

## CUARTA EVALUACIÓN DE TESTS EDITADOS EN ESPAÑA: FORMA Y FONDO

Paula Elosua<sup>1</sup> y Kurt. F. Geisinger<sup>2</sup>

<sup>1</sup>Universidad del País Vasco. <sup>2</sup>Buros Center for Testing

La evaluación de tests y divulgación de las memorias resultantes se ha implantando en el panorama internacional con el claro objetivo de mejorar el uso de los tests, y con ello la práctica profesional. Se trata de una herramienta que ofrece al usuario de tests una revisión independiente sujeta a procedimientos y criterios definidos. En España se han llevado a cabo cuatro ediciones del proyecto de evaluación liderado por el Consejo General de la Psicología (COP). Básicamente existen dos modelos de evaluación de tests, el modelo Buros y el modelo de la European Federation of Psychologists' Associations (EFPA), utilizado por el COP. Ambos, comparten objetivos y contenido, y excluyendo varias cuestiones de formato y gestión, el fondo en ambos modelos es el mismo. En este trabajo se ofrecen los resultados de la cuarta edición del proyecto de evaluación de tests del COP adecuándolos a las pautas generales que sigue un procedimiento de evaluación de tests, y mostrando los puntos críticos en cada fase del proceso.

**Palabras clave:** Tests, Uso de tests, Evaluación de tests, Psicometría.

Assessing the quality of tests and disseminating the test results has been implemented in the international arena with the aim of improving testing and professional practice. Assessing tests provides users with an independent review, which follows clearly defined procedures and criteria. In Spain, the Spanish Association of Psychology (COP) leads the "test evaluation" project, which has recently completed its fourth edition. Basically, there are two primary assessment models; the Buros model and the European Federation of Psychologists' Associations (EFPA) model, which has been adopted by the COP. The two models share the same objectives and content; excluding the evaluation format and the management system, they are similar in essence. In this paper, we present the results of the fourth edition of the Spanish test evaluation program following the general guidelines for assessing the quality of tests, and we explain the critical points affecting each phase of the process.

**Key words:** Tests, Tests use, Assessing tests quality, Psychometrics.

La evaluación de tests y divulgación de las memorias resultantes se ha implantando en el panorama internacional con el claro objetivo de mejorar el uso de los tests, y con ello la práctica profesional. En nuestro entorno el recurso fue implementado por la Comisión de Tests del Consejo General de la Psicología el año 2010 (Muñiz et. al., 2011), y con la publicación de este trabajo se han consolidado ya cuatro ediciones (Ponsoda y Hontangas, 2013; Hernández, Tomás, Ferreres y Lloret, 2014) de una valiosa herramienta profesional cuyos aspectos formales, procedimentales y de contenido más destacados examinaremos.

Básicamente existen dos modelos de evaluación de tests; el modelo propuesto por la *European Federation of Psychologists' Associations* (EFPA) (Evers et al., 2013), que es el adoptado por el Consejo General de la Psicología de España (COP), y el modelo americano cuyo máximo exponente es Buros (Carlson y Geisinger, 2012). Un análisis del contenido y procesos seguidos por uno y por otro, revela que si bien utilizan formatos de evaluación diferentes,

las afinidades entre ellos son mayores que las divergencias. Ambos comparten la intención y el punto de partida: la publicación de revisiones independientes sujetas a procedimientos bien definidos y criterios científicos suponen un impulso en el proceso continuo de mejora en la construcción/edición/uso de tests.

El procedimiento general de evaluación podría sintetizarse en 5 etapas: (a) elección de los tests a revisar; (b) identificación de revisores; (c) evaluación del test; (d) revisión y edición; y (e) comentarios de la editorial y memoria final. Cada una de ellas es relevante en el resultado final, y sobre cada una, penden amenazas que hay que controlar para garantizar la calidad de la memoria final, y con ello, que la evaluación cumpla con los objetivos perseguidos.

Antes de revisar los modelos evaluativos de tests, cabría recordar las diferencias estructurales entre los órganos que los gestionan. Buros es una organización independiente sin ánimo de lucro perteneciente a la Universidad de Nebraska, en la que trabajan 12 personas y cuenta con 8 post-graduados colaboradores. El proceso de revisión español descansa sobre un coordinador general nombrado al efecto por la Comisión de Tests del

Correspondencia: Paula Elosua. Facultad de Psicología. Av. Tolosa, 70. 20080 San Sebastian. España.

E-mail: paula.elosua@ehu.es



COP. El coordinador se encarga de la gestión administrativa y científica, si bien cuenta para la primera tarea con el apoyo del COP, y tiene libertad para nombrar un equipo de colaboradores que le asistan.

Buros tiene una dilatada experiencia en la evaluación de tests que se inició el año 1938 con la publicación del primer *Mental Measurement Yearbook* (MMY, Buros, 1938). Los MMY son compendios que recogen las memorias de evaluación de los tests que han sido gestionadas por este instituto. Hasta la fecha se han publicado 19 volúmenes, en los que se han revisado aproximadamente 10500 tests. BUROS edita además dos volúmenes relacionados con la evaluación de tests que se comercializan con los títulos *Tests in Print* (TIP) y *Pruebas Publicadas en Español* (PPE; Schlueter, Carlson, Geisinger, y Murphy, 2013). Son obras con un formato enciclopédico en las que se ofrecen listados descriptivos de tests disponibles en el mercado. El incremento de la población hispanohablante en los Estados Unidos impulsó la creación y publicación de PPE. El volumen recoge información sobre tests disponibles en los Estados Unidos que están íntegramente publicados en español o que aportan algún componente en este idioma. La edición actual incluye 422 entradas organizadas alfabéticamente por el nombre del test (Carlson y Gonzalez, 2015).

### ELECCIÓN DE LOS TESTS A REVISAR

La elección de los tests a evaluar se inicia de forma general, con la solicitud de revisión por parte de los editores. En España, las peticiones se tramitan a través de la Comisión de Tests del Consejo General de la Psicología (COP) (ver tabla 1; tests revisados en la 4<sup>a</sup> edición), que en la actualidad cuenta entre sus miembros con representantes de 3 editoriales. En Buros las peticiones se dirigen directamente a sus oficinas en la Universidad de Nebraska. Este instituto somete las solicitudes recibidas a un primer cribado con el objetivo de comprobar que los tests cumplen ciertas condiciones generales, entre las que se encuentran, (a) aportar datos psicométricos; (b) disponer de un manual de administración/corrección/interpretación; y (c) presentar un manual técnico. Sin estos requisitos los tests no se evalúan. La consecuencia directa de esta obligación es que los tests no comercializados no entran dentro del proceso de revisión. En España la comisión de tests tiene potestad para proponer la revisión de tests no comercializados. De hecho durante la segunda edición se evaluó un test no comercial; dado que la información y elementos adicionales que acompañan a una publicación en formato de artículo científico no son comparables, ni en forma ni en fondo, al material que compone un test comercializado por una editorial, la revisión

del test planteó necesidades adicionales al proceso de evaluación (Ponsoda y Hontangas, 2013).

Tras la elección de los tests, las editoriales remiten 3 ejemplares completos al órgano responsable de la revisión. El número de copias enviadas hace posible que el coordinador o la oficina de gestión, en su caso, dispongan de un juego de cada test que permitirá en todo momento contrastar la información de evaluación. En este punto del proceso Buros genera una ficha descriptiva sobre las características del test que incluye información sobre el título, objetivos, puntuaciones, población destinataria, tiempo de administración, acrónimo, editorial, fecha de publicación, autores, formato de administración y precio.

### IDENTIFICACIÓN DE REVISORES

El proceso de evaluación descansa sobre el juicio independiente y no sesgado de los revisores, en su capacidad y experiencia para analizar la documentación y los materiales de que consta el test, y para elaborar argumentos justificados e imparciales. La atinada identificación de profesionales aptos para esta labor es esencial.

El objetivo de la fase de identificación de revisores es conseguir para cada test a evaluar una pareja de especialistas con perfiles complementarios; uno con una formación más sólida en medida, y un segundo con un alto nivel de experiencia en el área que mide el test. Dispo-

TABLA 1  
LISTA DE TESTS EVALUADOS

Acrónimo	Test	Editorial
ABAS-II	Sistema de Evaluación de la Conducta Adaptativa	TEA Ediciones
BADyG/M-r	Batería de Aptitudes Diferencial y Generales. Nivel M Renovado	CEPE, S.L.
BETA	Batería para la Evaluación de los Trastornos Afásicos	Instituto de Orientación Psicológica EOS
BSI-18	Inventario Breve de 18 Síntomas	Pearson Educación
CECAD	Cuestionario Educativo-Clinico: Ansiedad y Depresión	TEA Ediciones
EHPAP	Evaluación de Habilidades y Potencial de Aprendizaje para Preescolares	Instituto de Orientación Psicológica EOS
PAIB-1	Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas	CEPE, S.L.
PECC	Prueba para la Evaluación de la Cognición Cotidiana	Instituto de Orientación Psicológica EOS
SCIP	Screening del Deterioro Cognitivo en Psiquiatría	TEA Ediciones
WMS-IV	Escala de Memoria de Wechsler -IV	Pearson Educación
WPPSI-IV	Escala de Inteligencia de Wechsler para Preescolar y Primaria	Pearson Educación



ner de ambos perfiles profesionales permite contar con fuentes de opinión construidas sobre diferentes perspectivas y experiencias. Sus aportaciones, en conjunto, ayudarán a obtener evaluaciones más comprensivas que incluyan aspectos formales, sustantivos y aplicados.

Según los criterios empleados por el COP y por Buros, los revisores han de acreditar experiencia en el uso/construcción de tests. Buros cuenta con una base de datos de 900 revisores entre los que selecciona los que considera más adecuados, y sobre los que impone ciertas restricciones para evitar sesgos; por ejemplo, dos revisores de un mismo test no han de pertenecer a la misma universidad, o editorial; y se intenta compaginar revisores experimentados con revisores más jóvenes. En España la labor de identificación y selección corre a cargo del coordinador, que a través de estudios bibliográficos intentará ajustar los perfiles necesarios para asegurar evaluaciones de calidad (tabla 2; revisores de tests en la 4ª edición). Como compensación al tiempo y esfuerzo dedicados en la revisión, Buros envía una copia gratuita del MMY en el que se ha participado; el COP ofrece cincuenta euros a cada revisor. En ambos casos los evaluadores se quedan en propiedad con el (los) ejemplar(es) del test revisado.

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

Revisores	Afiliación
Juan Antonio Amador	Universidad de Barcelona
Sonia Alfonso Gil	Universidad de Vigo
Constantino Arce	Universidad de Santiago de Compostela
Juan Ramón Barrada	Universidad de Zaragoza
Roberto Colom	Universidad Autónoma
Ana Delgado	Universidad de Salamanca
Eduardo Doval	Universidad Autónoma de Barcelona
Sergio Escorial	Universidad Complutense
Sara Fernández Guinea	Universidad Complutense
María Paloma González Castro	Universidad de Oviedo
Francisco Gutiérrez Martínez	Universidad Nacional de Educación a Distancia
José Antonio López Pina	Universidad de Murcia
Urbano Lorenzo Seva	Universitat Rovira i Virgili
Luis María Lozano	Universidad de Granada
Ignacio Montorio	Universidad Autónoma de Madrid
Juan Antonio Moriana	Universidad de Córdoba
Cristino Pérez	Universidad de Granada
Luz Pérez Sánchez	Universidad Complutense
Pedro Prieto Marañón	Universidad de La Laguna
Antonio José Rojas	Universidad de Almería
Bonifacio Sandín	Universidad Nacional de Educación a Distancia
Paula Samper	Universidad de Valencia

El contacto inicial con los revisores se lleva a cabo por email; en él se informa del proyecto de evaluación de tests, se invita a revisar un test concreto, se aconseja el rechazo de la revisión en situaciones de conflicto, y se fija un plazo de revisión orientativo de 3 meses. En la 4ª edición se pidió explícitamente a los revisores que firmaran una declaración de inexistencia de conflicto de intereses para asegurar la imparcialidad e independencia de la revisión.

Una vez aceptada la evaluación se envía el material a los revisores; los revisores realizan su trabajo con el mismo material que obtendría un usuario cuando compra el test. Junto al material cedido por la editorial, se envía el cuestionario de evaluación de tests (CET; Prieto y Muñiz, 2000), que en la tercera edición sufrió algunas modificaciones, y un documento con información aclaratoria sobre algunas de las cuestiones que figuran en el mismo (Hernández, Tomás, Ferreres y Lloret, 2014). Buros remite una guía sobre el proceso de revisión y las expectativas que se tienen respecto al mismo.

### EVALUACIÓN DEL TEST

La evaluación de los tests es la fase más delicada del proceso. Los revisores han de elaborar una memoria evaluativa del test basándose en el análisis detallado de la información aportada en el manual, y de los materiales que con él se acompañan.

La evaluación consiste en completar un cuestionario que guía al revisor sobre los aspectos más importantes a tener en cuenta durante su trabajo. Los formatos de los cuestionarios evaluativos utilizados por Buros y por el COP son diferentes. El primero trabaja sobre un formulario con ítems de ensayo ante los que se espera que el revisor exprese un juicio coherente e informativo. El formulario contempla 5 puntos a desarrollar: (a) descripción general del test; (b) construcción; (c) informe técnico con un análisis de la estandarización, fiabilidad y validez; (d) comentarios generales en los que expresan: (1) las fortalezas y debilidades del test; (2) referencias a investigaciones actuales que apoyen el modelo teórico sobre el que se ha construido el test, y (3) evidencias que respalden su uso; finalmente (e) un resumen corto en 6 o 7 frases a través de las cuales el revisor sintetiza sus conclusiones generales y recomendaciones sobre el uso correcto del test.

El cuestionario de evaluación de tests (CET; Prieto y Muñiz, 2000) utilizado por el COP combina respuestas cerradas y respuestas abiertas de ensayo. El CET está dividido en tres secciones: (a) descripción general del test; (b) valoración de las características del test y (c) valoración global del test. El primero se correspondería con una ficha técnica de descriptores sobre el test que ofrece información sobre



aspectos como el título, los objetivos, las puntuaciones o el tiempo de aplicación. Se correspondería, de forma general, con los descriptores que Buros prepara en la fase de selección de tests. El segundo apartado contiene 37 ítems de respuesta graduada (inadecuada, adecuada con carencias, adecuada, buena, excelente) con el propósito de cuantificar la calidad de los estudios de fiabilidad, validación y baremación. Incluye también ítems de ensayo sobre estos 3 últimos aspectos en los que el revisor genera un juicio fundamentado. La última parte del CET se destina a una única pregunta abierta en la que pide al revisor comentarios generales sobre el test que incluye fortalezas/debilidades y una síntesis de la información más relevante para el uso correcto del test.

Con independencia de que el número de secciones de los cuestionarios sea 3 o 5, ambos formatos, abordan los mismos puntos en la evaluación; lo cual, puede interpretarse en términos de validación concurrente. Siendo similares en contenido, cabría preguntarse si existen ventajas o desventajas en la utilización de uno u otro formato. La respuesta a esta cuestión vendría dada por los efectos que tiene la utilización de ambos tipos de ítem sobre cualquier test estandarizado. La naturaleza del ítem afecta entre otros aspectos, a la puntuación, a la representación de constructo y a la facilidad/dificultad de construcción de ítems (Haladyna, 2004). Los ítems de respuesta cerrada, siempre y cuando estén bien contruidos: (a) son versátiles; (b) aseguran la evaluación de los puntos significativos del dominio evaluado; (c) permiten una cuantificación rápida y eficiente; (d) son más sencillos de responder; y (e) permiten un análisis de ítems. Sin embargo, también presentan inconvenientes. En formatos de respuesta cerrados y evaluaciones complejas la riqueza y especificidad de la evaluación puede en alguna ocasión no estar perfectamente reflejada por las opciones de respuesta. Por ejemplo, en lo que al CET se refiere, la valoración de la calidad de una muestra de baremación se analiza de igual modo para un cuestionario de aptitudes intelectuales destinado a población comunitaria, que para un cuestionario diagnóstico que trabaja con grupos clínicos. En ambas circunstancias se analiza el número de participantes en la investigación, siendo valores difíciles de comparar en la medida en que provienen de poblaciones diferentes. Tal vez, en estas situaciones, pudieran añadirse alguna nota con referencia al tipo de muestra/población.

El CET añade a las ventajas de la evaluación por ensayos, una valoración cuantitativa sobre las características del test, que puede leerse como una instantánea de los puntos evaluados. Es un aspecto complementario de las preguntas de ensayo y añade valor a la evaluación, pero

no la sustituye ni la reemplaza. El hecho de ofrecer valores cuantitativos para cada uno de los apartados no legitima el cálculo de una media final para la valoración de la calidad de un test. La validación de un test ha de tener siempre en cuenta el uso propuesto de las puntuaciones (AERA, APA y NCMEA, 2014; Elosua, 2003), y la selección de un test por parte de un usuario potencial ha de considerar el contexto de aplicación; si la inferencia de las puntuaciones se relaciona con la predicción de un comportamiento futuro, por excelente que sea la calidad de los materiales o la muestra de baremación, por muy altas que sean las puntuaciones en esos apartados, será el aspecto relacionado con la validación predictiva uno de los primeros criterios a observar; y una media aritmética calculada sobre todos los aspectos evaluados no ofrece esta información.

Con respecto a los ítems del CET en las 4 ediciones llevadas a cabo hasta la fecha por el COP se han detectado algunos ítems cuya formulación actual genera dudas al evaluador y que como ha sido advertido convendría repasar (Hernández, Tomás, Ferreres y Lloret, 2014; Muñiz, Fernandez-Hermida, Fonseca-Pedrero, Campillo-Álvarez y Peña-Suarez, 2011; Ponsoda y Hontangas, 2013). Para garantizar una mejor comprensión del contenido de los ítems se elaboró en la tercera edición (Hernández et.al, 2013) un documento de apoyo que con algunas modificaciones también se incluyó en esta ocasión. Apostamos por esta línea de mejora del CET marcada por una revisión de algunos de los ítems y por la conveniencia de añadir enunciados explicativos tal y como lo hace la versión original del cuestionario de la EFPA (<http://www.efpa.eu/professional-development>). Para ello sería aconsejable, como señalan Ponsoda y Hontangas (2013), trabajar sobre una versión informatizada del CET que facilite su cumplimentación y permita añadir hipervínculos con enunciados explicativos.

## REVISIÓN Y EDICIÓN

Una vez recibidas las dos evaluaciones de cada uno de los tests, el proceso editorial continúa revisando los informes recibidos. En este punto Buros y el COP trabajan de forma diferente, lo cual es lógico teniendo en cuenta que las memorias finales que ofrecen ambos organismos son distintas. Buros publica para cada tests las dos revisiones efectuadas; el COP genera un único informe que sintetiza el trabajo de los dos revisores que analizaron el test.

Buros fija su atención en la verificación de la exactitud de los datos, y en cuestiones de estilo, con el objetivo de conseguir una memoria final acorde con su línea editorial. Una vez revisados y corregidos los textos publica en sus MMY las memorias remitidas por los revisores. Para la revisión factual Buros cuenta con 2 o 3 estudiantes de



doctorado que inspeccionan y corrigen aspectos como las referencias o el formato.

El COP elabora un único texto-memoria por cada test, que condensa la información cuantitativa a la par que resume las valoraciones ofrecidas por los dos revisores. Editar la evaluación significa básicamente revisarla y enriquecerla, asegurando que no existan errores sobre datos objetivos, y garantizando que el lenguaje utilizado sea homogéneo, correcto, adecuado y no hiriente. El editor analiza y corrige en su caso, el tono de la revisión, asegurando la correcta utilización de los términos psicométricos y la adecuación lingüística al perfil del usuario potencial; se evita el uso de tecnicismos o en determinados casos una jerga excesivamente metodológica. Sin negar el punto de vista del revisor, en el proceso de edición se eliminan comentarios negativos, y se amoldan las revisiones para que ofrezcan un resumen claro, y preciso sobre aquellas características del test que pueden ser importantes para el potencial usuario. La edición no altera, en natura, los juicios profesionales emitidos por los evaluadores, pero sí corrige cualificaciones poco descriptivas. El objetivo del proceso de revisión, no es resaltar las limitaciones o carencias de un test, sino ponerlas en un contexto adecuado.

En la elaboración de la memoria del test asumimos que el objetivo final de la revisión no es otorgar a un test una calificación absoluta que el usuario adopte como medida de calidad y justificación de uso, sino que defendemos una lectura crítica y no pasiva marcada por un criterio profesional en el que es importante el contexto de uso del test. Como señaló Thorndike (1999) las revisiones han de venir acompañadas de un juicio maduro y del conocimiento sobre la situación en la que quiere aplicarse el test; ambos juicios corresponden al usuario.

Una vez editadas las memorias Buros las envía a los revisores para que den su aprobación a las modificaciones introducidas. En el modelo seguido por el COP se obvia este paso porque las revisiones no se publican en su formato original.

#### COMENTARIOS DE LA EDITORIAL Y MEMORIA FINAL

Finalizado el proceso editorial los informes se envían a las editoriales de los tests evaluados para su revisión. Éstas, habitualmente envían los informes a los autores de los tests. Unos y otros, autores y editores redactan y remiten sus comentarios al coordinador. Esta etapa es importante para garantizar que las memorias no contengan errores o datos inexactos; en caso de que se detecten se corrigen antes de publicarlas.

Buros ha establecido que si en esta fase se solicitan modificaciones profundas, éstas se reenvían directamente al

revisor para su aprobación o rechazo. El COP no contempla esta opción.

El objetivo de esta fase es garantizar la precisión y exactitud de la información incluida en la memoria. Nuestro modelo, aún joven, no tiene pautas marcadas sobre el cometido de las editoriales en esta fase, y existe variabilidad en las respuestas de los editores; desde aquellos que se ciñen a la corrección de inexactitudes, hasta las que envían material complementario como artículos científicos o tesinas en las que se ha utilizado el test en cuestión. Sería importante establecer pautas para que el autor/editor fije su atención sobre la información factual; y centre sus comentarios en el material evaluado por el revisor.

Una vez corregidos los errores y analizadas las sugerencias, se emite la memoria final de cada uno de los tests, que formarán parte del MMY o se publicarán en la página del COP (<https://www.cop.es/index.php?page=evaluacion-tests-editados-en-espana>).

#### DISCUSIÓN Y CONCLUSIONES

Cada vez está más extendida la idea de que una buena práctica profesional incluye el uso de tests que hayan sido evaluados por organismos externos (Carlson y Geisinger, 2012; Evers, Sijtsma, Lucassen, y Meijer, 2010; Ponsoda y Hontangas, 2013) y que por tanto la evaluación de tests es una herramienta de apoyo en la labor profesional del psicólogo.

Si el objetivo de la evaluación es mejorar la práctica profesional cabría preguntarse qué información necesita un usuario potencial para considerar el uso de un test en una situación aplicada o de investigación. Esa es la cuestión de fondo en el proceso evaluativo, y de ella deriva la importancia de: (a) un manual claro y bien estructurado; (b) la buena calidad de los materiales; (c) la descripción técnica del test con información precisa sobre los estudios su construcción, fiabilidad, validación y baremación, (d) la necesidad de justificar el uso de las puntuaciones en un contexto concreto, y (e) las referencias a investigaciones recientes relacionadas con el modelo teórico sobre el que se ha construido el test.

Los puntos abordados por la evaluación de tests, independientemente de que el formato sea más o menos abierto, se convierten en un *gold standard* para la construcción/edición test. En tanto en cuanto se constituye en un criterio externo de evaluación, el CET ha de cumplir requisitos relacionados con su validez y ofrecer una representación completa del dominio evaluado. Pero además de la fuente de evidencia criterial, es importante considerar su aspecto consecuencial. La evaluación de



tests tiene consecuencias directas sobre todas las fases implicadas en la construcción/edición de tests, en tanto en cuanto induce mejoras en el proceso de construcción y estimula a las editoriales a adoptar criterios cada vez más rigurosos para la publicación de tests y manuales. Pero no es la única consecuencia. El impacto profesional de la evaluación de tests está teniendo efectos en la formación de grado y postgrado del psicólogo. Dada la importancia de la herramienta, y la necesidad de una lectura activa de las memorias de evaluación, la revisión de tests se está incorporando a los temarios de formación en psicología en las asignaturas relacionadas con la psicometría y la evaluación.

Otra consecuencia importante, especialmente en los Estados Unidos, es la relacionada con la imparcialidad de los tests y de las evaluaciones con referencia a diferentes grupos poblacionales; en este sentido, se anima a los revisores a que aporten observaciones sobre la idoneidad de la aplicación y uso de los tests en diferentes grupos, sean grupos étnicos, culturales, de capacidades diversas...

La aceptación de la revisión de tests por parte de la academia, de los autores, de los editores y de los usuarios es positiva. Las editoriales son aliadas en este proceso, y nos consta, tanto por su actitud como por su disposición, que convergen con las líneas marcadas por el COP.

El plan de revisión de tests se inscribe en un ambicioso

programa de proyectos para mejorar el uso de los tests (ver Elosua y Muñiz, 2013) que han sido diseñados y llevados a buen término en un momento definido claramente por un desarrollo de la psicometría al que comienzan a referirse en el exterior como escuela española de psicometría. Las cuatro ediciones llevadas a cabo hasta la fecha permiten concluir un balance positivo del proyecto. Pero la tarea no ha finalizado, la revisión de tests como herramienta y *gold standard* exige de un trabajo continuo de mejora, procedimental, formal y sustantiva, para poder cumplir con su objetivo: ofrecer información relevante sobre los tests disponibles al profesional de la psicología. En este sentido conviene recordar las anotaciones apuntadas en las ediciones anteriores que afectan a contenidos concretos y propuestas de mejora, que están ya siendo introducidas en nueva versión del CET.

Amparándonos en la todavía joven experiencia de revisión de tests del COP, consideramos que sería conveniente fijar una línea editorial clara que recoja y sistematice los objetivos y expectativas para cada una de las 5 etapas en las que hemos dividido la evaluación, los objetivos y expectativas. Es necesario formalizar un marco teórico y procedimental para cada uno de los agentes implicados (coordinador, evaluador, editorial/autor), que defina las funciones de cada uno de ellos en cada fase del proceso. Dado que estamos trabajando con un cues-

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

Características	TESTS										
	ABAS-II	BADyG/M-r	BETA	BSI-18	CECAD	EHPAP	PAIB-1	PECC	SCIP	WMS-IV	WPPSI-IV
Calidad de los materiales y Documentación	4,5	4	3	3	4,5	4	4	3,5	4	5	4,5
Fundamentación teórica	4	3	3,5	2	4	3	3	4	4	4,5	5
Adaptación española	4	NO PROCEDE	NO PROCEDE	3	NO PROCEDE	3	NO PROCEDE	NO PROCEDE	NO PROCEDE	4	5
Análisis de los ítems	--	4	--	--	5	--	--	3	3	--	3,5
Validez de contenido	3,5	3	3	3	4	2	3	3	3	4	5
Validez de constructo	5	3	4	2	5	3	2	3,5	4	4	4
Análisis del sesgo	--	--	--	--	--	--	--	--	--	--	2
Validez predictiva	--	3,5	2	--	--	3	4	2	3,5	4	3,5
Fiabilidad: consistencia interna	5	4,5	3,5	5	5	3,5	5	3	4	4	3,5
Fiabilidad: estabilidad	2,5	--	--	3,5	--	--	--	2,5	3,5	3	3,5
Fiabilidad: equivalencia	--	--	--	--	--	--	--	--	3	--	--
Baremos	4,5	4,5	2,5	3,5	4,5	2,5	3,5	2,5	4	4,5	4,5

Nota. -- No se aporta información.  
Escala de evaluación: 1 (Inadecuada) – 2 (Adecuada con carencias) – 3 (Adecuada) – 4 (Buena) – 5 (Excelente)



tionario, el objetivo sería estandarizar las condiciones de aplicación y asegurar que los objetivos de la evaluación y el procedimiento sean entendidos por todos. La necesidad de una línea editorial cobra especial relevancia en el modelo COP porque si bien la memoria final mantiene la intención y el tono aportado por los revisores, la forma se altera y esta circunstancia puede generar en algunas ocasiones desconcierto al profesional que llevó a cabo la revisión.

En conclusión, creemos que convendría que para futuras ediciones se definiera un equipo editorial plurianual que se encargara de redactar las líneas editoriales del COP en el proceso de revisión, y pautara los objetivos y expectativas para cada una de las etapas del proceso. Coincidimos con Carlson y Geisinger (2012) en que la revisión de tests ha de estimular avances hacia estándares profesionales de construcción de tests que primen el buen trabajo y desaprueben el pobre, al mismo tiempo que impulse a los autores y editores a elaborar manuales que incluyan información detallada sobre la construcción, estandarización, propiedades psicométricas y usos adecuados y no adecuados del test. Pero no puede dejarse atrás la incorporación continua de avances psicométricos para que la distancia entre la teoría y práctica deje de ser tan amplia (Elosua, 2012). Es claro que la psicología y la práctica profesional ya no son las mismas que al principio de su historia. Los modelos psicométricos han progresado, las teorías han evolucionado, y los problemas y necesidades de la práctica psicológica cambian y se adaptan a las necesidades de cada momento. Desde la publicación de los primeros tests conceptos como fiabilidad y validez se han enriquecido y los requisitos científicos, profesionales y éticos exigidos por los tests se han adaptado a nuevas necesidades (De Boeck y Elosua, in press). En esta evolución el proceso de evaluación continua de tests juega un papel relevante de estímulo y acicate para el acercamiento entre la teoría y la práctica profesional.

#### AGRADECIMIENTOS

Agradecemos a los revisores de la cuarta edición de tests y a las editoriales y autores de los tests evaluados su buena disposición, dedicación y profesionalidad.

#### REFERENCIAS

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological tests*. Washington DC: American Educational Research Association.

- Buros, O. (1938). *The 1938 Mental Measurements Yearbook*. Hingham Park, NJ: Gryphon.
- Carlson, J. F., y Geisinger, K. F. (2012). Test Reviewing at the Buros Center for Testing. *International Journal of Testing*, 12, 122-135.
- Carlson, J. F., y Gonzalez, S. E. (2015). Using Pruebas Publicadas en Español to enhance test selection. In K.F. Geisinger (Ed.), *Psychological testing of Hispanics (2<sup>nd</sup> ed.): Clinical and intellectual issues* (pp. 11-27). Washington, DC: American Psychological Association.
- De Boeck, P., y Elosua, P. (in press). Reliability and Validity: History, Notions, Methods, Discussion. En D. Iliescu (ed.). *International Handbook of Testing and Assessment*.
- Elosua, P. (2003). Sobre la validez de los tests. *Psicothema*, 15, 315-321.
- Elosua, P. (2012). Tests publicados en España: Usos, costumbres y asignaturas pendientes. *Papeles del Psicólogo*, 33, 12-21.
- Elosua, P., y Muñoz, J. (2013). Proyectos españoles para una mejora en el uso de los tests. *Revista Latinoamericana de Ciencia Psicológica*, 5, 139-143.
- 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., Muñoz, 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
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hernández, A., Tomás, I., Ferreres, A., y Lloret, S. (2015). Tercera evaluación de tests editados en España. *Papeles del Psicólogo*, 36, 1-8.
- Muñoz, 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 evaluación de tests editados en España. *Papeles del Psicólogo*, 24, 82-90.
- Prieto, G., y Muñoz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-71.
- Schlueter, J., Carlson, J. F., Geisinger, K. F., Murphy, L. L. (Eds.) (2013). *Pruebas publicadas en Español*. Lincoln, NE: Buros Center of Testing