EL VALOR DE UNA CERTIFICACIÓN CON GARANTÍAS: ELADE (EXAMEN EN LÍNEA PARA ACREDITACIÓN DE DOMINIO DE ESPAÑOL B1/B2) DEL CLM DE LA UGR

Alfonso Martínez Baztán, Aurora Biedma Torrecillas,
Lola Chamorro Guerrero, Sonia Sánchez Molero
Centro de Lenguas Modernas de la Universidad de Granada

RESUMEN

Este artículo plantea la necesidad de garantizar que los exámenes de acreditación así como sus certificaciones se realicen acorde con los estándares internacionales de validación, fiabilidad, transparencia y buenas prácticas. La sociedad actual exige procesos de certificación que doten de valor al examen para lograr un verdadero y duradero reconocimiento académico y social. Además de exponer la necesidad de demostrar garantías para el examen, se muestran estos aspectos en el eLADE del CLM de la UGR (Examen en línea para la Acreditación del Dominio del Español, examen binivel, de B1 y B2).

Palabras clave: evaluación, dominio, acreditación, ordenador, validación, español.

ABSTRACT

This article raises the need to ensure that accreditation exams as well as their certifications are carried out in accordance with international standards of validation, reliability, transparency and good practices. The current society demands certification processes that add value to the exam to achieve a true and lasting academic and social recognition. In addition to stating the need to demonstrate guarantees of the exam, these aspects are shown in the eLADE of the CLM of the UGR (on-Line for Spanish Domain Accreditation, bi-level exam, B1 and B2).

Keywords: testing, proficiency, accreditation, computer, validation, Spanish.

1. LA NECESIDAD DE OBTENER CREDIBILIDAD SOCIAL Y ACADÉMICA

La sociedad actual, tanto desde un punto de vista académico como laboral o meramente social o personal, exige a los evaluadores, cada vez con más determinación, garantías de que el proceso de elaboración de un examen, las notas y certificación (desde el planteamiento de su diseño hasta la elaboración de sus resultados y entrega a los candidatos y agentes sociales) siga los estándares más exigentes, acordes con la situación, objetivos e intenciones del examen y sus consecuencias académicas y sociales (ILTA, 2000; ALTE, 2001; EALTA, 2004; ALTE, 2008; CoE, 2002; CoE, 2009; CoE, 2011, ACLES, 2015).

Todas las pruebas pero especialmente las pruebas de certificación, por su alto impacto, deben perseguir una validez profesional y una credibilidad que haga que los agentes socio-laborales que vayan a utilizar sus resultados, y sus usuarios y candidatos confíen en ellas. Solo en la medida en que esto se logre serán de verdadera utilidad, serán duraderas y alcanzarán prestigio. Por estas importantes razones deben garantizar tres aspectos (sin olvidar el resto de formas de validez, fiabilidad y operatividad -Alderson, 1998, Bachman 1990-): a. la validez de apariencia, b. la transparencia en el proceso y entrega de datos, y c. el respeto a las intenciones, idiosincrasia y necesidades de los candidatos.

Creemos necesario recordar (Biedma, 2016; Biedma 2016b; Martínez Baztán, 2015) que un buen examen no es gratuito en sus intenciones y debe responder a una sensibilidad académica o social que lo sustente, en consonancia con las necesidades reales de uso que se vean adecuadamente conceptualizadas y reflejadas en el producto final que es el examen.

Mediante la validación de apariencia -(autenticidad según Bachman y Palmer, 1996)- se consigue credibilidad, en la medida en que si lo que se mide representa la vida real de los candidatos estos están mejor y más dispuestos a responder (Martínez Baztán, 2011). Tener en cuenta la idiosincrasia, necesidades e intenciones de los candidatos (y no solo las de las instituciones evaluadoras o los agentes sociales que demandan una certificación) ayuda a que los candidatos respondan con motivación y reflejen mejor sus habilidades (Fulcher, 2003, Shohami, 2001, Shohami, 2001b). Finalmente, hay que asegurarse de que la custodia de datos, envío de resultados, revisión, etc. sigan protocolos claros, aceptables e inteligibles para los candidatos.

El formato de nuestro examen facilita estas labores (Martínez Baztán, 2015),

pues entre las ventajas de los exámenes en línea (para desventajas ver Madsen, 1991) están, de un lado, que son más motivadores y facilitan una actitud más positiva y, de otro, simplifican la burocracia de envíos de materiales y de acceso a la información y a muestras de exámenes (Mubarak Pathan, 2012). Respecto a la motivación, esto es debido no solo a aspectos externos como la familiaridad que actualmente se tiene con los ordenadores como medio habitual de trabajo o porque posibilitan una gestión personalizada de la prueba, sino a hechos más relevantes como que se evita la dispersión en la resolución de tareas y se favorece la concentración. A este respecto cabe destacar que el formato en línea aumenta la autenticidad y la interacción real entre las tareas y los candidatos, al mejorar las situaciones comunicativas de la lengua en muchos aspectos (Douglas, 2010) o debido a que las habilidades de audio y vídeo resultan más fiables y auténticas (Winke, 2008).

Desde la publicación del MCER el debate está abierto. Ante el aumento de ofertas para necesidades de acreditación con fines diversos y la esperable y deseable proliferación de pruebas de distintas empresas e instituciones, las pruebas deben ganarse un prestigio basado en la validez, la transparencia y la honestidad de sus resultados, más allá de la pertenencia a una institución o marca empresarial. Con nuestro examen y la apuesta por la transparencia, la validación y la operatividad esperamos contribuir a mejorar los procesos de diseño de estos exámenes de español así como a que se reconozca el derecho de los usuarios y candidatos a la información.

2. TRANSPARENCIA EN LOS PROCESOS DE DISEÑO Y VALIDACIÓN DEL ELADE DEL CLM-UGR

La información y transparencia son premisas indispensables para que los candidatos (y los agentes sociales que van a acreditar o usar sus datos) sepan qué se les va a medir, mediante qué procesos y si ese es el modelo de examen que necesitan o están dispuestos a realizar. Por ello, hemos querido ser transparentes tanto en mostrar su elaboración como en hacer públicos los resultados estadísticos en la red, y en hacer evidente que el examen está claramente diseñado de acuerdo a estándares internacionales.

En este artículo presentamos información del prepilotaje de la última convocatoria del examen, mostrando que los procesos de su diseño son adecuados.

2.1 Descripción y especificaciones del examen

El eLADE -examen en Línea de Acreditación de Dominio del Español- es un examen que surgió ante la reciente necesidad de acreditar el dominio de lenguas tras la implantación del Plan Bolonia en Europa. La Universidad de Granada encargó a su Centro de Lenguas Modernas crear un examen binivel que acreditara los dos niveles más demandados, B1 y B2. Nació así un examen de acreditación en soporte de papel y presencial en 2010, que daría paso al actual eLADE en línea en 2013. Dada la fiabilidad y validez del examen se solicitó y consiguió la garantía de certificación de ACLES, la CRUE y CERCLES.

Al conocer la existencia de un examen en línea de español con estas garantías de fiabilidad y validez, aumentó la demanda de otras instituciones no universitarias y personas que necesitan por diferentes motivos acreditar su nivel de español. Por esta razón se amplió el uso y funciones del mismo, extendiendo sus fines más allá de la admisión a cursos universitarios y accediendo a nuevos ámbitos académicos y sociales.

Con respecto a la versión en papel, el actual eLADE ha introducido cambios en el diseño, el contenido y la validación. Los primeros están derivados del hecho de ser en línea, como la duración controlada, el orden de los audios, la variedad de formatos y otras mejoras en los aspectos de autenticidad y motivación. En cuanto al constructo y contenido, se ha renovado por completo la sección de Expresión escrita y los textos de lectura y audiovisuales se han ajustado aún más a los descriptores del Marco respecto a la tipología, temática, ámbito, registro y autenticidad. Dado que el examen pretende ser una prueba que refleje, en la medida de lo posible, situaciones reales de comunicación, se ha cuidado la selección de la tipología textual, la presentación y contextualización de los textos y, por otro lado, se ha puesto un gran interés en que sean motivadores e interesantes, así como en que las tareas sean idóneas a cada tipo de texto, de modo que se facilite que los candidatos puedan demostrar su dominio real en cada habilidad.

Por último, se han incluido análisis estadísticos que contribuyen tanto a validar el examen como a probar que este sirve efectivamente para medir los dos niveles requeridos, B1 y B2.

El examen consta de cuatro pruebas diseñadas para evaluar seis actividades comunicativas de la lengua: comprensión auditiva, comprensión lectora, expresión e interacción orales y expresión e interacción escritas. Tiene una

duración de 3 horas y 15 minutos y todas sus pruebas son en línea (la prueba de Expresión e Interacción oral se hace por Skype directamente con el candidato).

En la siguiente tabla se especifican el número de tareas que componen cada prueba y su duración:

EXPRESIÓN E INTERACCIÓN ORALES	**EXPRESIÓN E INTERACCIÓN ESCRITAS**	**COMPRENSIÓN LECTORA**	**COMPRENSIÓN AUDITIVA**
3 tareas de expresión e interacción orales	2 tareas de escritura	5 tareas de lectura	5 tareas auditivas y/o visuales
10-15 minutos	60 minutos	75 minutos	45 minutos
Skype/presencial en el CLM	En línea	En línea	En línea

Figura 1. Tareas y duración del eLADE

En cada prueba, las tareas están graduadas: aproximadamente la mitad corresponden al nivel B1 y la otra mitad al nivel B2. El número total de ítems de las pruebas, en la CA y en la CL, es de entre 38 a 42.

Puede verse la descripción completa del examen, sus especificaciones, tablas de análisis estadísticos de cada convocatoria y un ejemplo del mismo en el siguiente enlace: www.clm-granada.com/es/esp/examenes/elade/elade-info.html

2.2 Prepilotaje

En cada una de sus convocatorias oficiales, el eLADE se somete a entrevistas pre y post test con expertos, nativos y candidatos y a un prepilotaje con una muestra de unos 100 candidatos (Martínez Baztán, 2015). Con el análisis de las entrevistas y el prepilotaje se logra saber si las tareas e ítems se atienen a lo que se deseaba medir y ello permite rediseñar, modificar o desechar los elementos y aspectos más débiles y adecuar el diseño final a los objetivos y estándares de contenido, viabilidad y nivel (B1 y B2) establecido en las especificaciones. Para ello se someten a un análisis clásico (mediante SPSS -*Statistical Package for the Social Sciences*) y a otro de teoría de respuesta al item (IRT, mediante *Winsteps*). Así obtenemos diversas informaciones sobre el funcionamiento del

examen junto a otras sobre la habilidad de las personas participantes y su capacidad para desenvolverse ante las tareas propuestas (Green, 2013).

A continuación ofrecemos un cuadro con datos del prepilotaje de las habilidades de comprensión auditiva y lectora diseñadas para el eLADE de la última convocatoria de junio de 2017; datos sobre dificultad de los ítems, error de medida de cada prueba, confianza en que el examen en conjunto es apropiado para evaluar los dos niveles, discriminación de los ítems, correlación entre las pruebas y fiabilidad de cada una, que previamente explicamos:

1. Dificultad y nivel de cada ítem: la decisión del nivel así como de qué ítems son de uno u otro nivel se decide mediante el análisis estadístico (SPSS e IRT) y la valoración de un equipo de expertos en una sesión de fijación de puntos de corte mediante el método Angoff. Remitimos a la página web (www.clm-granada.com/pdf/elade/validacion_eLADE_junio_2014.pdf), donde se explica detalladamente el método (Cizek, 2007; Consejo de Europa, 2009), ampliado con datos de IRT (según Figueras, 2011).
2. Error de medida: ninguna medición es perfecta así que averiguamos sus niveles de error mediante procedimientos estadísticos.
3. Medida de separación de los ítems: al ser un examen binivel nos aseguramos de que efectivamente sirve para evaluar dos niveles diferenciados, para lo que necesitamos una medida de separación, obtenida con el análisis de IRT, superior a 3.
4. Discriminación: es la medida en que un ítem es respondido correctamente por las personas mejor preparadas y de forma errónea por las de menor competencia, debido a distintas causas. Los ítems aptos deben tener un mínimo de .250.
5. Correlación: es la medida que indica que entre dos pruebas se solapan algunos elementos, que se medirían, por lo tanto, de forma repetida e injusta (favoreciendo a quienes tuvieran esas habilidades medidas doblemente, frente a otras). Lo óptimo en exámenes de lengua es tener un índice menor a 0.7.
6. Fiabilidad: es la medida con la que nos aseguramos de que si se repitiera el examen se obtendrían los mismos resultados. Un buen examen de alto impacto necesita un índice de 0.9, al menos.

En esta tabla ofrecemos todos estos datos en el prepilotaje del eLADE de nuestra última convocatoria. Como se ve, las cifras en que nos movemos se ajustan a los estándares de calidad internacionalmente aceptados.

COMPRENSIÓN AUDITIVA	COMPRENSIÓN LECTORA	DATOS ESTADÍSTICOS	
43	40	Ítems	
B2: 20 B1: 23	B2: 17 B1: 21	Nivel	1.
5.7 puntos	5.4 puntos	Error de medida	2.
4.42	3.57	Separación	3.
entre .356 y .501	entre: .315 y .503	Discriminación	4.
.668	.668	Correlación CL-CA	5.
.95	.92	Fiabilidad	6.

Figura 2. Datos estadísticos de prepilotaje del eLADE para junio de 2017

2.3 Resultados y encuestas

El uso de este examen en sus convocatorias ha sido exitoso para las instituciones que lo utilizan, centros sede donde se realiza y usuarios y candidatos en general. Esto se demuestra por los análisis de las encuestas pre y post test realizadas en cada convocatoria, donde se comenta la satisfacción por el examen manifestando agrado por el contenido, operatividad y sencillez que ya hemos señalado en otras ocasiones (Martínez Baztán, 2015).

REFERENCIAS BIBLIOGRÁFICAS

Asociación de Centros de Lengua de Enseñanza Superior (ACLES), (2015), *Guía explicativa al formulario de solicitudes de acreditación de exámenes binivel*, [pdf], [en línea: 24/03/2017),www.acles.es/files/ckeditor/guiaexplicativa_al_formulario_binivel_2017-2018.pdf

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., 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. (2016). "eLADE: The University of Granada Bi-level B1/B2 Accreditation Exam. Construction and Validation Process", en *CASALC (Czech and Slovak Association of Language Centers) Review* 5, 1: 208-220.

Biedma, A., Chamorro, D., Martínez, A., Sánchez, A., Sánchez, S., Amador, C., Puertas, J., y Rodríguez, J. (2016b). "eLADE: Examen para la Acreditación de Dominio del Español de niveles B1 o B2 de la Universidad de Granada", en A. Illescas García y A. Perales Castro (eds.), *Actas del I Congreso Internacional de la lengua, la Cultura y el Ocio*. Granada: Editorial Universidad de Granada, 21-38.

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.

European Association for language Testing and Asessment (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.

Fulcher, G. (2003), *Testing Second Language Speaking*, London: Pearson/ Longman.

Green, R. (2013). *Statistical Analyses for Language Testers*, Palgrave McMillan.

International Language Testers Association (ILTA) (2000). *Code of Etics 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. (2011). *La evaluación de lenguas: garantías y limitaciones*, Octaedro, Barcelona.

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 [21/03/2017] en www.certiunijournal.com

Mubarak Pathan, M. (2012). *Computer Assisted Language Testing [CALT]: Advantages, Implications and Limitations*. En www.researchvistas.com, vol 1 (104).

Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. London: Longman.

Shohamy, E. (2001b). "Democratica assessement as an alternative". *Language Testing*, 18 (4), 373-392.

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.