llorián, 2012). En este sentido, se necesita que las instituciones que han desarrollado pruebas de EFA muestren cuáles han sido los criterios considerados para el diseño, elaboración y validación de estos instrumentos. Considerando que, en la actualidad, no existen pruebas de EFA reconocidas a nivel nacional e internacional que permitan medir las competencias de los estudiantes extranjeros en cuanto al uso del español en un ámbito académico (Mendoza, 2015). En consecuencia, se requiere de instrumentos en el ámbito del EFA que logren evaluar de manera válida y confiable el nivel de proficiencia de los estudiantes. Dichas propuestas deben adaptarse a los nuevos formatos de aplicación de pruebas de lenguas extranjeras, considerando las ventajas de los instrumentos automatizados (García y Navarro, 2008).

Teniendo en cuenta la problemática señalada, este trabajo tiene como objetivos: 1) Proponer un modelo de criterios evaluativos para el diseño de un prototipo de una prueba de EFA, en línea, para un nivel B2 y 2) Evaluar el prototipo de prueba de EFA, en línea, en términos de la validez de apariencia, de contenido y de una aplicación piloto a una muestra de estudiantes extranjeros de ELE.

2. La evaluación del Español con Fines Específicos Académicos (EFA)

En relación con la evaluación del español con fines académicos, no está claro lo que una prueba tendría que evaluar en este ámbito, dado que se necesita delimitar los contextos de uso que debería abordarse en las pruebas, los cuales tienen que dar cuenta de las distintas habilidades que necesitan demostrar los estudiantes en este ámbito. Es sabido que no se puede desarrollar un examen individualizado o dirigido a grupos restringidos, motivo por el cual se requiere de investigaciones que contribuyan en la determinación de los contextos de uso de lengua para fines académicos. Lo que hay a nivel investigativo en cuanto a elaboración y resultados de pruebas con fines específicos no cumple con los estánda-

1. El estudio que se presenta en este artículo se enmarca en el desarrollo de los aspectos metodológicos y empíricos del proyecto Fondecyt 114065.

res definidos en los códigos de práctica por las distintas sociedades de evaluación (Llorián, 2012).

En la evaluación del Español como Lengua Extranjera (ELE), a diferencia del inglés, hay un desarrollo muy incipiente en cuanto a la evaluación para fines específicos, especialmente, para el ámbito académico. En relación con la certificación del EFA, pocos estudiantes extranjeros de intercambio en universidades de países donde se habla el español presentan algún tipo de certificado que acredite su dominio del español con fines académicos. Esto refleja que los actuales sistemas de admisión universitaria no siempre exigen una certificación del español en este ámbito.

En síntesis, en la evaluación del EFA se necesita contar con un modelo evaluativo que delimite aquello que se va a evaluar, definir lo que se considera específico y proporcionar garantías que permitan evaluar de manera válida y confiable el Español como Lengua Extranjera con fines específicos Académicos (Llorián, 2012).

3. Metodología

En relación con la problemática de definir cuáles son los criterios que se debe considerar para elaborar un prototipo de una prueba de EFA válida y confiable, en este trabajo se planteó las siguientes preguntas de investigación:

1. ¿Qué criterios se debe considerar en el diseño de un prototipo de prueba de EFA válida y confiable?

2. ¿Cuál es el comportamiento del prototipo de prueba de EFA en una aplicación piloto con una muestra de estudiantes de ELE?

3.1. Tipo de estudio

Para responder a las preguntas planteadas, se llevó a cabo un estudio mixto de tipo exploratorio secuencial de dos etapas: 1) etapa cualitativa y 2) etapa cuantitativa.

3.2. Etapas del estudio

3.2.1. Etapa 1: Modelo de criterios evaluativos para el diseño de un prototipo de prueba de EFA en línea para un nivel B2

Con la finalidad de responder a la primera pregunta planteada, en esta primera etapa se revisó la literatura especializada en relación con las directrices que proponen los expertos en evaluación de lenguas extranjeras (LE) y los códigos de práctica de las sociedades internacionales de evaluación de LE. De esta manera, se definió el modelo de criterios evaluativos considerados necesarios para elaborar un prototipo de prueba de EFA válido y confiable, como se observa en la figura 1.

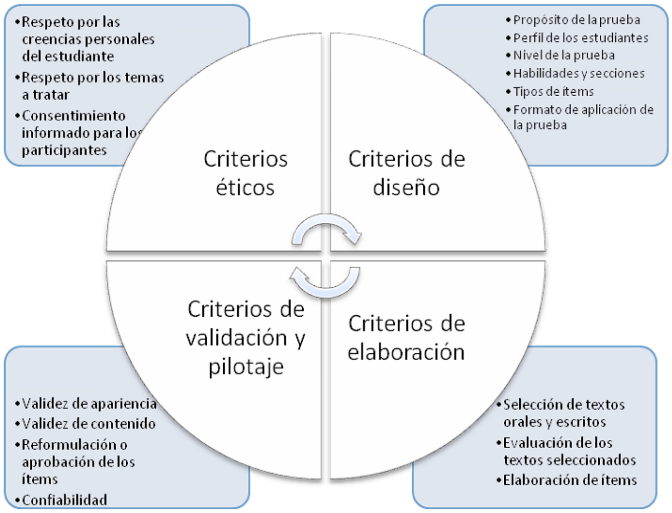


Figura 1: Modelo de criterios evaluativos para el diseño de un prototipo de prueba EFA.

En consecuencia, el modelo propuesto consta de cuatro grandes criterios, los que a su vez precisan una serie de directrices que permiten implementar un prototipo de prueba de EFA en línea para un nivel B2.

3.2.2. Etapa 2 Cuantitativa: Aplicación del modelo de criterios evaluativos en la elaboración del prototipo de prueba EFA en línea

En esta segunda etapa se presenta la operacionalización de cada uno de los criterios del modelo, en la elaboración y validación del prototipo de prueba en línea de nivel B2.

- **Criterios éticos y de equidad.**

  Los criterios éticos se consideraron en la selección de los temas que trataban los textos orales y escritos y las tareas de expresión oral y escrita, para ello se cauteló que las temáticas fueran propias de un ámbito académico y que no privilegiaran determinados contenidos disciplinares, pues lo que se evalúa está en el contexto de las habilidades lingüísticas y no del contenido temático. Asimismo, se tuvo en cuenta que los temas tratados en los textos y tareas no fueran polémicos ni discriminatorios.

- **Criterios de diseño.**

  De acuerdo con los criterios de diseño de la figura 1, el prototipo de prueba EFA es una prueba de proficiencia que evalúa lo que es capaz de hacer (habilidades) el estudiante en cuanto al uso del español en un contexto académico y está orientado a un nivel B2. Para definir el perfil de los estudiantes extranjeros de intercambio hacia los cuales se orienta este instrumento, se consideró los datos regis-

trados por el Programa de Español como Lengua Extranjera de la Universidad de Concepción (ELE-UdeC), que cada semestre atiende a estudiantes de intercambio. De acuerdo con estos datos, la representatividad de las distintas áreas del conocimiento es la siguiente: Letras y Educación (22%), Ingeniería (21%), Ciencias Políticas (16%), Economía (14%), Ciencias Físicas y Matemáticas (13%) y Ciencias Sociales (6%).

En relación con las habilidades lingüísticas evaluadas, se elaboró una tabla de especificaciones donde se definió de manera operativa lo que se entendió por cada una. A modo de ejemplo, se observa en la figura 2 que la expresión e interacción escrita se evalúa en función de la adecuación al género discursivo (en este caso, el ensayo), la coherencia, la corrección y el dominio de vocabulario que presenta el estudiante.

Expresión e interacción orales			
Se comunica con fluidez, precisión y eficacia con hablantes nativos sobre temas académicos. Expone con claridad temas académicos y argumenta sus opiniones y puntos de vista. Participa activamente de una entrevista y amplía y desarrolla sus ideas sobre temas académicos.			
Coherencia	***Fluidez***	***Corrección***	***Vocabulario***
Presenta un discurso claro, comprensible y bien estructurado, con un uso adecuado de mecanismos de cohesión. Si se extiende en su discurso puede mostrar cierta pérdida de control. Mantiene la conversación de manera adecuada, colaborando con el entrevistador.	Se expresa de manera fluida y a un ritmo bastante uniforme, aunque puede dudar mientras busca las palabras adecuadas, las pausas no son largas. Su pronunciación es claramente comprensible, aunque cometa algunos errores esporádicos debido a su	Muestra un alto dominio y control gramatical. Comete errores que no afectan la comprensión y casi todos los corrige.	Tiene un amplio dominio léxico que le permite expresarse adecuadamente sobre diversos temas especializados, sin evidenciar la búsqueda de palabras y utilizando oraciones complejas. Si comete algún error de selección léxica, este no afecta la comunicación.

Figura 2: Ejemplo de tabla de especificaciones del prototipo de prueba EFA.

Cada una de las habilidades se evalúa de manera separada, por lo que este prototipo de prueba se compone de cinco secciones: comprensión lectora, comprensión auditiva, competencia lingüística, expresión e interacción escrita y expresión e interacción oral.

En cuanto a los tipos de ítems, el prototipo de prueba cuenta con tres tipos de ítems distintos para evitar que el instrumento se oriente hacia un método o a un tipo de estudiante (Alderson *et al.*, 1998). Se consideró los ítems de corrección objetiva de verdadero/falso y de selección múltiple para las secciones de comprensión lectora y auditiva, y las tareas de respuesta abierta para evaluar la expresión oral y escrita.

Este prototipo de prueba se implementó en la plataforma *Moodle* (del inglés *Object-Oriented Dynamic Learning Environment*) con cuatro secciones, dado que la sección cinco, de expresión e interacción oral se aplica de manera presencial. Las opciones avanzadas de *Moodle* permiten el registro de los estudiantes y la programación de la prueba en cuanto a la fecha, la hora y la duración. El formato de aplicación de esta prueba representa un aporte a la evaluación del EFA, dado que permite obtener de manera rápida y eficaz los resultados sobre el nivel de proficiencia de los estudiantes en el EFA. Además, *Moodle* otorga seguridad en diversos aspectos:

restringe el número de intentos del estudiante, por lo que una vez que ha comenzado la prueba no puede detenerla; muestra el tiempo del que dispone el alumno y facilita la navegación. Además, la plataforma da la opción de desactivar el procesador de texto, lo cual permite acceder a muestras de lengua escrita auténticas. Sin duda, este formato promueve una interacción entre el estudiante y la prueba (Figueras y Puig, 2013) y permite dos formas de aplicación: presencial y a distancia.

- **Criterios de elaboración.**

Teniendo definido el diseño del prototipo de prueba se procedió a la búsqueda y selección de los textos orales y escritos, acorde con el perfil de los estudiantes. Se recopiló textos de diversas fuentes: revistas de divulgación académica y científica programas de radio, televisión y entrevistas a expertos. Los textos corresponden a fragmentos de artículos científicos, charlas y ponencias, por lo tanto son representativas las modalidades expositiva y argumentativa. La longitud de los textos escritos seleccionados fue entre 265 y 550 palabras y la de los orales, entre 1 y 3 minutos. En estos textos se utiliza un lenguaje y un estilo académico. En los audios se cauteló que hubiera una articulación y velocidad adecuadas para el nivel B2. En algunos casos, se hizo adaptaciones de los textos orales y escritos manteniendo su organización discursiva (coherencia y cohesión). De acuerdo con los criterios de diseño, se recopiló un conjunto de textos, de los cuales finalmente se seleccionaron diez textos orales y diez escritos, que se consideraron más adecuados en términos de pertinencia, contenido y formato para el nivel B2 de EFA. Estos diez textos fueron evaluados por un grupo de diez expertos (5 profesores de segundas lenguas con grado de Magíster en Lingüística, 2 estudiantes de Magíster en Lingüística Aplicada y 3 Doctores en Lingüística) a través de una escala Likert con el objeto de validar en qué medida estos materiales eran aptos para evaluar el nivel de proficiencia B2 de EFA. La evaluación de los textos consideró tres criterios: pertinencia, contenido y formato.

Acorde con el juicio de expertos, se seleccionó los tres textos que presentaban un 80% de la categoría muy de acuerdo, de la escala Likert, en los tres criterios definidos: pertinencia, contenido y formato. A partir de estos resultados se seleccionó tres textos para la comprensión lectora y tres para la comprensión auditiva, como se muestra en la tabla 1.

Sección	Géneros académicos y temas	Tipo de ítem	Longitud/ duración
Comprensión lectora	Texto 1: *Abstract* (Sistema de competencias sustentables para el desempeño profesional en ingeniería).	5 ítems verdadero/ Falso.	265 palabras.
	Texto 2: Fragmento de artículo científico (El aprendizaje de una segunda lengua).	5 ítems selección múltiple.	338 palabras.
	Texto 3: Reseña (La biodiversidad).	5 ítems selección múltiple.	517 palabras.

Comprensión auditiva	Audio 1: Exposición (Energía eólica).	5 ítems verdadero/ Falso.	1:22 minutos.
	Audio 2: Exposición (Proyecto Agua Copiapó).	5 ítems selección múltiple.	1:36 minutos.
	Audio 3: Exposición (Los biocombustibles).	5 ítems selección múltiple.	1:57 minutos.
Competencia lingüística	Ensayo (El cobre y el futuro de Chile).	20 ítems de selección múltiple.	227 palabras.

Tabla 1: Secciones del prototipo de prueba EFA en línea de nivel B2.

Este prototipo de prueba en línea tiene una duración total de 1 hora, 45 minutos. Las secciones de comprensión lectora, auditiva, competencia lingüística y de expresión e interacción escrita se rinden en el computador, y la expresión e interacción oral se realiza de manera presencial una vez que el estudiante ha terminado la prueba en línea. Cabe señalar que esta habilidad también podría evaluarse a través de *Skype* si el estudiante se encontrara en su país de origen o no pudiera asistir a rendir la prueba de manera presencial.

La sección de comprensión lectora se compone de tres textos y la sección auditiva de tres audios. En el primer texto de cada una de estas secciones se plantea cinco preguntas de verdadero y falso, y en los textos 2 y 3 se plantea cinco preguntas de selección múltiple por cada texto, por lo que en total cada sección cuenta con 15 ítems de corrección objetiva. La sección de competencia lingüística comprende un texto con veinte ítems de selección múltiple donde se evalúa los diferentes usos gramaticales definidos para el nivel B2 en el contexto del EFA. Los ítems de respuesta abierta se consideraron para evaluar la expresión e interacción escrita y oral. En la sección de expresión e interacción escrita, el estudiante debe escribir un ensayo. Para ello, se propone dos temas distintos, el alumno elige uno y escribe el texto según el formato entregado. En la expresión e interacción oral, el estudiante tiene que cumplir con tres tareas de una entrevista: 1) presentarse, 2) exponer sobre un tema a elección y argumentar un punto de vista sobre dicho tema y 3) conversar con el entrevistador sobre el tema expuesto. Cada una de estas tareas abiertas tiene 16 puntos, distribuidos en cuatro criterios en una rúbrica analítica.

Una vez que se contó con la selección de textos definitiva para el prototipo de prueba se procedió a la elaboración y redacción de los ítems e instrucciones acorde con los criterios de diseño del modelo.

- **Criterios de validación y pilotaje.**

En este modelo, el criterio de validación y pilotaje se considera en tres niveles: no experta, experta y mediante el pilotaje con una muestra de estudiantes. Por lo

tanto, se considera una validez de apariencia con la participación de un grupo de hablantes nativos; una validez de contenido a través del juicio de expertos, y el pilotaje con una muestra de estudiantes extranjeros de intercambio universitario.

4. Resultados del comportamiento del prototipo de prueba de EFA en línea para un nivel B2

En este apartado se da a conocer los resultados cuantitativos del prototipo de prueba de EFA, una vez que se sometió al proceso de validación de apariencia y de contenido.

4.1. Resultados de la validez de apariencia

En esta primera etapa de la validación participó un grupo de 9 estudiantes de la carrera de Traducción e Interpretación (en su primer semestre de universidad). Ellos respondieron el prototipo de prueba y luego respondieron un cuestionario sobre sus apreciaciones del instrumento. De acuerdo con Alderson *et al*., (1998) y Martínez (2011) es esencial realizar este proceso de validación.

A partir de los resultados obtenidos y con el objeto de observar el comportamiento del prototipo de prueba se elaboró una matriz con el número y porcentaje de aciertos obtenidos en las respuestas en las distintas secciones de corrección objetiva de la prueba. Se revisó aquellos ítems que habían tenido un bajo porcentaje de aciertos (igual o inferior al 56%). De esta forma, se mejoró el prototipo en cuanto a la redacción de los enunciados de los ítems y de algunas opciones de respuesta que eran ambiguas. Del mismo modo, se mejoró parte de la tarea oral, que como se había planteado no elicitaba que los estudiantes hablaran de desventajas sobre uno de los temas propuestos.

4.2. Resultados de la validez de contenido de los ítems del prototipo

En la validez de contenido del prototipo de prueba EFA en línea para el nivel B2 participó un grupo de ocho expertos, profesionales de lenguas (2 estudiantes de Magíster en Lingüística Aplicada, 1 Magíster en Lingüística, 3 candidatos al grado de Doctor en Lingüística y 2 Doctores en Lingüística). De manera individual, cada experto, evaluó a través de una escala Likert de 5 puntos (1: Muy en desacuerdo, 2: En desacuerdo, 3: Indiferente, 4: De acuerdo, 5: Muy de acuerdo) en qué medida los ítems de las diferentes secciones del prototipo de prueba en línea medían las dimensiones definidas en los criterios de elaboración en cuanto a las habilidades lingüísticas. Para ello, se les entregó una pauta y se les explicó en qué consistía la evaluación y se les proporcionó la lista de ítems en las distintas dimensiones. Los evaluadores leyeron la pauta y accedieron a la prueba en línea, con una clave. Enseguida, leían los textos y evaluaban si los ítems medían las distintas dimensiones.

La escala de evaluación entregada a los expertos contiene 35 ítems, distribuidos en las seis dimensiones que evalúa la prueba: dimensión 1: identificación de información literal o específica en textos especializados, dimensión 2: inferencia de las ideas principales de los párrafos, dimensión 3: inferencia de la idea global del texto, dimensión 4:

dominio gramatical nivel B2 de EFA, dimensión 5: modalidades expositiva y argumentativa, dimensión 6: género académico.

Para observar el grado de acuerdo entre los evaluadores se aplicó el coeficiente de acuerdo de Kendall, que arrojó 0.427 con un valor p inferior a 0,05. Esto significa que hubo acuerdo entre los evaluadores en cuanto a que el prototipo de prueba está evaluando seis dimensiones distintas, y los ítems miden lo que dicen medir.

4.3. Resultados del pilotaje del prototipo de prueba EFA

En este primer pilotaje del prototipo de prueba EFA en línea para el nivel B2 participó un grupo de 15 estudiantes de intercambio en la Universidad de Concepción, 8 hombres y 7 mujeres, cuyas edades fluctúan entre los 20 y 27 años. En cuanto a sus lenguas maternas, hay 4 de alemán, 4 de francés, 2 de sueco, 1 de italiano, 1 de checo, 1 de inglés, 1 de ruso y 1 de portugués. En relación con las áreas de estudio, en la muestra hay 9 estudiantes del área de Ingeniería, 2 de Administración y Economía, 1 de Ciencias Químicas, 1 de Ciencias Biológicas, 1 de Ciencias Políticas y 1 de Letras. La tabla 2 presenta los resultados obtenidos por la muestra.

N° sujeto	C. lectora 25%	C. auditiva 20%	C. lingüística 10%	E.I escrita 25%	E.I oral 20%	Total (100%)
1	25%	16%	8%	14%	12%	**74%**
2	22%	15%	9%	14%	16%	**75%**
3	18%	9%	7%	16%	14%	**64%**
4	15%	12%	9%	14%	8%	**57%**
5	17%	12%	9%	16%	16%	**68%**
6	18%	13%	8%	13%	15%	**67%**
7	10%	16%	7%	13%	14%	**60%**
8	17%	15%	8%	10%	13%	**62%**
9	18%	11%	9%	19%	16%	**72%**
10	20%	9%	6%	9%	4%	**49%**
11	12%	16%	8%	16%	11%	**63%**
12	17%	9%	7%	11%	11%	**54%**
13	13%	9%	8%	7%	13%	**50%**
14	20%	13%	7%	9%	12%	**62%**
15	17%	11%	7%	6%	10%	51%
Promedio	**17%**	**12%**	**8%**	**13%**	**12%**	**62%**

Tabla 2: Resultados obtenidos por la muestra de estudiantes de intercambio.

Para conocer el rendimiento real de cada sujeto se calculó el porcentaje de logro por cada sección, como se especificó en el criterio de diseño del modelo evaluativo. La comprensión lectora equivale a un 25% del total de la prueba, la comprensión auditiva a un 20%, la competencia lingüística a un 10%, la expresión e interacción oral a un 20% y la expresión e interacción escrita a un 25%.

En términos generales se aprecia que el mayor logro obtenido por el grupo fue en la competencia lingüística con un 8% del 10% exigido, seguido de la comprensión lectora con un promedio de un 17% y en la comprensión auditiva con un 12%, del 20% exigido. Por el contrario, las habilidades donde mostraron un bajo rendimiento fueron la expresión e interacción escrita con un 13% y la expresión e interacción oral con un 12%. Si bien la interacción oral no representa un problema en la vida cotidiana de un estudiante de nivel intermedio, que está en un periodo de intercambio, sí tiene implicancias en un contexto académico.

Teniendo en cuenta que las universidades chilenas exigen un 60% como mínimo para la aprobación de las asignaturas y que los estudiantes se insertan en este sistema educativo y evaluativo, en este prototipo de prueba se consideró un 60% de exigencia. Es decir, aquellos que muestren un 60% o más de logro en este prototipo de prueba podrían considerarse como estudiantes que presentan un nivel B2 de EFA, y por lo tanto, estarían en mejores condiciones para enfrentar las diversas tareas y actividades que demanda el ambiente académico.

Acorde con la exigencia del 60% para aprobar como estudiante de B2, se observa que de los 15 estudiantes, solo 10 están mejor preparados para insertarse a las distintas actividades que exige el ambiente académico, por ejemplo, comprender textos orales y escritos, atender y participar de las clases, hacer exposiciones, etc. Estos resultados están en concordancia con lo señalado por Pastor (2006, 2010), en cuanto a que los estudiantes que llegan a las universidades españolas no siempre presentan un nivel B2, y que sería el ideal que las universidades exigieran algún tipo de certificación, pues así se podría garantizar que el estudiante pudiera comprender y participar en las clases adecuadamente.

5. Conclusiones

Los resultados de este trabajo evidencian que el modelo de criterios evaluativos ha permitido elaborar un prototipo de prueba de EFA en línea para un nivel B2 con validez de apariencia y de contenido. Los datos cualitativos y cuantitativos de este prototipo de prueba de EFA sugieren los criterios y procedimientos necesarios para que, en un futuro, se diseñe y elabore una prueba de proficiencia de EFA que permita evaluar a estudiantes que requieran precisar su dominio en el Español como Lengua Extranjera con fines específicos Académicos.

En cuanto al pilotaje, se necesita continuar evaluando el prototipo de prueba, para tener mayor variabilidad en los resultados y poder determinar los niveles de dificultad y de discriminación que presentan los ítems. Del mismo modo se prevé aumentar el

número de ítems con el objeto de poder, a largo plazo, analizar los ítems desde una perspectiva de la Teoría de Respuesta al ítem (TRI).

El modelo propuesto resulta interesante, puesto que evidencia desde los procesos iniciales de planeación del prototipo de prueba en línea hasta las etapas finales de elaboración y validación, cuestión que los expertos reclaman a las instituciones que actualmente ofrecen exámenes de EFA (Llorián, 2012). No obstante, hay que señalar que es difícil proporcionar un modelo de elaboración y validación de un prototipo de prueba de EFA que considere todas las variables y pueda ser replicado sin modificaciones, dado que son múltiples los escenarios y realidades educativas, el modelo puede ser mejorado o adaptado en función de las necesidades del usuario.

Referencias Bibliográficas

ALDERSON, CH.; CLAPHAM, C. y WALL, D. (1998): *Exámenes de idiomas*, Madrid: Cambridge University Press.

FIGUERAS, N. y PUIG, F. (2013): *Pautas para la evaluación del español como lengua extranjera*, Madrid: Edinumen.

GARCÍA, J. y NAVARRO, C. (2008): "Automatizar los exámenes del Diploma de Español como Lengua Extranjera (DELE): ¿Utopía o realidad?", *Revista de Lingüística Teórica y Aplicada* 46 (2), 81-93.

LLORIÁN, S. (2012): "Los retos de la certificación del español con fines específicos", *Revista de Nebrija de Lingüística Aplicada,* 12 (6), 3-34.

MARTÍNEZ, A. (2011): *La evaluación de las lenguas: garantías y limitaciones*, Granada: Ediciones Mágina.

MENDOZA, A. (2015): "La selección de tareas de escritura en los exámenes de lengua extranjera destinados al ámbito académico", *Revista de Nebrija de Lingüística Aplicada, 18* (9).

PASTOR, S. (2006): "La enseñanza del español como lengua vehicular en contextos académicos", *MarcoELE*, n.º 2.

PASTOR, S. (2010): "Enseñanza de español con fines profesionales y académicos y aprendizaje por contenidos en contexto universitario", *Testi e Linguaggi*, 4, pp.71-88.