

## ¿SON FIABLES LOS RESULTADOS DE LOS DELE? EN BUSCA DE NUEVAS EVIDENCIAS EMPÍRICAS<sup>1</sup>

Abel CARDENAS  
abeljpn@aol.com

### INTRODUCCIÓN

El sistema de evaluación de los Diplomas de Español como Lengua Extranjera (DELE, de aquí en adelante), es probablemente el sistema de acreditación de dominio del español más utilizado actualmente en el mundo. En más de dos décadas de existencia el sistema ha experimentado importantes mejorías, sin lugar a dudas. El constante aumento del número de centros de exámenes (unos cuantos en sus inicios; más de 600 en la actualidad) y de candidatos tanto a nivel nacional en Japón como mundial (hasta llegar a 47171 en 2009, según cifras oficiales), y la reciente incorporación de los exámenes de nivel A1, A2 y C1, por ejemplo, son muestras de los esfuerzos que el Instituto Cervantes, el organismo principal responsable de los DELE, ha hecho por mejorar su sistema de certificación. Cabe preguntarnos, sin embargo, si estos esfuerzos por mejorar el sistema han realmente llegado a la toma e implementación de iniciativas o medidas para perfeccionar los aspectos más relevantes de un sistema de evaluación: la fiabilidad y validez de los resultados. ¿Son fiables y válidos los resultados de las pruebas de los DELE? Es una pregunta que las instituciones involucradas en el sistema de evaluación que nos ocupa no han respondido de una manera empírica.

Siguiendo con la línea de investigación iniciada hace años, el autor continúa con la búsqueda de *evidencias empíricas* de la fiabilidad de los resultados de los DELE. La fiabilidad, recordémoslo, es una de las cualidades principales que un instrumento de evaluación debe cumplir y se ha definido como “la medida en que los resultados de un examen son estables, consistentes y libres de error” (Saville, 2003: 69). El objetivo principal de este trabajo es, por lo tanto, reexplorar el grado de fiabilidad de los resultados de las pruebas objetivas de los DELE. Primero, realizaremos una revisión de la literatura, incluyendo los Estándares Mínimos de Calidad de ALTE y las publicaciones relevantes sobre los DELE. En la segunda parte, se presentarán los resultados de un *nuevo estudio* en el que se buscaron evidencias empíricas sobre el grado de fiabilidad o consistencia de los resultados de las pruebas objetivas del DELE de nivel inicial o B1. Posteriormente, se concluirá con un análisis de los problemas detectados y las medidas a tomar para que el sistema de evaluación DELE sea más fiable, válido y justo para los diferentes usuarios.

<sup>1</sup> Este trabajo se realizó con ayuda del 2009 Pache Research I-A-2 Fund de Nanzan University

## I. REPASO DE LA LITERATURA

### 1 “LO QUE DEBE SER”: LOS ESTÁNDARES MÍNIMOS DE CALIDAD DE ALTE

En el campo de la evaluación de las lenguas extranjeras existen actualmente diferentes asociaciones profesionales que han publicado una serie de estándares que las instituciones que ofrecen servicios de evaluación y certificación deben cumplir para garantizar la transparencia y calidad de sus sistemas. Como tanto el Instituto Cervantes, el organismo que encabeza el sistema DELE, como la Universidad de Salamanca, encargada de la elaboración y calificación de los exámenes, son miembros de ALTE (Asociación Europea de Evaluadores de Lenguas) y, por consiguiente, deben sujetarse a sus principios, conozcamos en qué consisten sus Estándares Mínimos de Calidad. Como podemos ver en el apéndice, ALTE fija 17 normas para establecer exámenes de calidad en cinco apartados: *confección de exámenes*, *administración y logística*, *puntuación y calificación*, *análisis de exámenes*, y *comunicación con los partícipes*. En el apartado de *Puntuación y Calificación*, por ejemplo, encontramos que en un examen de calidad:

11. La calificación es lo suficientemente precisa y fiable para la finalidad y el tipo de examen del que se trate.

12. Puede documentarse y explicarse cómo se lleva a cabo el proceso de calificación, cómo se calcula la fiabilidad y cómo se recogen y analizan los datos correspondientes al rendimiento de los calificadores de las pruebas de expresión oral y escrita.

Obsérvese que la importancia de explicar y documentar la fiabilidad está presente no sólo en estas normas, sino también en aquellas incluidas en las secciones sobre *Análisis de Exámenes* y la *Comunicación con los Partícipes* (estos apartados aparecen sombreados). Pasemos ahora a la literatura oficial para determinar de qué manera el Sistema de Evaluación DELE cumple con estos estándares.

### 2 “LO QUE OFICIALMENTE SE DICE”.

El tema de la fiabilidad no parece interesarle mucho a la institución responsable del sistema de evaluación. Es cierto que ya en 1992, el Ministerio de Educación y Ciencia de España garantizaba “*la calidad de los exámenes, así como la fiabilidad de los resultados*” (MEC, 1992:5) respaldándose en la experiencia que la Universidad de Salamanca había adquirido en la organización de cursos de ELE. También es cierto que en sucesivas ediciones de publicaciones oficiales se incluyen repetitivamente afirmaciones como las siguientes:

“El sistema de examen de los Diplomas de Español ofrece un servicio de

evaluación independiente, *altamente estable* en el que se combina la utilización de las más modernas tecnologías en la calificación automática y análisis de las pruebas objetivas con la experiencia de más de 2000 *especialistas* en todo el mundo para las pruebas no objetivas”

“La homogeneidad de las condiciones de celebración de las convocatorias de examen, las medidas de seguridad que se adoptan y los análisis de rendimiento permiten al Instituto Cervantes *garantizar la consistencia* de las pruebas convocatoria a convocatoria, centro a centro”.

(Instituto Cervantes, 2007: 5)

Sin embargo, estas aseveraciones nunca se han avalado con evidencias empíricas, es decir, con referencias o reportes de estudios de fiabilidad de los resultados de las diferentes pruebas que integran el sistema. ¿Cuál es el grado de fiabilidad de los resultados de cada una de las tres pruebas que se califican por corrección automática, es decir, las pruebas de *gramática y vocabulario, comprensión auditiva, y comprensión lectora*? ¿Cuál es el grado de fiabilidad de los calificadoros de las *pruebas de expresión oral y expresión escrita*? A pesar de que se afirma que, “El Instituto Cervantes, como miembro de ALTE, suscribe el código deontológico y los principios de buena práctica de esta asociación” (Instituto Cervantes, 2009: 2), es obvio que en el aspecto de fiabilidad de los resultados de los DELE la institución no está cumpliendo con los estándares mínimos que se han mencionado al principio.

**3 LO QUE OCURRE EN LA REALIDAD: LA FIABILIDAD DE LOS RESULTADOS DE LOS DELE**  
Desafortunadamente, el aspecto de la fiabilidad de los DELE tampoco ha interesado a los investigadores de nuestro campo de estudios. De hecho, los únicos trabajos empíricos sobre la fiabilidad de los resultados de los DELE que se han publicado hasta la fecha son las modestas contribuciones de este autor (Cárdenas, 2006, 2007, 2008a, 2008b).

En cuanto a la fiabilidad de los resultados de las pruebas objetivas, como se puede ver en la tabla No. 1, en una investigación empírica con estudiantes universitarios japoneses de ELE realizada durante los años 2004 y 2005, el autor encontró que las 3 pruebas objetivas de nivel intermedio y dos de las pruebas del nivel inicial presentaban un indicador de fiabilidad que variaba de 0.47 a 0.79 según el coeficiente alfa de Cronbach. Es conveniente recordar aquí que el coeficiente alfa de Cronbach, considerado por los expertos en evaluación como indicador de fiabilidad o consistencia de una prueba, varía de 0 a 1, y que el mínimo aceptable es 0.80 o incluso superior a 0.90 cuando se trata de exámenes de alto impacto de uso internacional (McNamara, 2000; Weir, 2005). Los bajos coeficientes de cuatro de las pruebas, específicamente las de comprensión auditiva

y comprensión de lectura tanto en el nivel inicial como en el nivel intermedio, por lo tanto, indican que los resultados se ven influenciados negativamente por factores ajenos a los conocimientos o habilidades de los candidatos en un gran porcentaje. Lo de la “garantía de fiabilidad”, “altamente fiable”, “altamente estable”, quedó cuestionado con los resultados de este primer estudio empírico.

	DELE INICIAL			DELE INTERMEDIO		
	Int. de Textos Escritos	Int. de Textos Orales	Conciencia Comunicativa	Comp. Lectora	Comp. Auditiva	Gramática y Vocabulario
Número de Estudiantes	104			121		
Número de Ítems	20	23	30	12	12	60
Puntuación Media	13.63	16.45	23.21	7.46	5.23	33.85
Desviación Típica	2.63	2.74	4.72	2.15	2.24	7.55
Coeficiente Alfa (Fiabilidad)	0.52	0.49	0.82	0.47	0.49	0.79

Tabla No. 1 La fiabilidad de los resultados de las pruebas objetivas de los DELE (Cárdenas, 2006, 2007)

## II. EN BUSCA DE NUEVAS EVIDENCIAS

Es obvio que los resultados de un solo estudio no deben tomarse como definitivos. Unos años después, por lo tanto, decidimos continuar con la búsqueda de más evidencias empíricas que rectificaran o corroboraran los resultados del primer estudio.

### 1 OBJETIVO DEL ESTUDIO

En el segundo estudio, nos propusimos buscar respuestas a dos preguntas:

(1) ¿Qué nivel de fiabilidad tienen los resultados de las pruebas de interpretación de textos orales, interpretación de textos escritos, y conciencia comunicativa del DELE Inicial o B1 al aplicarse a un grupo de estudiantes japoneses?

(2) ¿Indican los resultados mejorías en el sistema DELE al compararse con el estudio previo?

## 2 PARTICIPANTES

Nuestro estudio contó con la participación de 121 estudiantes universitarios japoneses, la mayoría del sexo femenino, cuya edad fluctuaba entre los 19 y 23 años. En el momento de hacer la investigación los participantes estaban a punto de terminar el segundo o el tercer año de la licenciatura en lo que normalmente en Japón se conoce como Estudios Hispánicos y habían cursado aproximadamente entre 436 y 570 horas de asignaturas de lengua española.

## 3 PROCEDIMIENTO

A los participantes en el estudio se les administró una versión oficial del examen para obtener el Diploma de Español como Lengua Extranjera de nivel inicial (B1) del 2004.<sup>2</sup> Las pruebas se administraron a fines de noviembre del 2007, bajo la supervisión del autor y otros colegas de la institución donde se realizó la investigación.

Como sucede en una administración oficial, las hojas de respuestas de las tres pruebas de corrección objetiva fueron calificadas con el uso de medios tecnológicos. Una vez que se obtuvieron los resultados básicos, se realizó un análisis de las propiedades psicométricas de las pruebas con ayuda de los programas informáticos *TiaPlus* o Test and Item Analysis (Cito, 2005) y *TAP* o *Test Analysis Program* (Brooks, 2003, 2005).

## 4 RESULTADOS

En la tabla No. 2 se puede ver un resumen del análisis de los resultados que se obtuvieron de la administración de las diferentes pruebas: el número de estudiantes, número de ítems o reactivos de cada prueba, puntuación media, etc. Obsérvese que la prueba de conciencia comunicativa es la que resultó más fácil para los participantes, ya que obtuvieron una puntuación media del 23.61 que representa casi el 80% del total de reactivos. La puntuación media más baja, se obtuvo en la prueba de interpretación de textos orales, con un promedio de 14.76 (el 67.1% del total de reactivos).

<sup>2</sup> Se decidió utilizar una versión diferente a la del primer estudio porque se trataba de una versión más reciente y que, en teoría, pertenecía al nuevo ciclo de pruebas que se estaban elaborando como parte del movimiento de reforma (Veáse Fernández 2004, Prieto et al 2004).

Resumen Estadístico de Resultados DELE Inicial 2007			
	Interpretación de Textos Escritos	Interpretación de Textos Orales	Conciencia Comunicativa
Número de Estudiantes	121		
Número de Ítems	20	22	30
Puntuación Media	15.69	14.76	23.61
Desviación Típica	2.84	3.15	3.86
Coeficiente Alfa (fiabilidad)	0.64	0.61	0.73

Tabla No. 2 Resultados de un nuevo estudio sobre la fiabilidad de los resultados de las pruebas objetivas de los DELE de nivel inicial (B1)

Pasemos a la última fila, para contestar a la primera pregunta del estudio y ver lo que más nos interesa, es decir, el nivel de fiabilidad de los resultados. En esta fila tenemos el coeficiente alfa de Cronbach que, como hemos indicado, nos revela la consistencia de las pruebas y se considera como muestra de fiabilidad de los resultados. Nótese que de las tres pruebas, la de conciencia comunicativa obtiene la mejor indicación de fiabilidad (0.73), seguida de la prueba de interpretación de textos escritos (0.64), y la de interpretación de textos orales (0.61).

Comparemos ahora las dos tablas para contestar a la segunda pregunta de la investigación, es decir, si los resultados indican mejorías en el sistema DELE al compararse con el estudio previo. Como se puede notar, al comparar los resultados de los dos estudios, se podría decir que sí hay indicaciones de una ligera mejoría en la fiabilidad de los resultados en dos de las pruebas. Si comparamos los coeficientes de fiabilidad obtenidos en la prueba de interpretación de textos escritos de la administración 2004-2005 con la misma prueba en este segundo estudio, notamos que hay un incremento de 0.12 en el nivel de fiabilidad de los resultados, es decir, de 0.52 a 0.64. Curiosamente, el mismo tipo de incremento o mejoría, se observa en la prueba de interpretación de textos orales que cambia de 0.49 a 0.61. La prueba de conciencia comunicativa, sin embargo, nos presenta un panorama diferente, ya que en el segundo estudio, el indicador de fiabilidad bajó de 0.82 a 0.73, lo que significa una disminución en el grado de consistencia de los resultados.

Aunque se reconoce que los coeficientes de fiabilidad obtenidos tanto en la prueba de interpretación de textos escritos como en la de interpretación de textos orales en este segundo estudio, sí son indicación de una ligera mejoría en el

sistema de evaluación DELE, no debemos pasar por alto un aspecto de gran preocupación: los coeficientes de fiabilidad de las tres pruebas *siguen siendo bajos o muy bajos*. Ya se ha mencionado que los expertos en evaluación consideran que el coeficiente de fiabilidad mínimo aceptable debe ser de 0.80 e incluso superior a 0.90 cuando se trata de un sistema de evaluación de alto impacto. Los coeficientes de 0.61, 0.64, y 0.73 obtenidos para las tres pruebas en la segunda investigación, por lo tanto, se consideran inaceptables para un sistema de evaluación de uso internacional como es el caso de los DELE.

Antes de pasar al último apartado de este trabajo, dado que los coeficientes de fiabilidad que se obtuvieron no eran satisfactorios, se decidió dar un paso más y analizar una de las posibles causas del bajo nivel de fiabilidad de las pruebas del DELE Inicial. De acuerdo con la literatura (Alderson, 1995; Hughes, 2003; Bachman, 2004; Brown, 2005), uno de los factores primordiales que intervienen en la fiabilidad es el nivel de calidad de los ítems o preguntas que se incluyen. ALTE (2003, 2005) por ejemplo, recomienda que los ítems apropiados deben tener un coeficiente de dificultad de 0.20 a 0.80<sup>3</sup>, y un índice de discriminación de 0.30 o más<sup>4</sup>. Como se puede notar en las tablas No. 3 y 4, desafortunadamente, 10, o sea la mitad de los 20 ítems de la prueba de interpretación de textos escritos no cumplen con estos requisitos mínimos. La situación con la prueba de interpretación de textos orales es todavía peor, ya que 13 de las 22 preguntas no se consideran adecuadas para ser incluidas en la prueba.

Este análisis, por lo tanto, sugiere que el bajo nivel de fiabilidad de los resultados podría deberse al uso de ítems o reactivos de baja calidad.

### III. CONCLUSIONES Y SUGERENCIAS

En resumen, los datos presentados en este trabajo, indican que aunque sí ha habido una ligera mejoría en el nivel de fiabilidad de los resultados de las pruebas de interpretación de textos orales e interpretación de textos escritos del nivel inicial o B1 de los DELE, los coeficientes de fiabilidad siguen sin alcanzar los estándares mínimos de calidad establecidos por ALTE y otros organismos o especialistas en el mundo de la evaluación. Las instituciones involucradas en el sistema de evaluación DELE tienen, por lo tanto, mucho camino por recorrer para poder lograr y *demonstrar con evidencias empíricas* que sus pruebas arrojan resultados fiables, y por ende, válidos para los diferentes usuarios del mundo.

<sup>3</sup> El coeficiente de dificultad representa el porcentaje de candidatos que contestan correctamente una pregunta

<sup>4</sup> El índice de discriminación nos indica que bien distingue o discrimina un ítem entre los candidatos de distintos niveles de capacidad

No. de Item	Coefficiente de Dificultad	Índice de Discriminación	Correlación Biserial Puntual	Ítems Problemáticos según criterios de ALTE (2003. 2005)
Item 01	0.70	0.45	0.40	
Item 02	0.55	0.55	0.49	
Item 03	0.92	0.15	0.20	X
Item 04	0.83	0.41	0.42	X
Item 05	0.79	0.38	0.40	
Item 06	0.60	0.39	0.32	
Item 07	0.69	0.61	0.52	
Item 08	0.77	0.46	0.40	
Item 09	0.75	0.46	0.50	
Item 10	0.68	0.32	0.32	
Item 11	0.94	0.13	0.25	X
Item 12	0.61	0.53	0.41	
Item 13	0.91	0.10	0.14	X
Item 14	0.94	0.13	0.27	X
Item 15	0.84	0.25	0.37	X
Item 16	0.70	0.48	0.43	
Item 17	0.80	0.30	0.37	X
Item 18	0.91	0.18	0.32	X
Item 19	0.89	0.12	0.22	X
Item 20	0.86	0.28	0.33	X

Tabla No. 3 Análisis de ítems de la prueba de interpretación de textos escritos

Para mejorar la fiabilidad de los resultados de las 3 pruebas de tipo objetivo, el autor sugiere las siguientes recomendaciones:

- 1ª. Realizar estudios periódicos de fiabilidad con candidatos de diferentes nacionalidades
- 2ª. Revisar la calidad de ítems y pruebas (coeficiente de dificultad, índice de discriminación, coeficiente de fiabilidad, etc.)
- 3ª. Modificar o rechazar ítems o pruebas que no reúnan los estándares mínimos
- 4ª. Proporcionar datos a los interesados en realizar investigaciones externas
- 5ª. Actuar con más transparencia y responsabilidad.

No. de Item	Coefficiente de Dificultad	Indice de Discriminación	Correlación Biserial Puntual	Ítems Problemáticos según criterios de ALTE ( 2003. 2005)
Item 01	0.38	0.13	0.12	X
Item 02	0.45	0.38	0.34	
Item 03	0.88	0.16	0.30	X
Item 04	0.93	0.14	0.30	X
Item 05	0.37	0.44	0.26	X
Item 06	0.96	0.10	0.17	X
Item 07	0.60	0.54	0.42	
Item 08	0.62	0.48	0.45	
Item 09	0.88	0.26	0.42	X
Item 10	0.51	0.46	0.45	
Item 11	0.60	0.68	0.55	
Item 12	0.90	0.19	0.39	X
Item 13	0.61	0.21	0.26	X
Item 14	0.75	0.22	0.22	X
Item 15	0.93	0.19	0.39	X
Item 16	0.70	0.42	0.41	
Item 17	0.65	0.38	0.30	X
Item 18	0.60	0.20	0.18	X

Item 19	0.73	0.41	0.43	
Item 20	0.64	0.21	0.20	X
Item 21	0.49	0.46	0.37	
Item 22	0.56	0.50	0.46	

Tabla No. 4 Análisis de ítems de la prueba de interpretación de textos orales

## BIBLIOGRAFIA

- ALDERSON, J. Charles, Caroline CLAPHAM y Diane Wall (1995): *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- ASSOCIATION OF LANGUAGE TESTERS IN EUROPE (2005): Materials for the guidance of test item writers [en línea] <http://www.alte.org/downloads/index.php?docid=89>
- BACHMAN, Lyle F. y Adrian S, PALMER (2004): *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- BORDON, María Teresa (2006): *La evaluación de la lengua en el marco de E/L2: Bases y procedimientos*, Madrid: Editorial Arco/Libros, S. L.
- BROOKS, Gordon P. (2003-2005): TAP, Test Analysis Program.
- BROWN, James D. (2005): *Testing in language programs: A comprehensive guide to English language assessment*, New York: McGraw-Hill ESL/ELT.
- BROWN, James D, y Thom HUDSON (2002): *Criterion-referenced language testing*. Cambridge: Cambridge University Press
- CARDENAS, Abel (2001): «La evaluación oral en los exámenes del Diploma de Español como Lengua Extranjera: una guía práctica para profesores y estudiantes», *Academia Language and Literature*, vol. 69, 25-67.
- CARDENAS, Abel (2005): «La evaluación de la proficiencia en español como lengua extranjera: introducción y contextos», *Academia Language and Literature*, vol. 77, 191-208.
- CARDENAS, Abel (2006): «La fiabilidad de los DELE y los candidatos japoneses: ¿un motivo de preocupación?», *Cuadernos CANELA*, vol. XVIII, 63-76.
- CARDENAS, Abel. (2007). «La fiabilidad de la prueba de comprensión auditiva en los DELE: de la "historia oficial" a la realidad», *Actas del XVII Congreso Internacional ASELE*: 395-408.

- CARDENAS, Abel. (2008a): «La formación y certificación del profesor de español como examinador oficial en los DELE: lo que falta por hacer». Ponencia presentada en el XIX Congreso Internacional de ASELE
- CARDENAS, Abel. (2008b). «No todo lo que brilla es oro" La fiabilidad de los resultados de la prueba oral en los DELE», Cuadernos CANELA XX:91-100.
- CITO, M. R. D. (2005). TiaPlus. Amhem: NL.
- EGUILUZ, Juan y Clara M. VEGA SANTOS (1996): «Criterios para la evaluación de la producción escrita», *Actas de Expolingua 1994-1995 (Cuadernos del tiempo libre)*, vol. 3, 75-94.
- EGUILUZ, Juan y Angel EGUILUZ (2004): «"La evaluación de la comprensión lectora», *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2)/lengua extranjera (LE)*, Madrid: SGEL, 1025-1041.
- EGUILUZ, Juan y Angel EGUILUZ (2004): «"La evaluación de la expresión escrita», *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2)/lengua extranjera (LE)*, Madrid, SGEL, 1005-1024.
- FERNANDEZ GONZALES, J. (2004): «"Nuevos modelos de diplomas de español como lengua extranjera"», *III Congreso Internacional de la Lengua*
- FULCHER, Glenn (2003): *Testing second language speaking*, Harlow: Pearson Education Limited.
- HUGHES, Arthur (2003): *Testing for language teachers*, Cambridge: Cambridge University Press.
- JONES, Neil (2001): «Reliability as one aspect of test quality», *ResearchNotes* vol. 4, 2-5.
- INSTITUTO CERVANTES (2005): Descripción general del sistema de evaluación DELE [en línea] [http://diplomas.cervantes.es/docs/ficheros/200602080001\\_7\\_11.pdf](http://diplomas.cervantes.es/docs/ficheros/200602080001_7_11.pdf)
- INSTITUTO CERVANTES (2007): *Descripción general del sistema de evaluación DELE*
- INSTITUTO CERVANTES (2009): *Manual para los examinadores. Nivel inicial, nivel intermedio, nivel superior*, Instituto Cervantes, Departamento de Certificación Lingüística, Dirección Académica.
- McNAMARA, Tim (1996): *Measuring second langage performance*, Harlow: Longman.
- McNAMARA, Tim (2000): *Language Testing*, Oxford: Oxford University Press.
- MINISTERIO DE EDUCACION Y CIENCIA (1992): *Guía para la obtención de los diplomas de español como lengua extranjera (D.E.L.E.)*
- MINISTERIO DE EDUCACION Y CIENCIA, IINSTITUTO CERVANTES Y UNIVERSIDAD DE SALAMANCA (1995): *Exámenes para la obtención de los Diplomas de Español como Lengua Extranjera del Ministerio de Educación y*

- Ciencia (D.E.L.E.). Instrucciones para examinadores orales.*
- MINISTERIO DE EDUCACION Y CIENCIA, INSTITUTO CERVANTES Y UNIVERSIDAD DE SALAMANCA (1995): *Examen para la obtención del Certificado Inicial de Español como Lengua Extranjera del Ministerio de Educación y Ciencia (D.E.L.E.). Instrucciones para examinadores orales.*
- PARRONDO RODRIGUEZ, José Ramón (2004a): «Aspectos éticos de la evaluación: impacto de la actividad evaluadora y códigos deontológicos de los examinadores», *Carabela*, vol. 55, 31-44.
- PARRONDO RODRIGUEZ, José Ramón (2004b): «Modelos, tipos y escalas de evaluación», *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2)/lengua extranjera (LE)*, Madrid: SGEL: 967-981.
- PISONERO DEL AMO, Isidoro y Alvaro GARCIA-SANTA CECILIA (1991): «Diplomas de español como lengua extranjera», *CABLE: revista de español como lengua extranjera*, vol. 7, 3-4.
- PRIETO HERNANDEZ, J. Miguel, Carmen DIEZ SANTOS, Charo DOMINGUEZ LOPEZ y Escolástico MARTIN MAESTRE (2004): «La elaboración de una prueba de nivel: la reforma de los DELE», *Carabela*, vol. 55, 85-140.
- SAVILLE, Nick (2003): «The process of test development and revision within UCLES-EFL», *Continuity and innovation: revising the Cambridge Proficiency in English examination 1913-2002*, Cambridge: Cambridge University Press, 57-116.
- SAVILLE, Nick. (2005): «Setting and monitoring professional standards: a QMS approach», *ResearchNotes*, vol. 22, 2-5.
- WEIR, Cyril. J. (2005): *Language testing and validation : an evidence-based approach*, Basingstoke: Palgrave Macmillan.

## APENDICE

### El Código de Práctica y los Estándares Mínimos de ALTE

#### Estándares mínimos para el establecimiento de perfiles de calidad en los exámenes de ALTE

##### CONFECCIÓN DE EXÁMENES

1. El examen está basado en un constructo teórico, por ejemplo un modelo de competencia comunicativa.
2. Puede describirse la finalidad y el contexto de uso del examen, así como la población para la que es adecuado.
3. Se facilitan criterios para la selección y formación de los creadores de las pruebas y se emplean juicios expertos tanto para la confección de exámenes y para su revisión.
4. Los exámenes son comparables en paralelo a través de las diferentes administraciones de los mismos en términos de contenidos, estabilidad, consistencia y puntos de corte.
5. Si se dice que el examen está vinculado a un sistema de referencia externo, por ejemplo el MCER, existen pruebas fehacientes de su alineación a este sistema.

##### ADMINISTRACIÓN Y LOGÍSTICA

6. Todos los centros que administran las pruebas son seleccionados de acuerdo con procedimientos establecidos claros y transparentes, y todos ellos tienen acceso a la normativa que regula estos procesos.
7. Los materiales de examen son remitidos en buenas condiciones y por medios de transporte seguros a los centros de examen, el sistema administrativo de exámenes facilita la seguridad y trazabilidad de todos los envíos de documentos y puede garantizarse la confidencialidad de todos los procedimientos del sistema.
8. El sistema de administración de exámenes cuenta con sistemas adecuados de apoyo, por ejemplo, línea de teléfono de incidencias, página web, etc.
9. Se protege de manera adecuada la seguridad y la confidencialidad de los resultados y de los certificados, así como de los datos correspondientes, en cumplimiento de la legislación vigente sobre protección de datos, y los candidatos cuentan con información sobre el modo de acceder a sus datos.
10. El sistema de exámenes facilita el acceso a los mismos a aquellos candidatos con necesidades especiales.

##### PUNTUACIÓN Y CALIFICACIÓN

11. La calificación es lo suficientemente precisa y fiable para la finalidad y el tipo de examen del que se trate.
12. Puede documentarse y explicarse cómo se lleva a cabo el proceso de calificación, cómo se

calcula la fiabilidad y cómo se recogen y analizan los datos correspondientes al rendimiento de los calificadores de las pruebas de expresión oral y escrita.

### **ANÁLISIS DE EXÁMENES**

13. Se recogen y analizan datos de una muestra adecuada y representativa de los candidatos y así se puede confiar en que su desempeño es resultado de las destrezas que se evalúan en el examen y no se encuentra influido por factores como lengua materna, país de origen, género, edad o extracción étnica.

14. Se recogen y analizan datos a nivel de ítem (por ejemplo para calcular la dificultad, discriminación, fiabilidad y errores estándar de medida del examen) de una muestra adecuada y representativa de los candidatos

### **COMUNICACIÓN CON LOS PARTICIPES**

15. El sistema de administración de exámenes comunica los resultados de los mismos a los candidatos y a los centros de examen (por ejemplo, escuelas) con prontitud y claridad.

16. Se facilita información a los partícipes sobre los contextos adecuados, la finalidad y el uso del examen, sobre sus contenidos y sobre la fiabilidad general de los resultados del mismo.

17. Se facilita información adecuada a los partícipes para ayudarles a interpretar los resultados y a utilizarlos de modo procedente.

© Association of Language Testers Code of Practice Working Group  
Translated by Instituto Cervantes

Fuente: <http://www.alte.org/downloads/index.php>

Nota: El sombreado es del autor de este trabajo