

La evaluación de la habilidad comunicativa específica en contextos académicos: la comprensión de las clases magistrales

Testing specific communicative ability in academic contexts: lectures comprehension

Susana Llorián González
Universidad Complutense de Madrid
sllorian@ucm.es

Resumen

La investigación que se describe en este artículo se origina a partir de la constatación de que los alumnos no hispanohablantes de programas de intercambio y los matriculados directamente en las universidades españolas, cada vez más numerosos, se encuentran con problemas para cursar sus estudios de grado y de posgrado, debido a barreras lingüísticas. En este escenario, los exámenes específicos de español, de admisión y de post admisión, se están convirtiendo en una necesidad, puesto que podrían discriminar a los alumnos capaces de seguir con éxito los estudios universitarios, por un lado, e impactar de forma positiva en los cursos de refuerzo y apoyo a los programas académicos, por otro. Después de que, en un estudio preliminar, se evidenciara que la comprensión de las clases constituye el principal obstáculo con el que se encuentran los alumnos de las escuelas de ingenierías y arquitectura en el contexto de la investigación, se desarrolla una prueba específica de comprensión de clases magistrales. En este artículo se describe la recogida de una serie de evidencias de validez, que refrenda la necesidad de que se administren pruebas específicas en el ámbito académico.

Palabras clave: comprensión auditiva académica; exámenes con fines académicos; comprensión de clases magistrales; pruebas de comprensión audiovisual

Abstract

In this article, we describe a research that originated in the corroboration of mounting difficulties experienced by university students who are not Spanish speakers belonging to undergraduate and graduate programs. Language barriers hindered students from studying effectively. In this scenario, specific admission and post-admission exams of Spanish, are becoming a necessity, since they could discriminate students who are able to successfully pursue university studies, on the one hand, and have a positive impact in courses of reinforcement and support to academic programs, on the other hand. In a preliminary study we found that lectures academic listening is the main obstacle to the engineering and architecture students in technical universities, the context of our research. For this reason, we developed a specific test of lectures listening. This article describes the collection of several validity evidences, which support the need to administer specific tests in the academic domain.

Key words: academic listening; academic testing; lectures listening; audiovisual listening tests; academic discourse listening

1. Introducción

España ha sido durante años país líder europeo en la recepción de estudiantes universitarios de intercambio internacional. Los programas de grado y posgrado de las universidades públicas y privadas del país reciben cada año miles de alumnos extranjeros, que precisan el español como medio, para acceder al conocimiento de las áreas de especialidad, y como fin, para incrementar su grado de dominio en nuestra lengua. Estas universidades ofrecen cursos de Español, cuyos objetivos, salvo excepciones muy contadas, no se orientan al uso del español en contextos académicos, por lo que constituyen un incentivo para el aprendizaje del español como fin, pero no para el español como medio.

En un estudio preliminar de naturaleza exploratoria que precede a la investigación en la que se focaliza este artículo, descubrimos que los alumnos extranjeros utilizan el español durante sus estudios universitarios, en combinación con el inglés, para realizar tareas académicas, en las que tienen que procesar textos, que se asocian a géneros discursivos propios del ámbito académico (Llorián, 2015). Entre las tareas que ejecutan en español, los informantes, incluidos los que presentan grados de dominio de nivel B2 en adelante, declaran tener especial dificultad para comprender las clases magistrales, en primer lugar, y la realización de presentaciones orales, en segundo lugar. Los informantes declaran que el resto de las tareas académicas que requieren comprensión auditiva se realizan en inglés. De este estudio preliminar se extrajo además la conclusión de que es necesario desarrollar exámenes específicos tanto de entrada como de postadmisión que puedan discriminar a los estudiantes que son capaces de emplear el español como lengua vehicular para acceder a los estudios universitarios y que impacten de forma positiva en el diseño de los cursos de español de apoyo universitario. Partimos de la base de que los exámenes de tipo generalista no pueden asumir este tipo de función, puesto que no evalúan la competencia comunicativa específica que se requiere en el ámbito académico.

Como consecuencia de los resultados de este estudio exploratorio, se decidió poner en marcha una investigación con los objetivos siguientes:

- Elaborar una prueba de examen específica con la cual se evalúe el aspecto que se evidencie en el estudio exploratorio previo como el más urgente y relevante para los alumnos de estas características: la comprensión de una clase magistral.
- Recabar una serie de evidencias de validez de contenido y compararlas las que se recojan de otra prueba que evalúe la comprensión auditiva en exámenes generalistas.
- Recoger evidencias de que los procedimientos que se incluyen en la prueba específica se corresponden con los que se dan en las situaciones de comunicación del ámbito académico, a diferencia de aquellos que se producen en los exámenes de tipo general.

- Utilizar análisis psicométricos con los que se pueda demostrar que los resultados de la prueba se pueden emplear para los fines con los cuales fue concebida, son fiables y presentan capacidad de discriminación, además de un índice de facilidad acorde a las capacidades de los candidatos.
- Comparar las relaciones que se dan entre diferentes mediciones de distintas facetas del mismo constructo: los distintos aspectos que mide la prueba y mediciones de otros aspectos relacionados (gramática y vocabulario y una prueba de otro examen que mida lo mismo).
- Demostrar que la estructura factorial de la prueba específica es el reflejo del constructo que se ha declarado que evalúa y que esta es diferente de la estructura de otras pruebas de tipo generalista.
- Aportar evidencia de que los alumnos que presentan un nivel inferior al nivel B2 de la escala del MCER no son capaces de resolver las tareas de una prueba que emula las que se dan en el ámbito académico.

A lo largo de las páginas siguientes, se exponen los fundamentos teóricos que subyacen a la investigación, los procedimientos metodológicos utilizados y los resultados obtenidos. Finalmente, se presentan sucintamente conclusiones y algunas de las líneas de investigación que quedan abiertas a partir de este trabajo.

2. Fundamentos

2.1. La clase magistral universitaria como género discursivo

La clase magistral se concibe en esta investigación como un género discursivo, por cierto, muy poco explorado en nuestra lengua. Nos encontramos ante un tipo de texto que, según las conclusiones del proyecto ADIEU (Vázquez, 2001, 2006), constituye una de las características más idiosincrásicas de la tradición universitaria española. A pesar de ello, en español, no ha sido objeto de investigaciones, hasta la fecha, en otros trabajos diferentes a este al que nos referimos. El proyecto no se incardina dentro de la era de estudios basados en corpus informatizados. Esta circunstancia, unida a que la investigación no se ajusta a los procedimientos empleados habitualmente en los trabajos localizados en la corriente del análisis del género, lleva a la conclusión de que este género como tal permanece aún al margen de las descripciones fundamentadas en estas dos corrientes. Esto no implica restar en absoluto interés a los resultados del proyecto, que realiza un análisis de los siguientes aspectos, a partir de un corpus de 30 clases del área de humanidades: estructura global, estrategias discursivas, recursos fónicos y componentes no verbales. Se tratan cuestiones como las relaciones lógico-semánticas, la modalización epistémica y afectiva, los procedimientos de cita, la interdiscursividad y las estrategias retóricas.

Uno de los trabajos pioneros sobre el género corresponde a Hansen y Jensen (1994). Para afirmar que este género presenta rasgos específicos recurren a los estudios de Brown y Yule (1983), según los cuales, el discurso oral es susceptible de segmentarse en unidades semánticas ideacionales, que a su vez encuentran correspondencia en unidades curvas entonativas, que se separan unas de otras con pausas. Estas unidades tienen una extensión desigual, que varían de un género a otro. En las clases magistrales la media es de 11 palabras, considerablemente más alta que en otros géneros orales menos formales como la conversación casual. Estamos ante textos de alta complejidad morfosintáctica,

en los que abundan, como suele suceder en estos casos, los grupos nominales, las oraciones de relativo, etc.

Una de las cuestiones más importantes que se ha puesto de manifiesto en los estudios sobre las clases magistrales es que se trata de un género multimodal. De hecho, hace tiempo que Jordan (1997) repara en que, en este evento se incluyen segmentos de discurso que se incluyen en las fotocopias, en la pizarra, en las transparencias, en las proyecciones, etc. Todo lo anterior se combina con el discurso del profesor, en una amalgama de textos que el alumno procesa de forma simultánea.

La función retórica principal de una clase consiste en transmitir información que resulte comprensible para el auditorio. Es necesario para ello que la estructura esté claramente marcada y que responda a unas convenciones claras, pese a que se han identificado, al menos en inglés, numerosos estilos de clase magistral, que varían de un área disciplinar a otra (Nesi, 2001), de un docente a otro y en virtud de variables de tipo contextual como el tamaño de las clases (Cheng, 2012), los recursos de apoyo, el grado de espontaneidad Goffman (1981: 165), los objetivos (Huckin y Olsen, 1990), etc.

El primer estudio que se centra en el análisis de la clase magistral como género es el de Young (1994), que se realiza con el fin de encontrar patrones comunes a las 72 clases que incluye en su corpus. De ahí extrae su modelo de estructura de las clases magistrales en dimensiones micro y macro. Para el análisis de esta última, fija una unidad de análisis, que denomina «fase» (*phase*). Estas fases se dan en el discurso no de forma lineal, sino discontinua y recurrente. De ellas depende que progrese el discurso dentro del evento comunicativo. Los marcadores del discurso son las evidencias del paso de una a otra fase. Young distingue dos tipos de fase: las fases que denomina *metadiscursivas*, a saber, la estructura del discurso (presentación de lo que se va a tratar), la conclusión y la evaluación del contenido, además de las que llama *no metadiscursivas*, es decir, la fase de interacción, la fase de contenido y la fase de ejemplificación. En español, disponemos en este ámbito de la descripción macroestructural que facilitan Ortega (Vázquez (coord.), 2001: 17-27), dentro del proyecto ADIEU. Este análisis presenta las cinco grandes secciones en las que se articula una clase magistral: el preámbulo, en que se da el contacto con los alumnos y tienen lugar los saludos, comentarios, etc.; el planteamiento del tópico general; la explicación, que es el núcleo de la clase y consta de un desarrollo del tópico, una ejemplificación y un cierre; la interacción que precede al cierre; el cierre propiamente dicho. Dentro del marco de este mismo proyecto, este autor realiza otro análisis en un nivel que se puede entender como microestructural, en el que se ocupa de las estrategias discursivas (funciones retóricas) que se dan en la clase magistral en el corpus objeto de estudio. Distingue las siguientes: preguntas, reiteración, contraste, argumentación, ejemplo, aclaración y resumen. Es necesario destacar el valor de este trabajo y la puerta que deja abierta para que las investigaciones basadas en corpus informatizados validen y afiancen todos estos hallazgos como sucede desde hace dos décadas en la lengua inglesa.

En esta lengua, el acceso libre a corpus específicos de textos que se corresponden con géneros académicos como el que nos ocupa ha facilitado este tipo de investigaciones centradas en las clases magistrales desde la perspectiva del género, cuyos procedimientos se podrían extrapolar al español. Además de los análisis del metadiscurso que toman el modelo de Hyland (2003, 2005, 2013) como punto de partida, hay una serie de estudios centrados en el análisis de patrones fraseológicos propios de este género (*lexical bundles*). Por ejemplo,

Nesi y Basturkmen (2006) investigan haciendo uso del corpus BASE para comparar la presencia que tienen en las clases del contexto británico con el norteamericano.

En definitiva, los análisis de las clases magistrales como género discursivo necesitan de investigaciones basadas en corpus informatizados, que den cuenta de las características macroestructurales, de las funciones retóricas y de los elementos del metadiscurso. Estos estudios se tienen que realizar en diferentes ámbitos, de forma que queden en evidencia las características de los distintos estilos de clase en virtud de variables como las áreas de conocimiento, los ámbitos disciplinares, los objetivos de las clases, el número de alumnos, etc. De este modo, se podrá determinar la forma en la que estos factores son determinantes en la comprensión de las clases en los diferentes contextos académicos.

2.2. La comprensión de las clases magistrales

Lynch (2011) realiza un análisis del estado de la cuestión sobre la comprensión auditiva académica. Entre otras apreciaciones, señala el bajo perfil que presentan los estudios realizados en torno al tema, que achaca a la dificultad implícita a la investigación realizada sobre la comprensión académica. Se debe a la cantidad y a la diversidad de variables que intervienen en estos procesos como la diversidad de acentos, la velocidad de emisión de los textos de entrada, etc. En relación con las operaciones de orden cognitivo, hay que reparar en las dificultades que implica acceder a los procesos internos, que solo es posible si se analizan las respuestas indirectas. A todo esto, se añade la complejidad de emular las tareas que se producen en el contexto académico, en particular, el de la clase magistral.

La comprensión auditiva que se da en este tipo de evento comunicativo se ha investigado con arreglo a metodologías que se incardinan en las dos líneas siguientes: una, que se sitúa en una dimensión etnográfica y otra que se centra en las actividades de comprensión académica en relación con la comprensión que se da en otros contextos. Nos interesa especialmente esta última, por las relaciones que establece con la investigación que se describe en este artículo. Destacamos los trabajos de Richards, (1983), Benson (1989, 1994) Flowerdew (1994) y Rost (2002, 2011).

Rost (Rosto, 2011) señala que la intencionalidad es el factor que resulta más determinante para la comprensión auditiva, puesto que el propósito con el que se realiza la comprensión es lo que orienta los procesos. En el contexto académico, los propósitos de la comprensión que se dan en eventos como las clases magistrales están regidos por los objetivos de las clases, es decir, añadir nuevos conocimientos a los ya que se ya se tienen o modificarlos (Chaudron, 1995). La comprensión auditiva en estos contextos está determinada, en primer término, por el texto de entrada, por las tareas de comprensión que se plantean y por los procesos internos que se producen.

En relación con el texto de entrada, los textos que se dan en los ámbitos académicos son de duración extensa (Bejar *et al.*, 2000). Además de esto, se caracterizan por transportar un enorme caudal de información de gran densidad informativa (Chaudron *et al.*, 1994). Por lo tanto, la carga cognitiva que se pone en juego durante los procesos de comprensión es más elevada que la que se da en otros contextos no académicos. Por lo demás, hablamos de un discurso planificado (Chaudron, 1995), que permite que la negociación del significado requiera un número reducido de intercambio de turnos de palabra. Esta

planificación determina, además, la longitud y la complejidad morfosintáctica y la densidad léxica. Nos hemos referido ya a la dimensión multimodal de este tipo de texto.

En cuanto a las tareas de procesamiento del texto, el conocimiento disciplinar es el principal factor desencadenante de los procesos de comprensión (Flowerdew, 1994). Las tareas se requieren que se seleccione la información relevante para ser incorporada al conocimiento ya existente implican un trabajo de criba permanente de la información que es susceptible de separarse de los datos menos importantes: anécdotas, incisos, etc. (Hansen y Jensen, 1994). Bejar *et al.* (2000) subrayan que las operaciones de inferencia superan en intensidad y en cantidad a las que se dan en los procesos de comprensión de lectura. (Rost, 1994), citado en Lynch (2011), repasa en esta misma cuestión para concluir en que, en el contexto de las clases magistrales, la comprensión depende tanto del que habla, que *empaqueta* la información, como del que escucha, que se encarga de *desempaquetarla*. En el proceso de *desempaquetamiento* de la información, es crucial el conocimiento previo del usuario. De este modo, un porcentaje importante del éxito académico no depende tanto de lo que el usuario sepa, sino de su capacidad de comprensión.

En relación con los procesos internos que subyacen a la comprensión, destacamos la importancia del trabajo de Buck, Tatsuoka, Kostin y Phelps (1997), desarrollado en el marco del examen TOEFL. Haciendo uso de la metodología de la regla del espacio, una aplicación de la teoría de respuesta al ítem, fueron capaces de identificar los patrones de respuesta entre los candidatos de definir los 15 rasgos o atributos medidos a través de las pruebas de examen. Diferencian entre rasgos no atributos de nivel superior (generalizables al resto de las destrezas) y rasgos de nivel inferior. Estos atributos se pueden asimilar a lo que se ha llamado tradicionalmente microdestrezas (habilidad para reconocer la tarea y determinar cuál es la información relevante, habilidad para emplear los ítems anteriores para seleccionar la información, habilidad para analizar rápidamente el texto oral en tiempo real, habilidad para procesar información densa, etc.). La perspectiva de las microdestrezas ha dado lugar a polémicas, por un lado, sobre la posibilidad de asimilar los procesos de comprensión auditiva y los de lectura y, por otro, acerca de si a la comprensión subyace un factor único o son varios que determinan este fenómeno.

Toda esta serie de estudios viene a demostrar, en resumidas cuentas, que el fenómeno de la comprensión auditiva que se da en el ámbito académico difiere del que se produce en otro tipo de contextos de orden general. Esto constituye un importante soporte teórico para nuestra investigación

2.3. El modelo de comprensión adoptado

El desarrollo de la prueba específica más arriba mencionada parte del principio, como se ha indicado, de que se evalúa con ella una habilidad comunicativa específica dentro del ámbito académico. La definición del constructo y la forma en la que este se operacionaliza, que se describen en el epígrafe siguiente, se fundamentan en los aspectos que se detallan a continuación.

En primer lugar, la comprensión se realiza sobre un tipo de *input*, que refleja los rasgos del discurso académico, tal y como lo han descrito Regueiro y Sáez (2013); Montolío *et al.* (2000) y Parodi (2007, 2005) y que se materializa en un texto que pertenece género discursivo de la clase magistral (Bejar *et al.*, 2000),

tal y como esta se da en la tradición española (Ortega, 2001; Vázquez, 2001a, b, 2006), en concreto, en el ámbito académico de las ingenierías y de la arquitectura. En segundo lugar, un modelo de comprensión que encaja con esta noción, la de género discursivo, lo cual se explica y justifica seguidamente. Se trata del modelo *construcción-integración* (*construction-Integrationmodel*), desarrollado por Kintsch y van Dijk (1978, 1983) y aplicado a la evaluación por McNamara y Kintsch (1996) y por Khalifa y Weir (2009), en el marco de los exámenes de Cambridge, sobre todo, para el examen IELTS, que certifica el grado de dominio en contextos académicos.

Para justificar que este modelo de *construcción-integración* se ajusta a la comprensión de los géneros académicos, en particular la de las clases magistrales, nos apoyamos, en un plano teórico, en las razones siguientes. En primer lugar, porque considera que los textos presentan dos tipos de estructura, a los que subyace una red de proposiciones semánticas: una estructura local (microestructura) y otra global (macroestructura), que se corresponden con la idea general del texto y se deriva de la anterior. En segundo lugar, porque el modelo parte de la idea de que la comprensión se produce en dos fases. Durante la primera fase, la de *construcción*, el que escucha focaliza la comprensión en el nivel de la estructura local, con lo que logra configurarse una idea aproximada del significado del texto, todavía incoherente, a partir del *input* textual que recibe, de las metas que se marca con la comprensión y del conocimiento previo que tiene acerca del tema del texto, en este caso concreto, de la materia o contenido. En esta primera fase, las operaciones que realiza son de reconocimiento y la comprensión, por lo tanto, es literal y se da en un plano superficial.

La segunda fase se denomina *integración*. Cuando se produce, el que escucha rechaza las construcciones inapropiadas que hizo en la fase previa en un nivel local, en favor de las que encajan en un todo coherente, es decir, en la macroestructura del texto. Es aquí donde se logra una comprensión profunda del significado global del texto y de su estructura. Para captar la idea general, es necesario que el usuario establezca una red de relaciones entre las ideas principales. La comprensión, en este caso, se realiza de un modo interpretativo, por lo que las operaciones cognitivas son principalmente de inferencia.

El modelo, por todo lo anterior, contempla dos niveles de comprensión: el primero, *basado en el texto*, a partir del cual el usuario se representa mentalmente la red proposicional del texto en los niveles de la macroestructura y de la microestructura; el segundo nivel, el *situacional*, consiste en que el alumno se represente el texto, de manera que la imagen mental que extrae de su significado esté ya integrada con el conocimiento que tiene sobre el tema del que trata y, en este caso, sobre el área de su especialidad académica y profesional. Por todo lo anterior, en términos muy generales, se podría partir de la base de que la mayoría de las actividades de comprensión que se dan en las clases magistrales se focalizan en un nivel global y son de tipo interpretativo.

2.4. La validación de la prueba específica

Para la validación de la prueba específica se ha empleado el modelo de Weir (2005), que se incardina en el paradigma de los procedimientos basados en la acumulación de evidencias. Estos se diferencian de los que se basan en la construcción de un argumento de validez (Le, 2011: 9). Se trata del modelo que se utiliza en la actualidad para validar la mayor parte de los exámenes de Cambridge ESOL.

La aplicación de este modelo a las pruebas de comprensión académica, en concreto, a la comprensión de clases magistrales, consiste en recoger pruebas, a lo largo de todo el ciclo de examen, de las diferentes facetas de la validez descritas en el modelo de Weir (2005), de manera que se pueda demostrar lo siguiente: que las tareas se adecúan a los estilos cognitivos, los esquemas afectivos, las motivaciones, etc. del perfil de los candidatos (características del candidato); que la prueba, y el contexto en la que esta se administra (validez contextual¹), se ajusta a las situaciones de comunicación en las que este se va a desenvolver en la "vida real" en cuanto a las instrucciones, al propósito de la escucha, al tipo de discurso o género del texto de entrada, las características del texto (tema, contenido, extensión, velocidad de emisión, sintaxis, léxico, etc.) o el formato de las tareas de la prueba; que el tipo de preguntas propicia que se pongan en marcha los mismos procesos cognitivos que se dan cuando el candidato se enfrenta a la comprensión de las clases magistrales y que el constructo de la evaluación se encuentre realmente representado, además de que la medición no se vea contaminada por factores ajenos este (validez basada en la teoría); que las calificaciones sean consistentes y que los ítems empleados tengan calidad psicométrica (validez de la calificación); que los índices de correlación entre pruebas que miden el mismo constructo sean aceptables; que el uso de los resultados de la prueba sirve para predecir el grado de comprensión de las clases magistrales que van a mostrar los candidatos, según su grado de dominio (validez referida al criterio) en las situaciones reales y va a impactar positivamente en la programación de cursos destinados, entre otros objetivos, a proporcionar apoyo a los futuros alumnos para comprender las clases (validez consecucional).

3. Metodología de investigación

3.1. El contexto de la investigación y la muestra de informantes

La investigación se desarrolla en el ámbito del servicio de idiomas de la Universidad Politécnica de Madrid (PROLINTER). Por lo tanto, los informantes son estudiantes de los programas académicos de grado y de posgrado de las áreas de las ingenierías y de la arquitectura. Se utiliza un criterio de muestreo de oportunidad o conveniencia, de tipo no probabilístico. Se trata del más frecuente en el ámbito de la Lingüística Aplicada (Dörnyei, 2007: 98-99). Este tipo de muestra se configura partir de las poblaciones cautivas de las organizaciones en las que se desarrolla la investigación. En el caso que nos ocupa, los resultados de las pruebas de evaluación que se deriven de la investigación aquí descrita impactarán directamente en el diseño del currículo institucional.

Se obtiene una muestra de 60 informantes, que se distribuyen en cuatro grupos, determinados por el nivel de los contenidos de los cursos en los cuales están matriculados. La oferta de los cursos se articula en virtud de los niveles de la escala del *MCER*. Se consideran los que se encuentran dentro la franja siguiente: B1+ y B2. La distribución responde a lo siguiente: curso B12A (19 informantes), B12B (19 informantes), B21A (11 informantes), B2A2 (11 informantes). La selección de los grupos de alumnos responde al nivel de la

¹La validez contextual es un concepto acuñado por Weir (2005: 56). Con él, este autor hace referencia a lo que tradicionalmente se conoce como validez de contenido. La argumentación de esta faceta de la validez se apoya en la media en la que las tareas que se proponen en el examen reflejan las que los candidatos realizan en la vida real.

misma escala sitúan la prueba las especificaciones y el panel de jueces expertos. Los alumnos matriculados en los cursos de niveles inferiores al B1 no se tienen en cuenta y se consideran únicamente aquellos que están en grupos de nivel B1 en adelante (hasta B2+). Se entiende, por supuesto, que el hecho de que estén matriculados en un curso de un determinado nivel de la escala del *Marco común europeo de referencia* (MCER en este artículo) no implica que presenten ese grado de dominio en el momento en que se realiza la investigación. De hecho, se comprueba, a partir de la administración de una prueba de *Gramática y vocabulario*, cedida por el Centro de Lenguas Modernas de la Universidad de Granada, que hay numerosos alumnos en los cursos relacionados con los niveles del MCER que están por encima o por debajo del nivel al que alude el nombre del curso. Esta prueba está debidamente calibrada y se vincula a los niveles del MCER a partir de la aplicación de procedimientos científicos.

Además de este parámetro de descripción, se tiene en cuenta, con vistas a identificar posibles sesgos introducidos por esta variable, la familia de la lengua materna de los alumnos. La casualidad da como resultado que, pese a tratarse de un muestreo de oportunidad, la mitad de los informantes tengan una lengua materna de origen romance y la otra mitad, no romance.

3.2. El diseño de la investigación y el desarrollo

La metodología de la investigación se basa en el modelo sociocognitivo de Weir (Weir, 2005). Le (2011: 9) identifica dos enfoques en los procesos de validación de pruebas de examen: el modelo de acumulación de evidencias y el de construcción de un argumento de validez. El modelo seleccionado (Weir 2005) se incardina en el primer paradigma y consiste en recoger pruebas que avalen las diferentes facetas de la validez utilizando diversos procedimientos. El modelo es componencial y conecta con las facetas tradicionales de la validez. Es el que se emplea en la mayor parte de los exámenes de la factoría Cambridge ESOL, en particular, en el examen IELTS (International English Language Testing System). La elección de este modelo responde a que el examen IELTS evalúa el uso de la lengua en contextos académicos y porque es el examen más investigado de la historia (Davies, 2008:91).

Para la recogida de los datos, se diseña una prueba con la que se pretende identificar a los alumnos que son capaces de comprender las clases magistrales y determinar en qué grado pueden hacerlo. A partir de los resultados de esta prueba y de un análisis de necesidades, podrán formularse los objetivos de los cursos de español, de apoyo académico a los programas de grado y de posgrado. Con el fin de demostrar que es necesario emplear pruebas específicas y que los resultados las pruebas de comprensión auditiva de los exámenes generalistas no son útiles para estos fines, se administra paralelamente, a los mismos informantes, en el mismo momento, una prueba de comprensión auditiva estándar de uno de estos exámenes.

El diseño de la prueba específica se basa, para describir el constructo, en el modelo de *habilidad comunicativa* Bachman y Palmer (2010). En este estudio, se adapta al uso de la lengua en contextos académicos.

El *conocimiento organizativo* comprende los siguientes: por un lado, el *conocimiento gramatical* –que, según el modelo, se compone del conocimiento de la sintaxis, del vocabulario y de la fonología–; por otro, el *conocimiento textual* (cohesión y organización retórica). El *conocimiento gramatical* en el contexto académico integraría el conocimiento implícito o explícito de los rasgos del discurso académico, según los han descrito, los autores más arriba

mencionados: léxico preciso y descontextualizado, alto grado de abstracción semántica, vocabulario semitécnico, en este caso, de los ámbitos de la ciencia y la tecnología, aunque no se tiene en cuenta la terminología específica de las áreas disciplinares; elaboración morfosintáctica, profusión de recursos para impersonalizar, índice elevado de marcadores y recursos subordinantes, etc.

Además del *conocimiento organizativo*, el modelo de Bachman y Palmer (2010) considera el *conocimiento pragmático*, que se compone, a su vez de los siguientes: el *conocimiento funcional* y el más importante cuando se trata de contextos académicos, es decir, el *conocimiento sociolingüístico*. En relación con este último, adquiere especial relevancia en concepto de género discursivo, a partir del cual se realiza el análisis de las clases magistrales al que nos hemos referido más arriba en relación con el modelo de comprensión que hemos adoptado. El conocimiento *sociolingüístico* abarca, además del género, los rasgos del registro, en este caso, formal, propio de los contextos académicos.

Junto con los tipos de conocimiento hasta aquí enumerados, se consideran 11 actividades comunicativas de comprensión propias del discurso académico: seguir instrucciones del profesor, identificar el propósito de la sesión de clase o de una parte de la sesión, identificar los tópicos y los subtópicos de la clase, resumir el contenido de la lección, relacionar el contenido de la clase con sesiones anteriores y posteriores, identificar información irrelevante (bromas, incisos, comentarios, etc.). Las actividades comunicativas se relacionan con los niveles y tipos de focalización en los que se centra la escucha, según el modelo de McNamara y Kintsch, (1996), Khalifa y Weir (2009): más o menos global, más o menos local; más o menos literal, más o menos interpretativo.

Este constructo se hace operativo en las tareas de la prueba específica de comprensión audiovisual, cuyas especificaciones se presentan a continuación de forma sucinta. Esta prueba específica, tal y como se ha diseñado, consta de 35 ítems discretos, de respuesta preseleccionada, que se ajustan a las modalidades de opción múltiple, de correspondencia múltiple y de reconstrucción de un esquema. Estos ítems se asocian a cuatro textos audiovisuales de entrada y se distribuyen en siete tareas. Cada una de estas tareas se focaliza una o varias actividades comunicativas especificadas. Por su parte, los textos de entrada reflejan los rasgos del género de la clase magistral y del discurso académico más arriba descritos y, finalmente, los ítems, las modalidades de comprensión contempladas en el modelo adecuado.

3.3. La recogida de evidencias

La recogida de evidencias de validez que se obtienen de la aplicación del modelo sociocognitivo de Weir (Weir, 2005) puede emplear una gran diversidad de procedimientos. Nosotros hemos seleccionado y diseñado los métodos que mejor se adecúan a los objetivos de la investigación que aquí se describe. Establecemos el punto de partida para la exposición de los procedimientos empleados en las preguntas que formula Weir (2005: 48) para describir los componentes de su modelo de validación de pruebas de examen: características físicas y fisiológicas de los candidatos, validez contextual (validez de contenido), validez basada en la teoría, validez de la calificación; validez consecucional; validez referida al criterio.

La respuesta a las características de los candidatos queda garantizada con el estudio preliminar, de naturaleza descriptivo-exploratoria, del que parte esta investigación (Llorián, 2015). En este estudio, se realiza un análisis documental, en el que se identifican los géneros a los que se asocian los textos que los

alumnos de los programas de grado y de postgrado de las facultades de ingeniería y arquitectura se ven en la necesidad de procesar utilizando el español lengua vehicular. Se obtienen además datos de otra fuente: los propios alumnos, que se someten a una encuesta, que indaga acerca las necesidades de uso del español en este contexto en relación con los géneros identificados, de la importancia que les confieren los informantes y del grado de dificultad que les atribuyen. Como ya se ha señalado, la clase magistral es el género que más problemas presenta para ellos y para el que necesitan servirse únicamente del español, puesto que, en el resto de las actividades que requieren comprensión auditiva, utilizan habitualmente el inglés.

En el marco del estudio adquiere especial relevancia el conjunto de evidencias relativas a la validez contextual. Se debe a que esta investigación se centra en el diseño de una sola de las pruebas de examen y a que, de momento, se ha desarrollado una forma piloto.

El procedimiento utilizado para recabar este tipo de evidencias es el de juicio experto. Se realiza a partir de la convocatoria de un panel de jueces, a los que se encomienda una serie de tareas, que se orientan a refutar la hipótesis nula (H_0), a favor de la alternativa (H_1): "El contenido de la prueba de comprensión audiovisual refleja facetas de la habilidad comunicativa, específicas de la comunicación académica".

Durante el desarrollo del panel, en primer lugar, se solicita a los jueces expertos que relacionen cada uno de los ítems de las tareas, tanto de la prueba específica como de la prueba general, con las actividades de comprensión de las clases magistrales que figuran en las especificaciones. En segundo lugar, se les pide que relacionen las tareas de las dos pruebas con el grado (+ o ++) del nivel (local o global) y el tipo (literal o interpretativo) de focalización que requiere la escucha. Los jueces tienen, en tercer lugar, que hacer un análisis de los textos de entrada de las tareas de ambas pruebas con los rasgos del discurso académico detallados en la descripción del constructo de la prueba y en las especificaciones. Para ello, se emplea un instrumento de análisis, mediante el cual los jueces atribuyen valores (del 0 al 3) a cada uno de los rasgos respecto al grado en el que estiman que estos se hallan presentes en los textos de entrada.

Con el fin de comparar los juicios emitidos respecto a las dos pruebas, se replica el método que utilizaron Bachman *et al.* (1995), citado en Bachman (2004, 272-275) para comparar el constructo de las pruebas del examen TOEFL (versión anterior al *Proyecto TOEFL 2000*) y las del examen First Cambridge Certificate.

En lo referente a la validez basada en la teoría, los procedimientos empleados se limitan a demostrar que la secuencia de pasos que se sugiere en las instrucciones de la prueba para que los alumnos resuelvan la tarea no introduce varianza irrelevante del constructo. En el intento de emular los procesos de escucha que se producen en las clases magistrales, se da la opción a los alumnos de que empleen la toma de notas como estrategia de apoyo a la comprensión. Con objeto de preservar la amenaza de los eventuales sesgos introducidos por el uso de esta estrategia, se pone en marcha un estudio a pequeña escala, en el que la toma de notas se considera variable independiente. El estudio consiste en comparar los resultados de los alumnos que deciden utilizar la toma de notas con los de aquellos que no lo hacen. Se aplica la prueba de T de Student de muestras independientes para corroborar que los resultados no se sesgan por influencia de la variable.

Esta investigación se propone, en lo que respecta a la validez de la calificación, comprobar la calidad psicométrica de los ítems de la prueba

específica diseñada *ad hoc* para la comprensión de las clases magistrales. La calidad psicométrica se preserva con un ensayo, en el que se llevan a cabo análisis en los cuales se aplican valores de estadística descriptiva básica, que se combinan con procedimientos definidos por la Teoría Clásica del Test, de la Teoría de Respuesta al Ítem del modelo de Rash de un parámetro. Se realizan asimismo cálculos sobre la fiabilidad. Una de las fuentes de evidencia más relevantes cuando se pretende demostrar que dos pruebas miden facetas diferentes del mismo constructo es la que se obtiene del análisis de la estructura factorial. En esta investigación se aplica a la prueba específica y a la prueba de comprensión general.

Finalmente, nos referimos a las evidencias de validez referidas al criterio, para las que se toma como sistema de referencia externo el que proporciona el *Marco común europeo de referencia*. Lo primero que se hace es relacionar la prueba específica con los niveles de la escala, aplicando los procedimientos del *Manual para relacionar exámenes con el MCER* del Consejo de Europa (MRE, en adelante). Seguidamente, se comparan los resultados obtenidos de la administración de las dos otras pruebas, Gramática y vocabulario y comprensión general, ambas relacionadas con el MCER, con los de la prueba específica. Todo ello conduce a verificar la hipótesis alternativa (H_1), relacionada con la validez referida al criterio, según la cual, los alumnos de los programas de grado y de posgrado que se encuentran por debajo del nivel B2 se encontrarían con serias dificultades para comprender las clases magistrales.

4. La presentación de los resultados

4.1. La validez contextual y la validez basada en la teoría

Se presentan, en primer lugar, los resultados relativos a la validez contextual. Son, como se ha visto, el producto del procesamiento de los datos que se obtienen de la aplicación del procedimiento de juicio experto en relación con lo siguiente: la identificación de las habilidades comunicativas académicas evaluadas en las tareas de las dos pruebas, el nivel y el tipo de focalización (más o menos local o global, más o menos literal o interpretativo) y la presencia de rasgos del discurso académico en los textos de entrada de la prueba específica y de la prueba general.

En lo referente a las actividades comunicativas, los jueces estiman que todas ellas son objeto de evaluación mediante las tareas de la prueba específica del ámbito académico. No hallan rastro de ellas en la prueba de tipo generalista, a excepción de *seguir instrucciones*, que se relaciona con una de las tareas.

Respecto al nivel y al tipo de focalización, los jueces presentan un alto grado de consenso en sus atribuciones. Se describen seguidamente los detalles de los juicios en relación con las tareas de la prueba.

Comenzando con el nivel de focalización, los jueces estiman, con un ciento por ciento de acuerdo, que dos de las siete tareas, la Tarea 1 (centrada en *seguir instrucciones, captar ideas para aplicarlas, identificar el propósito de la sesión de clase*) y la Tarea 7 (*relacionar el contenido de la clase con el de sesiones anteriores o posteriores; seguir indicaciones sobre avisos, sugerencias etc., respecto a plazos, tareas...; reconocer la información irrelevante: bromas, incisos, comentarios, etc.*), se focalizan en un nivel local en su más alto grado.

Existe asimismo consenso pleno en que otras dos tareas (la Tarea 2 y la Tarea 5) presentan el más alto grado de nivel de focalización global. La Tarea 2

se centra en *captar la idea principal* y la Tarea 5, en *resumir el contenido del texto de entrada*.

El acuerdo de los jueces disminuye ligeramente en cuanto a la estimación del nivel de focalización de las tareas 3 y 6, que sitúan en un grado más bajo (+) de nivel global. La Tarea 3, centrada en *identificar tópicos y subtópicos* y la Tarea 6, en *captar las relaciones y las jerarquías entre las ideas*. El consenso es bastante menor en relación con la Tarea 4 (*identificar las relaciones entre las unidades del discurso*). Para la mayor parte de los jueces se focaliza en un nivel literal, aunque dos de ellos le atribuyen un nivel de focalización global en diferente grado.

Siguiendo con el tipo de focalización, los jueces aprecian, con pleno consenso, que dos de las tareas (la 1 y la 7) presentan el más alto grado (++) de focalización literal, ambas centradas, como se ha visto, en actividades orientadas en seguir instrucciones, indicaciones, etc. y en localizar información concreta. Este tipo de focalización literal, aunque con menor grado y menor consenso entre los jueces, se atribuye a las tareas 3 y 4 (*identificar tópicos y subtópicos* e *identificar las relaciones entre las unidades del discurso* respectivamente).

El resto de las tareas se focalizan en un tipo de comprensión interpretativa, aunque el grado no es tan alto y el consenso entre los jueces no es pleno. Los resultados de la prueba de comprensión auditiva general son bastante claros. El acuerdo entre los jueces es pleno en cuanto a que, en la totalidad de las tareas, se produce un nivel de focalización local y un tipo de comprensión literal. Los juicios varían, no obstante, respecto al grado en ambos casos. Al comparar la totalidad obtenidos en cada una de las pruebas, aplicando el procedimiento descrito por Bachman *et al.* (1995), se aprecian diferencias significativas.

Nos centramos, para terminar con las evidencias de la validez contextual, en los resultados del análisis de los textos de entrada para localizar en ellos los rasgos del discurso académico señalados en la descripción del constructo. Los jueces muestran fuerte consenso en el hallazgo de todos los rasgos considerados en los textos de entrada de la prueba específica. Se obtiene pleno acuerdo respecto a que se trata de textos asociados a un género privativo del ámbito académico. Los rasgos que obtienen más puntuación y un considerable acuerdo entre los jueces se refieren al alto grado de planificación del discurso, al registro formal y a la especialización del tema. En esta misma línea se considera la elaboración sintáctica, la frecuencia de aparición de vocabulario técnico y preciso, así como la presencia de recursos de impersonalización de nominalizaciones, sobre todo, deverbales. Estos rasgos no están del todo ausentes en los textos de la prueba de comprensión auditiva general. De hecho, casi todos ellos se localizan el texto de una de las tareas, de carácter monológico, que se corresponde con el género de la conferencia. Aun así, la puntuación es considerablemente más baja. En todos los rasgos, sin excepción, las comparaciones entre las dos pruebas presentan diferencias significativas, en un rango entre 0.5 puntos y 2.2, siendo el valor máximo de las medias 2.4 y 0.7 el de las desviaciones típicas.

Cerramos este epígrafe con los datos de las diferencias que se dan entre los alumnos que se decantan por el uso de la toma de notas como estrategia, la gran mayoría, y los que deciden prescindir de ella. Los resultados de la prueba T de Student para muestras independientes indican que la *prueba de Levene* para igualdad de varianzas revela que el P valor (sig.), 0.721, es mayor que el valor crítico, $\alpha = 0.05$, por lo que se asumen varianzas iguales. La significación bilateral 0.185 es mayor que el valor crítico $\alpha = 0.05$. Esto quiere decir que las

diferencias entre las medias de los dos grupos, los que utilizan la toma de notas y los que no, no se pueden considerar como significativas y que la superioridad de uno respecto al otro puede deberse al azar.

4.2. La validez de la calificación y la validez de criterio

Los resultados de las tres pruebas administradas muestran una distribución normal. Todas las pruebas presentan asimetría y curtosis negativas, lo cual quiere decir que no resultan difíciles para los candidatos y que los valores no se agrupan en torno a la mediana, sino que están bastante dispersos. Esto se verifica con los valores de las desviaciones típicas y de las varianzas, en especial, en las pruebas de gramática y vocabulario y de comprensión académica. En el caso de la curtosis, al dividir $-0.764 / 0.608$ da como resultado -1.2 . Por lo que respecta a la asimetría, $-0.442 / 0.309$ es igual a -1.4 . Los dos resultados se mantienen entre los valores -2 y $+2$, por lo que se encuentran dentro de lo aceptable (Green, 2013: 52). En cuanto a la fiabilidad de la prueba de comprensión académica, los coeficientes Alpha, KR21 de Crombach y Beta presentan valores de 0.77. Pueden considerarse igualmente dentro de los límites de lo normal si tenemos en cuenta que, según Carr (2011, 304), en los exámenes de alto perfil estos valores deben estar por encima de los 0.7 puntos y oscilar en torno a 0.8.

En cuanto a los datos obtenidos de la aplicación la Teoría Clásica del Test y la Teoría de Respuesta al Ítem, nos limitamos a comentar los resultados de la prueba específica. Se observa que los valores son igualmente aceptables, aunque algunos de los ítems presentan índices de discriminación bajos (cinco en total), por lo que son susceptibles de eliminarse sin que esto afecte a la fiabilidad de la prueba. Por lo demás, ninguno de los valores se encuentra fuera de los parámetros de normalidad.

Una de las evidencias más importantes se deriva del análisis de la estructura factorial de la prueba específica. En este sentido, llama la atención que, pese al reducido tamaño de la muestra, el análisis factorial confirmatorio evidencia que la estructura factorial interna de la prueba refleja el constructo, tal y como se ha definido. Se buscan las cargas que asumen los ítems en relación con los factores que se hacen corresponder con las tareas de la prueba y se constata que las cargas pueden hacerse corresponder con las áreas en las que se focaliza la evaluación.

Por lo que respecta al conjunto de evidencias referidas al criterio, señalamos que, después de aplicar los procedimientos descritos en el *Manual para relacionar exámenes con el MCER* del Consejo de Europa (MRE, en lo sucesivo), en concreto, el *Item-descriptor Matching Method* para establecer puntos de corte, los datos resultantes de la aplicación de los procedimientos cualitativos revelan que, para la resolución de las Tareas 1 y 2, los informantes necesitan haber alcanzado un nivel B1, un nivel B1+ para la Tarea y nivel B2 para el resto. Estos resultados se cotejan con los datos resultantes del análisis psicométrico (MRE, 2009: 73-75; Cizek y Bunch, 2007: 193). El objetivo consiste en calcular el número de ítems correctos que se corresponden con el límite fijado entre las franjas B1+ y B2. Una vez establecidos los valores, el grueso de los informantes se aglutina en el nivel B2. Los alumnos de este grupo son los que obtienen los mejores resultados.

5. Discusión

Los datos cuantitativos y cualitativos relacionados con la validez contextual ponen de manifiesto que la prueba de comprensión específica evalúa las actividades comunicativas enumeradas en las especificaciones y que los ítems de las diferentes tareas proporcionan cobertura a todas ellas. Existe, por lo demás, variedad en cuanto a los niveles y los tipos de focalización de la comprensión. No obstante, en versiones sucesivas, sería conveniente incrementar los ítems focalizados en niveles de comprensión más globales y, sobre todo, más interpretativos. El resultado obtenido en relación con este aspecto se asimila bastante al que obtienen Weir *et al.* (2012) en relación con el examen IELTS. Los textos de entrada reflejan, por lo demás, los rasgos del discurso académico y las características de los géneros de este ámbito, en este caso, de las clases magistrales.

Más arriba quedó claro que la toma de notas como procedimiento sugerido para prestar asistencia a la resolución de las tareas de comprensión académica no introduce sesgos en los resultados. Es necesario añadir que esta fuente de evidencia no se ve, en ningún caso, colmada con estos resultados, sino que lo aconsejable es profundizar y emplear otros procedimientos que den acceso a los procesos internos empleados por los alumnos extranjeros para la comprensión de las clases en situaciones reales de comunicación y contrastarlos con los que ponen en marcha durante la resolución de las tareas de examen. El uso de cuestionarios y de protocolos de pensamiento en voz alta son los métodos más recomendados para estos fines. Hay que tener presente además que, durante la administración de la prueba, los alumnos que conformaban la muestra de informantes se encontraban en una situación real de evaluación. Si bien esto contribuye de forma muy positiva a las garantías de validez, también es cierto que el uso de estrategias se puede ver condicionado.

Por lo que respecta a la calidad psicométrica de los datos, ha quedado claro que los valores se mantienen dentro de la normalidad. A pesar de todo, y de haber realizado el correspondiente ensayo, cuatro de los ítems tuvieron que ser eliminados, por lo que la calidad psicométrica es todavía susceptible de mejora.

Una de las evidencias de mayor peso referida a la validez del constructo es la que se deriva del análisis factorial. En este caso, pese a que los datos corroboran las hipótesis formuladas en un principio, es necesario acoger los resultados con cautela. Aunque las cargas se vislumbran con claridad, es preciso tener en cuenta que la muestra de informantes no tiene un tamaño suficiente como para que se puedan extraer conclusiones contundentes.

Los datos revelan claramente que los alumnos que se sitúan por debajo del nivel B2 tendrán severas dificultades para enfrentarse a la comprensión de las clases magistrales. No obstante, hay que tener en cuenta que la vinculación de la prueba específica a la escala de niveles de referencia del MCER no está libre de problemas. Las dificultades se ponen de manifiesto desde la perspectiva del contenido y también con la validación psicométrica. En el primer caso, es complicado clasificar el contenido especializado de los textos de entrada, por ejemplo, en términos de "concreto" o "abstracto", que son los que se contemplan en los descriptores del MCER. Este es solo un ejemplo entre otros muchos, puesto que las escalas del documento no se refieren al uso de la lengua en contextos académicos o científicos ni aluden al concepto de género, que ha resultado básico en la definición del constructo de la prueba específica.

En definitiva, puede afirmarse que los datos, que se obtienen de distintas fuentes de evidencia, haciendo uso de diferentes procedimientos y herramientas, convergen. Pese a las ineludibles amenazas, que se han señalado al hilo de la exposición, prevalecen las numerosas garantías de validez que se han aportado.

6. Conclusiones

La acumulación de evidencias de la validez que se ha descrito hasta aquí viene a demostrar que la prueba específica de comprensión académica centrada en las clases magistrales discrimina a los alumnos que son capaces de procesar los textos que se corresponden con un género propio del contexto académico, lo cual significa que tiene la función de transportar conocimiento de tipo especializado. Se caracterizan además por emplear un registro formal y por el predominio de determinadas secuencias textuales y funciones retóricas, además de vocabulario semitécnico y rasgos morfosintácticos particulares como el predominio de un estilo nominal entre otros. El procesamiento requiere, además, actividades comunicativas específicas y procesos de comprensión, preferentemente, de tipo interpretativo, que se dan en niveles focalización global. Estas evidencias de validez basada en el contenido se ven refrendadas con los resultados de los análisis psicométricos.

Ha quedado igualmente demostrado que el procesamiento de los textos que se precisa en este contexto requiere, como mínimo, que se haya alcanzado un nivel B2. No obstante, el simple hecho de presentar este nivel no es suficiente como para predecir que los alumnos sean capaces de comprender clases magistrales.

En resumidas cuentas, es muy recomendable que las universidades receptoras de alumnos de programas de intercambio administren pruebas de examen específicas del ámbito académico, en concreto, pruebas de comprensión audiovisual, centradas en el género de las clases magistrales. Son necesarios más estudios en esta dirección, centrados en la comprensión de lectura y en actividades de expresión e interacción, que vengan a verificar que la administración de pruebas de tipo generalista no basta para identificar a los alumnos que son capaces de realizar las tareas que se requieren en los contextos académicos utilizando el español como lengua vehicular.

REFERENCIAS BIBLIOGRÁFICAS

- Bachman, L. (2004). *Statistical Analysis for Language Assessment*. Cambridge: Cambridge University Press.
- Bachman, L. y Palmer, A. (2010). *Language Assessment in Practice*. Oxford: Oxford University Press.
- Bachman, L., Davidson, F., Katherine, F. (1995). *An Investigation into the Comparability of Two Tests of English as a Foreign Language*. Cambridge: Cambridge University Press.

- Béjar, I., Douglas, D., Jamieson, J., Nissan, S., Turner, J. (2000). *TOEFL 2000. Listening Framework: a Working Paper*. Educational Testing Service Princeton, NJ.
- Benson, M. J. (1994). Lecture listening in an ethnographic perspective. En: Flowerdew, J. ed., *Academic English: Research perspectives* Cambridge: Cambridge University Press, pp. 181-198.
- Benson, M. J. (1989). The academic listening task: a case study. *Tesol Quarterly*, 23 (3), pp. 421-445.
- Brown, G. y Yule, G. (1983). *Discourse Analysis*. Cambridge University Press.
- Buck, G., Tatsuoka, K., Kostin, I., & Phelps, M. (1997). The sub-skill softlistening: Rule-space analysis of a multiple-choice of second language listening comprehension. En: Huhta, A. Kohonen V., Kurki-Suonio y Louma, S., eds., *Current developments and alternatives in language assessment*. Tampere: University of Jyväskylä.
- Chaudron, C. (1995). *Academic Listening. A Guide for the Teaching of Academic Listening*, Cambridge Cambridge University Press, 1995.
- Chaudron, C., Loschky, L., & Cook, J. (1994). Second language listening comprehension and lecture note-taking. En J. Flowerdew, ed., *Academic listening: Research perspectives*, Cambridge: Cambridge University Press, pp. 75-92.
- Cheng, S. W. (2012). "‘That’s it for today’: Academic lecture closings and the impact of classsize". *English for Specific Purposes*, 31, pp. 234-248.
- Council of Europe. (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)*. [En línea], disponible en: <http://www.coe.int/en/web/common-european-framework-reference-languages/relating-examinations-to-the-cefr>. [Acceso, 20 de febrero de 2018].
- Consejo de Europa. (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación*. Madrid: Anaya, Instituto Cervantes Ministerio de Educación. [en línea], disponible en: https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/. [Acceso, 20 de febrero de 2018].
- Carr, N. (2011). *Designing and Analyzing Language Tests*. Oxford: Oxford University Press.
- Cizek, G. J. y Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. New York: SAGE Publications Ltd.
- Davies, A. (2008). *Assessing academic english: testing english proficiency, 1950-1989: the ielts solution*. Cambridge: University Press.

- Dörnyei, Z. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford: University Press Oxford.
- Green, R. (2013). *Statistical Analyses for Language Testers*. Palgrave Macmillan, 2013.
- Goffman, E. (1981). *The Lecture, Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Hansen, C. y Jensen, C. (1994). Evaluating lecture comprehension, en J. Floverdew. *Academic Listening: Research Perspectives*. Cambridge: Cambridge University Press, pp. 241-268
- Huckin, T. N. y Olsen, L. A. (1990). *Technical Writing and Professional Communication for Nonnative Speakers of English*: McGraw-Hill.
- Jordan, R. (1997). *English for Academic Purposes. A Guide and Resource Book for Teachers*. Cambridge: Cambridge University Press.
- Khalifa, H. y Weir, C. J. (2009). *Examining Reading*. Cambridge: Cambridge University Press.
- Kintsch, W. y Van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 8 (5): 363-394.
- Le, Huong (2011). Developing a validity argument for the English placement fall 2010 listening test at Iowa State University, [en línea], disponible en: <http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1054&context=etd> [consulta: 20 de febrero de 2018].
- Llorián, S. (2015). La evaluación de la competencia comunicativa específica en el ámbito de las ingenierías y de la arquitectura. Tesis doctoral no publicada, Universidad Politécnica de Madrid.
- Lynch, T. (2011). Academic listening in the 21st century: reviewing a decade of research, *Journal of English for Academic Purposes*, 10(2), pp. 79-88.
- McNamara, D. y Kintsch, W. (1996). Learning from texts: effects of prior knowledge and text coherence, *Discours eProcesses*, 22(3), pp. 247-288.
- Montolío, E. Garachana, M. y Santiago, M. (2000). *Manual de escritura académica*. Barcelona: Ariel.
- Nesi, H., (2001). A corpus-based analysis of academic lectures across disciplines. En: Cotterill, J., Ife, A., eds., *Language Across Boundaries*, BAAL, vol. 16. London: Continuum, pp. 201-218.
- Nesi, H. y Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, 11, (3), pp. 283-304.

- Parodi, G. (2007). El discurso especializado escrito en el ámbito universitario y profesional: constitución de un corpus de estudio. *Revista Signos* 40(63), pp. 147-178.
- Parodi, G. (2005). *Discurso Especializado e Instituciones Formadoras*. Valparaíso: Ediciones universitarias de Valparaíso.
- Rost, M. (2011). *Teaching and Researching: Listening*. London, Routledge.
- Rost, M. (2002). *Listening Tasks and Language Acquisition*, London, Routledge.
- Rost, M. (1994). *Listening*. London, Longman.
- Richards, J. C. (1983). Listening comprehension: approach, design, procedure. *TESOL Quarterly*, 17(2), pp. 219-240.
- Regueiro, M. L. y Saéz, D. (2013). *El español académico*. Madrid: Arco Libros.
- Vázquez, G. (2006). Un análisis didáctico del discurso académico español como contribución a la movilidad estudiantil europea. En: *Actas del III Congreso Internacional del Español con Fines Específicos*, Consejería de Educación en Bélgica, Países Bajos y Luxemburgo, pp. 132-144. [En línea], disponible en: file:///C:/Users/Usuario/Desktop/Actas_III_CIEFE.pdf [acceso, 20 de febrero de 2018]
- Vázquez, G., ed. (2001a). *Guía didáctica del discurso académico escrito: ¿cómo se escribe una monografía?* Madrid: Edinumen.
- Vázquez, G., ed. (2001b). *El discurso académico oral: guía didáctica para la comprensión auditiva y visual de clases magistrales*. Madrid: Edinumen.
- Weir, C. J. (2005). *Language Testing and Validation*. New York: Pallgrave Mc Milan.
- Weir, C. J., Hawkey, R., Green, A., Unaldi, A. y Devi, S. (2012). The relationship between the academic reading construct as measured by IELTS and the reading experiences of students in the first year of study at a British university. En Taylor, L. and Weir, C. J. *Research in Reading and Listening Assessment*. Cambridge: Cambridge University Press.
- Young, L. (1994). University lectures—macro-structure and micro-features. En Flowerdew, J., *Academic Listening: Research Perspectives*, Cambridge, Cambridge University Press, 1994, pp. 159-176.