

ANÁLISIS DEL CONTENIDO Y EVIDENCIAS DE VALIDEZ DEL CONSTRUCTO EN LOS EXÁMENES DE ESPAÑOL CON FINES GENERALES Y ACADÉMICOS

CONTENT ANALYSIS AND CONSTRUCT VALIDITY EVIDENCES IN SPANISH TESTS WITH GENERAL AND ACADEMIC PURPOSES

SUSANA LLORIÁN GONZÁLEZ
Universidad Complutense de Madrid
sllorian@ucm.es

RESUMEN

La especificidad del contenido de los cursos y de los exámenes de lenguas extranjeras o segundas para usos en contextos académicos y profesionales ha sido tradicionalmente una cuestión controvertida. El género discursivo se ha convertido en la piedra angular de este segmento de la didáctica y en la clave principal de la especificidad (Belcher, 2009). Ninguno de los sistemas de certificación del español que se declaran válidos para acceder a programas universitarios considera la noción de género discursivo en la descripción de su contenido y tampoco conciben en términos de especificidad los procesos de comprensión de textos que se emplean en el contexto académico. En este panorama cabe preguntarse si se necesitan pruebas que evalúen una habilidad comunicativa específica para el ámbito académico o si la capacidad de los candidatos para desenvolverse lingüísticamente en estos contextos puede medirse con los exámenes generalistas. En este artículo se exponen los objetivos, los procedimientos metodológicos y los principales resultados de una investigación orientada a demostrar, a través de la aportación de evidencias de validez del contenido (validez contextual) y la validez del constructo (validez cognitiva), la necesidad de emplear exámenes específicos para discriminar a los candidatos capaces de cursar estudios universitarios en español, facultad de la que carecerían los exámenes certificativos generalistas. Se emplean procedimientos de juicio experto y análisis factorial confirmatorio en relación con las evidencias de validez de contenido y del constructo, respectivamente. Los hallazgos muestran diferencias significativas entre los resultados de pruebas de exámenes de tipo generalista y los de pruebas con constructo específico del ámbito académico.

Palabras clave: Exámenes certificativos, evaluación con fines académicos, validación de pruebas, comprensión auditiva académica.

ABSTRACT

The content specificity of foreign languages with academic and professional purposes courses and tests has traditionally been a controversial issue. The notion of genre has become the main key of specificity (Belcher, 2009). None of the Spanish certification exams systems that declare themselves valid for access to university study programs takes into account the idea of textual genre when defining their content; neither, the comprehension processes of academic text are not considered in specific terms in that kind of exam. In this scenario, it is necessary to ask whether there is a need for tests that assess a specific communicative ability to the academic field or the ability of candidates to function in these contexts can be measured with general tests. This paper presents the objectives, the methodological procedures and the main results of a research aimed at demonstrating, through evidences of content and construct validity, the need to develop specific academic tests that serve to identify the candidates capable of studying with Spanish as communication vehicle at universities, a faculty that general exams would lack. Expert judgment procedures and confirmatory factor analysis are used in relation to content and construct validity evidences. Research findings show significant differences between the results of tests with general purposes and those of tests with specific academic construct.

Keywords: High-stakes certification tests, academic purposes tests, tests validation, academic listening.

Recibido: 29/04/2019. *Aceptado:* 18/11/2019.

1. INTRODUCCIÓN

En el panorama actual de las universidades de los países de habla hispana, en el que imperan las políticas de internacionalización, por regla general, no se requieren certificados específicos que acrediten grados de dominio de español académico como filtro para admitir a los alumnos extranjeros en los programas de grado o de posgrado. Tampoco es habitual que se administren exámenes de elaboración propia, capaces de discriminar a los estudiantes que pueden cursar con éxito estudios universitarios empleando el español como vehículo de comunicación.

Esta descripción guarda perfecta coherencia con la situación general de la certificación del español como lengua extranjera o segunda, que está muy poco diversificada. A excepción del examen del *prototipo de prueba de multinivel EFA* (Ferreira, 2016), desarrollado en el marco de los Proyectos Fondecyt 1040500 (Ferreira, 2014) y 1180974 (Ferreira, 2018), (Comes *et al.*, 2016, p. 1715) y del *EXELAA (Examen de Español como Lengua Extranjera para el Ámbito Académico)*, a cargo de la Universidad Nacional Autónoma de México, la Universidad de Guadalajara y la Universidad de Costa Rica, no existen exámenes públicos de español con fines académicos. Lo más grave es que, tal y como lo señala Mendoza (2015),

ningún examen de español con estos fines cuenta con reconocimiento nacional o internacional. Por el contrario, en los países de habla inglesa operan sistemas tan importantes como el *IELTS* (*International English Language Testing System*), el *TOEFL* (*Test of English as a Foreign Language*) o el *PTE Academic* (*Pearson Test of English*), que tienen un enorme impacto en las universidades, en la industria de la enseñanza y aprendizaje de lenguas, no solo académico, sino también económico.

Además de lo anterior, es preciso no perder de vista lo siguiente: los sistemas certificativos de naturaleza generalista de español como lengua extranjera o segunda de amplio espectro de administración (*CELU, Certificado de Español: Lengua y Uso; DELE, Diploma de Español como Lengua Extranjera; SIELE, Servicio Internacional de Evaluación del Español como Lengua Extranjera; eLADE, etc.*) declaran en su documentación informativa, sobre todo en sus páginas electrónicas, que los resultados de sus exámenes sirven para este propósito, es decir, como filtro para acceder a las universidades, además de otros muchos fines.

El hecho de que los resultados de un solo examen se puedan utilizar para múltiples propósitos hace aflorar el principal problema, si no el único, de la evaluación certificativa: la validez. Según la concepción actual que de esta noción mantiene la máxima autoridad en esta materia, la American Psychological Association (APA), a través de sus *Standards for Educational and Psychological Testing* (APA, AERA, NCME, 2014), se trataría de demostrar, con apoyos anclados, por un lado, en la evidencia y, por otro, en la teoría que los resultados de los exámenes (aprobado o suspenso, apto o no apto, etc.) sirven —es decir, valen— para el propósito con el cual los exámenes han sido diseñados (admisión en programas de estudios, clasificación en los cursos de apoyo, acreditación de un grado de dominio, etc.).

En este escenario, surge la pregunta principal: «¿Se necesitan en español exámenes con fines académicos que impacten de forma positiva en los cursos de apoyo a *los programas universitarios o pueden cumplir esta función los exámenes generalistas?*». Para hallar la respuesta, es preciso definir primero cuál es el constructo de los exámenes con fines académicos; en definitiva, describir qué es lo que se evalúa realmente con ellos y contrastarlo con el constructo de los exámenes generalistas.

De esta pregunta se deriva el objetivo general que guiará el desarrollo del trabajo de investigación que aquí se describe: elaborar una prueba *ad hoc* cuyas tareas constituyan la operacionalización de un constructo específico del ámbito académico y demostrar, con evidencias de validez de diferentes fuentes, que este se diferencia en mayor o menor grado del que miden las pruebas de los exámenes generalistas. Una vez desarrollada la prueba específica, la investigación se centra en dar respuesta a las otras dos preguntas siguientes, que son una concreción de la anterior:

- *¿En qué medida el contenido de la prueba específica, que evalúa la habilidad comunicativa académica, se diferencia del de una prueba de un examen generalista?*

- ¿Se corresponde la estructura factorial de la prueba específica con el constructo se declara que subyace a su diseño?

La base teórica de la argumentación de la validez hay que buscarla en las dos áreas de conocimiento que confluyen en el planteamiento de la investigación: por un lado, los exámenes con fines específicos, en concreto, los que se desarrollan en el ámbito académico; por otro lado, la enseñanza del español académico en que el problema de la especificidad es una constante, que se soluciona progresivamente gracias a los descubrimientos de los últimos años propiciados por la asistencia de la metodología de corpus. Las investigaciones recientes apuntan a que esta especificidad reside principalmente en las necesidades de uso de la lengua y en el concepto del género discursivo (Woodrow, 2018, p. 11; Hyland y Shaw, 2016. p 4; Flowerdew, 2011, p. 124; Belcher, 2009, p. 29).

Por estas razones, se desarrolló una fase previa de la investigación, de tipo descriptivo exploratorio (Llorián, 2015), destinada a indagar en las necesidades de la población de interés, es decir, los estudiantes extranjeros de programas de intercambio, respecto a los géneros discursivos a los que se asocian los textos que procesan en español dentro del ámbito universitario. De este sondeo, emerge, entre otros, un hallazgo fundamental: el principal escollo con el que se encuentran los alumnos de los programas de intercambio radica en la comprensión de las clases magistrales. Estas percepciones en términos de dificultad de la comprensión de las clases se corresponden, en términos generales, con los resultados de las investigaciones de Pastor y Pandor (2017), centradas en alumnos estadounidenses y con lo referido en el estudio de Ferreira y Elejalde (2019) en que se señala que la mayoría de los estudiantes encuestados comprenden con cierto grado de dificultad las clases de los profesores. Por ello, la prueba de examen diseñada *ad hoc* sobre la que se desarrollará la investigación estará basada en la comprensión del género de la clase magistral, que consideramos como actividad comunicativa de la lengua de comprensión audiovisual, debido a la cantidad de *input* visual que aporta el texto de entrada y a su importancia (Geranpayeh y Taylor, 2013).

2. MARCO TEÓRICO

El soporte basado en la teoría al que alude la definición de validez de los estándares de la APA (APA, AERA, NCME, 2014), más arriba mencionado, hay que buscarlo, en este caso, en el despliegue de todas las áreas de conocimiento que convergen en esta investigación. En primer lugar, en los modelos que se han descrito en torno a la validación de exámenes. En segundo lugar, en el

constructo las pruebas específicas del ámbito académico. Este, en el caso de la comprensión, se define, por un lado, a partir de los rasgos que caracterizan el género discursivo de los textos de entrada y, por otro, de las peculiaridades que presentan los procesos de comprensión. Nos ocupamos seguidamente de cada uno de estos aspectos.

2.1. Los procesos de validación de pruebas de examen

En lo referente a los procedimientos de validación, según Xi y Sawaki (2017, p. 193), es necesario dejar clara la forma en que se conceptualiza la idea de validez, puesto que esta determina la finalidad y la naturaleza de las investigaciones y, por consiguiente, los métodos utilizados para recoger evidencias. Le (2011, p. 9) distingue dos paradigmas en los modelos validación de pruebas de examen. Ambos están fundamentados en la idea de la validez como concepto unitario aportada por Messick (1989), germen de la definición de la APA. El primer paradigma, por el que aquí nos decantamos, es el de acumulación de evidencias; frente a este, se sitúan los modelos que se basan en la construcción de argumentos de validez. La acumulación de evidencias, como su propio nombre indica, consiste en recabar pruebas, a lo largo de las diferentes etapas del ciclo de vida del examen, que confieran soporte a las distintas facetas en torno a las cuales se configura el concepto de validez.

Los principales representantes de este enfoque de acumulación de evidencias en el área de la evaluación de lenguas extranjeras y segundas son, por una parte, Bachman (1990) y, por otra, Weir (2005), en cuyo modelo sociocognitivo se incardinan los procedimientos metodológicos de nuestra investigación. Este modelo es el que se emplea en los trabajos de validación del examen más investigado de la historia de la certificación de lenguas extranjeras: el *IELTS* (Taylor y Weir, 2012; Geranpayeh y Taylor, 2013). El modelo sociocognitivo se describe someramente en el epígrafe siguiente, con énfasis en las fuentes de evidencia en las que se centra este artículo.

Otros sistemas de exámenes certificativos de inglés con fines académicos de la misma envergadura que el *IELTS* utilizan enfoques de validación basados en la construcción de argumentos. Es el caso del *TOEFL* (Chapelle, Enright y Jamieson, 2008) y también de trabajos de investigación como el de Aryadoust (2013) o el de Le (2011), que construye un argumento de validez referido a pruebas de clasificación utilizadas en universidades americanas. En fechas recientes, Bachman y Palmer (2010) y Knoch y Elder (2013) presentan sendos marcos de validación de pruebas dentro de este mismo paradigma.

2.2. Los exámenes de lenguas con fines académicos: exámenes universitarios de admisión y post admisión

El desarrollo de exámenes que evalúan la capacidad de los candidatos para desenvolverse en los contextos universitarios empleando una lengua extranjera o segunda como vehículo de comunicación tiene origen en el ámbito anglosajón y se debe a la afluencia masiva de estudiantes extranjeros a las universidades de países de habla inglesa, en concreto, el Reino Unido, los Estados Unidos de América, Sudáfrica, Australia y Nueva Zelanda. Entendemos por exámenes de admisión, siguiendo a Read (2015, 2016) y a Elder (2017) aquellos que se utilizan como filtro (Davies, 2008, p. 3) para permitir matricularse en las universidades únicamente a los alumnos competentes en la *habilidad comunicativa académica*. Los exámenes de post admisión, según esta misma línea terminológica, son los que se administran a los alumnos ya matriculados, con el fin de diagnosticar el grado de desarrollo de esta habilidad comunicativa académica y de tomar decisiones, en consecuencia, acerca de su posible canalización hacia cursos de apoyo o refuerzo lingüístico con programas de lengua con fines académicos. Weigle y Malone (2016, p. 609) dejan fuera de la categoría de exámenes que se utilizan para evaluar el aprovechamiento de los cursos que se desarrollan con estos fines. La escasa tradición con la que cuentan los exámenes de español con fines académicos, más arriba comentada, hacen que todavía no exista en las universidades españolas este tipo de distinción. Por consiguiente, seguimos la corriente anglosajona.

Davies (2008) fundamenta la especificidad de este tipo de exámenes en un constructo que consiste en la habilidad para utilizar, en un discurso adecuado, la variedad del inglés propia de los contextos académicos (Davies, 2008, p. 1). Esta variedad se daría en las manifestaciones orales y escritas de la lengua, y en los planos productivo y receptivo. Afecta a usos generales y a otros más específicos, propios de los diferentes campos de conocimiento disciplinar. Se trata de una habilidad aprendida durante el proceso de alfabetización académica. Read (2015) puntualiza y diferencia las variedades de lengua asociadas a áreas disciplinares (lenguas de especialidad) de la alfabetización académica. Esta sería, para él, común a todos los campos de conocimiento y se basaría en la capacidad del usuario para dominar una serie de funciones retóricas (definición, ejemplificación, digresión, etc.), además de la lista de palabras frecuentes del inglés académico (*Cox head academic wordlist*, Coxhead, 1998), que no tiene todavía ningún equivalente en español.

Esta forma de entender la especificidad de los exámenes con fines académicos afecta directamente al tema principal que aquí nos ocupa: la definición del contenido, y en relación con este, los procedimientos que se utilizan para la evaluación en estos exámenes. Elder (2017, p. 281), Weigle y Malone (2016, p. 611) y Wei-

deman, Patterson y Pot (2016, p. 179) llaman la atención sobre lo complicado que resulta representar el constructo de estos exámenes, debido a la complejidad que entraña el análisis de la lengua utilizada en estos contextos y los múltiples factores que determinan la variación, que se da además en diferentes grados: culturales, disciplinares, etc. Weideman *et al.* (2016, p. 181) señalan que el constructo de las pruebas debe operacionalizarse a través de especificaciones que lo conecten con las tareas de examen. Según estos autores, en estos exámenes tienen que predominar las tareas que reflejen el uso de la variedad de lengua que se utiliza en este tipo de contextos. A la dificultad de analizar y describir la lengua, se añade la interrelación que se produce entre las habilidades cognitivas generales, relacionadas con la lengua, y las específicas de este ámbito; por ejemplo, entre los procesos de comprensión auditiva de textos académicos y la capacidad para interpretar un gráfico.

En definitiva, es necesario definir un constructo específico, la *habilidad comunicativa académica*, que tenga en cuenta todo lo anterior: las peculiaridades de la lengua empleada en estos contextos, junto con los procesos y las habilidades implicados en la producción y en la recepción de textos orales, escritos y audiovisuales. Para ello se utilizó, para lo primero, de la noción de género discursivo y, para lo segundo, de un modelo de comprensión auditiva.

2.3. El género discursivo de los textos de entrada

El género discursivo se ha establecido como la unidad de análisis de los cursos de lenguas extranjeras para su uso en contextos académicos. No está tan claro, sin embargo, que se considere como base para la definición del contenido de los exámenes certificativos de este ámbito, tal y como constatan Ainciburu y Rodríguez (2015). El concepto de género cuenta con la ventaja de que, gracias a él, se puede dar cuenta de la forma en que los rasgos del discurso académico se ponen de manifiesto en los diferentes tipos de texto, a través de las convenciones adoptadas por las comunidades discursivas que los producen y que determinan también los patrones léxicos y gramaticales, las normas ortotipográficas, además de las funciones retóricas, el registro, la textura discursiva y la macroestructura.

El llamado discurso académico estaría conformado por el conjunto de textos que tratan temas propios de las distintas áreas disciplinares. Estos textos se caracterizan por la coocurrencia sistemática de unos rasgos lingüísticos concretos, cuya aparición variará en virtud de los grandes campos del conocimiento (técnico, humanístico, etc.), aunque en este sentido las investigaciones sobre el español, que precisan metodología de corpus, no se hallan tan desarrolladas como en inglés. Los trabajos recientes (Pastor y Ferreira, 2018; Pastor, 2016, 2010; Cademartori, Parodi y Venegas, 2006; Parodi, 2004, 2006, 2010 o Regueiro y Sanz, 2013, entre

otros) revelan, sin embargo, que existe una serie de rasgos inherentes a los géneros de los textos que se desarrollan en el ámbito académico dentro de los territorios de habla hispana. Los textos se caracterizan, por ejemplo, por el uso de un registro formal o por la ausencia de marcas que revelan la relación que establece el autor del texto por su contenido; asimismo, se evita la ambigüedad y se aboga por la precisión léxica y por la eliminación de redundancias; se observa también un alto grado de elaboración sintáctica, una preferencia por el orden canónico de los constituyentes de la oración, una presencia llamativa de nominalizaciones deverbales, etc., además de otros muchos rasgos. La especificación de las pruebas de examen de español con fines académicos requiere un análisis de la forma en que estos rasgos del discurso académico español se reflejan en los textos asociados a los géneros que son significativos para los candidatos.

Más arriba hemos esgrimido las razones por las cuales nuestra investigación se centra en el género discursivo de la clase magistral y en su comprensión. Vázquez (Coord.) (2001) es responsable del único trabajo de investigación que se realiza sobre este género en español, en el marco del proyecto ADIEU. Pese a la calidad del estudio, lo alejado de la fecha, la metodología empleada —que no utiliza corpus informatizados— y la restricción de la muestra de textos objeto de análisis —de escaso volumen y limitada a las humanidades— apelan a una necesidad imperiosa de desarrollo de corpus orales y a la consiguiente actualización de la investigación. El panorama que presentan los estudios de este género en lengua inglesa es radicalmente diferente. La existencia de corpus como *BASE (British Academic Spoken English)* o *MICASE (Michigan Corpus of Academic Spoken English)* ha permitido que se desarrollen multitud de investigaciones que facilitan un profundo conocimiento de este tipo de texto en la cultura anglosajona: macro- y microestructura, fraseología, funciones retóricas, estrategias discursivas, etc.

Lo más destacable del trabajo de Vázquez (Coord.) (2001) en relación con nuestra investigación es el análisis de la macroestructura de las clases (Ortega, 2001, pp. 17-27), en las que el autor diferencia cinco partes: *preámbulo*, *planteamiento del tópico general*, *explicación*, *interacción previa al cierre de la lección* y *cierre de la lección*. El estudio se completa con una clasificación de los modos de presentar la información: *información básica*, *información básica particular*, *información de fondo*, *información parentética*, *información resaltada*, *información conclusiva*, *información secundaria* y de carácter *interactivo*. Ortega (2001, p. 19) se detiene finalmente en las funciones retóricas características del género: *preguntas*, *reiteración*, *contraste*, *argumentación*, *ejemplo*, *aclaración*, *resumen*. Se ponen en relación con cada una de estas funciones los recursos lingüísticos y paralingüísticos que las caracterizan. Todos estos elementos se tienen en cuenta en la definición del constructo de la prueba específica desarrollada *ad hoc*.

2.4. Un modelo de comprensión auditiva específica del ámbito académico

Esta investigación toma como referencia para la descripción de la comprensión auditiva académica el modelo teórico de comprensión denominado *construcción-integración* (*construction-Integration model*), que se inspira en el planteamiento de Kintsch y Van Dijk (1978). La elección se justifica por los antecedentes de utilización en exámenes académicos. Las primeras aplicaciones a la evaluación se deben a McNamara y Kintsch (1996), primero, y a Khalifa y Weir (2009), más tarde. Más recientemente, Moore, Morton y Price (2012) se fundamentan en él para argumentar sobre la validez basada en la teoría y la validez del constructo de la prueba de comprensión de lectura del examen *IELTS* y analizan las tareas, en virtud de si los ítems que la componen se focalizan en el *nivel local* de la comprensión del texto o en el *nivel global*. Junto a este parámetro de análisis, establecen el de *interpretación literal* del contenido del texto o el de *interpretación inferencial*.

Kintsch (1998), en las primeras versiones, parte de la idea de que la comprensión de un texto se realiza en dos estadios. El primero, *construcción*, consiste en configurarse una representación del significado del texto, aproximada, en un plano local. Esta representación se produce a partir del *input* textual y de las metas de la comprensión, además del conocimiento previo del que escucha; en el segundo estadio, el de la *integración*, se rechazan las construcciones, realizadas en el nivel local, que resulten inapropiadas, en favor de las que encajan en un todo coherente. Del primer estadio brotan respuestas (*output*) desordenadas, incluso incoherentes, que se derivan directamente del texto de entrada. En esta fase, se ponen en juego el texto, el conocimiento del tema y la experiencia. La construcción del significado del texto de entrada que tiene lugar en esta fase se ve limitada por la memoria de trabajo en tiempo real. Hasta el estadio de *integración* no se logra una reconstrucción completa del significado global del texto y de su estructura.

Por lo demás, el modelo se fundamenta en el principio de que el significado del texto se compone una estructura semántica, es decir, una red de proposiciones que se articula en un predicado y varios argumentos. Esta red proposicional configura dos tipos de estructura en el texto: la microestructura (estructura local del texto) y la macroestructura (estructura global del texto), que se deriva de la anterior. Esta última representa el significado «concentrado», en forma resumida, es decir, la idea general. La captación de esta idea, que requiere la extracción de la esencia semántica y de las relaciones entre las ideas principales, implica la realización de una serie de operaciones de inferencia, más que de reconocimiento, y se logra en la fase de integración. El primer estadio genera un nivel de comprensión superficial que, sin embargo, impide la utilización de lo comprendido.

Apuntamos finalmente que el modelo *construcción-integración* requiere una re-

flexión en dos niveles de comprensión: el que se basa en el texto y el modelo de situación. El nivel basado en el texto es necesario para que el usuario se represente la red proposicional, en los planos macro- y microestructural, que residen en el propio texto. El modelo situacional, por su parte, es una representación del texto, que se encuentra ya integrada con el propio conocimiento sobre el tema o sobre el área de especialidad. Esta descripción del modelo teórico permite intuir la adecuación a la idea de género discursivo.

3. ANTECEDENTES

Nos detenemos brevemente en los antecedentes a nuestro trabajo, en concreto en los que se focalizan en la validación de pruebas de examen. En la misma línea de nuestra investigación, no podemos eludir el estudio de Bachman, Davidson y Ryan de 1995, por su trascendencia, en el que se comparaba el examen *TOEFL* (versión anterior a la reforma del año 2000), con el examen generalista *First Cambridge Certificate*. Pese a la expectativa en el sentido opuesto, se hallan más similitudes entre los dos exámenes de lo previsto. Aunque tenemos muy en cuenta la metodología empleada en este trabajo es preciso tener en consideración que la versión del examen *TOEFL* que se emplea es muy diferente de la actual y está muy alejada de presupuestos comunicativos. El resto de los trabajos comparativos entre sistemas generalistas y académicos se orientan a la argumentación de la validez predictiva. Es importante albergar cierta cautela en relación con ellos, puesto que son muchas y muy diversas las variables que influyen en los resultados académicos.

Al margen de estudios comparativos, son muchos los trabajos que se desarrollan con el fin de demostrar la validez de los procedimientos utilizados en exámenes de alto perfil y de amplio espectro de administración. El que se toma habitualmente como referencia es el que desarrollan Chappelle *et al.*, del año 2008, en el que se aportan diversas evidencias de validez de la versión actual del *TOEFL*. Pese a su reciente irrupción en el mercado, es llamativa la profusión de estudios destinados a demostrar la validez del examen *PTE* de Pearson. Read (2015) muestra cómo los exámenes locales post admisión de las universidades de habla inglesa son objeto de intensa investigación. Esta se orienta a la resolución de problemas de orden práctico, sobre todo, de tipo económico.

Señalamos por último que, en los trabajos en los que se extrae la estructura factorial de las pruebas o de los exámenes completos no existe consenso entre los investigadores, debido a la dificultad implícita en la interpretación de los factores. A esto se suman los problemas derivados de la definición de los constructos cuando se trata del uso de la lengua en contextos muy concretos, como ocurre con el académico.

4. METODOLOGÍA DE INVESTIGACIÓN

Para responder a las preguntas específicas de la investigación más arriba formuladas, nos propusimos los objetivos siguientes:

- Recoger evidencias basadas en el contenido de la prueba específica desarrollada *ad hoc*, que se puedan contrastar con las de una prueba generalista que evalúe el mismo aspecto.
- Recabar evidencia de que la estructura factorial de la prueba específica refleja los aspectos del constructo, tal y como se ha definido, que difiere en alguna medida con el de la prueba con fines generales.

4.1. Marco metodológico

La recogida de evidencias de diferente naturaleza se enmarca en nuestro trabajo en el modelo sociocognitivo del profesor Weir (2005). El modelo es de tipo componencial y se relaciona con las diferentes facetas de la validez. Los componentes que lo integran son los siguientes: características del candidato, validez basada en la teoría, validez contextual, validez de la calificación, validez referida al criterio, validez consecuencial. Los que adquieren relevancia para la consecución de nuestros objetivos, derivados de las preguntas de la investigación, son los siguientes: la validez contextual y la validez basada en la teoría.

Weir (2005, p. 56) alude, en relación con la validez contextual, a lo que tradicionalmente se conoce como validez de contenido. Se trata de la faceta más relevante en relación con el foco de este artículo. Las evidencias tendrían que servir para apoyar la demostración de que las tareas de examen son una representación de las que los candidatos necesitan realizar en los contextos reales cuando emplean la lengua como vehículo de comunicación. La correspondencia se demuestra comparando las variables que configuran las tareas, en particular, todo lo referente a las demandas cognitivas que requiere la ejecución: el tipo de discurso o género del texto de entrada, el canal de comunicación, la extensión del texto de entrada o de salida, la naturaleza de la información del texto, el conocimiento declarativo que tienen los candidatos sobre el tema, la relación entre el material de entrada y el material de salida, además de las variables referidas al interlocutor.

La validez basada en la teoría o validez cognitiva (Khalifa y Weir, 2009; Weir, Hawkey, Green y Unaldi, 2008; etc.) está relacionada con el constructo de la evaluación. Para garantizarla es necesario indagar en los procesos cognitivos que subyacen a la realización de las tareas y determinar en qué grado existen facto-

res ajenos al constructo que pudieran estar interfiriendo en la medición. En una prueba de comprensión en el contexto académico universitario, como ocurre con el caso que nos ocupa, será necesario que se plantee una escucha atenta y también selectiva, que diferirá de la que se pone en marcha en situaciones de ocio o entretenimiento (Green, 2017). Además de los objetivos de la escucha, es preciso que se emule de las situaciones reales de comunicación el tipo de procesos que se ponen en juego durante la comprensión.

4.2. Procedimientos de recogida de evidencias

Comenzamos describiendo los procedimientos de recogida de evidencias de validez contextual (análisis del contenido) empleados en la investigación. Estos se orientan a refutar la siguiente hipótesis nula (H_0), a favor de la alternativa (H_1).

- H_0 : El contenido de la prueba específica no refleja ninguna faceta de la habilidad comunicativa académica que la diferencie de la prueba general.
- H_1 : El contenido de la prueba de comprensión audiovisual refleja facetas de la habilidad comunicativa, específicas de la comunicación académica.

Consiste, en términos generales, en un análisis comparativo del contenido de la prueba de comprensión auditiva de un examen generalista y del contenido de prueba específica, desarrollada *ad hoc*, de comprensión académica. Se emplea el procedimiento de juicio experto, para lo que se convoca un panel de jueces expertos, constituido por siete integrantes que responden al siguiente perfil: se trata de profesores de diferentes contextos y de técnicos del Área Académica del Instituto Cervantes. Todos ellos con sólida formación en didáctica del español como lengua extranjera y dilatada experiencia, algunos en enseñanza y otros, en evaluación.

En el análisis comparativo de las dos pruebas los jueces realizan las siguientes operaciones, cuyos resultados reflejan en tres instrumentos de análisis creados específicamente para ello.

1. Se pide a los jueces que relacionen cada uno de los ítems de las dos pruebas con las diez habilidades que se declara que evalúan en las especificaciones de examen: *seguir instrucciones del profesor, captar la idea principal de la lección, identificar tópicos y subtópicos, relacionar el contenido de la clase con el de sesiones anteriores o posteriores*, etc. Las habilidades se proporcionan a los jueces en una lista numerada, que incluye una opción nula (*0. ninguna*), junto con un instrumento en el que tienen que asociar los ítems de las pruebas con el número de la habilidad que se supone que evalúa.

2. Se requiere del panel de jueces que realice un análisis sobre el tipo de focalización en el que se da la comprensión de los ítems (*literal* o *interpretativo*) y sobre el nivel de focalización de las tareas (*global* o *local*), según el modelo de construcción-integración (McNamara y Kintsch, 1996; Khalifa y Weir, 2009). Se toman como variables cuantitativas de escala los cuatro grados en los que se miden los elementos de los extremos de las líneas de continuo, a partir de los cuales se determinan el nivel (*local*, *global*) y el tipo (*literal*, *interpretativo*) de focalización de la comprensión.
3. El panel de jueces busca en los textos de entrada de las dos pruebas (comprensión audiovisual específica y comprensión auditiva) los rasgos del español académico que se identifican en la bibliografía de referencia: *grado de planificación del discurso*, *grado de especialización del tema*, *calificación del género (académico o no académico)*, *grado de denotación del vocabulario*, *grado de formalidad*, etc. En el instrumento que reciben tienen que señalar el grado (de 0 a 3) en que estiman que los textos de entrada de las dos pruebas presentan estos rasgos.

El procesamiento de los resultados se lleva a cabo siguiendo el proceso que describe Bachman (2004, pp. 272-275), a partir del precedente de Bachman *et al.* (1995), en el que se comparan el constructo de dos exámenes, generalista y académico, desde el punto de vista del contenido. Consiste en calcular la media de los juicios y la desviación típica. El criterio que utilizaron los autores es el siguiente: si la diferencia entre las medias de los juicios emitidos sobre cada uno de los exámenes es mayor que la desviación típica, se interpretará como una diferencia estadísticamente significativa. Seguidamente, se hace un análisis cualitativo de los resultados, que se comparan con las especificaciones de examen.

Por lo que respecta a las evidencias de validez basadas en la teoría, que se orientan a la validación del constructo, se obtienen de la realización de un análisis factorial confirmatorio. Este se aplica a la prueba general y a la prueba específica, con el fin de determinar si existen factores comunes o si se aprecian coincidencias. El análisis factorial exploratorio o confirmatorio se emplea con frecuencia para identificar los factores que subyacen a la actuación de los candidatos y confirmar así las hipótesis sobre estos de Xi y Sawaki (2017, p. 201). Por esta razón, como hipótesis alternativa (H_1) es que la estructura factorial de la prueba refleja la descripción del constructo.

4.3. Contexto de la investigación

La investigación se desarrolla en la Universidad Politécnica de Madrid, en el marco de los cursos del programa Prolinter. Se trata, por lo tanto, de alumnos de grado

y de posgrado del ámbito de la ciencia y la tecnología, de estudios de ingenierías y arquitectura.

Las tres pruebas que se describen más abajo se administran, en una sola jornada, a una muestra seleccionada con criterio de contingencia de 60 alumnos. Todos ellos reciben clases de español en cursos con fines generales de niveles B1, B1+, B2 y B2+. Una vez administrada la prueba de gramática y vocabulario que se emplea como instrumento calibrador, se constata que hay alumnos desde A2 hasta C1 (3,3% de cada uno de estos niveles). El resto se sitúan en las bandas intermedias, con la distribución siguiente: B1 (13,3%), B1+ (15%), B2 (55%), B2+ (10%). Se considera, en primera instancia, la variable de la lengua materna de los alumnos, pero su análisis se desecha más tarde, al comprobar que no introduce ningún tipo de sesgo.

4.4. Desarrollo

El diseño del proyecto de investigación prevé tres mediciones distintas, para cada una de las cuales se utilizará un instrumento diferente. La primera actuará como eje calibrador y medirá el nivel de gramática y vocabulario de los informantes a través de una prueba que se solicitó al Centro de Lenguas Modernas de la Universidad de Granada, que este cedió gentilmente. La razón por la cual se recurre a esta herramienta es que se trata de los pocos instrumentos que existen en español, a cuyo proceso de validación subyace un sólido proceso de investigación. La segunda medición se obtiene de administrar una prueba de comprensión auditiva estándar, tomada de un examen generalista de administración pública. Finalmente, se administra la prueba específica desarrollada *ad hoc*, con la que se pretende evaluar la competencia comunicativa empleada en el ámbito académico, de comprensión audiovisual, atendiendo al principio básico de autenticidad (Bachman y Palmer, 2010).

Como se ha adelantado más arriba, el constructo de la prueba específica se fundamenta en la comprensión del discurso académico de naturaleza especializada y en la clase magistral como género, lo cual implica tener en cuenta los rasgos gramaticales, léxicos y fonéticos, además de los discursivos (macro- y microestructura, funciones retóricas, secuencias textuales, registro, etc.). Para definir el constructo de la evaluación se toma como base el concepto de *habilidad lingüística comunicativa* de Bachman y Palmer (2010), que se ajusta al uso de la lengua en contextos académicos. Por lo que respecta a la comprensión, se parte, como se ha indicado, del modelo teórico de comprensión *construcción-integración*. Las tareas de la prueba que se diseñan a partir de este modelo se focalizan en las operaciones cognitivas, es decir, de una lista de habilidades generales y específicas, implícitas en la comprensión del texto académico.

Los cuatro textos de entrada son breves y tienen formato audiovisual. A cada uno de ellos se asocia una o dos tareas. La prueba se compone, en total, de siete tareas. Dos de los textos son grabaciones destinadas a *MOOC* (*Masive Online Open Courses*). Se opta por esta modalidad de texto, debido a que los *MOOC* reflejan los componentes de la clase magistral como género y constituyen unidades completas de significado, con la ventaja de su breve duración, que repercute favorablemente en la fiabilidad; el problema de los textos de entrada de las pruebas que emulan la comprensión de las clases magistrales auténticas es la extensión, con el consiguiente efecto fatiga. Los otros dos textos son fragmentos de clase reales grabadas, que se extraen de Internet. El primero de estos dos está centrado en la focalización en el tópico de la clase; el segundo, en la funcionalidad de una de las partes en que se articula la estructura de las lecciones: comienzos, cierres, etc. Por esta razón, los cortes no resultan determinantes.

Los formatos de los 35 ítems asociados a los textos son todos de respuesta pre-seleccionada. Esto se explica por las necesidades del contexto de administración de la prueba y del modelo de comprensión: se precisan resultados inmediatos, garantías máximas de fiabilidad y se cuenta habitualmente con un número elevado de candidatos.

5. PRESENTACIÓN Y DISCUSIÓN DE LOS RESULTADOS

Presentamos a continuación los resultados, como respuesta a las preguntas de investigación más arriba formuladas. Seguidamente, ofrecemos una breve discusión de los aspectos más relevantes y llamativos en relación con los dos tipos de evidencias de validez recabadas.

La primera pregunta se responde con las evidencias a la validez contextual, es decir, las que se desprenden del análisis del contenido. Los resultados se hacen corresponder con los tres procesos a los que se someten los jueces expertos.

El emparejamiento que realizan los miembros del panel entre la lista de las habilidades de comprensión académica y los ítems de las dos pruebas de comprensión: la prueba específica y la prueba generalista. Salvo una de estas habilidades (*reconocer información irrelevante a partir de bromas, incisos, etc.*), que obtiene una puntuación muy baja, las demás son claramente identificadas por los jueces en la prueba específica. Los jueces no reconocen relación alguna entre las habilidades de la lista y los ítems de la prueba generalista. Llama la atención que ni siquiera hallen relación directa entre los ítems habilidades de comprensión que pueden considerarse de orden general como *captar la idea general* o *resumir el contenido del texto*.

El segundo aspecto analizado es la presencia de rasgos de español académico en los textos de entrada de las dos pruebas. El procedimiento empleado para procesar los juicios, que se han cuantificado es el que emplean Bachman *et al.*, en el

trabajo de 1995, retomado por Bachman (2004). Según este, si la diferencia entre las medias es mayor que las desviaciones típicas, las diferencias entre las dos pruebas estimadas por los jueces son significativas (DS), como sucede con los datos procesados en nuestro trabajo. La Tabla I muestra los resultados de los juicios y el cálculo de las diferencias.

Tabla I. Resultados de los juicios y cálculo de las diferencias.

	Prueba específica		Prueba general		diferencia	
	media	d. típica	media	d. típica		
Discurso planificado	2,3	0,1	0,7	0,3	1,6	DS
Género académico	3	0	0,8	0,1	2,2	DS
Tema especializado	2,4	0,1	0,6	0,2	1,8	DS
Estilo impersonal	2,1	0,3	0,6	0,2	1,5	DS
Duración extensa	0,08	0,2	0,6	0,2	0,5	DS
Vocabulario técnico	2,2	0,4	0,5	0,1	1,7	DS
Vocabulario denotativo	2	0,2	0,5	0,1	0,5	DS
Léxico preciso	2,5	0,1	0,8	0,2	1,7	DS
Elaboración Sintáctica	2,3	0,2	0,8	0,1	0,5	DS
Registro formal	2,4	0,3	1	0,1	1,4	DS

Para el análisis centrado en el nivel y en el tipo de focalización de la comprensión, se emplea el mismo procedimiento. Si bien es cierto que, aun en la prueba de comprensión específica no son muchos los ítems que se focalizan en un nivel interpretativo, como corresponde a la comprensión académica, las diferencias son igualmente significativas, tal y como se muestra en la Tabla II.

Tabla II. Nivel y tipo de focalización, y cálculo de las diferencias.

	Prueba específica		Prueba general		diferencia	
	media	d. típica	media	d. típica		
Nivel de focalización	2,6	1,1	1,1	0,2	1,5	DS
Tipo de focalización	2,07	0,9	1,5	0,07	1,02	DS

En definitiva, el análisis del contenido de la prueba específica revela que está fundamentada en el concepto de género y que refleja los rasgos del discurso académico. Además de ello, a los procesos de comprensión subyacen tipos y niveles de focalización, además habilidades cognitivas propias del procesamiento de textos del contexto académico. La prueba del examen generalista no presenta ninguna de estas características.

La pregunta referida a la validez basada en la teoría alude a la estructura factorial de la prueba. Para responder a ella y eliminar la hipótesis nula se aplica el análisis factorial confirmatorio (Green, 2013). En la prueba específica se buscaron y encontraron los factores correspondientes a las facetas del constructo que subyacen a cada una de las tareas. En la prueba generalista resulta prácticamente imposible identificar más de un factor, puesto que las cargas no aparecen claramente concentradas. Presentamos, por lo tanto, el análisis de la estructura factorial de la prueba específica.

Las cargas de los nueve ítems de la primera tarea se agrupan claramente en torno a un factor. El ítem, número 10, asociado al mismo texto de entrada que los anteriores, aparece netamente diferenciado en el análisis factorial. Este resultado guarda coherencia con las valoraciones de los jueces expertos en el análisis del contenido. Los 9 primeros ítems se relacionan con un tipo de comprensión literal y, el último (ítem número 10), se juzga en un plano interpretativo. En cuanto a las habilidades, la mayor parte de los juicios apuntan a que casi todos los ítems del primer grupo (del 4 al 8) evalúan la habilidad *seguir instrucciones del profesor*. Los ítems 1 y 2 se centran en *comprender el objetivo de la clase*. El ítem número 10 se valora de forma diferente a los demás: se asocia a una focalización de la comprensión más global e interpretativa y a la habilidad de *captar la idea principal*.

Los tres ítems siguientes se corresponden con la tercera tarea y con el segundo texto de entrada. Las cargas factoriales están, en términos generales, claramente diferenciadas. Según la estimación de los jueces, estos tres ítems evalúan la habilidad de *identificar los tópicos y subtópicos de la clase*. En este caso, los juicios relativos a la focalización de la comprensión coinciden en considerarla más global e interpretativa.

La claridad disminuye con la tarea 4, en la cual los ítems distribuyen sus cargas en dos factores. En este caso, aunque el consenso no es pleno, la comprensión se valora como más local y literal. La distribución de las cargas se corresponde con la estimación de los jueces expertos en cuanto a las dos habilidades evaluadas: *identificar tópicos y subtópicos* e *identificar las relaciones de las ideas*.

Las cargas de los ítems asociados al tercer texto de entrada están agrupadas en torno a dos factores, que se corresponden con sendas tareas. La primera, centrada, según los jueces, en la actividad de *resumir el contenido*. La segunda, en la de *captar relaciones entre las ideas*. En este último caso no hay consenso pleno entre los jueces en cuanto al tipo y al nivel de focalización de la comprensión.

La última tarea, asociada al cuarto texto de entrada, consta de ocho ítems. La carga factorial, está un poco más dispersa, salvo en el caso de cuatro de los ítems, que los jueces asocian claramente a la habilidad de *relacionar el contenido de la clase con el de otras*. Hay tres ítems que los jueces asocian a esta misma actividad, con cargas que no se concentran en el mismo factor. El último ítem se diferencia de los anteriores no solo en cuanto a la carga factorial, sino también al análisis del contenido: la mayor parte de los jueces coincide en que evalúa la habilidad de *reconocer información relevante*.

En resumidas cuentas, el análisis factorial confirmatorio permite descartar la hipótesis nula. La correspondencia entre el análisis del contenido de la prueba y el resultado del análisis factorial es clara en el caso de la prueba de comprensión específica. No se pueden extraer conclusiones contundentes de la prueba de comprensión general en este sentido.

6. CONCLUSIONES

Las pruebas específica y general difieren desde el punto de vista del contenido: género de los textos de entrada, habilidades, focalización. Por lo demás, la estructura factorial de la prueba específica refleja que las cargas se distribuyen en factores que se pueden hacer corresponder con el contenido que subyace a las tareas de examen, tal y como este se ha descrito.

Las evidencias de validez procedentes de fuentes distintas demuestran que los exámenes con fines académicos son necesarios para los propósitos más arriba señalados: discriminar a los alumnos que son capaces de cursar estudios en las universidades de habla hispana empleando la lengua española como vehículo de comunicación. Estos exámenes deben evaluar una *habilidad comunicativa académica específica*.

El género discursivo desempeña un papel clave en el diseño de los exámenes con fines académicos. Es la base de su especificidad y la piedra angular del constructo de la evaluación. Los procesos de comprensión que se producen en el ámbito académico son de naturaleza específica. A ellos subyacen habilidades que difícilmente se ponen en juego en otros contextos. La focalización de la comprensión se produce con más énfasis en niveles y tipos más globales e interpretativos que en los contextos generales.

REFERENCIAS

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *Standards for*

- educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Ainciburu, Cecilia y Rodríguez, Patricia. (2015). La elección de géneros textuales en los exámenes de certificación de lenguas académicas. Las necesidades del alumno universitario y la evaluación implícita de competencias no lingüísticas. En Morimoto, Yucó; Pavón, María Victoria y Santamaría, Rocío (Eds.). *La enseñanza de ELE centrada en el alumno*. Madrid, España, ASELE.
- Aryadoust, Vahid. (2013). *Building a Validity Argument for a Listening Test of Academic Proficiency*. Cambridge, Cambridge Scholars.
- Bachman, Lyle. (1990). *Fundamental Considerations in Language Testing*. Oxford, UK.: Oxford University Press.
- Bachman, Lyle. (2004). *Statistical Analyses for Language Assessment*. Cambridge, UK.: Cambridge University Press.
- Bachman, Lyle y Palmer, Adrian. (2010). *Language Assessment in Practice*. Oxford, UK.: Oxford University Press.
- Bachman, Lyle; Davidson, Fred y Ryan, Katherine. (1995). *An Investigation into the Comparability of Two Tests of English as a Foreign Language*. Cambridge, UK.: University Press.
- BASE (British Academic Spoken English) Corpus. Disponible en: <https://warwick.ac.uk/fac/soc/al/research/collections/base/>.
- Belcher, Diane. (2009). *English for Specific Purposes in Theory And Practice*. Ann Arbor, MI, EE. UU.: University of Michigan Press.
- Cademartori, Yanina; Parodi, Giovanni y Venegas, René. (2006). El discurso escrito y especializado: caracterización y funciones de las nominalizaciones en los manuales técnicos. *Literatura y Lingüística*, 17, 243-265.
- CELU (Certificado de Español, Lengua y Uso). Disponible en: <http://www.celu.edu.ar/es/node/39>.
- Chapelle, Carol. A.; Enright, Mary. K. y Jamieson, Joan. M. (2008). Test score interpretation and use. *Building a Validity Argument for the Test of English as a Foreign Language*, 1-25.
- Comes, Claudia; Elejalde, Jessica; Ferreira, Anita; Gil, Ana María; López, Julián; Pastor, Susana; Pérez, María Isabel; Rodríguez, Alberto; Vine, Ana. (2016). *Investigación e Innovación Educativa en Docencia Universitaria. Retos, Propuestas y Acciones*. Alicante, España: Universidad de Alicante, Vicerrectorado de Calidad e Innovación educativa Instituto de Ciencias de la Educación.
- Coxhead, Averil. (1998). *Anacademic wordlist* (English Language Institute Occasional Publication No. 18). Nueva Zelanda: Victoria University of Wellington.
- Davies, Alan. 2008. *Assessing Academic English: Testing English Proficiency, 1950-1989: the IELTS Solution*. Cambridge, UK.: Cambridge University Press.
- DELE (Diploma de Español como Lengua Extranjera) del Ministerio de Educación y Formación Profesional del Gobierno de España y el Instituto Cervantes.

- Disponible en: <https://www.dele.org/espanol/>.
- eLADE (Examen en Línea de Acreditación de Dominio del Español) del Centro de Lenguas Modernas de la Universidad de Granada. Disponible en: <http://www.clm-granada.com/es/esp/examenes/elade/elade-que-es.html>.
- Elder, Catherine. (2017). Language Assessment in Higher Education. En Shohamy, Elana; Or, Lai G. y May, Stephen (Eds). *Encyclopedia of Language Education. Language Testing and Assessment*. Switzerland: Springer International Publishing.
- EXELAA (Examen de Español como Lengua Extranjera en el Ámbito Académico), Universidad Nacional Autónoma de México, Universidad de Guadalajara, Universidad de Costa Rica.
- Ferreira, A. (2014). Proyecto de investigación Fondecyt N° 1140651: el feedback correctivo escrito directo e indirecto en la adquisición y aprendizaje del español como lengua extranjera. Universidad de Concepción, Chile.
- Ferreira, A. (2016). Prueba de Multinivel con Fines Específicos Académicos, Proyecto Fondecyt 1140651 y Proyecto Fondecyt 1180974, Universidad de Concepción, Chile.
- Ferreira, A. (2018). Proyecto de investigación FONDECYT - CONICYT N°1180974: Diseño e implementación de un corpus escrito de aprendientes de ELE en formato computacional para el análisis de la interlengua. Universidad de Concepción, Chile.
- Ferreira, A., y Elejalde, J. (2019). Hacia un perfil lingüístico-comunicativo del estudiante de Español como Lengua Extranjera para fines Académicos. *Revista Nebrija De Lingüística Aplicada a La Enseñanza De Lenguas*, 13(27), 145-165.
- Flowerdew, Lynne. (2011). ESP and corpus studies. En Belcher, Diane; Johns, Ann M. y Paltridge, Brian (Eds.), *New directions in English for specific purposes research* (222-251). Ann Arbor, MI, EE.UU.: University of Michigan Press.
- Geranpayeh, Ardeshir y Taylor, Linda (Eds.). (2013). *Examining Listening: Research and Practice*, en *Assessing Second Language Listening, Studies in Language Testing*, volume 35, Cambridge: UCLES/Cambridge University Press.
- Green, Rita. (2013). *Statistical Analyses for Language Testers*. Nueva York, EE.UU.: Palgrave Macmillan.
- Green, Rita. (2017). *Designing Listening Tests: a Practical Approach*. Nueva York, EE. UU.: Palgrave Macmillan.
- Hyland, Kent y Shaw, Philip (Eds.). (2016). *The Routledge Handbook of English for Academic Purposes*. Londres, UK.: Routledge.
- Khalifa, Hanna y Weir, Cyril. 2009. *Examining Reading*. Cambridge, UK.: Cambridge University Press.
- Kintsch, Walter. (1998). *Comprehension: A Paradigm for Cognition*. Cambridge, UK.: Cambridge University Press.
- Kintsch, Walter y Van Dijk, Teun. A. (1978). Toward a model of text comprehen-

- sion and production. *Psychological Review*. 85 (5), pp. 363-394.
- Knoch, Ute y Elder, Katherin. (2013). A framework for validating post-entry language assessments (PELAS). *Papers in Language Testing and Assessment*. 2, pp. 48-66.
- Le, Huong Thi Tram. (2011). Developing a validity argument for the *English Placement Fall 2010 Listening Test* at Iowa State University. En *Iowa State University Digital Repository*. Disponible en: <http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1054&context=etd>.
- Llorián, Susana. (2015). *La evaluación de la competencia comunicativa específica en el ámbito de la ingeniería y la arquitectura*. Tesis doctoral. Madrid, España, Universidad Politécnica de Madrid. Disponible en: <http://oa.upm.es/39352>.
- Mendoza, Arturo. (2015). La selección de tareas de escritura en los exámenes de ELE para el ámbito académico. *Revista Nebrija de Lingüística aplicada*, 18(9). Disponible en: <http://www.nebrija.com/revista-linguistica/la-seleccion-de-las-tareas-de-escritura-en-los-examenes-de-lengua-extranjera-destinados-al-ambito-academico>.
- Messick, Samuel. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11.
- McNamara, Daniel y Kintsch, Walter. (1996). Learning from texts: effects of prior knowledge and text coherence. *Discourse Processes*. 22 (3), pp. 247-288.
- MICASE (Michigan Corpus of Academic Spoken English) Corpus. Disponible en: <https://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple>.
- Moore, Tim; Morton, Janne y Price, Steve. (2012). Construct validity in the IELTS academic reading test: a comparison of reading requirements. IELTS test items and in university study. *IELTS Collected Papers, 2: Research in Reading and Listening Assessment*. 2, pp. 120-150.
- Ortega, Jenaro. (2001). La lección magistral. En Vázquez, Graciela (Coord.) *El discurso académico oral. Guía didáctica para la comprensión auditiva y visual de clases magistrales*. Madrid, España: Edinumen, 15-42.
- Parodi, Giovanni. (2004). Textos de especialidad y comunidades discursivas técnico-profesionales: una aproximación basada en corpus computarizado. *Estudios Filológicos*. 39, pp. 7-36.
- Parodi, Giovanni. (2006). Discurso especializado y lengua escrita: foco y variación. *Estudios Filológicos*. 41, pp. 165-204.
- Parodi, Giovanni. (2010). *Alfabetización académica y profesional en el siglo XXI: leer y escribir desde las disciplinas*. Barcelona, España: Ariel/Academia Chilena de la Lengua.
- Pastor, Susana. (2010). Enseñanza de español con fines profesionales y académicos y aprendizaje por contenidos en contexto universitario. *Testi e Linguaggi*. 4, pp. 71-88.
- Pastor, Susana (Coord.). (2016). Español académico e inmersión universitaria del

- alumnado extranjero de la Universidad de Alicante. En J. D. Álvarez, M. T. Tortosa y N. Pellín (Eds.), 1737-1749. *Innovaciones metodológicas en docencia universitaria: resultados de investigación*, Alicante, España: ICE Universidad de Alicante.
- Pastor, Susana y Ferreira, Anita. (2018). El discurso académico en español como LE/L2: nuevos contextos, nuevas metodologías. *Journal of Spanish Language Teaching*. 5(2), pp. 91-101.
- Pastor, Susana y Pandor, Julian. (2017). El choque cultural académico del alumnado estadounidense en la universidad española. *RLA. Revista de Lingüística Teórica y Aplicada*. 55 (2), pp. 13-38.
- Read, John. (2015). *Assessing English Proficiency for University Study*. Londres, UK.: Palgrave Macmillan.
- Read, John. (2016). *Post-admission Language Assessment of University Students*. Switzerland: Springer International Publishing.
- Regueiro, María Luisa y Sanz, Daniel. (2013). *El español académico*. Madrid, España: Arco Libros.
- SIELE (Servicio Internacional de Evaluación de la Lengua Española) referida a las universidades y escuelas de negocio. Disponible en: <https://exámenes.cervantes.es/es/dele/preguntas-frecuentes>.
- Taylor, Linda y Weir, Cyril. (2012). *IELTS Collected Papers 2: Research in Reading and Listening Assessment*. Cambridge, UK.: Cambridge University Press.
- Vázquez, Graciela. (Coord.) (2001). *El discurso académico oral. Guía didáctica para la comprensión auditiva y visual de clases magistrales*. Madrid: Edinumen.
- Weideman, Albert; Patterson, Rebecca y Pot, Anna. (2016). Construct refinement in tests of academic Literacy. En Read, John (Ed.). *Post-admission Language Assessment of University Students*. Switzerland: Springer International Publishing.
- Weigle, Sara y Malone, Margaret. (2016). Assessment of English for academic purposes. en Hyland, Kent y Shaw, Philip (eds.). *The Routledge Handbook of English for Academic Purposes*. Londres, UK.: Routledge.
- Weir, Cyril. (2005). *Language Testing and Validation*. Nueva York, EE.UU.: Palgrave Mcmillan.
- Weir, Cyril; Hawkey, Roger; Green, Anthony y Unaldi, Aylin. (2008). The relationship between the academic reading construct as measured by IELTS and their reading experiences of students in their first year of study at a british university. *IELTS Research Reports*. 9, pp. 97-156.
- Woodrow, Lindy. (2018). *Introducing Course Design in English for Specific Purposes*. Londres, UK: Routledge.
- Xi, Xiaoming y Sawaki, Yasuyo. (2017). *Methods of test validation*. En Shohamy, Elana; Or, Lai G. y May, Stephen (Eds). *Encyclopedia of Language Education. Language Testing and Assessment*. Switzerland: Springer International Publishing.