



Artículo en prensa

SEXTA EVALUACIÓN DE TESTS EDITADOS EN ESPAÑA: RESULTADOS E IMPACTO DEL MODELO EN DOCENTES Y EDITORIALES

SIXTH TEST REVIEW OF TESTS PUBLISHED IN SPAIN: RESULTS AND IMPACT OF THE MODEL ON LECTURERS AND PUBLISHERS

María Dolores Hidalgo¹ y Ana Hernández²

¹Universidad de Murcia. ²Universitat de València

El cuestionario para la Evaluación de los Tests (CET; Prieto y Muñiz, 2000) y su revisión (CET-R; Hernández et al., 2016) se han venido aplicando sistemáticamente desde 2010, impulsado por la Comisión de tests del Consejo General de la Psicología del Colegio Oficial de Psicólogos. El objetivo es proporcionar información contrastada sobre la calidad de las pruebas a los profesionales, con el fin de mejorar el uso de los tests. El presente trabajo tiene un doble objetivo. El primero, presentar los resultados de la sexta evaluación de tests psicológicos y educativos, en la que se han revisado un total de 10 tests. El segundo, evaluar el impacto que la aplicación del CET/CET-R ha tenido durante estos años en dos agentes cruciales: las editoriales de tests, y los profesores universitarios encargados de formar a los futuros profesionales usuarios de tests. Los resultados de la sexta evaluación, así como los resultados de la encuesta para evaluar el impacto del CET/CET-R, se pueden considerar en general satisfactorios. Sin embargo, se identifican varios aspectos que son susceptibles de mejora.

Palabras clave: Tests, Uso de los tests, Evaluación de tests, Calidad de los tests, Impacto modelo evaluación de tests.

The Questionnaire for the Assessment of Tests (CET; Prieto & Muñiz, 2000) and the revised version of this questionnaire (CET-R; Hernández et al., 2016) have been applied systematically since 2010 by the Test Commission of the Spanish Psychological Association. The main goal is to provide practitioners with reliable information on the quality of the tests in order to improve test use. The aim of this paper is twofold. First, to present the results of the sixth review of psychological and educational tests, in which a total of 10 tests have been evaluated. Second, to assess the impact that the application of CET/CET-R has had over these years on two key agents: test publishers and university lecturers who are responsible for training future test users. Both the results of the sixth review and the results of the survey to assess the impact of CET/CET-R are satisfactory in general terms. However, some issues where there is room for improvement have been identified.

Key words: Tests, Use of tests, Test Review, Test quality, Impact test review model.

En los diferentes ámbitos de la Psicología, la medida y evaluación mediante tests suele ser una práctica habitual. Los resultados obtenidos de estas mediciones se utilizan con diferentes propósitos como para diagnosticar, tomar decisiones, evaluación, consejo o selección; pero, tal y como señalan los Standards de la APA (AERA, APA, y NCME, 2014) no todos los tests están bien desarrollados ni se utilizan del modo adecuado. Disponer de información sobre la calidad de un test y del uso adecuado de sus puntuaciones es clave en el desempeño profesional de la psicología independientemente del ámbito disciplinar que nos ocupe (clínica, educativo, organizacional,...). Desde hace unas décadas diferentes organizaciones profesionales de la Psicología así como asociaciones e instituciones han desarrollado y promovido el uso de guías o cuestionarios para evaluar la calidad de los tests psicológicos y

educativos (Evers, 2012; Geisinger, 2012; Hernández, Ponso da, Muñiz, Prieto y Elosua, 2016), con el objetivo de mejorar la calidad y el rigor de los tests. Más concretamente, y a partir del año 2010, la Comisión de Tests del Consejo General de la Psicología del Colegio Oficial de Psicólogos ha abordado la tarea de evaluación de los tests editados en España, con la finalidad de proporcionar información precisa y accesible acerca de la calidad de los tests en términos de sus características teóricas, prácticas y psicométricas, de modo que la misma pueda ser de utilidad a los profesionales en la toma de decisiones de uso de los tests de manera informada (Hernández, et al, 2016). Prieto y Muñiz (2000) propusieron un modelo de evaluación de tests que se materializó a través del Cuestionario CET (Cuestionario de Evaluación de Tests) y en 2016 este instrumento fue revisado por Hernández et al. (2016), dando lugar al CET-R. Durante este periodo de tiempo se han realizado seis evaluaciones de tests editados en España, incluyendo la referida a los resultados que se presentan en este artículo. En total suman 75 tests evaluados, cuyos informes se encuentran disponibles y se pueden descargar desde la web Colegio Ofi-

Recibido: 5 octubre 2018 - Aceptado: 9 noviembre 2018

Correspondencia: María Dolores Hidalgo. Departamento de Psicología Básica y Metodología. Universidad de Murcia. Campus de Espinardo. 30100 Murcia. España. E-mail: mdhidalg@um.es



Artículo en prensa

cial de Psicólogos (COP) en la entrada correspondiente a la Comisión Nacional de Tests del Consejo General de la Psicología (<https://www.cop.es/index.php?page=evaluacion-tests-editados-en-espana>). Un resumen de los resultados de la primera evaluación se puede consultar en Muñiz et al. (2011), de la segunda en Ponsoda y Hontangas (2013), de la tercera en Hernández, Tomás, Ferreres y Lloret (2015), de la cuarta en Elosua y Geisenger (2016) y de la quinta evaluación en Fonseca-Pedrero y Muñiz (2017), todos ellos publicados en la revista *Papeles del Psicólogo*. En especial, en la última revisión (anterior a la que aquí se presenta) se recoge la evolución histórica del proceso de evaluación de tests editados en España (Fonseca-Pedrero y Muñiz, 2017).

Todas estas evaluaciones han sido promovida por el Consejo General de la Psicología del COP como estrategia *informativa* (Muñiz, Hernández y Ponsoda, 2015), a fin de proporcionar a los usuarios de tests, información precisa y accesible sobre los mismos, y dar respuesta a una de las principales demandas de los profesionales de la psicología: disponer de información técnica y psicométrica que les ayude a la toma de decisiones basadas en evidencias (Elosua, 2012; Muñiz y Fernández-Hermida, 2010). Realizadas ya cinco rondas de evaluación, seis si consideramos la presentada en este trabajo, resulta de interés conocer el impacto que el modelo está teniendo en los diferentes ámbitos y profesionales de la Psicología. En este sentido, y aunque en el trabajo de Fonseca-Pedrero y Muñiz (2017) se recogen datos en términos del número de revisores y coordinadores que han participado en el proceso de revisión, así como del número de descargas de los informes disponibles en la página web del COP, se ha querido preguntar de forma directa sobre el impacto y uso de estos informes de calidad.

Teniendo en cuenta las consideraciones anteriores, los objetivos de este artículo son principalmente dos: 1) presentar los re-

sultados de la sexta evaluación de tests editados en España, y 2) analizar el impacto que ha tenido el modelo de evaluación de tests (CET) y su revisión CET-R en académicos (profesores universitarios que imparten Psicometría y Evaluación Psicológica) y en las editoriales de tests. Queda pendiente conocer el impacto por parte de los profesionales aplicados más allá de los aspectos ya considerados por Fonseca-Pedrero y Muñiz (2017), trabajo que será abordado en un estudio de mayor alcance en el que se recogerá también la opinión sobre los tests en general (en línea con los estudios realizados por Muñiz y Fernández-Hermida, 2000, 2010).

Así, en primer lugar, tras resumir el proceso de revisión seguido, se presentarán los resultados de la sexta evaluación de tests. En concreto se ha revisado un total de diez tests: cinco para la medida de aptitudes mentales y cognitivas e inteligencia, uno de evaluación de la depresión, uno para la evaluación de problemas en adolescentes (ansiedad, depresión, autoestima, riesgo psicosocial,...), uno de resolución de problemas matemáticos, cálculo y numeración, uno de procesamiento sensorial y otro para la evaluación de actitudes y comportamientos en pacientes médicos. En concreto, los instrumentos revisados han sido los siguientes: BADyG/E2-r, BADyG/S-r, CESPPO, BDI-FastScreen, MBMD, SENSORY PROFILE, Q-PAD, BAT-7, BPR y MATRICES (ver Tabla 1 para una descripción más detallada). En segundo lugar, se presentarán los resultados del impacto que tanto las evaluaciones realizadas hasta la fecha, como el propio cuestionario CET-R, está teniendo entre académicos de la Psicología. Por último, se presentan datos referidos al impacto que este modelo de evaluación está teniendo en las editoriales de tests.

SEXTA EVALUACIÓN DE TESTS EDITADOS EN ESPAÑA

Descripción del cuestionario

El cuestionario CET-R permite una valoración cualitativa y

TABLA 1
RELACIÓN DE INSTRUMENTOS DE MEDIDA ANALIZADOS EN LA SEXTA EVALUACIÓN DE TESTS

Acrónimo	Test	Editorial	Año Publicación
BADYG/E2-r	Batería de Aptitudes Diferenciales y Generales Renovado E2	CEPE, S.L.	2011
BADYG/S-r	Batería de Aptitudes Diferenciales y Generales	CEPE, S.L.	2011
BAT-7	Batería de Aptitudes de TEA	TEA ediciones	2015
BDI-FastScreen	Inventario de Depresión de Beck FastScreen para pacientes Médicos	PEARSON Education	2011
BPR	Batería de Pruebas de Razonamiento	TEA ediciones	2016
CESPRO	Batería para la Evaluación de las Estructuras Sintáctico-Semánticas que componen los enunciados de los problemas matemáticos y de la utilización de estrategias algorítmicas para su resolución	EOS	2016
MATRICES	MATRICES Test de Inteligencia General	TEA ediciones	2015
MBMD	Inventario Conductual de Millon para pacientes con diagnóstico médico	PEARSON Education	2014
Q-PAD	Cuestionario para la evaluación de problemas en adolescentes	TEA ediciones	2016
SENSORY PROFILE	Perfil Sensorial-2	PEARSON Education	2016

Artículo en prensa

cuantitativa de la calidad de los tests, y tiene como principal objetivo proporcionar a los usuarios de tests información precisa y accesible acerca de la calidad de los mismos.

El cuestionario se puede descargar desde la siguiente dirección web <http://www.cop.es/uploads/pdf/CET-R.pdf>. Además, junto al cuestionario propiamente dicho, se proporcionan unas instrucciones generales sobre la forma de responder el test y cumplimentarlo, así como un glosario de términos psicométricos.

Aunque las principales características del CET (Prieto y Muñoz, 2000) y CET-R (Hernández et al., 2016) se pueden consultar en los trabajos referenciados y en las diferentes publicaciones de las revisiones de tests llevadas a cabo con este modelo, vamos a presentar de manera resumida los diferentes aspectos de calidad psicométrica que se consideran. El CET-R incluye tres apartados de valoración del test diferenciados: 1) Descripción general del test, 2) Valoración de las características del test y 3) Valoración global del test. El primer apartado proporciona una breve descripción sobre la(s) variable(s) medidas, el área de aplicación, el número de ítems y/o escalas, soporte para la administración y corrección, la cualificación requerida para el uso del test de acuerdo con la documentación aportada, el tiempo de administración, e incluso sobre el precio de la prueba. En cuanto al segundo apartado se valoran tanto características teóricas del instrumento (fundamentación teórica, calidad de los materiales y documentación, bibliografía, desarrollo de los ítems), como aspectos más psicométricos como el análisis de ítems, las evidencias de validez, fiabilidad y la baremación e interpretación de puntuaciones. El último apartado (valoración global del test) requiere de una evaluación cualitativa del instrumento con especial énfasis en sus fortalezas y posibles debilidades.

Proceso de Evaluación

El procedimiento seguido en esta sexta evaluación es similar al establecido en las anteriores evaluaciones. En primer lugar las editoriales seleccionaron las pruebas a revisar, y la Comisión de Tests hizo suya dicha propuesta. En esta ocasión fueron seleccionadas 10 pruebas (una de la editorial EOS, dos de CEPE, tres de Pearson y cuatro de TEA). Además, en la Comisión, se planteó evaluar otras pruebas diferentes de las propuestas por las editoriales y que se usan en los colegios. De este modo, se propuso evaluar el test HABILMIND que evalúa habilidades de aprendizaje, razonamiento, lectura, etc. En segundo lugar, la coordinadora designada por la comisión de tests (primera autora del presente artículo), seleccionó un panel de expertos para el proceso de revisión, dos por test. Un experto con un perfil más técnico-psicométrico y otro con un perfil más teórico, con conocimientos y experiencia en los aspectos substantivos de la(s) variable(s) medida(s) por el test. Los aspectos éticos del proceso se tuvieron en cuenta. Así, se cuidó que los revisores no tuviesen relación directa con los autores de los tests, y que no tuviesen ningún conflicto de intereses. En cuatro casos, los revisores inicialmente seleccionados

rechazaron participar en el proceso por distintas razones justificadas, por lo que se tuvo que seleccionar a otro revisor. En la Tabla 2 se muestra el listado final de revisores que finalmente accedieron a colaborar como evaluadores expertos. Cabe destacar y apreciar la labor realizada y su participación en el proceso.

Las editoriales pusieron a disposición del Consejo General de la Psicología del COP, de manera gratuita, tres ejemplares completos de cada test. Desde el propio Consejo General de la Psicología se envió a cada pareja de revisores un ejemplar y el tercero se envió a la coordinadora. La tarea de los expertos, una vez recibido el ejemplar completo del test, consistía en aplicar el CET-R a la prueba asignada valorando cada uno de los aspectos teóricos, prácticos y psicométricos. Los revisores recibieron una cantidad simbólica de 50 euros (que algunos prefirieron declinar) además de un ejemplar gratuito del test.

Para cada test se elaboró un informe en el que las evaluaciones realizadas por los correspondientes revisores fueron integradas. Cuando había discrepancias entre los revisores, la coordinadora realizó una evaluación independiente a partir de los ejemplares proporcionados. A partir de toda esta información se asignaron las valoraciones cuantitativas y se realizó la valoración final. Siguiendo el procedimiento establecido en anteriores evaluaciones, el informe realizado de cada test fue remitido a las editoriales, a fin de que tanto los editores como los autores de los tests pudieran precisar, aclarar o comentar los aspectos recogidos en el informe de evaluación.

TABLA 2
REVISORES QUE HAN PARTICIPADO EN LA
SEXTA EVALUACIÓN DE TESTS

Nombre	Filiación
Maitte Barrios Cerrejón	Universidad de Barcelona
Isabel Benítez Baena	Universidad de Loyola
Marcelino Cuesta Izquierdo	Universidad de Oviedo
Beatriz Delgado Domenech	Universidad de Alicante
José Pedro Espada Sánchez	Universidad Miguel Hernández
Eduardo García Cueto	Universidad de Oviedo
José Manuel García Fernández	Universidad de Alicante
Arantxa Gorostiaga Manterola	Universidad del País Vasco
Georgina Guílera Ferré	Universidad de Barcelona
Francisco Pablo Holgado Tello	UNED
Pedro M. Hontangas Beltrán	Universidad de Valencia
Urbano Lorenzo Seva	Universitat Rovira i Virgili
Luis Manuel Lozano Fernández	Universidad de Granada
Laura Nuño Gómez	Hospital Clínic de Barcelona
José Luis Padilla García	Universidad de Granada
Óscar Pino López	Hospital General de Granollers
Antonio J. Rojas Tejada	Universidad de Almería
Inés Tomás	Universidad de Valencia
María Soledad Torregrosa Díez	Universidad Católica San Antonio de Murcia
Ana Vanesa Valero García	Universidad de La Rioja
Carme Viladrich	Universidad Autónoma de Barcelona

Artículo en prensa

Una vez recibidos los comentarios y aclaraciones por parte de las editoriales éstos fueron integrados en el informe final, modificando las evaluaciones cuando fueron justificadas. Por último, y previo a su publicación en la página web del COP, se envió de nuevo a las editoriales la propuesta de informe final. En la figura 1 se muestra un resumen del procedimiento de revisión por pares implementado.

Antes de pasar a los resultados de la evaluación, cabe señalar que el test HABILMIND no pudo ser finalmente revisado dado que los editores del test no pusieron a disposición del COP, tras varias solicitudes, copia gratuita del ejemplar del test.

Resultados

Los informes correspondientes a cada test que ha sido revisado en esta ronda de evaluación, tal y como hemos comentado, se pueden consultar y descargar en el siguiente enlace <https://www.cop.es/index.php?page=evaluacion-tests-editados-en-espana>. En este trabajo se presenta un resumen de los principales resultados obtenidos (ver tabla 3).

Se puede observar que las valoraciones fueron buenas, ya que en la mayoría de los ítems del cuestionario los tests fueron valorados con puntuaciones de 3,5 o superior (calificación de buena a excelente). Del total de tests y criterios evaluados solo en un test y para dos criterios encontramos puntuaciones inferiores a 3. Los aspectos mejor valorados y puntuados de los tests se refieren a los criterios de *Materiales y documentación y Fundamentación teórica*. El siguiente criterio mejor valorado fue el referido al análisis de la *fiabilidad como consistencia interna* y el referido a la calidad de los *Baremos e interpretación de puntuaciones*. Los tests evaluados proporcionaron diferentes evidencias de validez (contenido, interna y relación con otras variables), siendo valoradas entre buenas y excelentes. De los tests que fueron versiones adaptadas (la mitad de los evaluados) la valoración promedio del proceso de adaptación fue de 4,2 (entre buena y excelente). En general, el proceso de adap-

tación de estos tests se llevó a cabo siguiendo las Directrices de la Comisión Internacional de Tests (Hambleton, Merenda y Spielberg, 2005; International Test Commission, 2018; Muñiz, Elosua y Hambleton, 2013).

En comparación a las evaluaciones previas encontramos que en esta sexta evaluación un 30% de los tests aportan evidencias de fiabilidad en términos de la Teoría de Respuesta al Ítem, proporcionando información sobre la precisión del test en función del nivel de habilidad del sujeto. Además, en el 20% de los tests encontramos que se aportan evidencias de validez en términos de estudios de Funcionamiento Diferencial de los Ítems (DIF). Estos porcentajes son superiores a las evaluaciones previas, lo que sugiere que estas cuestiones son cada vez más tenidas en cuenta a la hora de publicar un test.

Por último, es interesante mencionar que la mayoría de los tests que se han revisado dispone de corrección automatizada por ordenador (90%) y que más de la mitad (60%) se pueden administrar de manera informatizada.

IMPACTO DEL CET EN PROFESORES Y EDITORIALES DE TESTS

Encuesta a los profesores de psicometría y evaluación psicológica y (psico)educativa

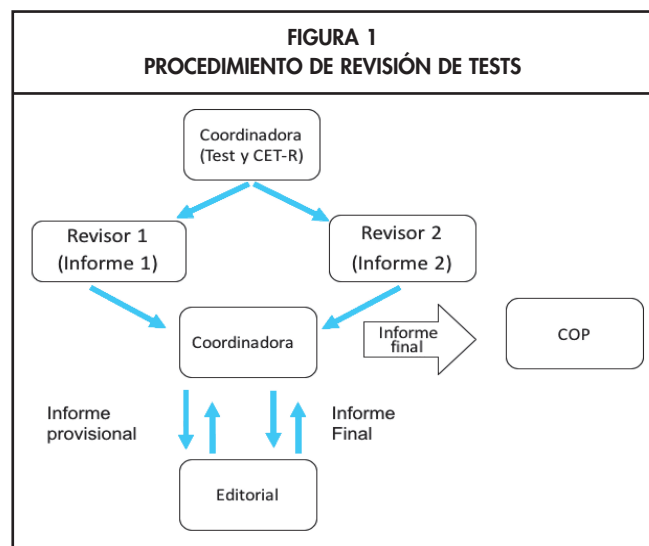
Para determinar el impacto que el modelo CET y su revisión (CET-R) estaban teniendo en la formación de los futuros profesionales de la psicología en materia de tests y evaluación mediante tests, se encuestó a los profesores de psicometría y de evaluación psicológica y (psico)educativa de los grados de psicología de las universidades españolas en el curso 2017-2018. El objetivo principal de esta encuesta era obtener información sobre si los docentes de los grados de psicología que imparten asignaturas de psicometría, evaluación psicológica y evaluación educativa o psicoeducativa, conocían el modelo CET, si lo utilizaban en sus clases y, de ser así, de qué forma. También resultaba de interés conocer las razones para no utilizarlo como herramienta de docencia, en caso de que así fuera.

Procedimiento y Participantes

A través de las páginas web de los grados de Psicología de las universidades públicas y privadas españolas, se identificaron un total de 316 profesores que fueron invitados a participar en el estudio mediante email. En el momento de la encuesta, el total de Universidades que impartían el título de Grado en Psicología eran 56, siendo el 53,57% universidades públicas (30) y el 46,43% privadas (26). Se garantizó el anonimato y el tratamiento confidencial de sus respuestas.

Tras enviar varios recordatorios, un total de 97 profesores (30,4%) respondieron a la encuesta. El 73% pertenecían a universidades públicas, la mitad aproximadamente (50,5%) eran varones y la antigüedad docente promedio era de 17,5 años (DT=11,1). Tras descartar las respuestas de dos participantes que indicaron que no impartían ninguna de las asignaturas de interés (a pesar de la información publicada en las páginas web de sus universidades), la mayoría (49,5%) eran profesores de Psicometría, seguidos del 35,8% que eran profesores de

**FIGURA 1
PROCEDIMIENTO DE REVISIÓN DE TESTS**





Artículo en prensa

Evaluación Psicológica. Solo dos participantes (2,1%) indicaron ser profesores de evaluación educativa o psicoeducativa. El resto (12,6%), indicaron que impartían varias de las asignaturas mencionadas o combinaban alguna de dichas asignaturas con otras más ajenas a la evaluación propiamente dicha. Las respuestas de los 95 profesores se consideran para dar los resultados globales. Con el fin de obtener información diferenciada según la asignatura principal de docencia se formaron dos grupos: quienes tenían como asignatura única o principal la psicometría (si impartían una segunda asignatura, ésta no tenía que ver con la evaluación propiamente dicha –por ejemplo, impartían estadística) y quienes tenían como única o principal una o varias asignaturas de evaluación –psicológica y/o (psico)educativa. Quienes impartían a la vez psicometría y evaluación no se incluyeron en los análisis diferenciales.

Resultados

Los resultados se presentan en la Tabla 4, tanto globalmente como diferenciados en función del tipo de asignatura impartida (psicometría vs. evaluación psicológica y/o (psico)educativa). En concreto, la mayoría de los profesores que respondieron a la encuesta (72,6%) afirmaron conocer el modelo CET, mientras que el 27,4% señalaron que no lo conoci-

an. El porcentaje de conocedores del modelo es significativamente superior para los profesores de psicometría en comparación con los de evaluación ($\chi^2=4,46$; g.l.=1; $p<,05$).

Del total de profesores que dicen conocer el modelo, la mayoría (63,8%) lo utilizan en sus clases en algún momento, sin que existan diferencias estadísticamente significativas en función del tipo de asignatura impartida. En general, el uso que se da mayoritariamente es informativo (en un 47,7% de los casos), para que los estudiantes sepan que el modelo y sus resultados les pueden ser útiles a la hora de tomar decisiones sobre el uso de un test, aunque también se da un uso más aplicado: o bien los estudiantes evalúan alguna(s) prueba(s) aplicando el modelo (37,1%) o bien revisan o trabajan con los informes disponibles en la página web del COP (15,9%). Aunque no se realizaron contrastes, debido al reducido tamaño de las muestras para confiar en los resultados de las pruebas ji-cuadrado, a nivel descriptivo se observa que la mayor diferencia porcentual en el uso que se da en clases de psicometría o evaluación, se encuentra en la evaluación de pruebas aplicando el modelo, que resulta ser una práctica más frecuente en la asignatura de evaluación que en la de psicometría.

Entre quienes no utilizan el modelo CET para la docencia, a pesar de conocerlo, la mayoría indica que no lo hace por falta

TABLA 3
RESUMEN DE LAS CALIFICACIONES DE LOS TESTS ANALIZADOS EN LA SEXTA EVALUACIÓN

Características	BADyG/E2-r (2011)	BADYG/S-r (2011)	BAT-7 (2015)	BDI-FS (2011)	BPR (2016)	CESPRO (2016)	MATRICES (2015)	MBMD (2014)	Q-PAD (2016)	SENSORY PROFILE-2 (2016)
Materiales y documentación	4,5	4,5	5	4,5	5	4	5	4,5	5	4
Fundamentación teórica	5	5	5	5	5	4	5	5	3,5	3
Adaptación	-	-	-	4	5	-	-	4	5	3
Análisis de los ítems	4	4	5	4	4,5	4	5	-	3	-
Validez contenido	5	5	4	4,5	-	4	5	4	4,5	2,5
Validez relación con otras variables	3,2	3	4,1	4	4,3	4,2	4,6	3,7	4	2,7
Validez estructura interna	4	4	5	5	5	4	5	4	4,5	-
Validez análisis del DIF	-	-	-	-	5	-	5	-	-	-
Fiabilidad: equivalencia	-	-	-	-	-	-	-	-	-	-
Fiabilidad: consistencia interna	5	4	5	4	5	5	5	3	5	3
Fiabilidad: estabilidad	-	-	-	4	3	-	5	3	3,5	3
Fiabilidad: TRI	-	-	4,5	-	-	5	5	-	-	-
Fiabilidad: inter-jueces	-	-	-	-	-	-	-	-	-	4
Baremos e interpretación de puntuaciones	4,7	4,7	4,3	4,5	4,5	4	5	4,5	4,7	3
Soporte	Papel y Lápiz Informatizado	Papel y Lápiz Informatizado	Papel y Lápiz Informatizado	Papel y Lápiz	Papel y Lápiz	Papel y Lápiz Informatizado	Papel y Lápiz*	Papel y Lápiz	Papel y Lápiz	Papel y Lápiz Informatizado (On-line)
Corrección automatizada por ordenador	Sí	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí	Sí

Notas: Las puntuaciones de la tabla están dadas en una escala de cinco puntos cuyos valores son: 1= inadecuado, 2= adecuado pero con algunas carencias, 3= adecuado, 4= bueno y 5=excelente. Cuando aparece un guión (-) significa que no se aporta información o no procede.

*Disponible una versión adaptativa informatizada



Artículo en prensa

de tiempo (56%), siendo esta la razón más aducida tanto por los profesores de psicometría como por los de evaluación.

Del total de profesores que no conocían el modelo hasta ser invitados a participar en el estudio, la mayoría señalan que, en el futuro, sí que lo utilizarán para la docencia (42,3%) o que probablemente lo harán (42,3%), mientras que un 15,4% todavía no lo sabe. Además, la mayoría (59,1%), especialmente entre los profesores de evaluación psicológica y/o (psico)educativa (73,3%), no tienen claro qué uso concreto le darían al modelo a la hora de impartir la asignatura, considerando necesario consultar con más detalle el modelo y las evaluaciones realizadas.

Encuesta a las principales editoriales de tests españolas

La calidad de los tests depende directamente de las editoriales que los publican y de su implicación para mantener unos estándares elevados. La implicación y buena disposición de las principales editoriales de tests en España (TEA Ediciones, Pearson, EOS y CEPE) ha sido esencial para implementar el proceso de aplicación del CET. También cabe destacar que las editoriales juegan un papel crucial a la hora de velar por que los tests tengan unos criterios de calidad adecuados y sean adquiridos únicamente por los profesionales cualificados. Para conocer la opinión de estas editoriales sobre el impacto del modelo CET en la calidad de los tests que publican, conocer

TABLA 4
IMPACTO DEL MODELO EN LA DOCENCIA POR TIPO DE ASIGNATURA - PSICOMETRÍA O EVALUACIÓN PSICOLÓGICA O (PSICO)EDUATIVA- Y GLOBAL

	Psicometría (N=52)		Ev. Psicológica y/o (psico)educativa(N=41)		Global ^a (N=95)	
	Sí	No	Sí	No	Sí	No
Conocimiento del modelo CET antes de la recepción del email solicitando participación [*]	80,8%	19,2%	61,0%	39,0%	72,6%	27,4%
Sí conocían el modelo CET	(N=42)		(N=25)		(N=69)	
Utilizan del modelo CET para la docencia	69,0%	31,0%	60,0%	40,0%	63,8%	36,2%
Uso concreto que se da del modelo ^b :						
Meramente informativo: para que los estudiantes sepan que el modelo y sus resultados les pueden ser útiles a la hora de tomar decisiones sobre el uso de un test	44,8%		53,3%	47,7%		
Aplicado: los estudiantes han evaluado alguna(s) prueba(s) aplicando el modelo	41,4%		20,0%	37,1%		
Aplicado: los estudiantes han revisado y/o trabajado en clase con los informes disponibles en la página web del COP)	10,3%		26,7%	15,9%		
Otros: rúbrica de trabajo	3,4%		0,0%	2,3%		
Razones para no usar el modelo ^b :						
Falta de tiempo		61,5%		60,00%		56,0%
No creo que aporte demasiado a la formación en la asignatura que imparto		0,0%		10,0%		8,0%
No es suficientemente exhaustivo		0,0%		0,00%		0,0%
Es demasiado exhaustivo		7,7%		10,0%		8,0%
Otras: no lo había pensado pero quizás lo use en el futuro, imparto otras técnicas de evaluación diferentes de los tests, lo utilicé una vez pero no despertó interés entre los alumnos, no es una competencia a trabajar en la asignatura		30,8%		20,0%		28,0%
No conocían el modelo CET	(N=10)		(N=16)		(N=26)	
Utilizarían del modelo CET para la docencia ^{bc}	Sí o probablemente	NS/NC	Sí o probablemente	NS/NC	Sí o probablemente	NS/NC
	70,0%	30,0%	93,7%	6,3%	84,6%	15,4%
Uso concreto que se daría al modelo ^b :						
Meramente informativo: para que los estudiantes supieran que el modelo y sus resultados les pueden ser útiles a la hora de tomar decisiones sobre el uso de un test	14,3%		13,3%		13,6%	
Aplicado: los estudiantes evaluarían alguna(s) prueba(s) aplicando el modelo	42,9%		13,3%		22,7%	
Aplicado: los estudiantes revisarían/trabajarían en clase con los informes disponibles en la página web del COP)	14,3%		0,0%		4,5%	
Todavía no lo sé	28,6%		73,3%		59,1%	

^ados profesores se excluyen de la comparación entre tipo de asignaturas, puesto que impartían ambos tipos (psicometría y evaluación)

^bLas pruebas ji-cuadrado indican que hay diferencias significativas en los patrones de respuestas de los profesores de ambos tipos de asignaturas

^cDebido al reducido tamaño de las muestras, no se realizan contrastes estadísticos ji-cuadrado para comparar patrones de respuesta en función del tipo de asignatura

^dLa opción "no" no fue marcada por ningún participante



Artículo en prensa

cómo mejorarían la aplicación del modelo y qué medidas tomarían para que las evaluaciones tuvieran un mayor impacto, se pidió la opinión de las cuatro editoriales mencionadas, obteniéndose respuesta de tres de ellas. En concreto, se plantearon cuatro preguntas (ver Tabla 5), resumiendo a continuación los aspectos más destacables de las respuestas obtenidas.

En general, cuando se les pregunta por el impacto del modelo CET en los procesos de desarrollo y adaptación de las pruebas, los comentarios son altamente positivos. Destacan que el modelo CET es tomado como input por autores y editores en el desarrollo y adaptación de las pruebas, así como para la edición del manual del test. De hecho, en los manuales, las evidencias que apoyan la calidad técnica de los tests se presentan de forma mucho más detallada durante los últimos años. Por ejemplo, algunas editoriales, además de los valores de fiabilidad de muestras y escalas, en el manual incluyen los promedios obtenidos, que es una información más sintética, así como su clasificación, según con los criterios de calidad establecidos por el modelo CET (bueno, excelente, adecuado, etc.). También valoran muy positivamente que el modelo se haya convertido en un sistema de mejora continua que refuerza los estándares de calidad que persiguen las editoriales. Asimismo, para el potencial usuario, supone una mayor garantía a la hora de seleccionar y utilizar una determinada prueba frente a otras. Un aspecto adicional interesante relacionado con el otro gran agente analizado en el presente estudio, los profesores universitarios, es que, según algunas de las editoriales participantes, las universidades han pasado de comprar tests más tradicionales, en algunos casos obsoletos, a comprar tests menos conocidos, pero ampliamente aplicados en el contexto aplicado, actualizados, y avalados por las evidencias de calidad técnica y psicométrica.

Sobre los aspectos del modelo y su aplicación que podrían modificarse para mejorar el proceso de desarrollo y publicación de los tests, se mencionan varios aspectos. Puesto que los manuales se hacen cada vez más detallados y exhaustivos, esto encarece el resultado y, en algunos casos, añade complejidad para el usuario del test. Una posibilidad es ceñirse a la información más básica en el manual, y publicar todas las evidencias detalladas en forma de materiales complementarios online. Estos materiales también serían considerados por la comisión a la hora de evaluar la calidad de las pruebas cuyos resultados son publicados en la página web del COP. En esta línea también se señala que se podrían priorizar distintos tipos de evidencias sobre la calidad del test en función del propósito del mismo y determinar qué aspectos son inexcusables para poder publicar un test y, en caso de que así sea, cuáles no. Aunque obviamente estos aspectos pueden variar dependiendo del propósito con el que se usa el test en última instancia. Finalmente, se señala la necesidad de incluir en el proceso de revisión otros tests que no son publicados por las editoriales más conocidas, que son las que tienen representación en la comisión de tests. De otra forma, los tests de estas últimas editoriales podrían resultar perjudicados en comparación con otros

tests que no son evaluados, pero que son ampliamente utilizados en distintos ámbitos, sin que, en algunas ocasiones, existan estudios empíricos rigurosos que avalen su calidad. Sólo los tests que son evaluados quedan expuestos a recibir críticas, por lo que se debería abrir la revisión a cualquier test que se esté usando profesionalmente. Esto contribuiría a mejorar la calidad de los tests que se desarrollan y adaptan, independientemente de la editorial que esté detrás. Según algunos editores supone un contrasentido que, editoriales con un alto bagaje en valorar y mejorar técnicamente las pruebas y dar luz o veto a las obras que pretenden publicarse, sea ahora la receptora de valoraciones, en algunos casos negativas, por parte de profesionales alejados del campo psicométrico más aplicado.

En cuanto a las medidas que podrían tomarse por la comisión para aumentar el impacto del CET y de las evaluaciones realizadas, se señalan varios aspectos. Además de la sugerencia de incluir en los procesos de revisión otros tests que no dependan de las editoriales con representación en la comisión de tests, se podría establecer una puntuación mínima o un porcentaje de criterios en los que se alcance la excelencia, que sirvan como punto de corte para recibir un sello de calidad por parte de la comisión. También se debería señalar con claridad los tests que no alcanzan suficientes garantías de calidad para ciertos propósitos. Finalmente, para que las evaluaciones tengan un impacto real en los profesionales, se destaca la necesidad de reducir la distancia entre los académicos (que son en su mayoría quienes participan en los procesos de evaluación de las pruebas, pero que a veces no han participado en procesos de desarrollo y adaptación de tests) y los profesionales aplicados (destinatarios últimos de las revisiones). Según se indica, la distancia entre lo académico y lo aplicado hace que el proceso sea visto como algo alejado a la realidad de los profesionales, como algo de carácter técnico que poco tiene que ver con las necesidades reales y la respuesta que dan las obras analizadas a esas necesidades. Para ello es necesaria la formación y actualización continua de los profesionales, para poder entender los aspectos de la evaluación más relacionados con los avances psicométricos y técnicos, pero también podría resultar conveniente aumentar el número de profesionales aplicados con conocimientos psicométricos contrastados, que participen en el proceso de revisión.

Estas últimas consideraciones están relacionadas con algunos de los posibles efectos contraproducentes del modelo y su aplicación. Dentro de los posibles efectos negativos, también se menciona que "más" no siempre es sinónimo de "mejor". Por ejemplo, para el caso de baremos, el modelo da una puntuación más alta cuando se utiliza un amplio rango de baremos en vez de un único baremo adecuado para la población objetivo. Según algunas editoriales esto podría llevar a introducir artificialmente baremos adicionales que pueden ser innecesarios o no convenientes con el objetivo de obtener una mejor puntuación en el CET. Además, también se puede entrar en la dinámica de dar información de todos los aspectos que se consideran en el modelo, incluso cuando no sean relevantes para



Artículo en prensa

el uso que se va a dar a la prueba. También se señala que algunos de los criterios son excesivamente exigentes en aplicaciones realistas (por ejemplo, los tamaños muestrales requeridos para la fiabilidad test-retest, especialmente en ciertos tipos de poblaciones). Finalmente, la subjetividad inherente a las valoraciones que se dan, especialmente las cualitativas, también se valora como un aspecto negativo en el proceso. Para matizar dicha subjetividad, en cierta medida, es crucial el papel del coordinador que integra las valoraciones de los distintos evaluadores de cada prueba, pero también introducir una mayor exigencia a la hora de justificar las valoraciones realizadas por los revisores.

Todas estas cuestiones deberán tenerse progresivamente en cuenta en futuras evaluaciones para continuar mejorando el CET, su impacto en la mejora de los tests y en el uso de los mismos por parte de los profesionales y académicos que imparten asignaturas relacionadas con los tests y la evaluación mediante tests.

CONCLUSIONES

En esta sexta evaluación de tests han sido evaluados con el CET-R un total de diez instrumentos tanto las valoraciones cuantitativas como cualitativas han sido buenas, ya que en la mayoría de los ítems del CET-R los tests fueron valorados con una calificación de buena a excelente. Los tests evaluados, en general, proporcionan evidencias de fiabilidad y validez del uso de las puntuaciones que se obtienen. Además, en esta evaluación de tests se ha constatado que casi un tercio de los instrumentos analizados habían abordado la precisión de la medida desde los modelos de TRI. El análisis de ítems usando TRI proporciona información valiosa que ayuda al desarrollo y depuración del instrumento en construcción. Las curvas características de los ítems (CCIs), así como, sus correspondientes funciones de información (FFIs) son herramientas gráficas que nos señalan el nivel de habilidad/rasgo donde los ítems son más precisos y nos orientan en la selección de aquellos que permitan alcanzar el mínimo necesario de calidad en términos de fiabilidad (Ames y Penfield, 2015; Hidalgo y

French, 2016). Esto es de especial utilidad en la toma de decisiones, cuando tenemos que ser precisos en un cierto nivel de habilidad que se establece como criterio de interpretación o punto de corte. El uso de estos modelos de medida tanto a nivel teórico como práctico nos permite obtener argumentos de validez útiles respecto al uso de las puntuaciones. Añadido a lo ya comentado podemos constatar importantes ventajas en el uso de estos modelos, siendo la base para la construcción de tests adaptativos (Olea, Abad y Barrada, 2010) y una evaluación más ecológica. Aunque son modelos que requieren de un tamaño muestral generalmente más elevado (Muñiz, 2010; Penfield, 2014), las ventajas mencionadas hacen que se considere muy positivo el incremento de los tests que emplean la TRI en comparación con las evaluaciones anteriores.

En esta sexta evaluación de tests editados en España, también se ha puesto de manifiesto que más de la mitad de los tests evaluados (60%) se pueden administrar de manera informatizada. Aunque parece que en el ámbito clínico y educativo se prefiere la administración de *papel y lápiz* frente a la informatizada (Antón, 2017), sería interesante indagar en los profesionales de la psicología sobre los motivos de estas preferencias, teniendo en cuenta las posibilidades de nuevos formatos de ítems que los avances en las tecnologías de la información y la comunicación nos proporcionan. También queremos hacer notar la importancia de compartir bases de datos de los estudios de fiabilidad y validez de los tests, y la necesidad de crear un repositorio con este tipo de datos a fin de poder acumular evidencias de validez y fiabilidad. Como apuntan Botella y Ortego (2010) compartir datos directos de las investigaciones redundaría en una acumulación más eficiente del conocimiento, que en el caso de la calidad métrica de los tests contribuiría a su mejora.

Por otro lado, en cualquier contexto (organizacional, educativo, sanitario, etc.) donde se haya implementado un sistema de garantía de la calidad, la evaluación del propio sistema de calidad es clave en el ciclo de mejora. En este sentido comentaremos algunos aspectos relativos a la implementación del CET-R que se han puesto de manifiesto en el proceso de evaluación de tests de esta sexta revisión. En primer lugar, se ha constatado la necesidad de facilitar la aplicación del CET-R, tanto para los revisores como para el propio coordinador, a la hora de integrar los resultados de las evaluaciones recibidas. En este sentido, informatizar el CET-R (aplicación computerizada) agilizaría la aplicación del mismo y permitiría, recogiendo algunas de las sugerencias de Fonseca-Pedrero y Muñiz (2017) incorporar más información técnico-psicométrica, ejemplos concretos de uso y/o video-tutoriales, que fueran accesibles on-line. La versión informatizada debería permitir obtener el resumen-valoración numérico para cada característica final contemplada en el CET-R. Al mismo tiempo podría ser de mayor utilidad como herramienta de autoevaluación para aquellos investigadores o profesionales inmersos en el proceso de desarrollo de un test.

TABLA 5

LISTADO DE PREGUNTAS PLANTEADAS A LAS EDITORIALES

Preguntas formuladas

1. Indica cuál ha sido el impacto del modelo de evaluación de la calidad de los tests en los procesos de desarrollo y adaptación de las pruebas psicológicas y educativas publicadas por tu editorial. Destaca los aspectos más positivos para la creación y adaptación de nuevas pruebas en tu editorial
2. Destaca qué aspectos del modelo y su aplicación podrían modificarse para mejorar la calidad de los tests en el proceso de desarrollo y publicación
3. Señala qué medidas podrían tomarse para aumentar el impacto del modelo y que la evaluación positiva de una prueba por parte de la comisión de tests sea interpretada por los usuarios como sello de calidad
4. Si piensas que el modelo está teniendo algún efecto contraproducente para las editoriales y usuarios, describe brevemente cuáles son dichos efectos



Artículo en prensa

En segundo lugar, se ha detectado algunas dificultades con la valoración referida al tamaño de la muestra. Por ejemplo, en las estimaciones de la fiabilidad como consistencia interna o tests-retest. Aunque las diferentes opciones de respuesta para estos ítems son claras, se encuentran problemas en integrar las valoraciones en aquellos instrumentos en el que existen diferentes versiones del cuestionario (niveles) en función de la edad de las personas evaluadas, y en los que las evidencias de fiabilidad se han obtenido en muestras de diferente tamaño para cada nivel, pero que no llegan a ser una muestra grande ($N > 500$). En el CET-R no queda explícitamente recogida la posibilidad de varios estudios con una muestra moderada y otras pequeñas. Por otro lado, resulta de especial dificultad valorar el tamaño de la muestra en los estudios de fiabilidad test-retest de un test cuando el test tiene diferentes versiones (niveles) y en las mismas se emplean distintos tamaños muestrales. La aplicación del CET-R se podría ver mejorada incluyendo ejemplos de aplicación que oriente al revisor en su proceso de evaluación. En tercer lugar, se recomendó que en el criterio 2.11.2.2.6 se modificara la graduación de la respuesta en las opciones más extremas (Buena, Muy buena y Excelente), dado que en este caso no se mantiene la graduación respecto a otras cuestiones del cuestionario (Buena y Muy Buena). En cuarto lugar, para aquellos tests evaluados en los que las evidencias de fiabilidad y/o relaciones con otras variables se hayan obtenido en varias muestras de diferente tamaño sería interesante considerar utilizar un estimador promedio del coeficiente de fiabilidad o del coeficiente de correlación ponderado por el tamaño muestral. Nos consta que algunas de estas cuestiones ya se están teniendo en cuenta en la séptima evaluación de tests.

Echando la vista atrás, la utilidad del CET-R como herramienta de evaluación de la calidad de los tests se ha ido poniendo de manifiesto a través de los ya 75 tests evaluados, y en algunos casos se ha sugerido que dicho modelo estaba teniendo un impacto positivo tanto el ámbito académico relacionado con la enseñanza de la Psicología, como en el de las editoriales de tests (e.g. Hernández et al., 2015, 2016). En este estudio se ha analizado de un modo más sistemático dicho impacto. Por una parte se han analizado los resultados de una encuesta realizada a profesores acerca del uso del CET (y su revisión CET-R) en la docencia de las asignaturas de Psicometría, Evaluación Psicológica y Educativa y disciplinas afines. Por otra se ha encuestado a los representantes de las editoriales en la comisión de tests, para conocer el impacto del modelo en la construcción/adaptación de tests, y conocer qué aspectos de su implementación podrían mejorarse.

En cuanto al impacto del CET/CET-R en el ámbito de la docencia universitaria en Psicología, aunque el porcentaje de profesores que respondieron a la encuesta sólo supone un 30.4% del total que fueron invitados a participar, los resultados muestran varios datos prometedores. En primer lugar, dos tercios de los profesores que respondieron a la encuesta, conocían el modelo de evaluación. Si bien este dato es positivo, todavía puede mejorarse la labor de difusión para que llegue

al 100% del profesorado. En este sentido la acción realizada al contactar a los profesores para preguntarles sobre el CET/CET-R puede contribuir a aumentar dicho porcentaje en el futuro. También es un buen resultado el hecho de que la mayoría de profesores que conocían el CET/CET-R sí que lo utilizan en sus clases en algún momento (63,8%). Esta es una forma de que los futuros profesionales conozcan el modelo y los resultados de las evaluaciones. Entre quienes no utilizan el modelo a pesar de conocerlo, la falta de tiempo es la razón más frecuentemente aducida. Finalmente, también tomamos como dato positivo el que, tras la encuesta, del total de profesores que no conocían el modelo hasta ser invitados a participar en el estudio, aproximadamente el 85% informan que lo utilizarán en sus clases, o bien seguro o bien probablemente.

En cuanto al impacto que el modelo CET/CET-R ha tenido en las editoriales de tests encontramos varios aspectos a destacar: (1) el modelo CET-R es tomado como una guía tanto por autores como editores en el proceso de desarrollo y adaptación del test, así como en la edición y confección del manual del test, y (2) el modelo se establece como un sistema de referencia de mejora continua para alcanzar los estándares de calidad de los tests. Sin embargo, (3) es necesario incluir en el proceso de revisión otros tests que no son publicados por las editoriales más conocidas y que son utilizados en distintos ámbitos profesionales. Y finalmente, (4) a fin de que las evaluaciones de los tests tengan un impacto real en los profesionales, se hace necesario reducir la distancia entre lo académico (lo técnico) y lo aplicado. Así, establecer estrategias eficaces de divulgación de los informes de evaluación de la calidad de los tests entre profesionales aplicados, se convierte en un reto a abordar por parte de la Comisión de Tests. Combinar la rigurosidad técnico-psicométrica de los informes de evaluación de los tests con actividades que hagan accesible ese conocimiento a los profesionales es una tarea pendiente, siendo necesaria la formación y actualización continua de los profesionales para poder entender los aspectos de la evaluación más relacionados con los avances psicométricos y técnicos. La divulgación del CET-R en ámbitos profesionales de intervención clínica y educativa (centros de salud mental, hospitales, centros educativos, etc.) a buen seguro será una aportación en esta línea y redundará en un mejor uso de los tests.

CONFLICTO DE INTERESES

No existe conflicto de intereses.

REFERENCIAS

- American Educational Research Association, American Psychological Association, y National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Ames, A.J., y Penfield, R.D. (2015). A NCME instructional module on item-fit statistics for item response theory models. *Educational Measurement: Issues and Practice*, 34, 39-48.



Artículo en prensa

- Antón, M. (2017). Mesa redonda "La salud de los test en España". Coordinadora: Ana Hernández. *III Congreso Nacional de Psicología*. Oviedo, 3-7 julio.
- Botella, J., y Ortego, C. (2010). Compartir datos: hacia una investigación más sostenible. *Psicothema*, 22, 263-269.
- Elosua, P. (2012). Tests publicados en España: usos, costumbres y asignaturas pendientes. *Papeles del Psicólogo*, 33, 12-21.
- Elosua, P., y Geisinger, K. (2016). Cuarta evaluación de tests editados en España: forma y fondo. *Papeles del Psicólogo*, 37, 82-88.
- Evers, A. (2012). The Internationalization of Test Reviewing: Trends, Differences, and Results. *International Journal of Testing*, 12, 136-156.
- Fonseca-Pedrero, E., y Muñiz, J. (2017). Quinta evaluación de tests editados en España: Mirando hacia atrás, construyendo el futuro. *Papeles del Psicólogo*, 38, 161-168.
- Geisinger, K. (2012). Worldwide test reviewing at the beginning of the Twenty-First century. *International Journal of Testing*, 12, 103-107.
- Hambleton, R.K., Merenda, P.F., y Spielberger, C. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Hernández, A., Ponsoda, V., Muñiz, J., Prieto, G., y Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 37, 192-197.
- Hernández, A., Tomás, I., Ferreres, A., y Lloret, S. (2015). Tercera evaluación de test editados en España. *Papeles del Psicólogo*, 36, 1-8.
- Hidalgo, M.D., y French, B. (2016). Una introducción didáctica a la Teoría de Respuesta al Ítem para comprender la construcción de escalas. *Revista de Psicología Clínica con Niños y Adolescentes*, 3, 1-9.
- International Test Commission (2018). ITC Guidelines for Translating and Adapting Tests (Second Edition). *International Journal of Testing*, 18, 101-134.
- Muñiz, J. (2010). Las teorías de los tests: Teoría Clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo*, 31, 57-66.
- Muñiz, J., Elosua, P., y Hambleton, R.K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25, 151-157.
- Muñiz, J., y Fernández-Hermida, J.R. (2000). La utilización de los tests en España. *Papeles del Psicólogo*, 76, 41-49.
- Muñiz, J., y Fernández-Hermida, J.R. (2010). La opinión de los psicólogos españoles sobre el uso de los tests. *Papeles del Psicólogo*, 31, 108-121.
- Muñiz, J., Fernández-Hermida, J.R., Fonseca-Pedrero, E., Campillo-Álvarez, A., y Peña Suárez, E. (2011). Evaluación de tests editados en España. *Papeles del Psicólogo*, 32, 113-128.
- Muñiz, J., Hernández, A., y Ponsoda, V. (2015). Nuevas directrices sobre el uso de los tests: investigación, control de calidad y seguridad. *Papeles del Psicólogo*, 36(3), 161-173.
- Olea, J., Abad, F.J., y Barrada, J.R. (2010). Tests informatizados y otros tipos de tests. *Papeles del Psicólogo*, 31, 97-107.
- Penfield, R.D. (2014). An NCME Instructional Module on Polytomous Item Response Theory Models. *Educational Measurement: Issues and Practice*, 33, 36-48.
- Ponsoda, V., y Hontangas, P. (2013). Second evaluation of tests published in Spain. *Papeles del Psicólogo*, 34, 82-90.
- Prieto, G., y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-72.