

Evaluación certificadora de la comprensión auditiva de ELE en el ámbito académico

Assessment of listening comprehension of Spanish as a foreign language in the academic context

SUSANA MARTÍN LERALTA
UNIVERSIDAD NEBRIJA
smartinl@nebrija.es

BEGA GARCÍA VIÚDEZ
UNIVERSIDAD NEBRIJA
bgarcia@nebrija.es

MARÍA CECILIA AINCIBURU
UNIVERSIDAD NEBRIJA
caincibu@nebrija.es

CRISTINA MARTÍNEZ SANZ
UNIVERSIDAD NEBRIJA
cmartine@nebrija.es

Resumen

La movilidad actual en el contexto universitario requiere replanteamientos para la didáctica de la lengua extranjera y para el diseño de fórmulas de evaluación válidas para certificar la competencia lingüística comunicativa requerida en las situaciones de interacción y el manejo de los géneros textuales orales y escritos del ámbito académico. El artículo parte de las habilidades de comprensión oral y las características de los textos orales propios del ámbito, para proponer una prueba evaluadora multinivel de español académico. Se presentan el procedimiento de validación de la prueba de comprensión audiovisual y conclusiones útiles para la evaluación y la didáctica de ELE en contexto académico.

Palabras clave: evaluación de la comprensión auditiva, español académico

Abstract

The current mobility in the university context requires new approaches for both the teaching and the certification assessment of the communicative linguistic competence in the foreign language. Students have to interact in academic situations and cope with specific oral and written textual genres of the academic field. Standpoint of this paper are the skills of oral comprehension required and the characteristics of the academic oral texts, with the aim of proposing a



multilevel test of academic Spanish. The validation procedure of the audiovisual test and some useful conclusions for the teaching and assessment of Spanish as a foreign language in academic context are presented.

Key words: listening comprehension assessment, Spanish as an academic language

1. Introducción

La importancia de la evaluación de las lenguas viene dada por su practicidad, la oportunidad política, el desarrollo en la enseñanza de idiomas y la lingüística aplicada y por la tendencia global del aumento de la movilidad (Green, 2012).

Los objetivos de la convergencia europea del Espacio Europeo de Educación Superior (EEES) abarcan varios aspectos, como recogen Gola y Miralles (2009). Según las autoras, la implementación de estas reformas precisa cambios que abarcan la programación curricular, la revisión de la metodología docente y la evaluación. Así, la enseñanza promulgada por el EEES se centra en las competencias del individuo, lo que entronca con los postulados del Marco Común Europeo de Referencia para las Lenguas (MCER) del Consejo de Europa (2001).

La competencia lingüística se constituye, en este escenario de movilidad internacional, en una de las más importantes, en tanto que resulta imprescindible para el acceso a los contenidos curriculares.

La lengua española es protagonista, junto con la lengua inglesa, de la creciente globalización de los procesos educativos y de los mercados de bienes y servicios, con una demanda en alza de productos culturales. En la sociedad del conocimiento, el principal soporte de lo que se sabe y de cómo se trasmite es la lengua, por lo que su aprendizaje y uso es cada vez más demandado.

De acuerdo con Vázquez (2005), el español con fines académicos constituye un pilar fundamental para el éxito de los programas de movilidad europea, en tanto que la conciencia de los diversos estilos lectivos, de los géneros académicos propios de las diversas tradiciones universitarias y de las características de los productos textuales empleados en el ámbito universitario, facilita la obtención de los créditos reconocibles a través de los sistemas de transferencia existentes.

En el caso del inglés, al implantarse el EEES resulta imprescindible para todos los universitarios poder acreditar un nivel de idioma al finalizar sus estudios de grado o solicitar el acceso a estudios de postgrado, y por ser actualmente la lengua internacional de la ciencia y de su transferencia y difusión. En el caso del español, porque la comunidad hispanohablante se está posicionando como la segunda de occidente, con sus 450 millones de hablantes y por la emergencia de varios de los países en los que es lengua oficial.

España, aunque no destaque en su demografía, sí lo hace con el prestigio de sus universidades, por lo que aumenta progresivamente el número de estudiantes extranjeros no nativos que eligen sus aulas. Si toda esta población universitaria debe alcanzar un dominio suficiente de español académico y si no existe una certificación de español académico universal, resulta de gran interés invertir esfuerzo investigador en generar un tipo de examen que dé respuesta oportuna, actual y eficaz a tal demanda.

En el curso académico 2017/18, según los últimos datos publicados en EDUCAbase¹ por el Ministerio de Educación y Formación Profesional y el Ministerio de Ciencia, Innovación y Universidades, las universidades españolas presenciales acogieron a más de 120 000 estudiantes internacionales para realizar diferentes modalidades de estudios en nuestro país, lo que representa el 9% del total del estudiantado. De ellos, 82.123 procedían de países no hispanohablantes (6,1% del total). A los estudiantes Erasmus, que ya elegían mayoritariamente nuestro país como destino, se ha sumado en los últimos años un notable incremento de los internacionales que acuden a realizar estudios universitarios completos, según el informe de la OCDE (Ministerio de Educación y Formación Profesional, 2018). Estos datos ponen de manifiesto la necesidad de un certificado acreditativo de la competencia en español académico para los denominados *alumnos incoming*.

Sin duda, el diseño y la implantación de una prueba de español académico que evalúe de manera válida y fiable el constructo deseado constituirá un gran valor para todos los agentes implicados (investigadores, alumnos, profesores, extranjeros e hispanohablantes).

En García Delgado et al. (2008) ya se destacaba la necesidad de superar la situación carencial del español como lengua de la ciencia y de la técnica. Es muy probable que un buen examen de español académico, debidamente difundido de manera universal, se convierta en un mecanismo eficaz para ayudar a superar tal carencia.

La CRUE (Conferencia de Rectores de la Universidad Española) reconoce “el valor estratégico que para las universidades españolas representa la evaluación y acreditación de competencias de sus estudiantes” y explica que el proyecto CertiUni, en colaboración con el Ministerio de Educación y la Confederación Española de Organizaciones Empresariales (CEOE), pretende responder al reto que supone el Espacio Europeo de Educación Superior, impulsando la formación y la certificación de las competencias transversales. Dentro de estas competencias transversales que se consideran en CertiUni, se encuentra la competencia en lenguas.

De acuerdo con Vázquez (2009:123), “Quienes se interesen por trabajar el discurso académico español y quieran apartarse de lo que es la clase magistral y la producción de trabajos escritos, se enfrentarán a un área poco investigada, la que se refiere a la evaluación”. El trabajo que aquí se presenta pretende avanzar en este asunto. Corresponde al diseño y validación de un examen multinivel de certificación de español académico que mide los niveles de dominio B1, B2 y C1 del Marco Común Europeo de Referencia (2001) en el uso oral y escrito de diferentes géneros propios de la universidad y de la ciencia. Concretamente, nos ocuparemos de la prueba de comprensión auditiva. El examen ha sido desarrollado por el Grupo de investigación en Lingüística aplicada a la enseñanza de lenguas extranjeras (LAELE) de la Universidad Nebrija para optar a la acreditación de ACLES

¹http://estadisticas.mecd.gob.es/EducaDynPx/educabase/index.htm?type=pcaxis&path=/Universitaria/Internacionalizacion/Estudiantes_Profesores_Extranjeros/&file=pcaxis

(Asociación de Centros de Lenguas en la Enseñanza Superior) como examen CertACLES, avalado por la CRUE.

Se decide evaluar la comprensión auditiva empleando un input audiovisual por el carácter multimodal de la destreza de comprensión oral en el ámbito académico. La prueba se articula en tres tareas comunicativas a partir de los géneros discursivos orales más frecuentes en las aulas universitarias españolas, a saber, las instrucciones para el seguimiento de una asignatura o la realización de un trabajo, la discusión o debate académico y la exposición magistral. La ejecución de estas tareas requiere del candidato la aplicación de las mismas habilidades y estrategias de comprensión oral que pondría en práctica en una situación comunicativa similar fuera del contexto de evaluación lingüística.

2. Competencia lingüística en el ámbito académico

2.1. Competencias en el EEES

Las competencias que los estudiantes deben adquirir al finalizar su trayectoria académica se clasifican en disciplinares (saber), profesionalizantes (saber hacer) y académicas (saber trascender).

Tal como señala Cruz (2009), en el sistema universitario español se ha llevado a cabo un profundo debate sobre la forma más adecuada para llevar a cabo estos principios a las especialidades académicas. La autora recoge la reflexión de José Ramón Alonso, Rector de la Universidad de Salamanca en 2009, publicada por el diario El País, quien concreta aquellos conocimientos comunes a todo aquel que comienza estudios superiores, independientemente de su especialidad; algo que no se había especificado anteriormente en los currículos tradiciones, como recuerda Cruz. De las diez capacidades establecidas por Alonso, ninguna se refiere específicamente a la comprensión oral. De hecho, la propia Cruz (ibídem: 381) indica que las destrezas orales han sido las más relegadas dentro del conjunto de competencias del estudiante universitario y que los contenidos lingüísticos universitarios olvidan frecuentemente el estudio y práctica de la lengua oral, aun siendo esta la variedad más utilizada.

En nuestro contexto de receptores de estudiantes internacionales en el sistema universitario español, la didáctica y evaluación deberán encaminarse a dotarles de los recursos lingüísticos e interculturales precisos para alcanzar estas competencias en el ejercicio de todas las destrezas lingüísticas. Nos encontramos, por tanto, ante la necesidad de enseñar y evaluar una lengua con fines específicos, académicos. La denominación de *lenguaje académico*, de acuerdo con Pastor Cesteros y Ferreira (2018), posee además la ventaja de mostrar una relación tripartita, la que se establece entre: (a) la utilización de la L2 como lengua vehicular en el ámbito académico; (b) la lengua de especialidad de la que se trate en función de los estudios cursados y, en su caso; y (c) el uso discursivo en el desempeño de la profesión hacia la que se encamina tal formación.

Veamos a continuación con más detalle las particularidades de este

lenguaje.

2.2. El lenguaje académico y los géneros discursivos del ámbito

En 1979 Cummins acuñó dos siglas esenciales, BICS y CALP, para entender cómo se adquiere una segunda lengua. Su distinción entre BICS (*Basic Interpersonal Communication Skills*) y CALP (*Cognitive Academic Language Proficiency*) pretendía dejar claro que el alumno debe ejercitar tanto las destrezas conversacionales, que le permiten mantener una conversación a diario, como las destrezas cognitivas académicas, que le facilitarán la construcción del pensamiento, y, por tanto, del conocimiento. Así pues, recientemente se ha empezado a considerar la importancia de incluir la enseñanza del lenguaje académico en los diferentes programas de enseñanza de lenguas.

Asimismo, Krashen y Brown (2007), basándose en la distinción inicial de Cummins, afirman que la competencia académica (*academic competence*) descansa en dos pilares fundamentales: por un lado, el conocimiento del lenguaje académico, que define como la lengua especial que se utiliza en el ámbito académico y en el mundo profesional; por otro, el conocimiento especializado de las distintas áreas de conocimiento (matemáticas, ciencias, historia, etc.). Además, incluyen las estrategias como un tercer componente en la competencia académica, ya que el uso de estrategias ayuda a la adquisición tanto del lenguaje académico como del aprendizaje de la materia en cuestión. Y concluyen: “We want our students to be able to use their second language for demanding tasks, for business, science, politics, etc. beyond carrying out daily conversation” (Krashen y Brown, 2007:1).

En esta investigación entendemos por lenguaje académico el uso específico de una lengua en el ámbito académico, que se concreta en los géneros discursivos que desarrollan tanto profesores como alumnos en el medio universitario y de acuerdo con sus respectivas funciones de enseñanza y aprendizaje. Comparte características con el lenguaje científico, ya que el ámbito académico es uno de los espacios fundamentales de generación y de transmisión del saber científico.

En lo sustancial, el discurso académico debe respetar, cuando hace públicos los resultados de las ciencias exactas tanto como de las aplicadas y de las especulativas, los rasgos comunes del lenguaje científico: universalidad, objetividad, denotación, verificabilidad, coherencia, adecuación, precisión, formalidad, claridad y elegancia expresivas. El científico, como el académico, son lenguajes propios de unos campos especializados de la actividad humana, por lo que hay tantos conjuntos diferenciados de dominios científicos como de actividades y conocimientos. Estas diferencias se reflejan también en el texto académico, en la terminología, en el distinto grado de formalización, en los criterios de verificación. La formalización en las ciencias exactas, por ejemplo, no coincidirá con la propia de los saberes especulativos; la terminología de las

experimentales será necesariamente diferente de la propia de las humanísticas. Por ello, el lenguaje académico debe cumplir con estas exigencias.

Por su parte, el concepto de género, como señala Parodi (2009), se ha discutido tan profusamente en los últimos años que el término se ha puesto de moda. En nuestro estudio adoptamos la visión tridimensional del autor cuando clasifica el género desde las dimensiones social, cognitiva y lingüística. Esto es así en tanto que los usuarios de la lengua poseemos “unos conocimientos previos y una dimensión lingüística que permite la interacción entre las redes sociales y las dimensiones cognitivas” (*ibídem*: 75).

Según Parodi, a lo largo de la vida vamos construyendo socio-constructivísticamente una representación cognitiva de los sujetos, participantes, entornos y contextos. Al tratarse de una representación dinámica, que va cambiando, podríamos pensar en enseñar una lengua en un contexto específico y para un fin didáctico particular.

De acuerdo con el autor (*ibídem*: 76), los géneros discursivos se pueden actualizar y caracterizar “a través de una serie de elementos: los propósitos comunicativos, el rol de los participantes, los rasgos léxico-gramaticales, los modos de organización del discurso y lo que algunos llaman *tramas*, que tienden a ocurrir mayoritariamente a través de un texto”.

Por lo tanto, podríamos concluir que los textos son fragmentos de lengua utilizados para llevar a cabo un acto o acontecimiento comunicativo de tipo discursivo, como propone el *Plan Curricular del Instituto Cervantes* (2006). En este sentido, coincidimos con Villar (2013:14) cuando afirma que

Los textos que pertenecen a un mismo género discursivo se han desarrollado históricamente en una comunidad de hablantes, ámbito social o profesional y comparten una misma forma de organización de la información y del discurso. Sin embargo, el uso de conocimientos lingüísticos y discursivos típicos de un género convencional, está estandarizado y establecido por la tradición.

En lo que se refiere al ámbito académico, Ainciburu (2010), añade que el discurso científico-académico se caracteriza por las dimensiones informacional y modalizadora: por una prosa densa en términos léxico-gramaticales y por rasgos que presentan modalización de las certezas expresadas en los textos.

Pero la definición de discurso académico requiere un apunte más. Como bien indica Vázquez (2009:119), no podemos referirnos a este discurso únicamente como el lenguaje de la ciencia, puesto que dejaríamos fuera una amplia esfera de acción que afecta al estudiantado. Esto hace que la autora vaya a una definición más amplia, al considerar como discurso académico “la comunicación cotidiana entre docentes y estudiantes en los ámbitos universitarios y las situaciones ritualizadas en las cuales nos encontramos”. Por ello la misma autora (2004: 1130) alude tanto a destrezas como a tareas propias de los ámbitos universitarios al referirse a la amplitud del término. Ambos elementos, destrezas y tareas, serán tenidos en cuenta en el diseño de la prueba de comprensión audiovisual de español académico del presente trabajo, como se irá mostrando más adelante.

El concepto está igualmente relacionado con el de *comunidad discursiva* (Regueiro y Sáez 2013: 15-16), lo que lleva a considerar las variables individuales de los agentes implicados, a saber, docentes y estudiantes, así como al contexto cultural concreto en que se encuentran.

En conclusión, si las convenciones del género están estandarizadas y son compartidas por los usuarios de una lengua y cultura, podemos suponer una necesidad de enseñar específicamente estas particularidades a hablantes de otras lenguas y culturas; en el caso que nos ocupa, nos referiremos a las convenciones del discurso académico en español, mediante una selección de textos que abarquen el mayor número de rasgos prototípicos del mismo, así como su uso para la resolución de tareas comunicativas representativas de las situaciones propias del ámbito académico español.

2.3. Lenguaje académico oral y habilidades de comprensión auditiva

El lenguaje oral en situaciones académicas es más complejo y difícil de procesar que el de las conversaciones diarias, tal como pone de manifiesto el estudio de Marx, Heppt y Henschel (2016). Esta hipótesis, que ya se había apuntado repetidas veces en estudios previos tanto empíricos como teóricos, ha llevado a lo largo de los años a postular que ambos tipos de comprensión podrían basarse en diferentes repertorios de subdestrezas, dando como resultado algunas definiciones de comprensión auditiva académica y de sus componentes. De la revisión bibliográfica previa, las autoras concluyen que el lenguaje académico oral se caracteriza por el dominio de géneros específicos, el manejo de estrategias de razonamiento y argumentación características del contexto académico, así como por poseer rasgos lingüísticos que requieren procesos cognitivos específicos, como la generalización o la abstracción.

Si nos atenemos al primero de los componentes, el dominio de géneros específicos, nos encontramos con una distinción entre los ligados a la enseñanza, a lo institucional y a la investigación según la clasificación de Villar (2013). En el primer grupo estarían las tutorías, consultas de cátedra y actividades de clase: clase magistral, seminario, taller, presentación académica oral y examen oral. Entre las segundas se encontrarían, entre otros, la conferencia inaugural del curso, el discurso de graduación o el de aceptación de una distinción, mientras que las de investigación científica abarcarían las conferencias, ponencias, comunicaciones, coloquios, defensas de tesis y presentaciones de proyectos.

En lo que respecta a los rasgos lingüísticos del lenguaje oral académico, Marx, Heppt y Henschel (2017) destacan el tipo de léxico y la densidad léxica, los rasgos sintácticos empleados para la segmentación oracional, la coherencia y el modo de representación, y los rasgos discursivos correspondientes a los turnos de palabra, las formas monologales, los tipos de texto específicos a la disciplina y los estándares estilísticos.

El procesamiento de este tipo de lenguaje requiere unas habilidades

específicas que vienen determinadas tanto por la habilidad lingüística del estudiante, como por el contenido, la estructura y el estilo del discurso académico (Ayardoust, Goh y Lee, 2012). Entre las habilidades de escucha, se encuentran, entre otras, distinguir información relevante de irrelevante, procesar oraciones extensas, procesar conceptos abstractos y comprender el significado del mensaje cuando existe un alejamiento del contexto del referente enunciado.

Estas habilidades han de aplicarse a la resolución de tareas en las que el input es principalmente multimodal y donde se integran diferentes destrezas comunicativas para procesarlo. Así, lo más habitual es que los estudiantes escuchen a interlocutores a los que están viendo (apoyo en la gestualidad), en un discurso oral que suele estar basado en documentación escrita compartida previamente con los oyentes (lectura anterior a la escucha para asegurar unos conocimientos de partida), y en el que el orador emplea recursos para comunicar más efectiva y didácticamente su mensaje: proyección de diapositivas con textos, imágenes, sonidos, vídeos que constituyen otra input auditivo, etc. El estudiante receptor debe por tanto ser capaz de escuchar, leer y escribir (toma de notas) al mismo tiempo para procesar adecuadamente el input, así como de preparar una respuesta oral cuando el mensaje requiera una interacción por su parte.

A la hora de diseñar un examen de evaluación válido será preciso, por una parte, que las tareas respeten la autenticidad situacional propia del desenvolvimiento de la destreza de comprensión auditiva en el ámbito académico, que el input sea fiel reflejo de los rasgos del lenguaje académico tanto a nivel macroestructural, como lingüístico y, por otra, que las preguntas de comprensión eliciten los mismos procesos de comprensión auditiva que se requerirían en la situación académica real de recepción de ese mensaje, de cara a poder medir la competencia del candidato en las habilidades de escucha en ese contexto. A continuación se recogen las habilidades objeto de evaluación en nuestra prueba multinivel.

3. Diseño y validación de una prueba multinivel de evaluación de comprensión auditiva de español académico

3.1. Especificaciones de la prueba evaluadora

3.1.1. Descriptores de las habilidades de comprensión evaluadas

Las especificaciones son un “conjunto de documentos con la información sobre lo que mide el examen y cómo lo mide y con toda la información relevante para quienes necesitan diseñarlo y usarlo” (Martínez Baztán, 2011:50). Como señalan Figueras y Puig (2013), es necesario disponer de un documento de especificaciones de examen que incluyan pautas para la elaboración de instrumentos de evaluación válidos y fiables, a través de los cuales los candidatos demuestren que son capaces de comprender y de expresar en situaciones de comunicación de la vida real.

Siguiendo las recomendaciones de Bachman y Palmer (2010), Figueras y Puig (2013), así como el Manual del Consejo de Europa (2009) para relacionar exámenes con el Marco común europeo de referencia para las lenguas, se han redactado unas especificaciones concretas para nuestro examen de español académico multinivel (niveles B1/B2/C1) de español en el ámbito académico.

De las tres opciones que recogen Bachman y Palmer (1996: 121), a la hora de definir el constructo de nuestro examen se ha optado por excluir de la definición del constructo el conocimiento del tema, para considerar únicamente la habilidad lingüística del candidato. Así, nos enfocaremos en los componentes de la habilidad comunicativa lingüística enunciados por Douglas (2000:35) basados en el conocimiento de la lengua (gramatical, textual, funcional y sociolingüístico) y la competencia estratégica (de evaluación, establecimiento de objetivos, planificación y control de la ejecución). Igualmente, se han seguido las recomendaciones para anclar el examen de español académico al *Marco Común Europeo de Referencia para las Lenguas*, recogidas en *CEFR: Manual for Test Development and Examining within the CEFR* (ALTE y Consejo de Europa, 2011) y *Manual for Relating Language Examinations to the CEFR* (Consejo de Europa, 2009).

Los objetivos específicos de nuestra prueba de comprensión auditiva resultan de esta manera:

- Nivel B1
 - Comprender ideas principales, información concreta e instrucciones en un discurso claro y en lengua estándar que trate temas cotidianos del ámbito educativo (Comprensión auditiva en general).
 - Identificar tanto el mensaje general como los detalles específicos en un discurso claro y en lengua estándar que trate temas cotidianos del ámbito educativo (Comprensión auditiva en general).
 - Comprender una conferencia o una charla sencilla en el ámbito universitario y claramente estructurada sobre un tema que resulte familiar (Escuchar conferencias y presentaciones).
- Nivel B2
 - Comprender discursos extensos en el ámbito académico y líneas complejas de argumentación siempre que el tema sea razonablemente conocido, no haya un ruido excesivo de fondo ni predomine un uso idiomático de la lengua, y el desarrollo del discurso se facilite con marcadores explícitos.
 - Comprender las ideas principales de conferencias, charlas, informes y otras formas de presentación académica lingüísticamente complejas (Escuchar conferencias y presentaciones).
 - Comprender declaraciones sobre temas tanto concretos como abstractos e identificar el contenido de la información, así como los puntos de vista y las actitudes del hablante.
- Nivel C1

- Comprender discursos extensos sobre temas abstractos y complejos del ámbito académico que sobrepasan su especialidad, aunque tenga que confirmar algún que otro detalle.
- Comprender con relativa facilidad la mayoría de las conferencias, discusiones, debates e instrucciones complejas del ámbito educativo.

3.1.2. Tipos de tareas de comprensión e ítems

Se proponen tareas de comprensión audiovisual propias de diferentes situaciones académicas. Se tienen en cuenta las recomendaciones de Buck (2001: 206-207) para diseñar un test de comprensión auditiva académica, de manera que las tareas incluyen los tres principales tipos de discurso observados en este contexto: instrucciones, discusión y conferencia/clase magistral.

Para ello se utilizan textos de características discursivas diferentes y con preguntas de distinta dificultad, de manera que permitan medir los diferentes niveles de dominio lingüístico. Se trata de una prueba de corrección objetiva en su mayor parte (80% del valor total de la prueba) y de corrección por criterios en una actividad de toma de notas y resumen (20%); para esta última se establece una rúbrica de evaluación.

Se emplean los siguientes tipos de textos:

- a) Explicaciones e instrucciones sobre asuntos habituales de la vida académica.
- b) Debate o discusión académica sobre temas de interés general y con vocabulario mayoritariamente no especializado.
- c) Fragmento de una conferencia o clase magistral sobre un tema de interés general y con vocabulario mayoritariamente no especializado.

En concreto, se miden las siguientes microhabilidades de comprensión oral:

- a) Escuchar para identificar la idea general y discriminar ideas secundarias.
- b) Escuchar para localizar información específica.
- c) Escuchar para identificar detalles relevantes.
- d) Escuchar para inferir el significado del enunciado por el contexto o por el conocimiento del mundo y del tema.
- e) Escuchar para inferir puntos de vista de los hablantes o información implícita.
- f) Escuchar el fragmento de una conferencia para tomar notas y resumir su contenido.

Se utilizan los siguientes tipos de ítems en las actividades de la prueba:

- a) Preguntas de elección múltiple (para comprensión analítica; identificar información, opiniones y postura del autor).
- b) Preguntas de relacionar fragmentos del texto con pequeños resúmenes o enunciados (para comprensión global e interpretación del contenido).

- c) Preguntas de Verdadero/Falso (para medir la comprensión de implícitos y de puntos de vista/voces del texto/actitudes del emisor).
- d) Preguntas de rellenar huecos con palabras textuales del discurso.
- e) Tomar notas y elaborar un resumen del contenido de un fragmento de discurso académico.

3.1.3. Número de tareas en relación con el input y número total de ítems

En el presente examen se denomina “tarea” al conjunto de ítems que miden la comprensión de un mismo input. Se proponen 4 tareas a partir de 3 vídeos cuya duración oscila entre 3 y 5 minutos. Todos los vídeos empleados corresponden a actividades académicas auténticas: fragmentos de instrucciones, debates y clases magistrales que han tenido lugar realmente en el ámbito universitario.

Vídeo 1_Tarea 1: Preguntas de respuesta cerrada (corrección objetiva) a partir de un texto oral sobre instrucciones o procedimientos propios del ámbito universitario.

Vídeo 2_Tarea 2: Preguntas de respuesta cerrada (corrección objetiva) a partir de un texto oral sobre un debate o discusión en el ámbito académico (interacción académica formal).

Vídeo 3_Tarea 3.1.: Toma de notas y resumen (corrección por criterios-integración de la destreza de expresión escrita, que no se evalúa en este caso) de un fragmento de una presentación académica oral (conferencia, lección magistral).

Vídeo 3_Tarea 3.2.: Preguntas de respuesta cerrada (corrección objetiva) a partir de un fragmento de una presentación académica oral (conferencia, lección magistral).

Los vídeos 1 y 2 se proyectan dos veces para completar las tareas 1 y 2 respectivamente. El vídeo 3 se proyecta dos veces para completar la tarea 3.1. (toma de notas y resumen) y una tercera vez para realizar la tarea 3.2.

La prueba está constituida por 60 ítems de respuesta cerrada repartidos en tres tareas más una tarea de toma de notas y resumen, de corrección por criterios. Las tres tareas de corrección objetiva tienen un peso total del 80% de la calificación de esta prueba, mientras que la de corrección por criterios tiene un valor del 20%. La duración estimada de la prueba es de 60 minutos.

3.1.4. Instrucciones de la prueba

Las instrucciones son claras y se formulan en español, en un nivel igual o menor que el nivel evaluado. El enunciado de la tarea anticipa el género discursivo y temática del texto oral para facilitar la activación de esquemas del oyente que favorezcan la comprensión auditiva. Asimismo, se deja tiempo para la lectura previa de las preguntas y se aclara cuántas veces se proyectará el vídeo para que el candidato pueda planificar mejor la escucha dirigida.

Ejemplo: *A continuación va a escuchar las instrucciones de una profesora sobre el desarrollo de una asignatura a través del Campus Virtual de la Universidad.*

En primer lugar, lea las preguntas. Después tendrá ocasión de ver el vídeo dos veces para completar sus respuestas según el contenido del vídeo.

3.2. Validación de la prueba multinivel de comprensión auditiva

3.2.1. Validación de contenido por expertos

Para el diseño de la prueba piloto se seleccionaron un total de 12 vídeos, de los que finalmente se eligieron 3 para someter a validación. Se cuidó especialmente que los redactores de ítems no trabajaran en ningún momento con transcripciones de los vídeos, de manera que se asegurase una medición válida de las habilidades de comprensión auditiva. Para la selección de vídeos y la elaboración de tareas se tuvieron en cuenta las recomendaciones de Buck (2001) y de Martín Leralta (2011). Se utilizó material audiovisual propiedad de la Universidad Nebrija que contaba con cesión de uso de derechos de imagen. Se elaboraron inicialmente 100 ítems, de los que 26 fueron descartados al revisarlos de manera conjunta por el equipo de redactores, con lo que quedaron 74 ítems para el proceso de validación.

En primer lugar se realizó el análisis de la validez de contenido de los ítems con expertos de las diferentes tareas del examen, de acuerdo con las indicaciones de Martínez Baztán (2011: 82-83). Se contó para ello un panel de 6 jueces: tres expertos en la didáctica de la destreza evaluada, uno en lenguaje académico y dos en evaluación de la lengua. Los jueces también estimaron a priori la dificultad de los ítems de corrección objetiva, datos que permitieron aplicar posteriormente el método Angoff (Angoff, 1971) para establecer los puntos de corte del examen multinivel, como se explicará más adelante.

3.2.2. Pilotaje de la prueba y análisis de los ítems

Tras introducir las mejoras sugeridas por los expertos, se llevó a cabo el pilotaje de los ítems con 68 aprendices de ELE del Centro de Estudios Hispánicos de la Universidad Nebrija de diferentes niveles de dominio lingüístico: 13 estudiantes de A2, 26 de B1, 8 de B2, 12 de C1 y 9 de C2. El proceso se realizó de la siguiente forma:

- a) Determinación de la matriz de resultados del test en una hoja de cálculo Excel y volcado de los resultados al programa jMetrik. Aunque la prueba oficial consta de 60 ítems de CAV, se pilotaron 73 ítems para poder eliminar aquellos que no demostraran una calidad suficiente en el proceso de test. A partir de los exámenes realizados por los alumnos se obtuvo una tabla Excel. Los datos de esta matriz se incorporaron al programa jMetrik a través de un fichero csv intermedio junto con las

claves de respuestas de la prueba. El test se plantea como una serie de ítems dicotómicos con una sola respuesta correcta, de manera que cada ítem toma los valores 0 para respuesta incorrecta y 1 para correcto.

- b) Análisis de los ítems a través del comando “Item Analysis” de jMetrik. Se obtuvieron los índices de dificultad y discriminación de cada ítem, así como la desviación estándar:

Item	Option (Score)	Difficulty	Std. Dev.	Discrimin.
cav_ta_ia	Overall	0,2500	0,4362	0,2657
cav_ta_i2	Overall	0,1324	0,3414	0,5678
cav_ta_i3	Overall	0,1765	0,3841	0,5842
cav_ta_i4	Overall	0,1618	0,3710	0,5940
cav_t1_i5	Overall	0,7794	0,4177	0,1921
cav_t1_i6	Overall	0,6324	0,4857	0,2614
cav_t1_i7	Overall	0,3971	0,4929	0,2507
cav_t1_i8	Overall	0,3529	0,4814	0,3001
cav_t1_i9	Overall	0,6618	0,4766	0,2881
cav_t1_i10	Overall	0,3088	0,4654	0,1948
cav_t1_i11	Overall	0,2941	0,4590	0,1769
cav_t1_i12	Overall	0,7500	0,4362	0,2732
cav_t1_i13	Overall	0,6618	0,4766	0,2577
cav_t1_i14	Overall	0,7647	0,4273	0,3040
cav_t1_i15	Overall	0,3971	0,4929	0,1548
cav_ta_ia6	Overall	0,0000	0,0000	NaN
cav_ta_ia7	Overall	0,4265	0,4982	0,2716
cav_t1_i18	Overall	0,3235	0,4713	0,2349
cav_t1_i19	Overall	0,8529	0,3568	0,2078
cav_t1_i20	Overall	0,1176	0,3246	0,1836
cav_t1_i21	Overall	0,5735	0,4982	0,3415
cav_t1_i22	Overall	0,8088	0,3962	0,1579
cav_t1_i23	Overall	0,8088	0,3962	0,1875
cav_t1_i24	Overall	0,4706	0,5028	0,3682
cav_t1_i25	Overall	0,2500	0,4362	0,2295
cav_t1_i26	Overall	0,2794	0,4520	0,1443
cav_t2_i1	Overall	0,3235	0,4713	0,3244
cav_t2_i2	Overall	0,8676	0,3414	0,1308
cav_t2_i3	Overall	0,4706	0,5028	0,2628
cav_t2_i4	Overall	0,6471	0,4814	0,2826
cav_t2_i5	Overall	0,3676	0,4857	0,5543

Tabla 1. Índices de dificultad y discriminación de cada ítem

Asimismo, se obtuvo la siguiente estadística general del test:

TEST LEVEL STATISTICS	
Number of Items = 73	
Number of Examinees =	68
Min =	14,0000
Max =	62,0000
Mean =	34,3676
Median =	31,5000
Standard Deviation =	11,4992
Interquartile Range =	16,7500
Skewness =	0,5464
Kurtosis =	-0,3914
KR21 =	0,8744

Tabla 2. Estadística general del test

Mediante el comando “Item deleted reliability” se realizó un análisis de fiabilidad del test. En este análisis se usan los modelos de Guttman's L2, Coefficient Alpha, Feldt-Gilmer, Feldt-Brennan y Raju's Beta y se obtuvo la siguiente tabla:

RELIABILIY ANALYSIS			
Method	Estimate	95% Conf. Int.	SEM
Guttman's L2	0,9120	(0,8794, 0,9393)	3,4120
Coefficient Alpha	0,9039	(0,8684, 0,9338)	3,5646
Feldt-Gilmer	0,9082	(0,8743, 0,9368)	3,4838
Feldt-Brennan	0,9079	(0,8738, 0,9365)	3,5163
Raju's Beta	0,9039	(0,8684, 0,9338)	3,5646

Tabla 3. Análisis de fiabilidad de la prueba

Además, se obtuvo la estimación de fiabilidad del test cuando un determinado ítem ha sido eliminado del análisis, con lo que se permitió la elección de los ítems sobrantes que convenía eliminar para mejorar la fiabilidad del examen.

c) Aplicación de la Teoría de Respuesta al ítem mediante el modelo de Rasch y establecimiento de los valores de dificultad de los ítems beta (β) y habilidad de los alumnos (θ), así como los parámetros Infit y Outfit de ítems y alumnos para el cribado de la muestra.

Los análisis anteriores confirmaron que todos los ítems cumplían con los requisitos previos para someterlos a un análisis mediante el modelo de Rasch utilizando una aproximación mediante estimación de máxima verosimilitud conjunta.

Los parámetros que jMetrik define como WMS (Weighted Mean Squared Fit) y UMS (Unweighted Mean Squared Fit) son más conocidos como INFIT y OUTFIT respectivamente, siendo Std. WMS y Std. UMS los valores de WMS y UMS estandarizados.

El software utilizado también proporciona para cada alumno los parámetros similares a los obtenidos para los ítems. Entre estos parámetros cabe destacar el correspondiente a la habilidad del alumno (θ) y sus parámetros Infit y Outfit de definición similar a los ítems. Valores altos del parámetro Outfit pueden ser resultado de respuestas acertadas aleatoriamente por alumnos con un bajo nivel y valores altos del parámetro Infit indican que los ítems no están funcionando correctamente para el nivel de alumno objetivo. El valor esperado del estadístico Infit es 1. Por convención se considera que los valores superiores a 1.3 indican desajuste en muestras con menos de 500 casos, 1.2 en muestras de tamaño medio (entre 500 y 1000 casos) y 1.1 en muestras con más de 1000 casos (Smith, Schumaker y Bush, 1998). Para Outfit también se considera que el valor esperado es 1. Valores muy

inferiores pueden indicar un sobreajuste de los datos al modelo y valores muy superiores a 1 implican mucho ruido no recogido en el modelo.

A partir del análisis de nivel de dificultad e índice de discriminación de los ítems y de la información proporcionada por los parámetros Infit y Outfit, junto con la valoración realizada de ellos por el panel de expertos, se decidió eliminar los siguientes 13 ítems:

Item	Difficulty	Std. Error	WMS (Infit)	Std. WMS	UMS (Outfit)	Std. UMS
cav_t1_i15	0,43	0,26	1,15	1,42	1,18	1,42
cav_t1_i22	-1,83	0,33	1,05	0,3	1,3	0,93
cav_t1_i23	-1,74	0,32	0,97	-0,11	0,95	-0,05
cav_t2_i2	-2,2	0,37	0,99	0,03	1,18	0,55
cav_t2_i9	-1,65	0,31	1,06	0,39	1,13	0,51
cav_t2_i10	0,02	0,26	1,21	2,32	1,22	1,86
cav_t2_i13	-0,05	0,25	1,32	3,42	1,43	3,38
cav_t2_i15	-0,21	0,27	1,29	3,06	1,38	2,74
cav_t2_i17	-2,34	0,38	1,13	0,53	1,42	0,99
cav_t2_i23	0,78	0,28	1,43	3,11	1,46	2,67
cav_t2_i25	-0,25	0,26	1,23	2,55	1,34	2,49
cav_t3_i21	0,13	0,26	1,29	2,92	1,38	2,97
cav_t3_i23	-0,84	0,3	1,22	1,86	1,31	1,45

Tabla 4. Ítems descartados

3.2.3. Establecimiento de los puntos de corte y determinación de sensibilidad y especificidad de la prueba

El siguiente punto en el proceso fue nivelar a los alumnos de acuerdo a las notas obtenidas y las puntuaciones de corte establecidas por el panel de expertos del Método Angoff (Angoff, 1971). Se trata de uno de los denominados *métodos de juicio* dentro de los métodos existentes para establecer puntos de corte y se basa en el contenido del reactivo (Hambleton, 1980 en Leyva, 2011).

Para aplicarlo se empleó el panel de jueces de la validación de contenido y se les pidió la estimación a priori de la dificultad de cada ítem. Se consensuó la probabilidad de que un candidato de cada nivel de dominio de la lengua responda correctamente una pregunta de cada uno de los niveles que mide el examen, con lo que se obtuvo la siguiente matriz de probabilidades de respuesta correcta:

Nivel del candidato	Dificultad del ítem				
	A2	B1	B2	C1	C2
A2	70%	45%	15%	0%	0%

B1	85%	70%	45%	15%	0%
B2	95%	85%	70%	45%	15%
C1	100%	95%	85%	70%	45%
C2	100%	100%	95%	85%	70%

Tabla 5. Matriz de probabilidades de respuesta correcta

El panel de expertos determinó para la destreza de comprensión audiovisual una puntuación mínima del 54% (32 puntos sobre 60) para obtener el nivel B1, 74% (44 puntos sobre 60) para obtener el nivel B2 y 87% (52 puntos sobre 60) para obtener el nivel C1. De este modo se obtiene la siguiente tabla para cada uno de los 68 participantes en la validación:

Código participante	Nivel obtenido	Vsum	Theta
NN_001	A2	19	-1,256
NN_002	A2	14	-0,472
NN_003	A2	19	-1,360
NN_004	A2	30	-0,451
NN_005	A2	26	-0,637
NN_006	A2	28	-0,605
NN_007	A2	22	-1,093
NN_008	A2	17	-0,565
NN_009	A2	27	-0,683
NN_010	A2	17	-1,023
NN_011	A2	26	-0,763
NN_012	A2	25	-0,792
NN_012	A2	28	-0,605
NN_014	A2	21	-0,746
NN_015	A2	28	-0,605
NN_027	A2	19	-1,360
NN_028	A2	22	-1,093
NN_029	A2	28	-0,599
NN_031	A2	25	-0,756
NN_032	A2	31	-0,375
NN_033	A2	24	-0,921
NN_034	A2	29	-0,521
NN_039	A2	28	-0,605
NN_040	A2	30	-0,429
NN_041	A2	26	-0,763
NN_042	A2	28	-0,605
NN_045	A2	28	-0,301
NN_022	A2	30	-0,397
NN_038	A2	30	-0,419

NN_049	A2	31	-0,138
NN_050	A2	25	-0,774
NN_051	A2	22	-0,766
NN_058	A2	20	-1,073
NN_059	A2	27	-0,657
NN_006	B1	35	-0,074
NN_020	B1	34	-0,149
NN_021	B1	39	0,224
NN_023	B1	33	-0,224
NN_024	B1	44	0,602
NN_030	B1	32	-0,299
NN_035	B1	35	-0,074
NN_036	B1	44	0,602
NN_037	B1	42	0,449
NN_044	B1	32	-0,246
NN_046	B1	36	0,000
NN_026	B1	42	0,498
NN_052	B1	37	0,170
NN_053	B1	40	0,342
NN_055	B1	41	0,419
NN_062	B1	44	0,602
NN_065	B1	41	0,436
NN_066	B1	43	0,525
NN_067	B1	42	0,449
NN_069	B1	34	-0,149
NN_076	B1	37	0,120
NN_079	B1	44	0,602
NN_025	B2	46	0,758
NN_054	B2	49	1,063
NN_060	B2	52	1,259
NN_061	B2	49	1,001
NN_063	B2	47	0,941
NN_072	B2	46	0,758
NN_074	B2	52	1,340
NN_068	C1	62	2,341
NN_073	C1	59	1,960
NN_075	C1	57	1,739
NN_078	C1	58	1,847
NN_080	C1	59	1,960

Tabla 6. Nivel obtenido por cada candidato

Por lo tanto, la habilidad media (θ) de cada grupo de alumnos en la destreza de comprensión audiovisual resultó así:

Grupo	Theta
A2	-0,712
B1	0,219
B2	1,017
C1	1,970

Tabla 7. Habilidad media de cada grupo de candidatos según nivel de dominio

A partir de estos valores de habilidad media (θ) de un alumno s de un determinado nivel y dificultad (β) de los de un ítem i se puede determinar la probabilidad esperada de acertar dicho ítem i por parte de un alumno s mediante la fórmula del modelo de Rasch:

$$P_{is} = e^{(\theta_s - \beta_i)} / 1 + e^{(\theta_s - \beta_i)}$$

Calculando todas estas probabilidades se obtuvo una tabla para todos los participantes. Con el valor de la habilidad media (θ) de cada grupo de alumnos y la dificultad (β) de cada ítem se realizó una simulación mediante el Método de Montecarlo, con Microsoft Excel, para obtener la distribución de puntuaciones obtenida por una muestra aleatoria de 1.000 alumnos de cada nivel en la prueba, similar a la muestra del pilotaje. Los valores se representan en la siguiente gráfica:

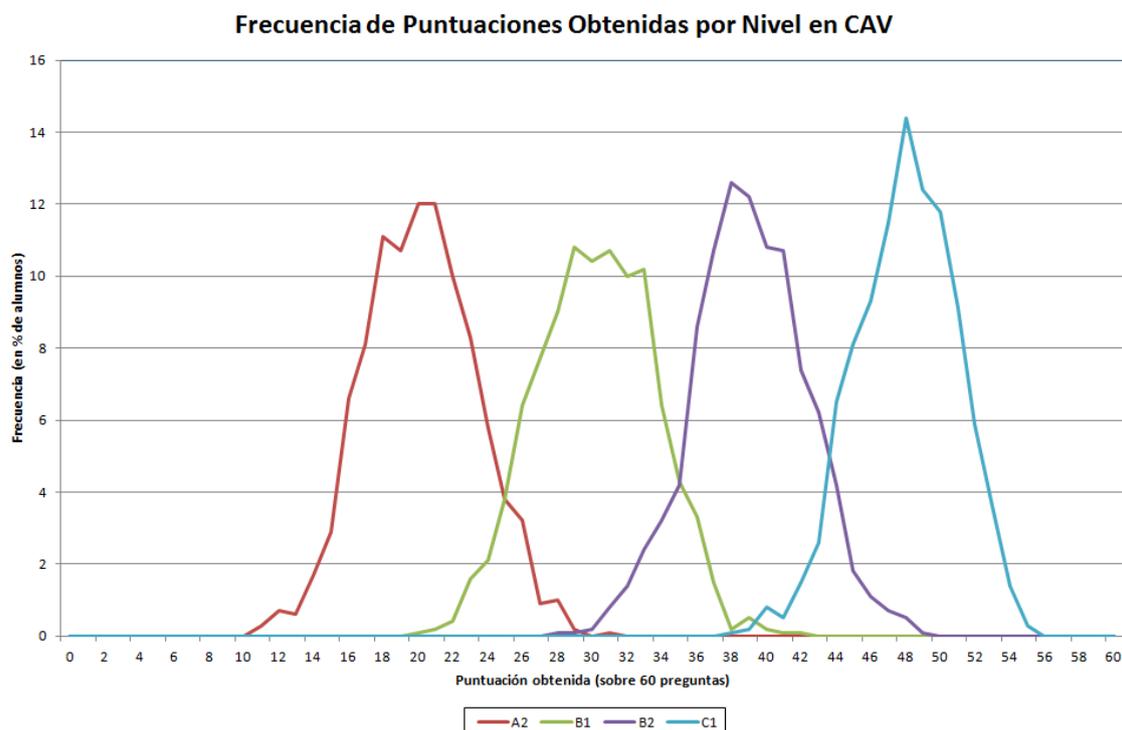


Gráfico 1. Frecuencia de puntuaciones obtenidas por cada grupo de alumnos según nivel

A partir de los resultados obtenidos para cada grupo se realizó un análisis mediante curvas ROC para determinar la sensibilidad y especificidad del test entre los distintos niveles de conocimiento que mide.

Tomamos de Hamel (2008) la explicación sobre el concepto y modo de interpretación de las curvas ROC, que son gráficos bidimensionales que representan visualmente el rendimiento y la compensación de rendimiento de un modelo de clasificación. Para construirlas se requieren dos métricas de rendimiento, la tasa de verdadero positivo (tpr) y la tasa de falso positivo (fpr). Los clasificadores mapeados en un gráfico ROC se pueden clasificar según su distancia al punto de "rendimiento perfecto". El algoritmo ROC utiliza la clasificación de instancia para barrer a través de diferentes umbrales de decisión desde el valor de clasificación máximo hasta el valor de clasificación mínimo en incrementos minados con anterioridad. Los valores de clasificación se normalizan típicamente a valores entre 0 y 1 (aparte, el umbral de decisión por defecto para la mayoría de los clasificadores se establece en .5 si el valor de clasificación expresa el valor real de probabilidad de que la instancia se clasifique como verdadera). En cada incremento de umbral, el desempeño del modelo se calcula en términos de las tasas positivas verdaderas y falsas y se grafica. Esto traza una curva de izquierda a derecha (rango máximo a mínimo) en el gráfico ROC. Esto significa que la parte izquierda de la curva representa el comportamiento del modelo bajo umbrales de decisión altos (conservador)

y la parte derecha de la curva representa el comportamiento del modelo bajo umbrales de decisión bajos (liberal).

Sobre las distribuciones de resultados de nuestra prueba se establecieron el espacio ROC y las curvas ROC.

En primer lugar, para la distribución de calificaciones obtenida se determinaron diferentes notas de corte y para cada nota de corte se establecieron el número de alumnos de ese nivel aprobarían o suspenderían la prueba. Así, por ejemplo, para una nota de corte muy baja tal como 20 puntos el 41.6% de los alumnos de A2 no pasaría la prueba (Verdaderos Negativos VN) y el 58.4% pasaría (Falsos Positivos FP). Por otra parte el 0.3% de los alumnos de nivel B2 no pasaría el examen (Falsos Negativos FN) y el 99.7% sí pasaría (Verdaderos Positivos VP).

A partir de los valores VN, FP, FN y VP se obtienen las ratios VPR (Razón de Verdaderos Positivos), FPR (Razón de Falsos Positivos) y ACC (Precisión o exactitud):

$$\mathbf{VPR} = \mathbf{VP} / (\mathbf{VP} + \mathbf{FN})$$

$$\mathbf{FPR} = \mathbf{FP} / (\mathbf{FP} + \mathbf{VN})$$

$$\mathbf{ACC} = (\mathbf{VP} + \mathbf{VN}) / (\mathbf{VP} + \mathbf{FP} + \mathbf{VN} + \mathbf{FN})$$

Con los valores de VPR y FPR obtenidos se puede representar en un gráfico una curva (curva ROC). La geometría de esta curva nos indicará la calidad de este test como calificador de alumnos entre niveles A2 y B1. Como parámetro de calidad se considera el área encerrada bajo la curva AUC.

A continuación se incluyen las curvas ROC para la separación de niveles A2/B1, B1/B2 y B2/C1 de la prueba de comprensión auditiva:

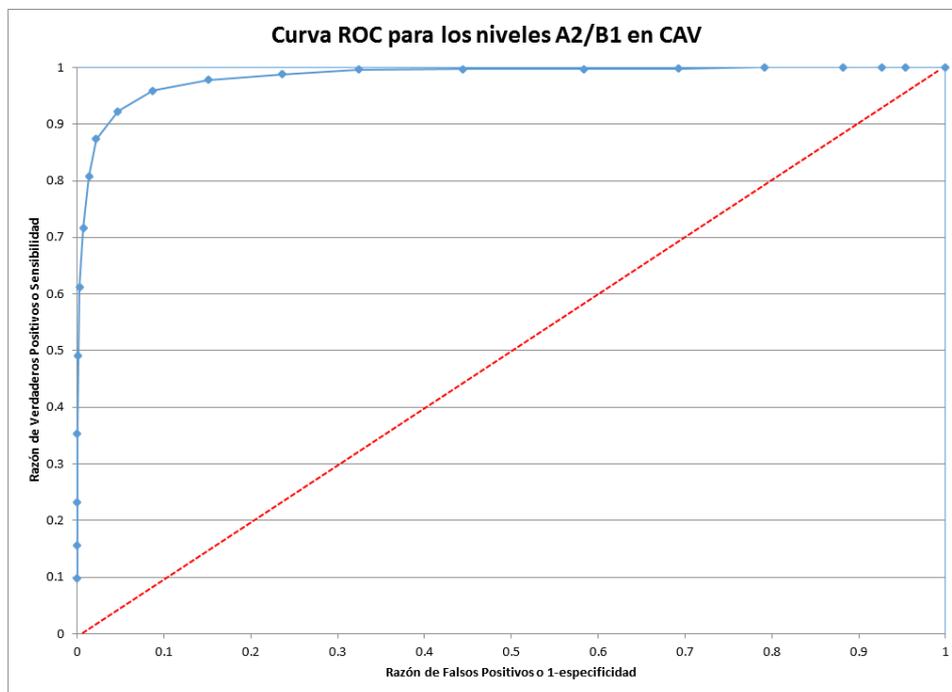


Gráfico 2. Curva ROC para los niveles A2/B1

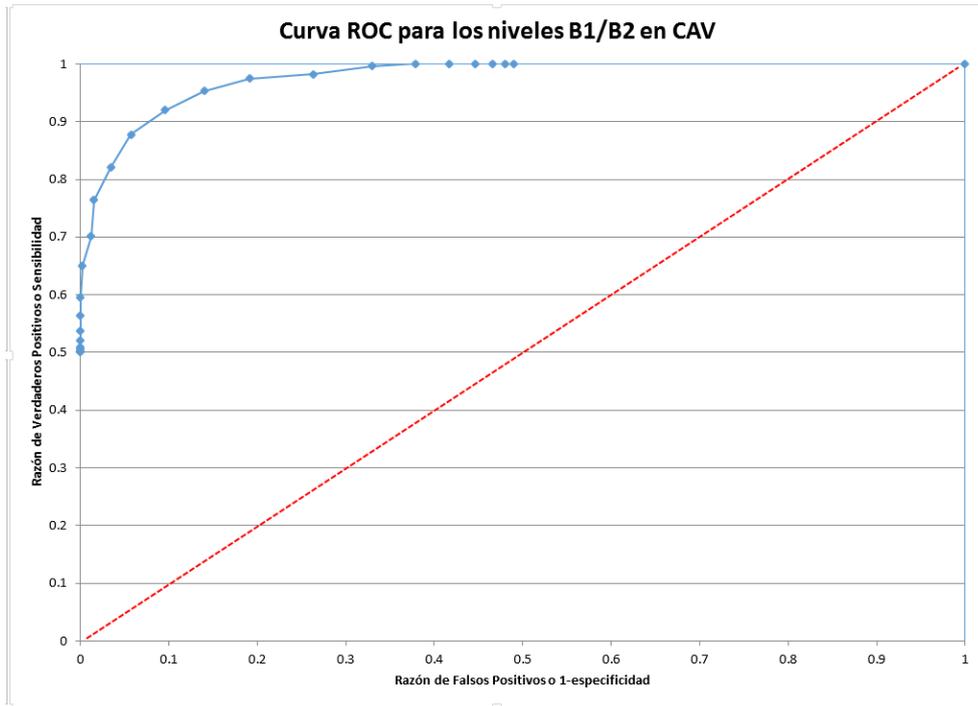


Gráfico 3. Curva ROC para los niveles B1/B2

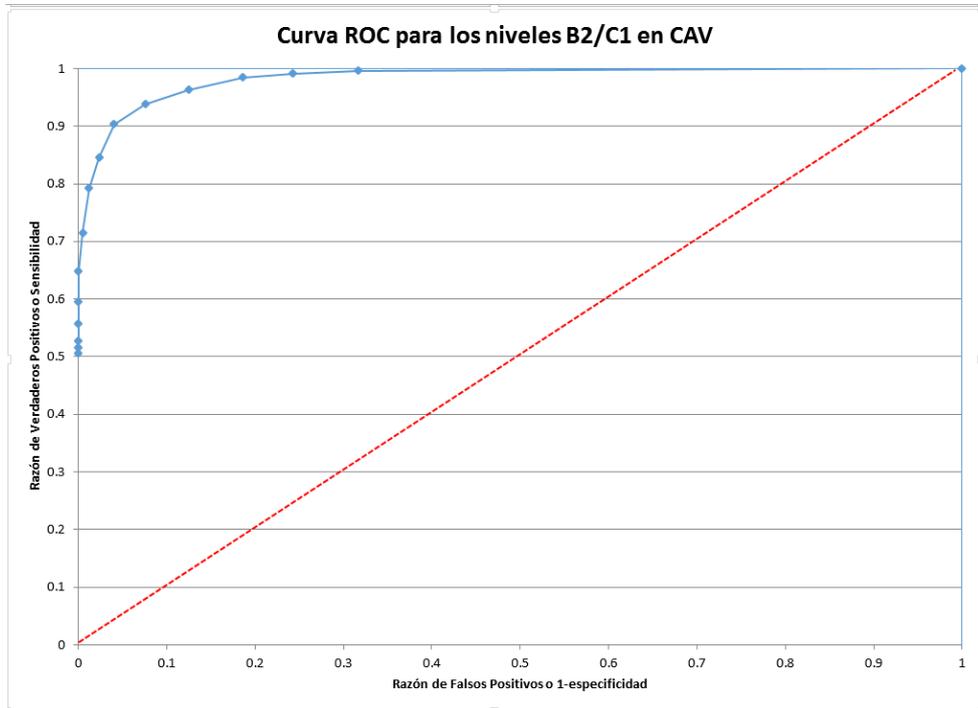


Gráfico 4. Curva ROC para los niveles B2/C1

Como conclusión general del análisis ROC realizado, se puede concluir que la prueba tiene una calidad de diagnóstico entre los distintos niveles buena o muy buena.

3.2.4. Rúbrica de evaluación para la tarea de respuesta abierta

Tal como se ha indicado anteriormente, la tercera tarea de la prueba consiste en la elaboración del resumen de un fragmento de conferencia o lección magistral que se ve dos veces. Se trata de una tarea de respuesta abierta y corrección por criterios. Los criterios de evaluación son (1) la comprensión de la idea principal y las ideas secundarias de un fragmento de conferencia o lección magistral; (2) la reelaboración del contenido a partir de la interrelación entre las ideas, siguiendo el hilo argumental del discurso y (3) la interpretación de la enseñanza que pretende transmitir el orador. Se trata de una tarea que mide la capacidad del candidato para aplicar dos destrezas comunicativas de manera integrada (comprensión oral y expresión escrita), propia del ámbito académico e imprescindible para el adecuado desenvolvimiento en el mismo. La tarea se califica utilizando la siguiente escala de descriptores de diez puntos:

- **Calificación 1:**
El candidato basa su resumen en una única palabra identificada en el input, que no es especialmente relevante interpretar su contenido, y elabora a partir de ella un nuevo mensaje que se aleja completamente de la intención del orador.
- **Calificación 2:**
El candidato identifica algunas palabras pertinentes para la elaboración del significado pero no es capaz de seguir el hilo del discurso, con lo que su reproducción es demasiado imprecisa para transmitir la idea principal del mensaje.
- **Calificación 3:**
El candidato comprende algunas ideas del mensaje, pero no es capaz de establecer adecuadamente la relación entre las principales y las secundarias, con lo que no logra interpretar adecuadamente la intención del orador.
- **Calificación 4:**
El candidato identifica la idea principal y algunas ideas secundarias relevantes pero no es capaz de deducir de ellas la enseñanza que el interlocutor pretende transmitir.
- **Calificación 5:**
El candidato identifica la idea principal y se aproxima a la enseñanza que el orador pretende transmitir, normalmente empleando las palabras

clave del mensaje, aunque presenta confusión en la interpretación de las aportaciones de las ideas secundarias a su tesis.

- **Calificación 6:**
El candidato muestra una comprensión adecuada de las ideas principales del mensaje en tanto que puede reelaborarlas con sus propias palabras para reproducir un mensaje propio, aunque la interpretación de la enseñanza del orador resulta imprecisa.
- **Calificación 7:**
El candidato identifica la idea principal del mensaje y gran cantidad de detalles secundarios (lo que a veces se evidencia en la toma de notas que precede al resumen), que utiliza para una reproducción bastante elaborada del mensaje, aunque la relación de estos detalles con la tesis del orador no es totalmente acertada para interpretar en su totalidad la enseñanza que este pretende transmitir.
- **Calificación 8:**
El candidato identifica la idea principal y las secundarias, y las relaciona adecuadamente siguiendo el hilo argumental del interlocutor, hasta llegar a una interpretación de la enseñanza que se pretende transmitir, que resulta correcta aunque poco detallada.
- **Calificación 9:**
El candidato integra correctamente las ideas principales y secundarias, respetando el hilo argumental del interlocutor y elaborando un mensaje propio en el que demuestra una comprensión en profundidad porque abstrae y generaliza a partir de los ejemplos y detalles del mensaje. La interpretación de la enseñanza que se pretende transmitir resulta adecuada aunque le falta algún detalle, como pueda ser la conclusión a la que llega el orador con su intervención.
- **Calificación 10:**
El candidato elabora un mensaje propio y bien estructurado con el que demuestra una comprensión en profundidad de las ideas, la jerarquía entre ellas y la intención del orador. La interpretación de la enseñanza que se pretende transmitir resulta completa y detallada.

3.2.5. Procedimiento de calificación de la prueba

Cada ítem de las tareas de corrección objetiva (tareas 1, 2 y 3.2.) se califica con 1 punto. La tarea 3.1., de toma de nota y resumen, se evalúa de 1 a 10 según los descriptores de calificación anteriormente recogidos.

La ponderación dentro de la prueba es la que sigue: las tareas de corrección objetiva tienen un peso del 80% mientras que la tarea de respuesta abierta y corrección por criterios equivale al 20% restante.

El candidato recibe finalmente la puntuación obtenida y la calificación correspondiente a la misma según una escala de 0-10. Al tratarse de un examen multinivel, los puntos de corte para la parte de corrección objetiva de la prueba de comprensión audiovisual se establecen en cada convocatoria.

4. Conclusiones

Las tareas y situaciones propias del ámbito académico requieren un uso de la lengua específico que no siempre se corresponde con el que recogen los exámenes de certificación lingüística empleados por las universidades para regular el acceso de los estudiantes extranjeros.

Desde algunos países hispanohablantes se están desarrollando propuestas de evaluación del lenguaje académico como la de Vine y Ferreira (2015). No obstante, por el momento, en el caso del sistema universitario español carecemos de un examen de certificación de español académico que facilite este trámite en la gestión de la movilidad internacional. Un examen de estas características redundaría en un impacto positivo sobre el sistema de enseñanza/aprendizaje y sobre la integración del propio usuario de la lengua en el país de acogida, puesto que los estudiantes se prepararían para alcanzar las competencias que les resultarán útiles en su estancia universitaria (seguimiento de clases magistrales, interacción con profesores y compañeros, discusión académica, etc.).

El proyecto aquí presentado pretende cubrir este vacío. Con esta finalidad, se ha diseñado y validado un examen de español académico multinivel. En lo que respecta a la prueba de comprensión audiovisual, en este trabajo se han presentado los procedimientos seguidos para la selección del input, la redacción de las tareas y la validación de las mismas por expertos, seguida de una validación con una muestra de informantes no nativos (N=68).

El diseño de la prueba ha pretendido asegurar la evaluación de la destreza en el desempeño de tareas de comprensión oral similares a las que los estudiantes desempeñan en su quehacer diario en la Universidad española. La selección de los inputs y del tipo de preguntas ha estado guiada por los procesos de comprensión auditiva que deben activarse para dar respuesta a estas tareas.

De cara a preparar a los estudiantes y futuros estudiantes *incoming* sería conveniente una didáctica que tenga en cuenta estas mismas pautas. De acuerdo con Newton et al. (2018: 1), la enseñanza basada en destrezas comunicativas es el núcleo de la instrucción de lengua con fines académicos,

lo que no restaría valor a la instrucción sistemática en el contenido lingüístico, ya que esta debe integrarse en la enseñanza basada en destrezas con el fin último de ayudar a los estudiantes a desarrollar las habilidades, estrategias y conciencia metacognitiva necesarias para tener éxito en sus estudios.

Se ha optado por un input audiovisual para una mayor autenticidad situacional. Consideramos que el apoyo visual y documental, así como la toma de notas, deben tenerse muy en cuenta a la hora de diseñar tanto actividades de enseñanza como de evaluación de la comprensión auditiva para el ámbito académico, pues como indica Carrell (2007):

Taking lecture notes is widely accepted as a useful strategy for augmenting student attention and retention of academic discourse. Notetaking is intuitively appealing to the lecture-listener and is generally viewed as a way to facilitate the process of learning and remembering lecture material.

Como suele suceder con los exámenes de alto impacto, es de prever que la implantación de un examen de estas características influya en la metodología y materiales de enseñanza para la preparación del mismo. Una prueba evaluadora de corte comunicativo-articulada en tareas de uso de la lengua que tiene en cuenta los procesos de escucha propios del contexto real al que se orienta, esperamos que redunde en una enseñanza del español para universitarios que facilite a los estudiantes internacionales su desempeño académico en español.

4. BIBLIOGRAFÍA

- AINCIBURU, MARÍA CECILIA (2010), «Las comunicaciones ASELE. Un género textual entre estilo profesional y académico», en *Actas XXI Congreso Internacional de la ASELE: Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español L2-LE*. Salamanca, 2010.
<http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/21/21_0069.pdf> (10/05/2019).
- ANGOFF, WILLIAM H. (1971), «Scales, norms and equivalent scores», en R. L. Thorndike (ed.) *Educational Measurements*. Washington, DC: American Council on Education.
- ASSOCIATION OF LANGUAGE TESTERS IN EUROPE (ALTE) Y CONSEJO DE EUROPA (2011). *Manual for Language Test Development and Examining for Use with CEFR*. Strasbourg: Council of Europe.
- ARYADOUST, VAHID; GOH, CHRISTINE; LEE, ONG KIM (2012), «Developing an academic listening self-assessment questionnaire: A study of modeling academic listening», en *Psychological Test and Assessment Modeling*, Vol. 54(3), pp. 227-256.

- BACHMAN, LYLE; PALMER, ADRIAN (1996), *Language testing in practice*. New York: Oxford University Press.
- BACHMAN, LYLE; PALMER, ADRIAN (2010), *Language Assessment in Practice. Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
- BUCK, GARY (2001), *Assessing Listening*. Cambridge: Cambridge University Press.
- CARRELL, PATRICIA L. (2007), *Notetaking Strategies and Their Relationship to Performance on Listening Comprehension and Communicative Assessment Tasks*. New Jersey: ETS.
- CONSEJO DE EUROPA (2001), *Marco común europeo de referencia para las lenguas: enseñanza, aprendizaje y evaluación*. Madrid: Ministerio de Educación y Anaya.
- CONSEJO DE EUROPA (2009), *Manual for Relating Language Examinations to the Common European Framework of Reference for Languages*. Strasbourg: Council of Europe.
- CRUZ MOYA, OLGA (2009), «El español con fines académicos en el marco del EEES: las competencias del estudiante», en *Actas XX Congreso Internacional de la ASELE: El español en contextos específicos: enseñanza e investigación*. Comillas, 2009. <http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/20/20_0376.pdf> (10/05/2019).
- CUMMINS, JIM (1979), «Cognitive/academic language proficiency, linguistic interdependence, the optimal age question and some other matters». *Working Papers on Bilingualism*, nº 19, pp. 121-129.
- DOUGLAS, DAN (2000), *Assessing English for Specific Purposes*. Cambridge: Cambridge University Press.
- FIGUERAS CASANOVAS, NEUS; PUIG SOLER, FUENSANTA (2013), *Pautas para la evaluación del español como lengua extranjera*. Madrid: Edinumen.
- GARCÍA DELGADO, JOSÉ LUIS; ALONSO, JOSÉ ANTONIO; JIMÉNEZ, JUAN CARLOS (2008), *Valor económico del español*. Barcelona: Ariel y Fundación Telefónica.
- GOLA, VERÓNIQUE Y MIRALLES ANDRÉS, NURIA (2009), «Los ambientes híbridos de aprendizaje en la educación superior», en *Actas XX Congreso Internacional de la ASELE: El español en contextos específicos: enseñanza e investigación*. Comillas, 2009. <http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/20/20_0470.pdf> (10/05/2019).
- GREEN, ANTHONY (2012), *Exploring Language Assessment and Testing. Language in Action*. London: Routledge.
- HAMBLETON, R. K. (1980), «Test score validity and standard setting methods». En R. A. Berk (Ed.), *Criterion-Referenced Measurement: The State of the Art*. Johns Hopkins University Press: Baltimore.
- HAMEL, LUTZ (2008), «Model Assessment with ROC Curves». En *Encyclopedia of Data Warehousing and Mining*, 2nd Edition. Hershey: Information Science Reference, pp. 1316-1323.

- INSTITUTO CERVANTES (2006), *Plan Curricular del Instituto Cervantes. Niveles de referencia para el español*. Madrid: Biblioteca Nueva.
- KRASHEN, STEPHEN; BROWN, CLARA LEE (2007), «What is academic language proficiency? Research Papers», en *STETS Language and Communication Review*. Singapore Tertiary English Teachers' Society. <http://www.sdkrashen.com/articles/Krashen_Brown_ALP.pdf> (10/05/2019).
- LEYVA BARAJAS, YOLANDA EDITH (2011), «Una reseña sobre la validez de constructo de pruebas referidas a criterio», en *Perfiles educativos*, Vol. XXXIII, nº 131, pp. 131-154.
- MARTÍN LERALTA, SUSANA (2011), «Certificación lingüística de nivel inicial para inmigrantes en contexto laboral: ejemplo de una prueba del examen DILE», en *Lengua y Migración*, nº 3(1), pp. 89-104.
- MARTÍNEZ BAZTÁN, ALFONSO (2011), *La evaluación de las lenguas. Garantías y limitaciones*. Granada: Editorial Octaedro Andalucía.
- MARX, ALEXANDRA; HEPPT, BIRGIT; HENSCHER, SOFIE (2017), «Listening comprehension of academic and everyday language in first language and second language students», en *Applied Psycholinguistics*, nº 38(3), pp. 571-600.
- MINISTERIO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL (2018), *Panorama de la Educación. Indicadores de la OCDE. Informe español*. Madrid: Secretaría General Técnica.
- NEWTON, JOHNATAN M.; FERRIS, DANA R.; GOH, CHRISTINE C.M.; GRABE, WILLIAM; STOLLER, FREDRICKA L.; VANDERGRIFT, LARRY (2018), *Teaching English to Second Language Learners in Academic Contexts*. New York: Routledge.
- PARODI, GIOVANNI (2009), «Corpus, discurso y géneros: español en contextos académicos y profesionales», en *Actas XX Congreso Internacional de la ASELE: El español en contextos específicos: enseñanza e investigación*. Comillas, 2009 <http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/20/20_0065.pdf> (10/05/2019).
- PASTOR CESTEROS, SUSANA; FERREIRA CABRERA, ANITA (2018), «El discurso académico en español como LE/L2: nuevos contextos, nuevas metodologías», en *Journal of Spanish Language Teaching*, nº 5:2, pp. 91-101.
- REGUEIRO, M. LUISA; SÁEZ RIVERA, DANIEL (2013), *El español académico. Guía práctica para la elaboración de textos académicos*. Madrid: Arco Libros.
- SMITH, SM; SCHUMACKER, RE; BUSH, MJ. (1998), «Using item mean squares to evaluate fit to the Rasch model», en *Journal of Outcome Measurement*, nº 2(1), pp. 66-78.
- VÁZQUEZ, GRACIELA (2004), «La enseñanza del español con fines académicos», en J. Sánchez Lobato e I. Santos Gargallo (eds.), *Vademécum para la formación de profesores. Enseñar español como segunda lengua y lengua extranjera*. Madrid: SGEL, pp. 1129-1147.

- VÁZQUEZ, GRACIELA (ed.) (2005), *Español con fines académicos: de la comprensión a la producción de textos*. Madrid: Edinumen.
- VÁZQUEZ, GRACIELA (2009), El discurso académico español: entre la identidad académica y la globalización del conocimiento, en *Actas XX Congreso Internacional de la ASELE: El español en contextos específicos: enseñanza e investigación*. Comillas, 2009 <http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/20/20_0117.pdf> (10/05/2019).
- VILLAR, CLAUDIA M. (2013), *Las presentaciones académicas orales en ELE de estudiantes alemanes: un análisis macrotextual, discursivo y contextual del género en nativos y no nativos*. Tesis doctoral. Madrid: Universidad Antonio de Nebrija.
- VINE JARA, ANA; FERREIRA CABRERA, ANITA (2015), «Propuesta de un modelo para una prueba con fines específicos académicos en ELE», en *Literatura y Lingüística*, nº 33, pp. 369-390.