

Descripción y análisis del examen online de acreditación de dominio de español de los niveles B1 y B2: el eLADE

Description and analysis of online exam on accreditation of Spanish language in levels B1 and B2: the ELADE

Alfonso Martínez¹, Aurora Biedma¹, Adolfo Sánchez¹, María D. Chamorro¹, Sonia Sánchez¹

¹ Centro de Lenguas Modernas. Universidad de Granada, España

alfonsom@ugr.es , abiedma@ugr.es , adolfos@ugr.es , lola.chamorroguerrero@yahoo.es , soniasm@ugr.es

RESUMEN. La evaluación de lenguas por ordenador es una posibilidad cada día más real consecuencia tanto del desarrollo de las nuevas tecnologías como del uso creciente de estas en ámbitos de aprendizaje. En este artículo se presenta el examen eLADE del Centro de Lenguas Modernas de la Universidad de Granada, examen innovador por tratarse de un examen de acreditación de dominio de español en línea. En él se exponen las circunstancias de su puesta en marcha en el año 2012, el proceso de diseño y producción en línea, los protocolos de validación y estandarización que se siguen, así como el desarrollo de la cuarta y más reciente convocatoria. El objetivo es tanto contribuir a la difusión del primer examen en línea de dominio de español como promover la máxima transparencia respecto a todos los procesos relacionados con los exámenes, en consonancia con lo establecido en los protocolos internacionales de diseño y validación de exámenes de lenguas. Su administración y los resultados obtenidos en las cuatro convocatorias realizadas hasta el momento han sido excelentes, mostrando los usuarios sociales y académicos gran interés y manifestando su agrado con este nuevo formato de administración y asesoramiento que les facilita la labor de evaluar, administrar, gestionar y acceder a un certificado de nivel de español válido, fiable, operativo y acreditado.

ABSTRACT. The increasing role of technology in today's society and its use in learning environments enable more than ever the use of computers in language testing. This article outlines the eLADE test by the Centre of Modern Languages of the University of Granada. This is an innovative test since it is a fully online accreditation exam of Spanish. We describe the circumstances that led to its kickoff in 2012, its academic and technical development, and the validation and standardization protocols followed, together with information on the latest fourth exam session. This article has a two-fold aim. It aims both at disseminating the information about the first validated online test of Spanish proficiency and at promoting test-development transparency in accordance with the international good-practice testing protocols. As it is here described, the results of the exam sessions have been so far excellent. Moreover, test-takers and institutions have shown great interest in the exam, claiming that both its online administration and facilitation protocols make it easier for them to assess, manage and obtain a Spanish proficiency certificate which is valid, reliable, functional, and fully accredited.

PALABRAS CLAVE: Examen, Dominio, Acreditación, Español, En línea, B1/B2, eLADE.

KEYWORDS: Test, Proficiency, Accreditation, Spanish, On line, B1/B2, eLADE.

1. Introducción: la evaluación de lenguas por ordenador

La evaluación de lenguas por ordenador y, como consecuencia de los desarrollos tecnológicos que se suceden día tras día, la evaluación en línea a través de la web, es una posibilidad cada día más real, consecuencia tanto del desarrollo de las nuevas tecnologías como del uso creciente de estas en ámbitos de aprendizaje. Como tal, debe ser tenida en cuenta como una herramienta que “no puede ser razonablemente rechazada -la tecnología se está usando y va a continuar usándose- (...) (sino) como un ineludible aspecto de la evaluación de lenguas” (Douglas, 2010, p. 139) o “como una evolución natural en la práctica de la evaluación” (Dunkel, 1999, p. 77). Incluso, y debido tanto a las posibilidades que se abren en cuanto a formatos, mejoras en los aspectos de autenticidad y motivación y a una suma de características provenientes tanto del propio campo de la evaluación de lenguas como de las tecnologías, está llamada a provocar bien una radical transformación en las prácticas de evaluación (Chalhoub-Deville, 2010), o cuando menos notables mejoras en los sistemas de evaluación, acompañadas de otras tecnologías y software que ya vienen siendo usados en la evaluación de lenguas, como SPSS (Statistical Package on the Social Sciences) e IRT (por sus siglas en inglés, Teoría de Respuesta al Ítem). Estas herramientas sirven para hacer cálculos y estadísticas sobre ítems así como para diseñar exámenes en formatos nuevos que permiten el control adaptativo de las preguntas, la autocorrección o el envío inmediato de resultados; permiten incluso corregir y calificar pruebas de respuesta abierta como la expresión escrita y la expresión oral, mediante muestras comparativas, campos léxicos o algoritmos complicados, imposibles de hacer manualmente.

Todo ello conlleva ventajas y desventajas de dos tipos, de índole humana y tecnológica (Brown, 1997), que deberán tenerse en cuenta y analizarse en cada caso. No obstante, debido a la facilidad de uso de las tecnologías y al creciente manejo de estas entre profesores y estudiantes, hay que seguir recordando la necesidad de usarlas con responsabilidad y profesionalidad, dentro de lo que exige una evaluación con garantías, válida, fiable y justa (Martínez Baztán, 2011) pues hace falta saber que “no todos los profesores que crean y suben exámenes para uso público son conscientes de muchas de las cuestiones de la evaluación de lenguas. Adoptar, de entre todos los modelos existentes, un examen de lenguas es una tarea que requiere cuidado, incluso si este solo tiene el propósito de usarse para funciones de bajo impacto” (Winke, 2008, p. 358).

Este artículo repasa someramente algunos de los aspectos innovadores de esta práctica, con sus ventajas y algunos problemas todavía no resueltos, y se centra en la explicación del diseño y uso de un nuevo examen de certificación de español administrado enteramente por ordenador a través de la web 2.0, discutiendo sus resultados y otros aspectos que deben tenerse en cuenta en el diseño de este tipo de exámenes.

Desde que en 1935 IBM utilizó el modelo 805 de computador (Mubarak Pathan, 2012, p. 31) para calcular notas objetivas en exámenes en Estados Unidos con el fin de facilitar la intensa labor y reducir el costoso trabajo de calificar millones de exámenes que se hacían cada año, la necesidad de evaluar a gran cantidad de personas de forma económica y eficaz no ha dejado de crecer. Posteriormente, con el desarrollo de la tecnología acaecido a partir de los años 80 esta labor se ha hecho mucho más sencilla. Hoy en día podemos citar numerosos exámenes para evaluar lenguas, en diferentes formatos, progresivos (que presentan las tareas y preguntas moduladas en virtud del nivel de cada respuesta dada por el candidato) o lineales, y con diferentes fines. Entre los que destacan por su popularidad y prestigio figuran el TOEFEL iBT (que mide la habilidad de hablantes no nativos de inglés para desenvolverse en tareas académicas de nivel universitario utilizando el inglés), BULATS (diseñado para evaluar el dominio de inglés para empleadores de negocios, solicitantes de empleo y candidatos para cursos de lengua de inglés de negocios), Basic English Skills Test (BEST) Plus (para medir las destrezas de expresión oral y de comprensión auditiva de estudiantes de inglés en los Estados Unidos), el COMPASS ESL Placement Test (para evaluar el conocimiento del inglés americano en estudiantes de inglés en escuelas e instituciones americanas de bachillerato), el Versant English Test (para medir habilidades de expresión oral en inglés de hablantes no nativos de inglés), o el Pearson Test of English (PTE) Academic (para evaluar el dominio de inglés de estudiantes internacionales en contextos académicos).

Y entre los existentes en español, además del eLADE que nos ocupa en este artículo, podemos señalar (ver

Winkel, 2008, p. 359) el STAMP (Standards-based Measurement of Proficiency), diseñado por el Center for Applied Second Language Studies (CASLS) de la Universidad de Oregón, que mide las destrezas de lectura, escritura y expresión oral según la escala de niveles de ACTFL inicial bajo, intermedio bajo e intermedio medio (lo que viene a ser A1, A2 y B1, Martínez Baztán, 2008) o el desarrollado entre cinco instituciones americanas (ACTFL, Brigham Young University, CAL, Defense Language Institute, y la Universidad de San Diego) de un sistema llamado Computer-Assisted Screening Tool (CAST).

No por la inevitabilidad del desarrollo tecnológico debemos olvidar que es necesario continuar investigando sus usos y resultados (Brown, 1997), sus ventajas y posibles problemas. Entre las ventajas (según Mubarak Pathan, 2012), destacan que los exámenes administrados por ordenador son más consistentes y fiables porque todos los candidatos reciben exactamente la misma información y debe realizar las mismas acciones sin influencias externas del medio en que las realizan, es más motivador y facilita una actitud más positiva, y simplifica la burocracia de envíos de materiales, reenvíos de exámenes una vez realizados, acceso a la información y a ejemplos o muestras de exámenes. Otros autores señalan que el uso adecuado de herramientas multimedia puede mejorar la autenticidad de los materiales y la interacción real con ellos, ya que puede ayudar al uso de situaciones comunicativas de la lengua en muchos aspectos (Douglas, 2010, p. 118) o que las habilidades de audio y vídeo son más válidas y auténticas (Winke, 2008).

Y entre los problemas que señala la literatura al uso, destacan principalmente los relacionados con aspectos de la validación y fiabilidad, especialmente en exámenes progresivos (Brown, 1997): la adecuada selección de ítems (problema que no es ajeno a los exámenes tradicionales hechos en papel), los sistemas para establecer los puntos de corte y de hacer llegar estos de forma inteligible a los usuarios, la comparación entre los resultados de un examen en papel y otro realizado por ordenador y el grado en que el formato provoca la medición de diferentes constructos (Winke, 2008), la posibilidad de que el conocimiento del uso de la tecnología influya en la obtención de la calificación (Madsen, 1991) y por tanto nuevamente un problema de validez de constructo, o la duda de si estos exámenes miden la destreza en el uso de ordenadores o la habilidad lingüística.

Para terminar esta introducción sobre los exámenes por ordenador y sus aspectos más significativos, vamos a señalar algunas de sus principales características y situar el eLADE dentro de ellas: a. la direccionalidad permite diseñar exámenes que ofrezcan los ítems de forma lineal, o progresiva, esto es, dependiendo de la respuesta dada a un ítem previo (el eLADE es un examen lineal); b. el formato de envío puede ser simplemente por ordenador o a través de la web (eLADE se realiza en línea a través de la web); c. la densidad de empleo de las herramientas multimedia, que hacen que el examen utilice un solo medio o los integre en formato multimedia (eLADE es multimedia); d. los objetivos de las tareas, bien sean mediante tareas independientes o que integren habilidades (eLADE utiliza tareas independientes); e. el sistema de calificación, humano o automático (eLADE utiliza ambos según las destrezas); f. el riesgo del examen, de alto, medio o bajo impacto (eLADE es de riesgo medio pues sus usos no conllevan consecuencias de alta responsabilidad); g. el propósito, diseñando exámenes relacionados con un currículo o no (eLADE es un examen de dominio no relacionado con un currículo sino con niveles de dominio de la lengua); h. los tipos de respuesta, seleccionada o abierta (eLADE tiene ambos tipos de respuesta en sus tareas); e i. los tipos de tareas utilizadas (eLADE utiliza varios tipos de tareas que se describen más adelante en este artículo).

2. Justificación del examen eLADE, examen en línea de acreditación de dominio de español

La creciente demanda para acreditar el conocimiento de las lenguas más habladas en el mundo hace necesaria la ampliación y diversificación de las ofertas de exámenes de dominio reconocidos por las instituciones académicas y laborales. Estos exámenes deben adaptarse a diferentes contextos y a las necesidades de los distintos usuarios potenciales, atendiendo a los nuevos intereses tanto de los usuarios como de las instituciones, que además deben velar por su buen uso, validez, operatividad y fiabilidad.

Con el Plan Bolonia, que permite la movilidad a los estudiantes europeos, y con la reciente implantación

del Programa Erasmus Plus, dirigido no solo a Europa sino a otros países, la demanda de exámenes de acreditación de dominio de español ha aumentado considerablemente, extendiéndose por todo el mundo y haciendo necesarias nuevas herramientas que posibiliten un buen acceso a tests fiables, validados, de sencilla administración y operativos.

La Universidad de Granada se cuestionó estos aspectos ante la importante demanda de los estudiantes de los diversos programas que deseaban seguir estudios de Grado o Posgrado y que procedían de muy diversas partes del mundo (Biedma et alii, en prensa). La universidad se enfrentaba al problema de la certificación adecuada de dominio suficiente de lengua española para seguir dichos estudios o para conceder becas a los estudiantes de otros países, para lo cual estableció, al igual que el resto de las universidades europeas, unos criterios de nivel que debían acreditarse debidamente. Algunos estudiantes venían avalados por el examen DELE del Instituto Cervantes pero la mayoría no tiene acceso a sedes de examen y otros presentan certificados de instituciones no oficiales sin garantía de acreditación, por lo que la Universidad se encontraba con el problema de conceder las becas de movilidad sin saber si el alumno cumplía los requisitos lingüísticos. De ahí que el Vicerrectorado de Relaciones Internacionales de la Universidad de Granada encargara en el año 2010 a su Centro de Lenguas Modernas (CLM) la elaboración de un examen binivel que acreditara B1 o B2 de español, tarea que desarrolló el Equipo de evaluación y acreditación del CLM, expertos en evaluación de lenguas, siguiendo los estándares y protocolos más exigentes de diseño y validación de exámenes internacionales y según las directrices del MCER, EALTA, ALTE y los manuales del Consejo de Europa (CoE, 2002; CoE, 2009; CoE, 2011; ILTA, 2000; ALTE, 2001; EALTA, 2004). Las primeras convocatorias de este examen se realizaron de manera presencial en papel en el Centro de Lenguas Modernas de la Universidad de Granada y desde 2013 se procedió a realizarlo en línea con el fin tanto de que se pudieran presentar personas desde cualquier parte del mundo sin necesidad de desplazarse, como de aprovechar las ventajas de este tipo de examinación, facilidad de administración, buena experiencia de uso, motivación, validez de apariencia y fiabilidad. Se puede acceder a toda la información en la web del CLM: http://www.clm-granada.com/html/c_espanol/esp/examenes/examen_acreditacion/examen_acreditacion.htm, donde hay también muestras del examen.

La dificultad de hacerlo presencial, dadas las características de procedencia de los alumnos implicados, nos llevó a realizar una prueba de dominio en línea que cumpliera con los requisitos y garantías necesarias para una buena acreditación. De esta manera los estudiantes podían certificar su nivel de idioma para poder asegurar la beca antes de venir a España desde su propia universidad o desde la sede oficial que los propios estudiantes propusieran según su ubicación. Este mismo problema, junto a la organización de sus convocatorias, envíos de exámenes, reenvío de resultados y notas, protocolos, etc., se plantea en el resto de las universidades españolas por lo que se decidió solicitar la certificación de garantía de la Asociación de Centros de Lenguas de Enseñanza Superior (ACLES) y de esta forma facilitar a las mismas el uso del eLADE como vía de acreditación para sus estudiantes de movilidad. El examen, diseñado y validado de acuerdo con normas internacionales y anclado al MCER, como se explicará a lo largo de este artículo, está acreditado y aceptado por esta asociación, así como por CERCLES (2012) y por la CRUE (2014) y a él se pueden sumar cuantas organizaciones e instituciones lo deseen, haciéndose sede examinadora según los requisitos establecidos en la página web (www.clm-granada.com/html/c_espanol/esp/examenes/examen_acreditacion/examen_acreditacion_solicitud_sede.htm).

Al conocer la existencia de un examen en línea con estas garantías de fiabilidad y validez, aumentó la demanda de otras instituciones no universitarias, tales como centros de enseñanza secundaria, otros centros de idiomas, así como personas en general interesadas en acreditar su nivel de lengua por la facilidad que ofrece el formato en línea del examen. Debido a esto, se amplió el uso y las funciones del examen -siempre siguiendo estrictos protocolos de diseño y validación internacional- haciendo de él un examen de acreditación del dominio de español en general, con fines más allá de la admisión a cursos universitarios, ampliando los contextos de uso y los contenidos tanto para ámbitos académicos como sociales y convirtiendo el test en uno de uso abierto a todo tipo de usuarios.

Este artículo presenta la cuarta edición del eLADE y ofrece cuantos datos son necesarios para mostrar sus protocolos de diseño y validación, además de datos estadísticos de validez y fiabilidad (obtenidos en pilotajes y mediante análisis de SPSS e IRT), para que la comunidad científica y educativa en general los valore, puedan ser tenidos en consideración por los usuarios y agentes sociales y para dotar al examen de la credibilidad y aceptación social y académica que consideramos merece, ganando la confianza de los usuarios e instituciones.

3. Descripción general del examen

3.1. Pruebas, tareas, ítems y criterios de evaluación

El eLADE es un examen binivel en línea que acredita el dominio del candidato en los niveles B1 y B2, por el momento. No obstante, nuestros planes se encaminan a implementarlo para niveles A2/B1 y B2/C1, manteniendo siempre su diseño de examen binivel y sus estándares de diseño y validez. Las habilidades que abarca y mide son la comprensión auditiva/audiovisual, la comprensión lectora, la expresión e interacción escritas y la expresión e interacción orales, según quedan definidas en el MCER.

El examen consta de cuatro pruebas diseñadas para evaluar seis actividades comunicativas de la lengua: la comprensión lectora, la comprensión auditiva, la expresión e interacción escritas, y la expresión e interacción orales. Tiene una duración de 3 horas y 15 minutos. Todas sus pruebas contribuyen a alcanzar el nivel, B1 o B2.

En la siguiente tabla se especifican el número de tareas que componen cada prueba, así como su duración.

| Comprensión auditiva | Comprensión lectora | Expresión e Interacción escrita | Expresión e Interacción orales |
|--------------------------------------|---------------------|---------------------------------|--|
| 5 tareas auditivas y/o audiovisuales | 5 tareas de lectura | 2 tareas de escritura | 3 tareas de expresión e interacción orales |
| 45 minutos | 75 minutos | 60 minutos | 10-15 minutos |
| En línea | En línea | En línea | Por Skype o presencial en el CLM |

COMPRESIÓN AUDITIVA (CA): 45 minutos

La Comprensión auditiva consta de 5 tareas auditivas y/o audiovisuales. Según su contenido, al menos dos son de B1 y dos de B2. La duración de los fragmentos es de entre 2 y 5 minutos.

Los fragmentos están extraídos de noticias, reportajes, conferencias, cuñas publicitarias, en los que se describe, argumenta, expone o se dan instrucciones. El registro de lengua es estándar (informal, formal y/o académico).

En total las tareas -siguiendo las normas de ACLES- constan de 25-40 ítems (de 5 a 8 por tarea). Cada grabación se escucha dos veces.

Los tipos de ítems de las tareas de comprensión son: a) respuesta alternativa o múltiple, de tres o cuatro opciones; b) relacionar enunciados con partes de la audición; c) respuesta corta, con un máximo de cinco palabras.

COMPRESIÓN LECTORA (CL) : 75 minutos

La Comprensión lectora incluye 5 tareas. Según su contenido, al menos dos son de B1 y dos de B2. En total consta -siguiendo las normas de ACLES- de entre 25-40 ítems (de 5 a 8 por tarea) y con textos que (siguiendo a ACLES) pueden variar de extensión entre 250 y 900 palabras según sean de B1 o B2.

Los textos elegidos son tanto del ámbito privado como público, escritos en un registro estándar (informal, formal y/o académico). En la medida de lo posible se intenta que respondan a géneros textuales diferentes - correos, artículos, editoriales, reseñas, anuncios, guías, instrucciones, entre otros- y a diversas funciones de la lengua y que presenten variedad temática.

Los tipos de ítems para las tareas de comprensión lectora son: a) respuesta alternativa o múltiple, de tres o cuatro opciones; b) respuesta corta con un máximo de cinco palabras; c) respuesta verdadera (V), falsa (F) con justificación de la respuesta; d) reintegración de un fragmento a un texto (insertar en su ubicación original una palabra o un fragmento de texto previamente extraído); d) relacionar o emparejar (relacionar elementos de dos listas formadas por textos cortos y por un enunciado).

EXPRESIÓN E INTERACCIÓN ESCRITAS (EIE): 60 minutos

La prueba de Expresión e Interacción escritas consiste en la redacción de 2 textos, el primero de ellos puede realizarse con un nivel B1, el segundo requiere nivel B2. La extensión de las tareas (siguiendo a ACLES) es de entre 200 y 350 palabras para B1 y entre 300 y 450 palabras para B2.

El candidato debe producir dos tipos de texto: el primero, una carta o e-mail (informal o neutro), o una narración; y el segundo, una carta o e-mail formal, un artículo, un informe o un ensayo. Cada género textual exige un específico tipo de discurso -descriptivo, narrativo, expositivo y argumentativo (este último para B2).

EXPRESIÓN E INTERACCIÓN ORALES (EIO): 10-15 minutos

La prueba de Expresión e Interacción orales se compone de tres tareas: las dos primeras diseñadas para evaluar el nivel B1 y la tercera el nivel B2.

La tarea 1 consiste en una entrevista, de entre 3-5 minutos, sobre cuestiones personales relacionadas con el candidato (país de origen, trabajo, estudios, gustos, etc.). La tarea 2, de entre 4-6 minutos, se plantea como una interacción en la que el candidato debe ser capaz de desenvolverse para resolver un conflicto o solicitar ayuda para solucionar un problema. La tarea 3 es una presentación o monólogo, de entre 4-6 minutos, sobre un tema controvertido, en el que el candidato debe exponer y argumentar tanto su punto de vista como analizar someramente el asunto y evaluar los pros y contras al respecto, y que da pie a una posterior discusión con el entrevistador de unos 3-5 minutos de duración. Esta tarea no es improvisada, sino que el candidato dispone de 10 minutos para prepararla antes de iniciar la sesión de evaluación oral. Para realizar las tareas 2 y 3, el candidato puede elegir entre tres opciones: en la tarea 2, al azar, y en la 3, para facilitar que pueda acercarse a un tema de su interés o especialidad, se elige un tema entre tres alternativas.

CRITERIOS DE EVALUACIÓN

Para conseguir la acreditación en un nivel, hay que superar el 50% de los ítems establecidos para cada nivel (B1 o B2), así como un total de un 75% en su conjunto, compensando un mayor dominio en unas pruebas frente a las otras. El nivel de dificultad de cada ítem se establece mediante un análisis de IRT que da el valor de dificultad de cada uno así como su error de medida, lo que permite decidir qué ítems y con cuáles de ellos se establece cada nivel de dominio de la lengua. Una vez establecidos estos ítems se procede a una sesión de fijación de puntos de corte para decidir el número mínimo exacto de ítems (o la nota mínima) con que superar cada prueba y el examen en su conjunto.

Para evaluar las pruebas de EIE y de EIO, siguiendo las recomendaciones del MCER se elabora una escala de calificación en la que consideran los siguientes criterios: para medir el grado de dominio en la destreza de EIE: 1. cumplimiento de la tarea, 2. coherencia, 3 corrección y 4. alcance y dominio de vocabulario; y para la EIO los mismos 4 criterios además de 5. fluidez y 6. capacidad de interacción. La nota final se elabora mediante la opinión de dos evaluadores.

3.2. Descripción de los niveles que evalúa

A continuación se describen brevemente las escalas y descriptores utilizados para la confección del examen, en cada una de las actividades de la lengua que se miden, para el nivel B1 y para el B2, siempre de acuerdo con el MCER.

COMPRESIÓN AUDITIVA: para el nivel B1 se espera que el candidato comprenda las ideas principales de un discurso en lengua estándar, pronunciado con claridad, que trate sobre temas de la vida cotidiana; para el nivel B2, el candidato debe comprender temas tanto concretos como abstractos, así como temas profesionales o académicos. En ambos casos, los candidatos deben ser capaces de comprender conversaciones entre hablantes nativos, así como escuchar conferencias o presentaciones, avisos o instrucciones, y retransmisiones y material grabado. La diferencia entre B1 y B2 a este respecto es la extensión del discurso y el grado de complejidad de las ideas expresadas, si bien en todo caso este debe presentar una estructura explícita (para marcar el límite en el que empieza la competencia de C1, que no la requiere). En cuanto a la comprensión de material audiovisual, ambos candidatos han de entender la mayoría de programas de televisión y películas, aunque el B1 requiere una articulación lenta y clara y un mayor apoyo en la imagen.

COMPRESIÓN LECTORA: para el nivel B1 se espera que el candidato pueda entender textos sencillos, de carácter general o de su especialidad, con un grado de comprensión satisfactorio; en el nivel B2 debe manejarse con soltura en una gran variedad de textos, al disponer de un amplio vocabulario de lectura, si bien su comprensión puede verse limitada ante una lengua no estándar, con modismos no habituales. El examen intenta medir que el candidato sea capaz de adaptarse a la situación, a diferentes tipos de texto y a diferentes objetivos de lectura: orientarse, buscar información, entender instrucciones.

EXPRESIÓN E INTERACCIÓN ESCRITAS: para B1 se espera que el candidato escriba textos sencillos sobre temas cotidianos o de su interés, con un grado adecuado de cohesión, así como distintos tipos de textos (cartas, informes, notas breves, mensajes, formularios, etc.); para el B2 se exige que escriba textos de mayor complejidad, no solo de carácter general sino sobre más específicos o de su especialidad en los que argumente su punto de vista, sintetice o reelabore información a partir de elementos procedentes de varias fuentes.

EXPRESIÓN E INTERACCIÓN ORALES: para el nivel B1 se espera que el candidato pueda sostener con una cierta fluidez un pequeño monólogo en el que describa temas de su interés, relate experiencias o argumente de forma sencilla. En la interacción oral, debe poder intercambiar información sobre temas habituales y disponer de un repertorio lingüístico sencillo para enfrentarse a situaciones cotidianas. Se espera asimismo que comprenda a un interlocutor nativo, tanto en conversaciones formales como informales. En el nivel B2, la diferencia atañe al tema, que se extiende a una amplia variedad de temas acordes a su interés o de su especialidad, y al grado de detalle y complejidad que se pide a su exposición, con argumentos y ejemplos relevantes sobre las ideas expuestas. En la interacción con hablantes nativos el candidato de B2 debe poder mantener una conversación con bastante fluidez y naturalidad, sin que se perciba tensión entre los interlocutores.

Los detalles de todos estos aspectos pueden verse publicados en la página web ya citada.

4. Desarrollo del examen y pilotaje

Todo examen debe someterse a criterios y protocolos estrictos de validación que demuestren que evalúa lo que pretende. Igualmente, un examen administrado por ordenador tiene la responsabilidad de demostrar su fiabilidad y validez, con independencia de la credibilidad y validez aparente que los usuarios le otorguen. El examen eLADE se somete durante todo el proceso de planificación y diseño a protocolos y directrices de buenas prácticas para asegurar: a) su adecuación y anclaje al MCER, b) que es justo y fiable, c) que posee el nivel apropiado y d) que evalúa lo que se pretende, la acreditación del dominio del español en los niveles B1 o B2 (Alderson, 1998; Bachman, 1990; Bachman, 1996; Bachman, 2004).

Para las pruebas de CA y CL, el diseño y planificación de tareas e ítems es sometido a análisis críticos con expertos, nativos y candidatos similares al candidato meta y se siguen los procesos establecidos para asegurar la validez de contenido y apariencia de las tareas: elección de los textos de partida, audiciones y textos escritos, mapeo en grupo de expertos para identificar la información relevante sobre la que diseñar los ítems, diseño de las tareas y revisión de los ítems. Posteriormente se someten a un pilotaje formal estadístico de análisis de dificultad, error de medida, discriminación y fiabilidad, entre otros factores, según SPSS e IRT.

Las pruebas de EIE y EIO se elaboran partiendo de un análisis de contenido y funciones según el contexto de uso de la lengua y las habilidades necesarias en los niveles B1 y B2. Las tareas se confeccionan considerando que su realización sea operativa para los candidatos que las tengan que realizar, poniendo especial cuidado y atención para que no se deslicen elementos que favorezcan o penalicen a ningún tipo de candidato. Todas ellas se pilotan con grupos de candidatos similares al candidato meta, expertos y nativos para comprobar tanto la adecuación de los textos y su nivel, como la claridad y capacidad de motivación de las instrucciones, así como para ver si provocan el lenguaje esperado (escrito u oral) y en calidad y cantidad suficientes como para poder extraer de los datos de discurso obtenidos lo necesario para elaborar una buena y justa calificación, basada en pruebas (no en meros indicios) de existencia de las habilidades, al aplicar las escalas y criterios determinados al efecto.

El proceso se completa con revisiones y análisis críticos en sesiones de familiarización y homogeneización de criterios, con el fin de poder mantener los parámetros adecuados y eliminar en la medida posible las normales subjetividades de este tipo de pruebas. En este proceso se generan tablas de recogida de datos y de calificación, escalas y criterios de calificación con doble corrección, sistemas de revisión de notas, así como análisis para determinar la fiabilidad intra- e intercorrectores/evaluadores, para las pruebas abiertas.

5. La administración del examen

El eLADE se administra en línea en centros de cualquier lugar del mundo que tengan un acuerdo para tal fin con la Universidad de Granada, así como en la sede del CLM de dicha universidad.

En el caso de que no haya ninguna sede próxima al lugar de residencia de un candidato, este puede ponerse en contacto con cualquier institución social o académica y que esta solicite ser sede acreditada cumplimentando el protocolo existente al efecto en nuestra página web ya citada. Los requisitos para ello son los siguientes: disponer de los recursos técnicos que exigen la realización del examen –especificados en la guía de administración- y firmar el acuerdo de confidencialidad por el que –entre otras cosas- toda entidad examinadora se compromete a no revelar ninguna información sobre el contenido de cada examen. Cada centro examinador debe asimismo designar a un responsable de administración que siga los protocolos de actuación antes, durante y después de la implementación del examen para asegurarse de la identidad del candidato así como de que no usen diccionarios y que lo realizan individualmente. No es necesario que la persona responsable hable español, pero sí que tenga unos conocimientos básicos de inglés, a fin de que pueda gestionar convenientemente la administración del examen, puesto que los protocolos de actuación están redactados en español e inglés.

Al matricularse se asigna a cada centro examinador o sede una fecha para que sus candidatos hagan la entrevista oral –mediante Skype- en un día y hora concretos. Al responsable de administración del examen se le envían, junto con el material, protocolos de actuación detallados sobre cómo debe gestionar cada una de las pruebas (tiempos, instrucciones sobre los recursos técnicos, instrucciones al candidato, etc.). Asimismo, debe comprobar la identidad de los candidatos y asegurarse de que no utilizan dispositivos, diccionarios o materiales adicionales.

6. Diseño del examen en línea: Aspectos y resultados técnicos del eLADE

El eLADE está desde un principio previsto para ser administrado a nivel mundial, por lo que el reto fue construir una plataforma que permitiera reproducir los contenidos del examen en ordenadores que necesitaran pocos recursos y que fuera fácilmente accesible desde cualquier dispositivo con acceso a internet.

Se ha desarrollado en una plataforma propia, construida a partir de herramientas comunes de la Web 2.0. buscando siempre la mayor compatibilidad posible con cualquier dispositivo que nos pudiéramos encontrar "al otro lado".

La seguridad se ve incrementada al ser necesaria una doble acreditación para entrar al examen. Por una parte, la página web de acceso es secreta hasta unos días antes del examen. Conocida la dirección de acceso, para poder entrar en la plataforma, se necesita tener el usuario y la clave de la sede. Una vez dentro se pueden hacer las comprobaciones de compatibilidad pertinentes, pero es imposible acceder al examen hasta que el candidato introduzca sus propios datos de acceso.

La plataforma está montada sobre un servidor de autenticación segura que refuerza aún más la consistencia de la herramienta.

Los vídeos están codificados conforme al protocolo "webm" que asegura una gran compatibilidad entre navegadores, ya que la decodificación la hace el propio navegador sin ser necesario cualquier tipo de descompresor externo.

Debido a las características del alfabeto español, que incluye letras o signos de puntuación que no podemos encontrar en teclados de otros países, hemos implementado una pequeña herramienta que permite al usuario introducir tales caracteres.

Por otra parte, y para evitar posibles errores, las teclas de función de los teclados son inhabilitadas remotamente.

Todos los datos de respuesta de los candidatos son salvaguardados en una base de datos a medida que el candidato los va introduciendo en su ejercicio en línea. Al estar todo el examen temporizado y monitorizado sabemos, en caso de cualquier problema técnico, todos los datos introducidos hasta el momento y en qué punto habría que retomar el examen en caso de ser necesario.

7. Datos de validación del examen

Ya hemos señalado (ver también Biedma, en prensa) que el eLADE se somete a un proceso de análisis de validación llevado a cabo con las pruebas, sus ítems y el examen en conjunto, con el fin de elaborarlo de forma adecuada a los niveles que pretende medir y a sus objetivos, utilizando SPSS, de IBM, y un análisis de IRT, con Winsteps. Con ellos, además de la fiabilidad, obtenemos datos como medias, discriminación, correlación entre pruebas, nivel de cada ítem, índice de dificultad de cada ítem con independencia de la población con la que se haya probado, errores de medida del examen en su conjunto y de cada uno de los ítems, número de niveles que es capaz de diferenciar, etc. A continuación ofrecemos los más relevantes de estos datos.

FIABILIDAD Y DISCRIMINACIÓN DE LOS ÍTEMS Y PRUEBAS

Fiabilidad

La fiabilidad es la seguridad con la que un test mide siempre de la misma manera lo que pretende estar midiendo, con la confianza de que en la obtención de la nota no influyan otros elementos ajenos a la habilidad lingüística que se quiere evaluar. Ello nos asegura que lo que evalúa se mantiene estable y solo afectado por incrementos de la habilidad lingüística. La fiabilidad se obtiene mediante programas informáticos y su resultado mediante SPSS es siempre entre -1 y $+1$, siendo $+1$ la mejor medida, lo que supone que si el test se repitiera con los mismos candidatos, obtendrían exactamente los mismos resultados si estos no hubieran aprendido nada nuevo. Las cifras aceptables para asegurar que una prueba tiene una buena fiabilidad no deben estar (según SPSS) por debajo de $.7$ (0.7), siendo mejores, como hemos dicho, cuanto más próximas están al 1 .

En la convocatoria de junio de 2015 la fiabilidad fue de .844 para la comprensión auditiva y de .892 para la comprensión lectora.

Discriminación

La discriminación es la capacidad que posee un ítem para diferenciar entre candidatos de más nivel de habilidad lingüística y candidatos de menor habilidad. Son ítems buenos aquellos que solamente son bien respondidos por los candidatos que poseen la habilidad y el nivel suficiente y, al contrario, son mal respondidos por los candidatos que no poseen suficiente nivel o habilidad. Esto no es así siempre, pues puede suceder que, por ambigüedad en el diseño de los ítems, por un mal diseño, o por otras razones, haya ítems que sean bien contestados bien por candidatos que no poseen el nivel adecuado, o que los fallen los que, sin embargo, sí poseen el nivel. En estos casos el ítem no sirve para decir quiénes son los que tienen mejores habilidades y los que no: el ítem no es bueno y debe desecharse. Una buena discriminación (en datos de SPSS) está entre .250 y 1.00, siendo mejor cuanto más alta sea (Green, 2013). Los ítems con los que finalmente se componen nuestros exámenes se sitúan en unos índices de entre .250 y .800.

Con los resultados del pilotaje y análisis de ítems y tareas de SPSS y IRT (fiabilidad, discriminación, error de medida, etc.) se reforma y diseña finalmente el examen, seleccionando los ítems que tienen el nivel apropiado para conseguir un test binivel (B1 y B2) suficientemente significativo y con los que se puedan obtener datos con los que elaborar una calificación final justa. Usualmente elaboramos unos 100 ítems de CA y CL y tras el pilotaje se seleccionan entre 25-40 de los mejores para cada destreza, desechándose los que no poseen el nivel de dificultad adecuado, los que no tienen una capacidad de discriminación óptima, los que tienen un error de medida inaceptable o los que no ayudan a mantener la fiabilidad del conjunto de la prueba. Con ellos se confeccionan las pruebas de CA y CL.

ESTABLECIMIENTO DE LOS NIVELES QUE EVALÚA EL EXAMEN

Para asegurarnos de que el nivel de los ítems y tareas finales de las pruebas y el examen en su totalidad evalúa el dominio tanto de candidatos de B1 como de B2 nos basamos en los datos del análisis de IRT, un método que, entre otras cosas, calcula el nivel y el valor de error de medida de cada ítem, con independencia de la población con que se haya probado.

Los datos obtenidos mediante IRT nos permiten decidir finalmente varias cuestiones importantes para la composición del examen: el nivel de cada ítem, su margen de error de medida, su poder de discriminación, la cantidad de niveles que se pueden evaluar mediante esa batería de ítems, la unidad de la prueba (o su capacidad para evaluar las facetas deseadas, eliminando los ítems que se salgan del modelo general del test), así como el establecimiento de los posibles puntos de corte entre los dos niveles medidos (tarea esta que, en última instancia, se realiza por un grupo de expertos en una sesión de fijación de puntos de corte utilizando su conocimiento y experiencia así como los datos estadísticos).

El nivel de los ítems y su relación con los niveles del MCER se establece mediante el siguiente cuadro (North y Jones, 2009), donde se considera que los ítems para evaluar el nivel B1 deben tener un "logit" de entre -1.23 y 0.72 (0.72 ya es B2) y para evaluar el B2, entre 0.72 y 2.80 (2.80 ya es C1). Para el diseño final del examen se seleccionan aquellos que tienen el nivel adecuado y un nivel de error de medida aceptable.

| Level | Cut-off | Range on logit scale |
|---------|---------|----------------------|
| C2 | 3.90 | |
| C1 | 2.80 | 1.10 |
| B2+ | 1.74 | 1.06 |
| B2 | 0.72 | 1.02 |
| B1+ | -0.26 | 0.98 |
| B1 | -1.23 | 0.97 |
| A2+ | -2.21 | 0.98 |
| A2 | -3.23 | 1.02 |
| A1 | -4.29 | 1.06 |
| Tourist | -5.39 | 1.10 |

CORRELACIÓN ENTRE LAS DISTINTAS PRUEBAS

Una vez diseñadas las pruebas con los ítems apropiados, también comprobamos si la confección del examen sirve para evaluar realmente habilidades diferenciadas entre sí (CA, CL, EIE y EIO). Para esto hallamos sus coeficientes de correlación: la medida en que en dos pruebas hay elementos comunes que, por lo tanto, estarían siendo evaluados varias veces y pesarían más en la nota final del examen. Si lo que se desea diseñar son pruebas diferentes (una que evalúe la CA y otra la CL, p. e.), responder a estas exige poner en funcionamiento elementos cognitivos, estrategias y conocimientos diferentes. En tal caso la correlación debe ser baja, entre 0.4 y 0.6. En nuestro caso, para considerar como definitivas las pruebas de CA, CL, EIE y EIO solo aceptamos -siguiendo la norma- una correlación menor de 0.7, lo que es prueba de que las pruebas miden habilidades diferentes. Por ejemplo, la correlación de la convocatoria de junio de 2015 entre las pruebas de CA y CL fue de .685.

ESTABLECIMIENTO DE PUNTOS DE CORTE Y NOTA FINAL

En las pruebas de CA y CL, los puntos de corte, es decir, el número de ítems que se deben responder correctamente para superar el examen en cada nivel, se establecieron siguiendo el método Angoff (Cizek, 2007; Consejo de Europa, 2009), ampliado con datos de IRT (según Figueras, 2011). Este método consiste en decidir el nivel de todos los ítems de una prueba por consenso entre un grupo de expertos: la cuestión es ver si un candidato mínimamente aceptable del nivel requerido para el examen podría responder correctamente a cada uno de ellos o no. El proceso se lleva a cabo mediante tres rondas de análisis y opiniones intentando acercar posiciones entre los expertos a los que se pide opinión. De cada una de las rondas se extrae una cifra (la media de las opiniones dadas por el conjunto de expertos) que es el punto de corte propuesto. Una vez establecido, en la primera ronda, un inicial punto de corte se ofrece a los expertos otros datos de interés que pueden hacerles variar de opinión y acercar sus posiciones intentando llegar a un acuerdo (número de aprobados según ese punto de corte, dificultad de cada ítem según los análisis estadísticos, etc.) y se procede a una nueva ronda para volver a establecer el punto de corte. En nuestro caso se utiliza una tercera ronda con datos de IRT (al modo en que se hace en Figueras, 2011). En las pruebas de EIE y EIO, como se ha mencionado ya, la calificación se realiza por el sistema de doble corrección. En la EIO uno de los evaluadores tiene el papel de entrevistador. Los evaluadores llevan a cabo sesiones periódicas de estandarización de criterios.

Las certificaciones se presentan al candidato con la calificación final de APTO en B1 o APTO en B2, explicitando el nivel que ha alcanzado en cada una de las destrezas por separado. Si el candidato ha obtenido un nivel B1 con alguna destreza de nivel B2, se le guarda la calificación de B2 de esa destreza durante dos convocatorias por si quiere presentarse a una nueva convocatoria para conseguir un certificado del nivel superior en el resto de las destrezas.

8. Principales resultados del uso de este examen en línea

El uso de este examen en sus sucesivas convocatorias ha sido exitoso tanto para las instituciones que lo utilizan, como la Universidad de Granada, como para los centros sede donde se realiza y los usuarios en general y candidatos de examen. Esto se demuestra por los comentarios recogidos en las encuestas periódicas realizadas en cada convocatoria, donde, en general, se comenta la satisfacción por el uso de las herramientas en línea, manifestando agrado por su operatividad y sencillez.

Además de los resultados estadísticos ya expuestos en el apartado anterior, vamos a reseñar aquí otros resultados igualmente interesantes procedentes de los cuestionarios que completan los candidatos o usuarios al terminar el examen. De todos los elementos que se evalúan en las entrevistas a usuarios, cabe destacar el hecho de que el uso de elementos multimedia para evaluar la habilidad audiovisual es considerado como muy positivo. Asimismo, en relación al uso de audio y vídeo en línea, los usuarios candidatos no señalan problemas de operatividad ni audición en comparación con lo que suelen manifestar cuando se realizan pruebas audiovisuales en papel. Este hecho -en consonancia con datos ya señalados en la literatura sobre computer testing (Mubarak Pathan, 2012 o Winke, 2008)- se debe a que el uso individual de los medios, añadido a la necesaria audición con auriculares, hace que las condiciones sean óptimas e iguales para todos los usuarios, frente a las variables que se ocasionan en métodos tradicionales donde las condiciones de audición pueden cambiar de un espacio a otro e incluso de un puesto en un aula a otro más alejado del foco de voz, cercano a una ventana, etc. Con ello la fiabilidad de la prueba gana, y la validez de apariencia del examen también, como lo manifiestan numerosos encuestados.

Otro elemento que también es valorado positivamente es la posibilidad por parte del candidato de gestionar el tiempo de que dispone. Fundamentalmente se elogia en las tareas de la prueba de comprensión lectora, donde se selecciona libremente el texto que se desea leer y al que se responde en cualquier momento durante sus 75 minutos de duración, pudiendo volver atrás y adelante en cualquier momento tanto para cambiar de texto como para responder o corregir las respuestas. Igualmente se señala la cierta tranquilidad que produce el hecho de controlar el tiempo restante para cada tarea y prueba, y especialmente -señalan algunos candidatos- la posibilidad de realizar una pausa al terminar las pruebas de CA y CL y antes de comenzar la de EIE, que está guiada asimismo mediante un temporizador si bien no supone una penalización si se excede el límite de tiempo y que se puede acortar permite acortar el tiempo si se desea comenzar antes la siguiente prueba.

En este aspecto todavía nuestro examen en línea puede mejorarse haciendo que esta gestión del tiempo se extienda a cada prueba y tareas por separado, para que los usuarios se administren el examen con mayores cotas de libertad, en concreto para seleccionar independientemente las audiciones, extremo este que no se permite (al contrario que en la prueba de comprensión lectora), pero es un asunto a día de hoy controvertido pues la apariencia de realidad y de validez que presenta choca con una posible pérdida de fiabilidad del examen, que debe ser estudiada antes de tomar la decisión.

Finalmente, dentro de los resultados más comentados por los candidatos, nos encontramos con el hecho positivo de ser un examen binivel y la posibilidad de guardar el resultado de una nota de nivel mayor que la obtenida en el certificado, con la posibilidad de presentarse al nivel superior (B2) sabiendo que ya tiene una parte de esa prueba superada; esto se guarda durante dos convocatorias. El hecho de ser binivel garantiza a casi todos los candidatos la obtención de un certificado, ya sea de B1 o de B2 (salvo a los que se quedan en nivel A2 o menos y no obtienen ninguna certificación). Esto es valorado por muchos candidatos. Además, al recibir el certificado se anota por separado la calificación de cada habilidad, por lo que quienes no obtienen el

nivel deseado conocen si en alguna habilidad son B1 o B2. Esto otro, conservar la calificaciones superiores a la obtenida en el certificado durante dos convocatorias es valorado muy positivamente en las encuestas. Igual opinión expresan aquellos candidatos que obtuvieron un certificado de B2 inesperadamente, por aspirar solamente al de B1.

En cuanto a los administradores de las pruebas y los centros meta de estas, cuando se utiliza como medio de admisión, hasta el momento no han manifestado disconformidades ni comentarios de mejoras relevantes. Sí podemos decir, en cuanto a la Universidad de Granada y el Centro de Lenguas Modernas, principales usuarios finales de este examen por el momento, que están muy satisfechos tanto por su operatividad como por sus resultados. Lo principal ha sido el cambio ofrecido a los estudiantes que antes debían venir a España sin saber si finalmente obtendrían un nivel suficiente para cursar los estudios que les interesaban y ahora ya vienen admitidos y seguros de su certificación. Otro aspecto reseñable ha sido el ahorro en tiempo que supone administrar los exámenes en línea pudiendo acceder a los materiales rápidamente y a sus resultados de forma fiable y rápida. Es por esto que se ha planteado seguir diseñando, utilizando e incluso implementando este examen, para evaluar desde A1 hasta C1, con fines más allá de la admisión a cursos y fines académicos. La forma de realizar estas implementaciones, ya se ha señalado anteriormente, conservaría el mismo formato binivel, pero es un tema que está en estudio y sometido a investigaciones. Igualmente, de la satisfacción por los resultados, da cuenta el compromiso de la Universidad de Granada y su CLM por seguir dedicando tiempo, materiales, personal y esfuerzo económico a un proyecto muy necesario y con muchas posibilidades de desarrollo. La existencia de este nuevo examen en línea ha cumplimentado adecuadamente y ha satisfecho todas las expectativas respondiendo de forma óptima a una necesidad social y académica que puede beneficiar a todas las universidades españolas.

9. Discusión

En la actualidad el examen eLADE de la Universidad de Granada, nacido para cubrir las necesidades de acreditación específicas para los programas de movilidad, se ha ampliado para evaluar a todo tipo de personas interesadas en acreditar un nivel de lengua tanto por razones académicas como sociales o personales. El hecho de ser un examen en línea y de fácil acceso, para el que no es necesario acudir a una sede determinada que implique un desplazamiento del candidato, con un formato sencillo y atractivo, y validado técnicamente con los estándares internacionales de evaluación de lenguas, lo convierten en una herramienta útil, fiable y necesaria para la comunidad educativa así como para el público en general.

Gracias al desarrollo tecnológico tenemos la oportunidad de facilitar a casi cualquier persona la oportunidad de obtener una acreditación de nivel de lengua sin desplazarse demasiado de su lugar de residencia. Y gracias a la profesionalidad y transparencia con que se desarrolla, el hacerlo de forma válida y fiable. Ello nos invita a pensar en la futura acreditación de otros niveles de dominio, A1, A2, C1 y C2, y de este modo contribuir a la expansión de la lengua española con un instrumento como el eLADE.

No obstante, diseñar, validar y administrar un examen por ordenador requiere de numerosos medios técnicos, humanos y económicos. También es necesario investigar los resultados e implementar ciertas mejoras, siempre que se disponga de las herramientas adecuadas para ello. En este sentido y a pesar del éxito ya señalado y a la satisfacción por el uso del eLADE, hay algunos aspectos que pueden mejorarse o sobre los que hay que realizar análisis en espera de posibles mejoras o reformas. Debemos seguir planteándonos la extensión de los niveles de dominio de la lengua que evaluamos mediante el eLADE ampliándolo a otros niveles que se están demandando en centros, institutos o, simplemente, por usuarios independientes, desde A2 a C2. En este sentido y debido a los buenos resultados de validación y de satisfacción obtenidos así como a las deficiencias observadas respecto a la fiabilidad de otros exámenes progresivos o la dificultad de los análisis estadísticos necesarios para validarlos adecuadamente, seguimos apostando por evaluar los nuevos niveles mediante un examen binivel que sería de A2/B1 y de B2/C1, en principio, utilizando los mismos bancos de ítems para los niveles comunes de cada examen.

Hay otros elementos que se podrían plantear de cara a nuevas convocatorias, como hacer un uso más fuerte de las posibilidades del formato por ordenador, siempre debidamente acompañadas de análisis que muestren su validez y eficacia. Entre ellos estaría explorar diferentes formatos de ítems, p. ej. desplazando elementos en el escritorio o pantalla del ordenador mediante el ratón (fotos, palabras, etc.) en tareas de emparejamiento o de completar huecos, que podrían dar juego a la dinámica, hacer un manejo más sencillo y aumentar la motivación y satisfacción de uso. También, como hemos señalado anteriormente, se puede ir hacia una gestión más autónoma o completamente autónoma (habría que estudiar los resultados) del uso del tiempo y del orden en que cada candidato elija realizar cada tarea -CA, CL y EIE- o volver atrás y adelante en ellas. Esto puede redundar en una mejora de la autenticidad -según el comportamiento libre de cada usuario- y en la apariencia de realidad del examen. Finalmente, es importante comparar resultados de los modelos del examen en línea y en papel, siguiendo el camino trazado por otras investigaciones (Winke 2008) -algo que en el CLM podemos comparar fácilmente con diferentes grupos de estudiantes- así como analizar el importante factor de en qué medida la habilidad en el uso de ordenadores influye en los resultados del examen o en las notas (Madsen 1991). Este hecho puede parecer innecesario por el momento, dado que la mayoría, si no todos los candidatos presentados al eLADE, han sido usuarios de ordenadores que no han manifestado problema alguno al respecto. No obstante, si continua aumentando el número de usuarios puede variar el grado de formación y destreza en el uso de tecnologías y darse el hecho de que haya quienes no estén habituados. Las diferencias de este hecho deberían poder ser detectadas para solucionar el problema o intentar minimizar sus consecuencias. Por ejemplo, habrá que analizar si se llegan a dar diferencias entre el uso por nativos tecnológicos frente al resto de usuarios, algo que puede tener relevancia en el futuro.

Y por último, y mientras sigamos empleando un modelo binivel, hay que intentar que el establecimiento de los puntos de corte y la forma de explicar la obtención de la nota final de Apto/No Apto sea adecuada e inteligible por los usuarios, dificultad que se apunta en la literatura sobre exámenes por ordenador (Brown 1997) y que nosotros también hemos detectado.

Respecto a la forma de calificar, vamos a seguir utilizando un sistema mixto, humano y tecnológico pues, especialmente en las pruebas abiertas de EIE y EIO, nos parece lo más adecuado para asegurarnos un constructo y una validez en las calificaciones de dominio de estas habilidades que pueda dar autenticidad y credibilidad al examen.

El planteamiento y análisis de todos estos aspectos entra dentro de los habituales protocolos de validación de exámenes que empleamos en el CLM de la UGR si adoptamos, como hemos hecho y pretendemos mediante este artículo, un enfoque de evaluación de lenguas garantista (Martínez Baztán, 2011) y transparente que nos asegura que evaluamos lo que pretendemos y lo que nuestros usuarios y agentes sociales demandan. La Universidad de Granada está satisfecha con los resultados e implicada en la constante mejora de las herramientas.

Esperamos contribuir con este artículo a la necesaria difusión del examen y a obtener la credibilidad social y académica así como el crédito de los usuarios y agentes sociales que lo deseen utilizar para que nos sigan ofreciendo su confianza como evaluadores.

Cómo citar este artículo / How to cite this paper

Martínez, A., Biedma, A., Sánchez, A., Chamorro, M.D., y Sánchez, S. (2015). Descripción y análisis del examen online de acreditación de dominio de español de los niveles B1 y B2: el eLADE. *Certiuni Journal*, 1, 52-66. Consultado el [dd/mm/aaaa] en www.certiunijournal.com

Referencias

- ALDERSON, J.Ch., CLAPHAM, C., y WALL, D. (1998). Exámenes de idiomas. Elaboración y evaluación, Madrid: Cambridge University Press. [Título original: (1995): *Language Test Construction and Evaluation*, Cambridge: Cambridge University Press].
- ASSOCIATION OF LANGUAGE TESTERS IN EUROPE (ALTE) (2001): *Principles of Good Practice for ALTE Examinations*, [pdf], [en línea: 21/12/2007], www.alte.org/attachments/files/good_practice.pdf
- BACHMAN, L.F. (1990). *Fundamental Considerations in Language Testing*. Oxford, Oxford University Press.
- BACHMAN, L. (2004). *Statistical Analyses for Language Assessment*. Cambridge, Cambridge University Press.
- BACHMAN, L., y PALMER, A. (1996). *Language Testing in Practice*. Oxford, Oxford University Press.
- BIEDMA, A., CHAMORRO, D., MARTÍNEZ, A., SÁNCHEZ, A., SÁNCHEZ, S., AMADOR, C., PUERTAS, J., y RODRÍGUEZ, J. (en prensa), "eLADE: The University of Granada Bi-level B1/B2 Accreditation Exam. Construction and Validation Process", CASALC (Czech and Slovak Association of Language Centers) Review.
- BIEDMA, A., CHAMORRO, D., MARTÍNEZ, A., SÁNCHEZ, S., AMADOR, C., PUERTAS, J., y RODRÍGUEZ, J. (en prensa), "eLADE: Examen para la Acreditación de Dominio del Español de niveles B1 o B2 de la Universidad de Granada" en Actas del I Congreso FLECOS, Granada.
- BROWN, J.D. (1997). Computers in language testing: Present research and some future directions. *Language Learning and Technology* 1(1), 44-59. <<http://llt.msu.edu/vol1num1/brown>>
- CHALHOB-DEVILLE, M. (2010). Technology in standardized language assessments. In R. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (2ª ed., pp. 511-26). Oxford, England: Oxford University Press.
- CIZEK, G., y BUNCH, M. (2007). *Standard Setting*. Sage Publications, Thousand Oaks, California.
- CONSEJO DE EUROPA (2002). Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación, Madrid: Ministerio de Educación, Cultura y Deporte; Instituto Cervantes. [Título original (2001): *Common European Framework of Reference for Languages Learning, teaching, assessment*, Strasbourg: Council of Europe].
- CONSEJO DE EUROPA (2009). *Relating Language Examinations to the CEF, Manual Preliminary Pilot Version. Introduction and Feedback to the pilot phase*, [pdf], Strasbourg: Council of Europe.
- CONSEJO DE EUROPA (2011). *Manual for Language test Development and Examining*, [pdf], Strasbourg: Council of Europe.
- DOUGLAS, D. (2010). *Understanding language testing*. London, England: Hodder Education.
- DUNKEL, P. A. (1999). Considerations in developing or using second/foreign language proficiency computer-adaptive tests. *Language Learning and Technology*, 2(2), 77-93.
- EUROPEAN ASSOCIATION FOR LANGUAGE TESTING AND ASSESSMENT (EALTA) (2004). *Directrices de EALTA para una buena práctica en el desarrollo de pruebas y la evaluación de lenguas*, [en línea 5/12/2007], <http://www.ealta.eu.org/documents/archive/guidelines/Spanish.pdf>
- FIGUERAS, N., KAFTANDJIEVA, F., y TAKALA, S. (2011). Relating a Reading Comprehension Test to the CEFR Levels: A Case of Standard Setting in Practice with Focus on Judges and Items. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 69, 4, (November / novembre), 359-385.
- GREEN, R. (2013). *Statistical Analyses for Language Testers*, Palgrave MacMillan.
- INTERNATIONAL LANGUAGE TESTERS ASSOCIATION (ILTA) (2000). *Code of Ethics for ILTA*, [en línea: 2/6/2012], www.iltaonline.com/images/pdfs/ILTA_Code.pdf
- MADSEN, H. S. (1991). Computer-adaptive testing of listening and reading comprehension: The Brigham Young approach. En: P Dunkel (ed.), *Computer-assisted language learning and testing: Research issues and practice* (pp. 237-57). New York: Newbury House.
- MARTÍNEZ BAZTÁN, A. (2008). *La evaluación oral: una equivalencia entre las guidelines de ACTFL y algunas escalas del MCER* (PDF) (tesis doctoral). Universidad de Granada. ISBN 978-84-338-4961-8
- MARTÍNEZ BAZTÁN, A. (2011). *La evaluación de lenguas: garantías y limitaciones*, Octaedro, Barcelona.
- MUBARAK PATHAN, M. (2012). *Computer Assisted Language Testing [CALT]: Advantages, Implications and Limitations*. En www.researchvistas.com, vol 1 (104).
- NORTH, B., y JONES, N. (2009): *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). Further Material on Maintaining Standards across Languages, Contexts and Administrations by exploiting Teacher Judgment and IRT Scaling*, [pdf], Strasbourg: Council of Europe, [en línea: 20/2/2014], http://www.coe.int/t/dg4/education/elp/elpreg/Source/Publications/Standards_language_EN.pdf
- WINKE, P., y FEI, F. (2008). Computer-Assisted Language Assessment. En: N. Van Deusen-Scholl and N. H. Hornberger (eds), *Encyclopedia of Language and Education*, 2nd Edition, Volume 4: Second and Foreign Language Education, 353-364. 2008 Springer Science+Business Media LLC.