



TERCERA EVALUACIÓN DE TESTS EDITADOS EN ESPAÑA

THIRD EVALUATION OF TESTS PUBLISHED IN SPAIN

Ana Hernández, Inés Tomás, Adoración Ferreres y Susana Lloret

Metodología de las Ciencias del Comportamiento e IDOCAL*. Universitat de València

Con el fin de proporcionar información técnica sobre los tests editados en España, y siguiendo el proceso de evaluación de algunos de los tests más empleados en nuestro país, el presente artículo presenta los resultados de la tercera evaluación de tests llevada a cabo desde la comisión de tests del Colegio Oficial de Psicólogos. En concreto se han evaluado un total de 11 tests. Como en las dos evaluaciones previas, los tests han sido evaluados por dos expertos mediante el Cuestionario para la Evaluación de los Tests (CET) propuesto por Prieto y Muñiz (2000), si bien en la presente edición, y con motivo de la reciente revisión del modelo de evaluación elaborado por la Federación Europea de Asociaciones de Psicólogos Profesionales, se han aplicado algunas modificaciones que son pertinentemente señaladas. Para cada test se presentan los resultados sobre la calidad de los materiales y su documentación, la fundamentación teórica, la adaptación española (si procede), el análisis de ítems, las evidencias de validez recogida, la fiabilidad de sus puntuaciones, y la calidad de los baremos. Se ofrecen tanto resultados cuantitativos sobre estos aspectos como información cualitativa que resalta los puntos fuertes del test y los aspectos a mejorar. Por otra parte también se presenta información sobre el proceso de revisión y algunas cuestiones a mejorar para futuras ediciones.

Palabras clave: Tests, Uso de los tests, Evaluación de tests, Propiedades psicométricas.

In order to provide technical information on the tests published in Spain, and continuing with the evaluation of some of the tests that are most frequently used in this country, this paper describes the results of the third evaluation of psychological and educational tests carried out by the Test Commission of the Spanish Psychological Association. In total, 11 tests were evaluated. As in the previous evaluations, each test was evaluated by two experts who responded to the Questionnaire for the Assessment of Tests -CET- (Prieto & Muñiz, 2000). However, considering that the European Federation of Psychological Associations has recently revised the European model -which was partially based on the CET model-, for this third evaluation we have introduced a number of changes that are described in this paper. In addition, for each test, results are presented regarding the quality of the documentation and materials, the theoretical foundation, the Spanish adaptation, the item analysis, the coverage of the validation studies, the reliability, and the norms. Both quantitative and qualitative information is provided, highlighting the main strengths and weaknesses of the test. Finally, we also present a number of recommendations and suggestions for future test evaluations.

Key words: Tests, Test use, Test evaluation, Psychometric properties.

Los tests son una herramienta básica en la evaluación psicológica, y ayudan al profesional a tomar decisiones que pueden tener importantes consecuencias para las personas. Por ello es necesario garantizar que las propiedades psicométricas de los tests son adecuadas y que son usados por profesionales competentes. Una de las estrategias informativas que está llevando a cabo el Colegio Oficial de Psicólogos (COP) para conseguir ambos propósitos y mejorar el uso de los tests consiste en proporcionar información contrastada

sobre las características teóricas, prácticas y psicométricas de los mismos, que ayuden a los profesionales a tomar las decisiones adecuadas y a hacer un uso correcto de las pruebas. De hecho, los psicólogos desde hace tiempo demandan este tipo de información técnica (ver Muñiz et al., 2001; Muñiz y Fernández-Hermida, 2010)

En este contexto, la comisión de tests del COP, siguiendo la estela de otros países como Holanda y Reino Unido, puso en marcha en 2010 el proceso de evaluación de los tests editados en España. En concreto, la evaluación se realizó mediante el Cuestionario de Evaluación de Tests (CET) (Prieto y Muñiz, 2000) que sirvió de inspiración, junto con otros modelos europeos, al modelo de evaluación propuesto por la comisión de tests de la Federación Europea de Asociaciones de Psicólogos Profesionales (EFPA) (Evers et al., 2013). Una de las características más importantes del CET es que permite

Correspondencia: Ana Hernández Baeza. Departamento de Metodología de las Ciencias del Comportamiento. Facultad de Psicología. Universitat de València. Av. Blasco Ibáñez, 21. 46010 Valencia. España. E-mail: Ana.Hernandez@uv.es

.....
*Instituto Universitario de Investigación en Psicología de los Recursos Humanos, del Desarrollo organizacional y de la Calidad de Vida Laboral



ofrecer tanto información cuantitativa como cualitativa sobre la calidad psicométrica del test evaluado, así como de la calidad de sus materiales y documentación.

Siguiendo este modelo, en 2011 se publicaron los resultados de la primera evaluación de tests editados en España (Muñiz, Fernández-Hermida, Fonseca-Pedrero, Campillos-Alvarez y Peña-Suarez, 2011), donde se valoraron un total de 10 tests (en concreto, las pruebas WAIS-III, WISC-IV, MCMI-III, MMPI-2-RF, 16PF-5, PROLEC-R, EFAI, NEO-PI-R, EVALUA, e IFG). Dos años más tarde se publicaron los resultados de la segunda evaluación (Ponsoda y Hontangas, 2013), en la que se valoraron un total de 12 tests (en concreto, las pruebas BAI, BAS-II, BDI-II, CEAM, CompeTEA, EPV-R, ESCOLA, ESPERI, Merrill-Palmer.R, PAI, RIASRSIT, y WNV). Los informes detallados de los tests revisados pueden consultarse en la página web del COP (sección de comisión de tests, apartado de evaluación de tests en España). Asimismo, el resumen de los resultados obtenidos y del proceso llevado a cabo en ambas evaluaciones, puede consultarse en los trabajos de Muñiz et al. (2011) y Ponsoda y Hontangas (2013), publicados en esta misma revista.

En este artículo se presentan los resultados de la tercera evaluación de tests, en la que se han valorado un total de 11 tests (en concreto, las pruebas BCSE, BECOLE, BOHEM-3, BOHEM-3 Preescolar, CESQT, ECLE, ESQUIZO-Q, IECI, SOC, TRAUMA y WAIS-IV) (ver Tabla 1). En primer lugar, y antes de presentar los resultados, se informa de algunas modificaciones llevadas a cabo, tanto en el proceso como en el propio cuestionario de evaluación CET. Si bien el proceso de revisión ha sido

similar al seguido en ediciones previas en términos generales, es necesario tener en cuenta estas pequeñas modificaciones a la hora de comparar los resultados con los obtenidos en evaluaciones previas. En segundo lugar, se presentan los resultados generales de los tests evaluados, destacándose sus puntos fuertes y los aspectos que son mejorables. Finalmente, a partir de los resultados y de nuestra experiencia en el proceso de revisión, se realizan algunas recomendaciones que pueden resultar útiles para las evaluaciones venideras.

PROCESO DE EVALUACIÓN

En esta tercera evaluación el proceso general coincide básicamente con el seguido en las anteriores evaluaciones, aunque existen algunas diferencias que son convenientemente señaladas. En primer lugar, mientras que en evaluaciones previas la comisión de tests seleccionó las pruebas a revisar, en esta ocasión las editoriales seleccionaron las pruebas que querían someter a revisión, y la propuesta fue aceptada de forma unánime por la comisión de tests. Como se ha señalado anteriormente se seleccionaron un total de 11 pruebas (tres de la editorial de EOS, cuatro de Pearson y cuatro de TEA). En segundo lugar, y al igual que en evaluaciones previas, el equipo coordinador designado por la comisión de tests (liderado por la primera autora del presente artículo), seleccionó a un conjunto de revisores, para asignar dos por test. Un revisor con un perfil más técnico-psicométrico y otro con un perfil más teórico, experto en los aspectos sustantivos de la variable medida por el test. Se procuró que los revisores no tuviesen relación directa con los autores de los tests, y que no tuviesen ningún conflicto de intereses (de hecho en la carta de invitación a participar en el proceso se les indicaba que no debían participar si dudaban de que pudiesen hacer una valoración objetiva). En algunos casos, los revisores inicialmente seleccionados rechazaron participar en el proceso por distintas razones justificadas, por lo que se tuvo que seleccionar otro revisor. El listado final de revisores participantes se muestra en la Tabla 2. Desde aquí agradecemos sinceramente a todos los revisores su respuesta positiva y su implicación en el proceso.

Los editores pusieron a disposición del COP tres ejemplares completos de cada test. El COP envió uno a cada revisor y el tercero al coordinador, en este caso coordinadora. En ocasiones anteriores sólo se enviaban los tests a los revisores, sin embargo, en esta ocasión, si-

TABLA 1
RELACIÓN DE TESTS EVALUADOS

BCSE	Test Breve para la evaluación del estado cognitivo
BECOLE	Batería de Evaluación Cognitiva de la Lectura y la Escritura
BOHEM-3	Test Boehm de conceptos básicos
BOHEM-3 PRESCOLAR	Test Boehm de conceptos básicos - 3 Preescolar
CESQT	Cuestionario para la Evaluación del Síndrome de Quemarse por el Trabajo
ECLE	Evaluación de la comprensión lectora.
ESQUIZO-Q	Cuestionario Oviedo para la Evaluación de la Esquizotipia
IECI	Inventario de Estrés Cotidiano Infantil
SOC	Escala de Dificultades de Socialización de Cantoblanco
TRAUMA	Test de Resistencia al Trauma
WAIS-IV	Escala de inteligencia de Wechsler para adultos-IV



guiendo la recomendación hecha por Ponsoda y Hontangas (2013), se envió un tercer juego a coordinación, para facilitar la tarea. A los revisores, además del test, se les entregó una cantidad simbólica de 50 euros (que algunos prefirieron declinar). Como en ocasiones anteriores, la tarea de los revisores consistía en aplicar el CET y valorar mediante éste la prueba asignada. Sin embargo, considerando que el modelo de evaluación de tests elaborado por la comisión de tests de la EFPA había sido recientemente revisado, la comisión de tests inspeccionó este nuevo modelo y sus cambios y decidió incluir algunos ítems nuevos en el modelo CET, y/o modificar algunas cuestiones concretas. El modelo EFPA revisado puede consultarse en inglés en la página web <http://www.efpa.eu/professional-development>, (ver la sección dedicada a *Assesment*). Asimismo, en el artículo de Evers et al., (2013), publicado en la revista *Psicothema*, puede consultarse un resumen en inglés de las principales modificaciones llevadas a cabo. Para el caso del CET, los principales cambios aplicados en esta edición son los que se describen a continuación.

Por lo que se refiere al apartado de descripción general del test, se añadieron algunas áreas generales de contenido de la variable medida por el test, así como algunas áreas nuevas de aplicación (ítems 1.11 y 1.12 del CET original publicado por Prieto y Muñiz, (2000)). Asimismo, para el ítem referido a la transformación de las puntuaciones (ítem 1.21 del CET original), se añadió la aclaración de que la transformación normalizada se refería a puntuaciones obtenidas mediante normalización aplicada mediante la tabla de la curva normal, mientras que la no normalizada se refería a puntuaciones estandarizadas obtenidas mediante transformaciones lineales. Por ello, cuando sólo se ofrecen percentiles, ninguna de las dos opciones era pertinente en esta nueva versión del CET, debiéndose indicar que no era aplicable. Las puntuaciones percentiles aparecen detalladas ahora al describir el tipo de escala (ítem 1.22 del CET original), diferenciándose distintos tipos de percentiles (centiles, quintiles y deciles), así como puntuaciones estandarizadas y sus derivadas (decatipos, estaninos, T, etc). Finalmente, en el ítem referido a la documentación aportada por el editor, se añadió la opción "Información técnica complementaria y actualizaciones" (ítem 1.26 del CET original).

Por lo que se refiere a la valoración de las propiedades del test, también se realizaron algunas modificacio-

nes. En primer lugar, al valorar las evidencias de validez de constructo (ítem 2.10.2.1 del CET original) se añadieron otras opciones a las ya existentes: las correlaciones con otros tests, y el análisis de invarianza/funcionamiento diferencial del ítem (DIF); asimismo, a la opción de diseño experimental existente se le añadió el diseño cuasi-experimental. Además, se añadieron 3 ítems para evaluar explícitamente los resultados de las diferencias entre grupos (pudiendo ser estos naturales o experimentales), los resultados del análisis de la matriz multirrasgo-multimétodo, y los resultados del análisis factorial. En cuanto al apartado de fiabilidad, en el ítem 2.11.1 del CET original, referido a los datos aportados sobre fiabilidad, a las opciones ya existentes se añadió la de "Cuantificación del error mediante TRI (Función de información u otros)". Asimismo se añadió un ítem para evaluar la adecuación del tamaño de las muestras al cuantificar el error mediante TRI y otro ítem que informara de los coeficientes de consistencia interna presentados. Finalmente, por lo que se refiere al apartado de baremos, se incluyó un ítem para valorar la actualización de dichos baremos.

Por otra parte, teniendo en cuenta las recomendaciones dadas en las evaluaciones previas (ver Muñiz et al., 2011 y Ponsoda y Hontangas, 2013), para esta tercera

TABLA 2
REVISORES QUE LLEVARON A CABO LA
EVALUACIÓN DE LOS TESTS

Revisor	Afiliación
Francisco José Abad García	Universidad Autónoma de Madrid
Amelia Catalán Borja	Centre de Psicologia Clínica i Formativa, Valencia
Paula Elosua Oliden	Universidad del País Vasco
Antonio M. Ferrer Manchón	Universidad de Valencia
Adoración Ferreres Traver	Universidad de Valencia
Eduardo Fonseca-Pedrero	Universidad de La Rioja
Maité Garaigordobil Landazabal	Universidad del País Vasco
José Manuel García Montes	Universidad de Almería
Luis F. García Rodríguez	Universidad Autónoma de Madrid
Remedios González Barrón	Universidad de Valencia
Giorgina Guilera Ferré	Universidad de Barcelona
M ^o Dolores Hidalgo Montesinos	Universidad de Murcia
Susana Lloret Segura	Universidad de Valencia
Sonia Mariscal Altares	UNED
Isabel Martínez Sánchez	Universidad de Castilla-La Mancha
José Carlos Núñez Pérez	Universidad de Oviedo
Julio Olea Díaz	Universidad Autónoma de Madrid
José Luis Padilla García	Universidad de Granada
Hermínia Peraita Adrados	UNED
Jesús Pérez Hornero	Hospital de Conxo, Santiago de Compostela
Ingeborg Porcar Becker	Universidad Autónoma de Barcelona
Patricia Recio Saboya	UNED



evaluación, la comisión de tests decidió enviar a los revisores unas instrucciones generales sobre la forma de cumplimentar el CET, con el fin de reducir algunas ambigüedades y estandarizar en mayor medida el proceso. Estas instrucciones generales se presentan en el Anexo 1. Asimismo, se proporcionó un glosario de términos psicométricos que sirviese de guía y recordatorio, especialmente dirigido a los revisores con un perfil más teórico.

Una vez recibidas las evaluaciones de los revisores de cada test, el equipo coordinador integró ambas revisiones generando un informe por test. Cuando había discrepancias entre los revisores, el equipo coordinador realizó una evaluación independiente a partir de los materiales proporcionados. Asimismo, en función de la característica evaluada, dio un peso diferencial a ambas evaluaciones en función del perfil de los revisores. Con ello se otorgó una puntuación definitiva y se realizó la valoración final. Como en evaluaciones anteriores los informes fueron entonces remitidos a los editores, para que tanto ellos como los autores tuvieran la oportunidad de aclarar y matizar algunos de los comentarios de los revisores y, en definitiva, exponer su punto de vista. Estas aclaraciones y puntos de vista fueron integrados en el informe final, modificando las evaluaciones cuando se consideró que estaba justificado. Desde aquí queremos destacar la profesionalidad de los autores y editores a la

hora de responder ante la revisión y a su compromiso para introducir algunas de las mejoras sugeridas para futuras impresiones y ediciones de los manuales.

RESULTADOS DE LA EVALUACIÓN Y CONCLUSIONES

Los principales resultados obtenidos para los 11 tests evaluados pueden consultarse en la Tabla 3. Cabe destacar que en ningún caso hubo formas paralelas de los tests evaluados, por lo que este aspecto (fiabilidad: equivalencia) no está incluido en la tabla. Si se considera que los ítems son valorados mediante una escala de 5 puntos (1= inadecuado, 2=adecuado pero con algunas carencias, 3= adecuado, 4= bueno, 5= excelente), y que valores promedio iguales o mayores a 2.5 son considerados adecuados, se observa que, para la gran mayoría de características evaluadas, la gran mayoría de los tests son, como mínimo adecuados, y en muchos casos buenos (a partir de 3.5) e incluso excelentes (a partir de 4.5). A través de los distintos tests, las puntuaciones medias otorgadas para cada una de las características evaluadas en ningún caso estuvieron por debajo del punto medio 3 (adecuado), siendo el promedio más alto en esta evaluación de 4.1 (bueno).

Así pues, se puede concluir que las pruebas evaluadas tienen una calidad razonable. Lógicamente hay cierta variabilidad, y en general todas las pruebas tienen pun-

**TABLA 3
RESUMEN DE LAS CALIFICACIONES DE LOS TESTS EVALUADOS**

Características	Tests												Medias evaluaciones		
	BECOLE	TRauma	ECLÉ	CESQT	ESQUIZO-Q	IECI	SOC	BCSE	Boehm-3	Bohem-3	Preescolar	WAIS-IV	3ª	2ª	1ª
Calidad de los Materiales y documentación	3	4	3	5	4.5	3.5	4.5	4.5	4.5	4	5	4.1	4.3	4.4	
Fundamentación teórica	3.5	2	3	4	4.5	3.5	4.5	2.5	4	3	5	3.6	4.0	4.2	
Adaptación Española	-	-	-	-	-	-	-	3.5	4	3	5	3.9	4.3	4.3	
Análisis de los ítems	3.5	3	-	4	4.5	4	4	-	-	-	3	3.7	3.9	3.6	
Validez de contenido	-	3	2	4	3.5	4.5	4.5	2.5	3.5	4.5	3.5	3.6	3.5	4.3	
Validez de constructo	3.5	3.5	3.5	4	4	4	4	3	4	3	5	3.8	4.1	3.6	
Análisis del sesgo	-	-	-	-	4.5	-	4	-	-	-	-	4.3	5	-	
Validez predictiva	3.5	-	3.5	3.5	-	2.5	4.5	-	-	-	4	3.6	3.6	3.6	
Fiabilidad: consistencia interna	4	4	4	3	3.5	3	3.5	-	3	3.5	4.5	3.6	4.0	4.5	
Fiabilidad: estabilidad	-	-	-	3	-	3	1	3	4	4	3	3.0	4.2	3.8	
Baremos	4	4	4.5	4	4.5	4	3.5	4	3	3	5	4.0	3.4	3.5	

Notas: Ningún test tenía formas paralelas, por lo que esta forma de evaluar la fiabilidad no ha sido incluida en la tabla. Las puntuaciones de la tabla están dadas en una escala cuyos 5 valores son: 1= inadecuado, 2=adecuado pero con algunas carencias, 3= adecuado, 4= bueno, 5= excelente. Cuando aparece un guión (-) significa que no se aporta información o no procede.



tos fuertes y otros aspectos que pueden ser mejorados. En concreto, los aspectos que más frecuentemente requieren atención a través de los distintos tests son la fundamentación teórica presentada, la validez de contenido y la estimación de la fiabilidad como estabilidad. Este último aspecto es el que tiene el promedio más bajo, con una puntuación de 3. En las evaluaciones completas, que pueden consultarse en la página web del COP: www.cop.es, en la sección de la comisión de tests, se puede encontrar la información que justifica las puntuaciones otorgadas y cómo podría mejorarse cada aspecto, bien a partir de información adicional, ya que en algunos casos la información no es completa o suficientemente clara, bien mediante la realización de estudios complementarios o el incremento del tamaño de las muestras utilizadas. También llama la atención la falta de información y análisis relacionados con el análisis de ítems, la validez predictiva, la fiabilidad como estabilidad, y el análisis del sesgo o DIF. Toda esta información debería ir añadiéndose progresivamente. Si en algún caso no fuese pertinente obtener información sobre alguno de estos aspectos (por ejemplo por referirse a variables que se espera cambien en cortos periodos de tiempo), se podría hacer explícito en el manual.

En la Tabla 3 se presentan también las medias de las puntuaciones otorgadas para las distintas características consideradas a través de las distintas evaluaciones. Sin embargo cabe recordar que, estas medias no son directamente comparables en muchos casos, ya que, en esta tercera evaluación, se han añadido algunas instrucciones más concretas sobre cómo evaluar ciertos aspectos, y se han añadido algunos ítems (para fiabilidad, validez de constructo y baremos) evaluándose así nuevos aspectos. Sí queremos destacar que va aumentando el número de tests que realizan análisis de sesgo y funcionamiento diferencial de los ítems. Si bien tan solo se realizan estudios de DIF para dos de los 11 tests, en las evaluaciones previas publicadas en 2011 y 2013 fueron ninguno y uno, respectivamente. En esta última revisión también se observa una reducción de la proporción de tests adaptados seleccionados para su evaluación, incrementándose el número de tests originales aplicables al contexto español. En tercer lugar, cabe destacar el comienzo de la utilización de los modelos de Teoría de Respuesta a los Ítems (TRI) en alguno de los tests evaluados. Finalmente, es interesante señalar que, en algunos casos y para los tests más recientes, los manuales

siguen con bastante detalle los apartados del modelo CET, por lo que creemos que este proceso de evaluación de tests en España, al enviar un mensaje claro a las editoriales sobre qué criterios de calidad se exigen, puede estar teniendo un impacto en la mejora de la información presentada en los manuales y, consecuentemente, en la calidad de los mismos.

MIRANDO AL FUTURO

Hasta la fecha, en España se han evaluado un total de 33 tests a través del modelo CET (Prieto y Muñiz, 2000). El objetivo es continuar con el proceso de revisiones hasta que, idealmente, la práctica totalidad de tests hayan sido revisados, como sucede en otros países como Holanda (ver Evers, Sijsma, Lucassen, y Meijer, 2010). Sin embargo, los avances tecnológicos y psicométricos de los últimos años, así como la experiencia acumulada en el proceso de revisión durante estos años, hacen aconsejable tanto la revisión del modelo, como la introducción de algunos cambios en el proceso de revisión, si bien algunos de estos últimos pueden ser debatibles.

Por lo que se refiere al modelo CET, ya se señaló anteriormente que en esta tercera evaluación de tests se han modificado o incorporado algunos ítems a partir de la reciente revisión del modelo de evaluación de tests que propuso la comisión de tests de la EFPA. Sin embargo, todavía es necesario realizar una revisión más profunda del modelo europeo, e ir incorporando cambios que permitan evaluar con detalle, entre otras cosas, la administración informatizada de tests, la evaluación a distancia mediante Internet, la calidad de los informes automatizados, la aplicación de la Teoría de Respuesta a los ítems, los tests referidos a criterio y la tipificación continua ("continuous norming"). En este CET revisado, también se deberían tener en cuenta las sugerencias aportadas por Muñiz et al. (2011), para añadir cuestiones evaluativas referidas a la bibliografía, y al proceso de traducción/adaptación del test (ver también Elosúa, 2012). Esperamos contar en breve con un modelo CET revisado que permita evaluar estos aspectos en las próximas evaluaciones.

Por lo que se refiere al proceso de evaluación en sí, y relacionado con la aplicación del modelo, creemos que sería conveniente diferenciar entre la información que no se aporta porque no es aplicable al test (por ejemplo, información del proceso de adaptación cuando se trata de un test construido y baremado en muestras españolas),



de la información que no se aporta, a pesar de que sí resultaría informativa y relevante para evaluar la calidad del test. Esta falta de diferenciación puede enviar el mensaje equivocado a las editoriales sobre la importancia de aportar toda la información disponible, aunque los resultados no resulten del todo adecuados. El test, en su evaluación, podría salir mejor parado en su puntuación final si no se presenta la información, lo que iría totalmente en contra del espíritu de estas evaluaciones. Sin embargo, no debemos olvidar dos cosas. Por una parte, el psicólogo profesional también puede tener en cuenta esta diferenciación a la hora de tomar sus decisiones (no es lo mismo que no se ofrezca la información porque no procede, a que no se ofrezca siendo relevante). Por otra, el proceso de validación y la evaluación de la calidad métrica del test es un proceso continuo, que no está completamente cerrado en el momento en que se publica el manual. Nuevas ediciones de los manuales podrían ir aportando nuevos estudios e información. Coincidimos con Ponsoda y Hontangas (2013) en que esto podría incrementar el precio del manual, poniendo en peligro su comercialización. Una vía alternativa sería que esta información adicional fuera publicada por las editoriales en sus páginas web conforme se fueran realizando nuevos estudios.

Siguiendo con el proceso de aplicación del CET, cabe señalar que, a pesar de las instrucciones adicionales proporcionadas a los revisores (ver Anexo 1), todavía parecen existir algunas ambigüedades y aspectos poco claros al aplicar el modelo. Aquí resumimos algunas de las cuestiones que creemos mejorables. En primer lugar, en el apartado de Análisis de ítems, se pide una valoración de la información psicométrica de los ítems y no de si son más o menos correctos formalmente. Pero no se entiende siempre así por los revisores. Asimismo, en ocasiones, a pesar de que el análisis de ítems no aparece cómo apartado diferenciado dentro del manual, sí se realiza un análisis de ítems que pasa desapercibido para algunos revisores. Por ello es importante señalarlo en el CET, para que los revisores no marquen la opción "no se aporta información en la documentación" cuando ésta aparezca detallada dentro de otras secciones como el análisis de validez o de fiabilidad como consistencia interna. En segundo lugar, otra cuestión que creemos no está suficientemente clara es a qué se refiere el ítem sobre "la calidad de los tests empleados como criterio o marcador". Parece que, a veces, se interpreta

como la adecuación de la selección del criterio en función de las teorías sobre el constructo, y otras, como la calidad métrica de los tests empleados para medir el criterio. Además, en vez de aparecer dentro del apartado de validez de constructo, quizás sería más adecuado que este ítem apareciera en el apartado de validez criterial (concurrente y predictiva). En tercer lugar, y por lo que se refiere a los análisis de sensibilidad y especificidad que permiten evaluar la capacidad diagnóstica del test, en ocasiones los resultados son presentados como diferencias entre grupos (dentro del apartado de validez de constructo) y otras como validez predictiva. Se debería de proporcionar información explícita sobre el apartado en el que se debe incluir este análisis. En cuarto lugar, y siguiendo con validez de constructo, al evaluar la mediana de las correlaciones del test con otros tests similares la tarea no está completamente clara. Puesto que los manuales incluyen en muchas ocasiones multitud de correlaciones con distintos tests y criterios dentro de una misma tabla, se debería aclarar o proporcionar ejemplos sobre qué correlaciones deberían considerarse. Finalmente, el apartado "Bibliografía básica acerca del test aportada en la documentación" es ambiguo. Mientras que algunos revisores evalúan su adecuación y actualización, otros se limitan a mencionar algunas de las referencias presentadas. Quizás sería bueno que estas cuestiones (y otras planteadas en las evaluaciones anteriores), se introdujesen como aclaraciones en el apartado correspondiente del CET para facilitar una aplicación más estandarizada. Sin embargo, también es responsabilidad de las editoriales el presentar la información de la forma más clara posible, con apartados claros que permitan localizar la información y presentando hipótesis claras de validación.

Por lo que se refiere a los revisores, se puede debatir sobre si las valoraciones de los perfiles diferenciales seleccionados para cada test (uno más técnico-psicométrico y otro más experto en el constructo de interés) deberían ponderarse de forma diferencial en función de cuál sea el aspecto concreto evaluado en el CET. En cualquier caso, creemos que es crucial que el equipo coordinador siga disponiendo de un juego completo del test, para poder realizar una mejor integración de las valoraciones de los revisores teniendo en cuenta su perfil, y también para responder a los comentarios de las editoriales ante el informe provisional elaborado tras la revisión. Otro aspecto a tener en cuenta es que los revi-



sores con un perfil más psicométrico, en ocasiones pueden ser demasiado exigentes, y pedir información y análisis que el psicólogo profesional que va a usar el test probablemente no entienda. Creemos que estos análisis e información son importantes y pueden aportar evidencia crucial sobre la calidad psicométrica del test, por lo que una posible solución sería, de nuevo, publicar los estudios más complejos con toda la información técnica en la página web de la editorial, aumentando así la transparencia. Los revisores, cuando tuvieran en cuenta esta información, deberían hacer comentarios evaluativos dirigidos a los profesionales que pueden no tener conocimientos sobre los últimos avances psicométricos.

Sin embargo, y en relación al último aspecto mencionado, los conocimientos psicométricos de los profesionales, no debemos olvidar que es responsabilidad de todo buen usuario de tests el seguir formándose e informándose sobre los nuevos avances psicométricos. De hecho, algunos países tienen sistemas de acreditación que intentan garantizar que los usuarios de tests tienen suficientes conocimientos y competencias para el uso adecuado de las pruebas, e incluso algunos países, como Reino Unido, han puesto en marcha un sistema de acreditación europea que añade valor a esta acreditación. Cabe señalar que el COP forma parte del comité europeo de acreditación de usuarios de tests (la TUAC, Test User Accreditation Committee, representado por la primera autora) aunque, por el momento, no ha iniciado ninguna acción para poner en marcha este proceso. Si se tiene en cuenta que, los psicólogos que respondieron al último estudio sobre actitudes hacia los tests (Muñiz y Fernández-Hermida, 2010) admitieron que la formación recibida en el grado de Psicología podría no ser suficiente para la correcta utilización de la mayoría de los tests, y teniendo en cuenta que el conocimiento psicométrico avanza de tal forma que la distancia con la práctica aplicada es hoy mayor que nunca (Elosúa, 2012), creemos que podría ser interesante abrir esta vía de acreditación en España, lo que podría motivar a los profesionales a reciclarse y ponerse al día en los avances psicométricos.

En resumen, todavía estamos muy lejos de haber evaluado la totalidad de tests editados en España. Sin embargo creemos que vamos por el buen camino. Aunque todavía hay muchos aspectos mejorables, creemos que el proceso de evaluación iniciado por el COP hace ya unos años funciona bien en términos generales y está, poco a poco, teniendo una incidencia en la mejora de los tests y

los manuales. También es importante resaltar los resultados de los tests evaluados a lo largo de las distintas ediciones, que son en general satisfactorios. De hecho algunos de ellos tienen valoraciones altamente positivas. Quizás sería bueno que la comisión de tests otorgara menciones de calidad a los mejores tests, como reconocimiento y acicate hacia la excelencia.

AGRADECIMIENTOS

Queremos agradecer a los miembros de la Comisión de Tests del COP que han participado en las reuniones y en el proceso, por toda su colaboración y ayuda. En concreto a José Ramón Fernández-Hermida, Miguel Martínez, Milagros Antón, Pablo Santamaría, Viviana Gutman, Frederique Vallar y, en especial a José Muñiz y Vicente Ponsoda, coordinadores de las evaluaciones anteriores, por su continua ayuda. Queremos reiterar también nuestro agradecimiento a los revisores, por su buena disposición y profesionalidad.

NOTAS

La primera autora quiere aclarar que, a pesar de coincidir en nombre con una de las autoras de la adaptación de algunos de los tests evaluados, no se trata de la misma persona ni tiene ninguna relación profesional con la editorial.

REFERENCIAS

- Elosua, P. (2012). Tests publicados en España: Usos, costumbres y asignaturas pendientes. *Papeles del Psicólogo*, 33, 12-21.
- Evers, A., Sijtsma, K., Lucassen, W. y Meijer, R. R. (2010). The Dutch review process for evaluating the quality of psychological tests: History, procedure and results. *International Journal of Testing*, 10, 295-317.
- Evers, A., Muñiz, J., Hagemester, C., Høstmælingen, A., Lindley, P., Sjöberg, A. & Bartram, D. (2013). Assessing the quality of tests: Revision of the EFPA review model. *Psicothema*, 25, 283-291
- Muñiz, J., Bartram, D., Evers, A., Boben, D., Matesic, K., Glabeke, K., Fernández-Hermida, J.R., y Zaal, J. (2001). Testing practices in European countries. *European Journal of Psychological Assessment*, 17, 201-211.
- 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.

Ponsoda, V. y Hontangas, P. (2013). Segunda evalua-

ción de tests editados en España. *Papeles del Psicólogo*, 24, 82-90.

Prieto, G. y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-71.

ANEXO 1

OBSERVACIONES A TENER EN CUENTA POR LOS REVISORES AL RESPONDER AL CET

El CET se ha aplicado en dos procesos de revisión y en ambos hemos visto que su cumplimentación plantea algunas dificultades. En los artículos en los que se informa de ambas revisiones (publicados en la revista *Papeles del Psicólogo*, en 2011 y 2013, en los volúmenes 32(2) 34(2) respectivamente), se detallan las dificultades encontradas.

Las líneas que siguen pretenden dar alguna pista de cómo resolverlas. Si tienes dudas, no dudes en planteármelas.

- 1) EL CET es un test y, como tal, sus preguntas y opciones no deben modificarse. En las ediciones pasadas ha ocurrido alguna vez que el revisor, al no encontrar la opción de respuesta que estaba buscando, ha modificado alguna de las existentes. Debes, por tanto, responder valiéndote de las opciones que el CET ofrece. Si en alguna ocasión no encuentras la que estás buscando, elige la más similar. EL CET pide justificaciones solo de algunas respuestas, pero no de la mayoría. De todos modos, cuando se pide, por ejemplo, "comentarios sobre la validez en general", se esperan las principales justificaciones de las puntuaciones dadas a todas las preguntas sobre validez.
- 2) En principio, no se debiera dejar ninguna pregunta sin contestar. Para algún tipo especial de test, el CEP pudiera no ser del todo apropiado. Por ejemplo, la estrategia de interpretación de las puntuaciones que se considera en CET son los baremos, cuando algunas pruebas (por ejemplo, las escalas clínicas) utilizan otras estrategias de interpretación alternativas, que no están expresamente recogidas en el CET. En ese caso, se pueden dejar en blanco las preguntas correspondientes a "baremos" y responder en "comentarios sobre baremos" de los procedimientos de interpretación alternativos y de su calidad, aportando las justificaciones oportunas.
- 3) En los tests adaptados, otro asunto relevante es qué peso dar a los estudios realizados con el test original y a los estudios realizados en el proceso de adaptación. Nuestra posición en este asunto es que deben tenerse en cuenta todos los estudios aportados, si bien parece razonable que se les dé más relevancia y peso en la evaluación a los que se aporten en el proceso de adaptación.
- 4) En los tests comercializados se espera que el revisor realice su revisión a partir de la documentación que se le entrega. Ha ocurrido en el pasado alguna vez que el manual omite alguna información que el revisor puede considerar relevante para evaluar la calidad de la prueba. En ese caso, lo apropiado es que el revisor pida dicha información al coordinador, quien, a su vez, pedirá a la editorial la información requerida y verá si es posible obtenerla y en qué condiciones.
- 5) Si se ha de revisar una batería o más de un test (por ejemplo, el test normal y su versión abreviada), se pueden seguir dos estrategias. La que indica el CET es rellenar tantos CETs como tests haya que revisar. La que resulta menos costosa y también posible, si tiene sentido, sería utilizar solo un CET, dejando constancia donde corresponda de los diferentes resultados obtenidos por los distintos tests.

En las dos primeras revisiones hemos advertido que no todos asignamos el mismo significado a los términos psicométricos que el CET usa. Algunos términos que inducen a error se comentan a continuación. Cuando se pregunta en el apartado de Análisis de ítems por su calidad, se pide una valoración de la información psicométrica que el manual ofrece de los ítems y no de si su lectura nos parece que están bien o mal redactados, por poner un ejemplo. Algo similar ocurre cuando se pregunta por validez de contenido. En realidad se quiere saber qué comprobaciones se aportan sobre si el test evalúa las partes relevantes del constructo de interés