

REVISIÓN DEL MODELO PARA EVALUAR LA CALIDAD DE LOS TESTS UTILIZADOS EN ESPAÑA

ASSESSING THE QUALITY OF TESTS IN SPAIN: REVISION OF THE SPANISH TEST REVIEW MODEL

Ana Hernández¹, Vicente Ponsoda², José Muñiz³, Gerardo Prieto⁴ y Paula Elosua⁵

¹Universidad de Valencia. ²Universidad Autónoma de Madrid. ³Universidad de Oviedo.

⁴Universidad de Salamanca. ⁵Universidad del País Vasco

Para usar adecuadamente los tests, es necesario que los profesionales cuenten con información rigurosa de su calidad. Es por ello que, desde hace unos años, se viene aplicando el modelo español de evaluación de la calidad de los tests (Prieto y Muñiz, 2000). El objetivo de este trabajo es actualizar y revisar dicho modelo, con el fin de incorporar las recomendaciones hechas en sus aplicaciones, y para incorporar los avances psicométricos y tecnológicos que se han producido durante los últimos años. El modelo original fue revisado en varias fases, y la revisión originalmente propuesta fue revisada por un conjunto de expertos, lo que dio lugar a la versión final que se describe en este trabajo. Se espera que la aplicación del modelo revisado y la publicación de los resultados correspondientes, contribuya a seguir mejorando el uso de los tests y, con ello, la práctica profesional de la Psicología.

Palabras clave: Uso de tests, Modelos de evaluación, Calidad de los tests.

For professionals to use tests adequately, they must have rigorous information about the quality of the tests. This is why during the last years, the Spanish test review model (Prieto & Muñiz, 2000) has been applied. The goal of this paper is to update and revise this model in order to incorporate the recommendations given after the original model was applied and, to incorporate the latest psychometric and technological innovations. The original model was revised following different steps, and the revised proposal was reviewed by a number of experts. After incorporating their suggestions, we arrived to the final version, which is described in this paper. With the application of the revised model, and the publication of the corresponding results, we expect to continue improving the use of tests, and consequently, the professional practice of Psychology.

Key words: Test use, Review models, Test quality.

Como bien es sabido, los tests son una herramienta básica para la práctica profesional de la Psicología, ya que pueden ser de gran utilidad sea cuál sea el área de especialización profesional: social, educativa, clínica, deportiva, jurídica, organizacional, etc. En España, algunos datos que avalan que los psicólogos utilizan los tests como herramienta básica en sus quehaceres diarios, los encontramos al acudir a la encuesta diseñada por la EFPA (*European Federation of Psychological Associations*) para conocer las opiniones de los psicólogos sobre el uso de los tests. Cuando se preguntó a los psicólogos españoles colegiados por el grado en que utilizaban habitualmente tests en su desempeño profesional, las medias obtenidas en las distintas áreas (psicología clínica, organizacional, educativa y otras), estuvieron cercanas a 4 en una escala de 5 puntos (Muñiz y Fernández-Hermida, 2010). Todavía se obtuvieron puntuaciones superiores en los ítems en los que se reconoce que los tests constituyen una excelente fuente de información si se combinan con otros datos, y que, utilizados correctamente, son de gran ayuda para el psicólogo.

Sin embargo, también es sabido que, para que los tests sean una herramienta realmente útil, estos deben tener calidad y rigor demostrados. Y los psicólogos, como usuarios de tests, deben ser

competentes y tener información contrastada que les permita elegir los tests que sirvan con rigor psicométrico a su propósito. En este sentido, los psicólogos españoles, en la encuesta antes citada, manifestaron necesitar más información (revisiones independientes, investigaciones, documentación, etc.) sobre la calidad de los tests editados en nuestro país.

Es en este contexto donde surgen los modelos de evaluación de la calidad de los tests y, en concreto, el modelo español (Prieto y Muñiz, 2000). Estos modelos tienen en común que definen una serie de criterios de calidad teóricos, prácticos y/o psicométricos, y son evaluados siguiendo un procedimiento estandarizado para, posteriormente, dar a conocer los resultados de la evaluación. El objetivo último es claro: proporcionar a los usuarios de tests información precisa y accesible acerca de la calidad de los tests disponibles. Entre los modelos propuestos, y sus correspondientes procesos de aplicación, destacan el seguido por el *BUROS Center for Testing* americano y el propuesto por la Federación Europea de Asociaciones de Psicología (EFPA), junto con modelos locales como el holandés, el británico, y, como no, el modelo español.

El modelo español fue impulsado por el Colegio Oficial de Psicólogos (COP), y culminó con la publicación del CET (Cuestionario de Evaluación de Tests) en el año 2000 (Prieto y Muñiz, 2000). Sin embargo, el CET no se aplicó hasta años más tarde, con el primer proceso de evaluación de tests promovido por el COP y su comisión de tests, el cual finalizó en 2011 (ver Muñiz, Fernández-Hermida, Fonseca-Pedrero, Campillo-Álvarez y Peña-Suarez, 2011). Desde entonces, el modelo, con al-

Correspondencia: Ana Hernández. IDOCAL y Departamento de Metodología de las Ciencias del Comportamiento. Facultad de Psicología. Universitat de València. España.

E-mail: Ana.Hernandez@uv.es



gunas pequeñas modificaciones, se ha aplicado de forma sistemática hasta la fecha, siempre a través de evaluadores independientes. Los resultados obtenidos se han dado a conocer a través de la página web del COP. Los resúmenes de los cuatro procesos de evaluación llevados a cabo hasta la fecha, así como sus principales resultados, se han recogido en los trabajos de Muñiz et al. (2011), Ponsoda y Hontangas (2013), Hernández, Tomás, Ferreres y Lloret (2014), y Elosua y Geisinger (2016).

Como resultado de estas evaluaciones, y gracias a la experiencia acumulada en la aplicación del modelo, se han ido proponiendo y aplicando una serie de recomendaciones para mejorar tanto el proceso de evaluación de tests como el modelo mismo (ver Elosua y Geisinger, 2016; Hernández et al., 2014; Muñiz et al., 2011; y Ponsoda y Hontangas, 2013). En lo que se refiere al modelo, algunas de estas recomendaciones tienen que ver con la necesidad incorporar los avances psicométricos y tecnológicos que se han producido durante los últimos años. De hecho, el modelo europeo de la EFPA ha sido recientemente revisado y actualizado (Evers et al., 2013) con el fin de incorporar dichos avances. El objetivo es acercarlos a la práctica profesional del psicólogo y, con ello, contribuir a reducir el salto que existe frecuentemente entre investigación y práctica profesional (Elosua, 2012).

Teniendo en cuenta estas consideraciones, el objetivo del presente trabajo es presentar la revisión y actualización del CET. Para esta revisión, junto con las recomendaciones realizadas en los distintos procesos de evaluación en que se ha aplicado el CET, se ha tenido muy en cuenta el modelo EFPA recientemente revisado y actualizado (Evers et al., 2013).

Con la aplicación de este modelo revisado (CET-R), se espera aumentar la claridad y riqueza de información proporcionada en el proceso de evaluación. La consiguiente publicación de los resultados obtenidos con el modelo revisado contribuirá a diseminar información más exhaustiva y rigurosa sobre la calidad de los tests y sus puntos débiles. Con ello esperamos continuar contribuyendo a la estrategia informativa iniciada por el COP hace años, con el fin de mejorar el uso de los tests y, consecuentemente, la práctica profesional de la Psicología.

A continuación, en primer lugar, se presentan brevemente los principales modelos de evaluación existentes en el ámbito internacional y, en segundo lugar, se presenta el modelo CET original. En tercer lugar, se describe el proceso seguido en la revisión del modelo y se destacan las principales novedades del modelo revisado (CET-R), para terminar con unas breves conclusiones.

ALGUNOS MODELOS INTERNACIONALES DE EVALUACIÓN DE TESTS

Los pioneros en presentar sistemáticamente información sobre la calidad de los tests fueron los norteamericanos, a través de BUROS, que es un instituto de evaluación asociado a la Universidad de Nebraska. Fue en 1938 cuando se publicó la primera edición de la serie *Buros' Mental Measurements Yearbooks*, con los resultados de las evaluaciones realizadas. Esta serie ha continuado publicándose de manera periódica hasta la fecha (Buros, 1938; Carlson, Geisinger, y Jonson,

2014). Cabe destacar que, desde hace unos años, y dado al incremento de hispano-hablantes en Estados Unidos, BUROS tiene publicaciones especiales dedicadas a los tests publicados en español (ver, por ejemplo, Carlson y González, 2015). El proceso y las características de la evaluación de BUROS, así como las similitudes y diferencias entre la evaluación llevada a cabo por BUROS y la que se ha realizado hasta la fecha al aplicar el CET, pueden consultarse en Elosua y Geisinger (2016).

Centrándonos en Europa, los progresos en esta área han sido impulsados principalmente por los colegios de psicólogos correspondientes, a través de sus comisiones de tests. Los primeros en realizar evaluaciones sistemáticas de los tests y publicar los resultados de dichas evaluaciones fueron los holandeses. En concreto, la primera publicación se hizo en 1969 (NIP, 1969). El modelo entonces utilizado ha sido revisado en cinco ocasiones, realizándose en 2009 la última revisión (Evers, Braak, Frima, y Van Vliet-Mulder, 2009). Una descripción más detallada de la historia, procedimiento y resultados obtenidos de las evaluaciones realizadas por los holandeses puede consultarse en el artículo de Evers, Sijtsma, Lucassen y Meijer (2010). A Holanda le siguió el colegio de psicólogos de Gran Bretaña, si bien es cierto que bastantes años después. Aunque comenzaron aplicando su propio modelo en la década de los 90 (ver Bartram, 1996; Bartram, Lindley y Foster, 1990; Bartram, Lindley, y Marshall, 1992; Bartram, Anderson, Kellett, Lindley y Robertson, 1995; Bartram, Burke, Kandola, Lindley, Marshall, y Rasch, 1997), en los últimos años han adoptado el modelo de la EFPA, propuesto en 2002 (ver Bartram, 2002), que precisamente estuvo basado en los modelos locales propuestos por holandeses, británicos y españoles. Además de los británicos (e.g. Lindley, 2009), el modelo EFPA ha sido aplicado durante los últimos años por noruegos y alemanes (ver, por ejemplo, Nielsen, 2009 y Moosbrugger et al., 2009, respectivamente).

Sin embargo, como se ha mencionado anteriormente, los nuevos desarrollos en el ámbito de la evaluación psicológica y educativa han motivado que el modelo de la EFPA haya sido revisado en profundidad recientemente (ver Evers, et al., 2013). Esta versión revisada permite evaluar los tests de forma integral. De modo similar al CET, en una primera parte se describe la prueba de forma exhaustiva, y en una segunda parte se lleva a cabo una evaluación cuantitativa de las características psicométricas de la prueba. En ambas partes la información descriptiva y cuantitativa se complementa con comentarios cualitativos que enriquecen la evaluación. Sin embargo, en comparación con el CET, el modelo revisado de la EFPA añade secciones que permiten evaluar con detalle aspectos relacionados con los últimos avances tecnológicos y psicométricos; entre otros, la administración de test por internet, la elaboración de informes automatizados, o la aplicación de la Teoría de la Respuesta a los Ítems (TRI).

La revisión del modelo CET, el CET-R, incluye algunos de estos nuevos aspectos, pero no todos. Sólo se han incorporado aquéllos que se han considerado más adecuados para el contexto español, intentando a la vez mantener la parsimonia del modelo CET y facilitar la comparabilidad entre las nuevas evaluaciones y las realizadas hasta la fecha mediante el modelo original.



EL CUESTIONARIO DE EVALUACIÓN DE TESTS (CET): MODELO ORIGINAL Y PROCESO DE APLICACIÓN

El CET fue diseñado fundamentalmente para evaluar tests contruidos desde la Teoría Clásica de los Tests, y está estructurado en tres apartados. El primero, centrado en la descripción técnica de la prueba, presenta 31 ítems relativos al nombre de la prueba, autor, constructo medido, ámbito de aplicación, etc. El segundo apartado se ocupa de la evaluación técnica de las características del instrumento. Se incluyen ítems referidos a la calidad de los materiales y la documentación, las instrucciones e ítems, la fundamentación teórica, la adaptación/traducción (si el test ha sido construido originalmente en otro país), el análisis de los ítems, el estudio de la validez (diferenciando validez de contenido, de constructo, predictiva y el análisis del Funcionamiento Diferencial de los Ítems (DIF)), el estudio de la fiabilidad (diferenciando formas paralelas, consistencia interna y test retest), y los baremos. En total este apartado incluye 32 ítems cerrados que, en su mayoría, se responden mediante una escala de respuesta de cinco categorías ordenadas en función de la calidad de la característica evaluada. También cuenta con varios ítems abiertos en los que se solicita una justificación razonada de las respuestas a los ítems cerrados para cada una de las características principales evaluadas (validez, fiabilidad y baremos), así como la descripción de los procedimientos de selección de las muestras empleadas para evaluar la calidad psicométrica del test, y de los criterios empleados a la hora de evaluar la validez predictiva. Finalmente, en el tercer y último apartado, se solicita una valoración global del test y un resumen de los dos primeros apartados, que queda plasmado en una ficha técnica.

El proceso de evaluación de tests mediante el CET comienza con la selección, por parte de la comisión de tests del COP, tanto de las pruebas a evaluar como del coordinador o coordinadora que se ocupará de gestionar el proceso de evaluación. Para cada prueba seleccionada, el coordinador escoge dos evaluadores, que trabajan de forma independiente: uno experto en psicometría y otro experto en el campo profesional de evaluación al que el test va dirigido. Los evaluadores no deben tener relación directa con los autores de los tests, ni manifestar un conflicto de intereses que cuestione la objetividad en la valoración. El coordinador es el encargado de integrar las valoraciones de ambos evaluadores en un informe final. Si no existiera un acuerdo sustancial entre los evaluadores, se podría acudir a un tercer evaluador. El informe generado se envía al autor y/o al editor del test para que efectúen las observaciones y aclaraciones oportunas y puedan aportar información complementaria. Finalmente, tras realizar las modificaciones pertinentes, el informe se hace público a través de la página web del COP.

Ya en la primera aplicación del modelo, Muñoz et al., (2011) pusieron de manifiesto la necesidad de mejorar las instrucciones para la cumplimentación del modelo, ya que no todos los evaluadores parecían seguir los mismos criterios al responder a algunos de los ítems y, algunos de ellos, no se interpretaban correctamente. Esta necesidad fue corroborada por Ponsoda y Hontangas (2013) en la segunda evaluación. Por ello, y teniendo en cuenta las sugerencias realizadas, a partir de la tercera

evaluación se dieron unas instrucciones adicionales que aclararan con más detalle lo que se esperaba de los revisores al responder al cuestionario, con el fin de reducir algunas ambigüedades y estandarizar en mayor medida el proceso (ver Hernández et al., 2014).

En cuanto al cuestionario mismo, en los dos primeros procesos de evaluación se apuntaron recomendaciones sobre la inclusión de ciertas cuestiones. Si bien muchas de las sugerencias no se incluyeron en las siguientes evaluaciones a la espera de realizar una revisión más profunda del modelo que tuviera en cuenta el modelo revisado y actualizado de la EFPA –de la que nos ocupamos en el presente trabajo–, sí que se incluyeron aclaraciones para algunos ítems, así como algunas cuestiones nuevas sobre validez de constructo, estimación de la precisión mediante TRI o la actualización de baremos, entre otras (ver detalles en Hernández et al., 2014). Las instrucciones adicionales y pequeñas modificaciones del CET se mantuvieron en la cuarta evaluación de tests (Elosua y Geisinger, 2016), en la que de nuevo se puso de manifiesto la necesidad de realizar una revisión más profunda del modelo.

EL NUEVO CUESTIONARIO DE EVALUACIÓN DE TESTS (CET-R): DESCRIPCIÓN DE LOS PRINCIPALES CAMBIOS

Partiendo del CET original con las pequeñas modificaciones aplicadas por Hernández et al. (2014), los dos primeros autores trabajaron en una propuesta inicial de CET-R que, por una parte, resolviera los problemas de interpretación todavía observados en las evaluaciones realizadas y, por otra, incorporara algunos de los avances psicométricos y tecnológicos producidos durante los últimos años. Se procedió en cuatro fases. En primer lugar, se revisaron las sugerencias realizadas por los coordinadores de las distintas ediciones de evaluación de tests llevadas a cabo, introduciendo las modificaciones e instrucciones correspondientes. En segundo lugar, revisamos el modelo actualizado de la EFPA (Evers et al. 2013), añadiendo las cuestiones que consideramos más adecuadas para el contexto español, además de algunas otras que consideramos especialmente relevantes. Así pues, en la propuesta inicial se incluyeron criterios de evaluación sobre ciertas estrategias de validación, otras formas de evaluar la fiabilidad, o la interpretación de puntuaciones referidas a un criterio. Sin embargo, hemos dejado fuera la evaluación exhaustiva de cuestiones como la administración informatizada de tests, la evaluación a distancia mediante Internet, o la calidad de los informes automatizados, aunque sobre esta última cuestión sí se ha añadido un ítem abierto para que se evalúe la calidad del informe, además de mantener el ítem que ya existía describiendo el tipo de informe. Tampoco se ha incluido una evaluación exhaustiva de la aplicación de la TRI (sólo hay dos ítems evaluativos referidos a la precisión y a la adecuación del tamaño de las muestras cuando se aplica TRI), o la tipificación continua (aunque sobre esto hay una pregunta también). Estos aspectos fueron excluidos, o no evaluados exhaustivamente por varias razones. En primer lugar, al menos por ahora, la mayoría de los tests editados en España no requieren la consideración de estos aspectos. Además, queríamos evitar un cambio drástico respecto del CET original, con el fin de facilitar la comparabilidad con los resultados obtenidos en evaluaciones previas, y con el fin de



mantener un número razonable de ítems que facilite la tarea de los evaluadores.

Cabe señalar que todos los cambios se realizaron manteniendo, en términos generales, la estructura, apartados y manera de calificar del CET original (aunque en algunos casos se realizaron aclaraciones adicionales sobre lo que suponía el criterio de excelencia y, por tanto, la máxima calificación).

La propuesta inicial fue revisada en profundidad por el resto de autores del presente trabajo y se generó una nueva versión del CET-R. Esta nueva versión fue revisada por once expertos cualificados concedores del CET, los cuales se relacionan en la Tabla 1. Una de las sugerencias de cambio más frecuente fue la de abandonar la clasificación tradicional de tipos de validez que mantenía el CET (y también la primera versión del CET-R) y adoptar la terminología sobre validez de los nuevos estándares de la AERA, la APA y el NCME (2014). Por ello, es este apartado de validez el que ha sufrido un mayor cambio respecto del CET original y respecto del nuevo modelo de la EFPA. Según la terminología de los estándares de la APA, AERA y NCME (1999, 2014) no se valida el test, sino las interpretaciones o usos concretos que se hacen de sus puntuaciones. Por ello, en vez de seguir la clasificación tradicional de tipos de validez propia de las normas APA de 1985 (AERA, APA, NCME, 1985), y diferenciar entre validez de contenido, de constructo, y predictiva, en el CET-R, se recogen tres fuentes de evidencias de validez: las basadas en el contenido, las basadas en las relaciones con otras variables (con otro test que mida el mismo o un constructo relacionado, con un criterio que se pretende predecir, etc.), y las basadas en la estructura interna del test (como, por ejemplo, evaluando la estructura factorial). En realidad, lo importante es que en la documentación y el manual del test se recojan evidencias que apoyen la validez del uso de las puntuaciones, independientemente de que se hable de validez de constructo, o de evidencias de validez basada en la estructura interna del test (anteriormente considerada “validez de constructo”), por ejemplo. De hecho, el modelo actualizado de la EFPA sigue utilizando la clasificación tradicional. Sin embargo, hemos creído que la actualización del CET debía incorporar las recomendaciones de los estándares internacionales actuales.

Tras realizar los ajustes y modificaciones pertinentes a partir de las sugerencias de los expertos, esta nueva versión fue presentada a la Comisión de Tests del COP, llegando así a la versión final. Esta versión, junto con las instrucciones de cumplimentación que se incluyen en el cuestionario con el fin de aumentar la claridad y la estandarización en el proceso de evaluación, se pueden descargar en la página web del Consejo General de Psicología de España (<http://www.cop.es>), en el apartado de la comisión de tests (o directamente en <http://cop.es/n>).

Como el modelo original, el CET-R está estructurado en tres apartados. El primero, centrado en la descripción técnica de la prueba, presenta ahora 28 ítems. Es prácticamente igual al original con la salvedad de que en algunos ítems se incluyen aclaraciones adicionales y/o se modifican o añaden alternativas de respuesta. Además, dos de los ítems del CET original, los ítems 1.20 y 1.21, referidos a las escalas utilizadas y puntuaciones transformadas, respectivamente, se unen en uno, y otros ítems, como por ejemplo el referido a la presentación de la bibliografía básica aportada, se eliminan.

El segundo apartado se ocupa de la evaluación técnica de

las características del instrumento. Se incluyen 55 ítems, 9 sobre cuestiones generales, uno sobre análisis de ítems, 20 sobre validez, 15 sobre fiabilidad y 10 sobre baremos e interpretación de puntuaciones. A los ítems iniciales del CET, referidos a la calidad de los materiales y documentación, la fundamentación teórica, etc., se añade un ítem referido al desarrollo de los ítems (cuando se trata de un test original y no de un test adaptado). Asimismo, se diferencia entre calidad de las instrucciones para quienes han de responder al test, y para quienes han de administrarlo y corregirlo, y se añade un ítem que evalúa la calidad de las referencias aportadas. La sección de validez es la que más cambios presenta, tal y como se ha indicado anteriormente. Junto a las evidencias basadas en el contenido, se evalúan las evidencias basadas en las relaciones con otras variables, diferenciándose las evidencias basadas en las relaciones entre las puntuaciones del test y otras variables (evidencia convergente, discriminante, basada en las diferencias entre grupos, etc.) y las evidencias basadas en las relaciones entre las puntuaciones del test y un criterio (lo que sería la validez predictiva en el modelo CET original). Asimismo, se evalúan las evidencias basadas en la estructura interna, incluyéndose en este punto tanto el análisis factorial como el análisis de DIF. Finalmente, se introduce un ítem que recoge si el manual informa de las posibles acomodaciones que se deban introducir en la administración del test para la correcta evaluación de personas con limitaciones o diversidad funcional. En cuanto a la sección de fiabilidad, en la evaluación de los coeficientes de equivalencia (formas paralelas) se añade un ítem sobre la evaluación del cumplimiento de los supuestos de paralelismo, y al considerar los coeficientes de consistencia interna, se añaden coeficientes basados en el análisis factorial. También se incluyen tres preguntas (dos evaluativas y una meramente descriptiva) referidas a la cuantificación de la precisión mediante TRI y dos preguntas referidas a la evaluación de la fiabilidad inter-jueces. En cuanto a la sección de baremos e interpretación de puntuaciones, la interpretación normativa incluye una pregunta sobre la tipificación continua, que permite obtener baremos más precisos con grupos más reducidos (e.g. Evers et al., 2010), y otra sobre la

TABLA 1
LISTADOS DE EXPERTOS QUE PARTICIPARON EN LA REVISIÓN DE LA PRIMERA VERSIÓN DEL CET-R

Nombre	Filiación
Constantino Arce	Universidad de Santiago de Compostela
Rosario Martínez-Arias	Universidad Complutense de Madrid
Roberto Colom	Universidad Autónoma de Madrid
Ana Delgado	Universidad de Salamanca
Eduardo Fonseca	Universidad de La Rioja
María Dolores Hidalgo	Universidad de Murcia
María José Navas	UNED
Julio Olea	Universidad Autónoma de Madrid
José Luis Padilla	Universidad de Granada
Pablo Santamaría	TEA Ediciones
Carme Viladrich	Universidad Autónoma de Barcelona



actualización de los baremos. Asimismo, se incluyen cuatro preguntas sobre la interpretación de puntuaciones referida a un criterio, especialmente aplicables a algunos tipos de tests (por ejemplo, educativos o clínicos). Cabe señalar que, para todas las secciones, en algunos ítems, se añaden aclaraciones sobre cómo se debe responder o lo que supone el cumplimiento del criterio de excelencia. Además, como en el CET original, se incluyen preguntas abiertas que permitan razonar las puntuaciones asignadas a los ítems cerrados y otras cuestiones descriptivas y evaluativas que pudieran resultar relevantes.

Finalmente, en el tercer y último apartado, se continúa solicitando una valoración global del test, y un resumen de los dos primeros apartados que queda plasmado en una ficha técnica.

CONCLUSIONES

Creemos que la aplicación del modelo CET propuesto por Prieto y Muñiz (2000) ha tenido un impacto positivo en muchos ámbitos durante los últimos años. En primer lugar, ha proporcionado a los usuarios de tests información técnica contrastada sobre la calidad de algunos de los tests disponibles (casi 50 hasta la fecha), con el fin de que puedan decidir sobre su elección. Pero además, en segundo lugar, la aplicación del CET ha contribuido a mejorar los procesos de construcción y edición de tests. A lo largo de las distintas evaluaciones realizadas, hemos observado que, cada vez más, los manuales recogen explícitamente la mayoría de los criterios de evaluación del CET, e incluyen información detallada sobre los procesos de construcción del test y su estandarización, la calidad psicométrica de sus puntuaciones y los usos adecuados –e inadecuados– del test. Finalmente, tenemos constancia de que el CET está teniendo un impacto en la formación de los futuros psicólogos, ya que los profesores de psicometría, en muchos casos, utilizan este modelo en sus clases, guiando a los alumnos de forma práctica en los aspectos básicos de la evaluación de la calidad psicométrica y técnica de los tests.

Reconocer lo anterior no resulta contradictorio con aceptar que, tras más de 15 años desde la publicación del CET, los conceptos de fiabilidad y validez se han enriquecido, y los requisitos científicos y profesionales exigidos por los tests se han ido adaptando a nuevas necesidades (De Boeck y Elosua, en prensa). Por ello era necesaria una revisión del modelo CET que permitiera incorporar los avances producidos, para seguir introduciendo mejoras en el uso de los tests por parte de psicólogos y educadores, e indirectamente, para seguir mejorando los procesos de construcción y edición de tests en nuestro país. Esta revisión se plasma en el CET-R, que será utilizado en la quinta edición de evaluaciones de tests que, impulsada por el COP, se ha puesto en marcha recientemente.

La publicación de los resultados obtenidos en las evaluaciones de tests, es una de las estrategias informativa que sigue el COP con el fin de mejorar el uso de los tests y, con ello, la práctica profesional del psicólogo. Pero no es la única. El COP, junto con la EFPA y la ITC (International Test Commission), de las que es miembro, llevan a cabo acciones y proyectos muy variados con el fin de mejorar el uso de los tests. Las distintas acciones y proyectos se enmarcan dentro de dos estrategias complementarias: una más restrictiva y otra informativa (para una información más detallada ver Muñiz y Bartram, 2007;

Muñiz y Fernández-Hermida, 2010; y Muñiz, 2012). La estrategia restrictiva agrupa al conjunto de acciones llevadas a cabo para restringir el uso de los tests a los profesionales que están realmente preparados para hacerlo. La estrategia informativa agrupa las iniciativas encaminadas a difundir información sobre la práctica de los tests con el fin de disminuir la probabilidad de que se haga un mal uso de las pruebas. En este sentido se han desarrollado códigos éticos y deontológicos (e.g. EFPA, 2005; Fernández-Ballesteros et al., 2001) y directrices sobre el uso de los tests, destacando los estándares técnicos de la AERA, APA y NCME (2014), así como las numerosas directrices elaboradas por la ITC: las directrices generales para el uso de los tests, (ITC, 2001), las directrices para la traducción y adaptación de los tests de unas culturas a otras (Hambleton, Merenda y Spielberger, 2005; Muñiz, Elosua y Hambleton, 2013), las directrices sobre el uso de tests informatizados, las directrices profesionales sobre la selección de tests y cómo proceder cuando los tests quedan obsoletos, las directrices sobre la seguridad en los tests, sobre el control de calidad de los tests, o sobre el uso de los tests en investigación. Lo más importante de las tres últimas está recogido en el trabajo de Muñiz, Hernández y Ponsoda (2015). Todas ellas están disponibles en la página web de la ITC y, muchas de ellas, están traducidas al español y son accesibles a través de la página web del Consejo General de Psicología de España (<http://www.cop.es>) en el apartado de la comisión de tests. Una última estrategia informativa que merece atención es la norma ISO-10667, que regula todo el proceso de evaluación de personas en contextos laborales. Para una revisión más detallada de todas las acciones que se llevan en España para mejorar el uso de los test se puede consultar Elosua y Muñiz (2013).

La evaluación de los tests publicados en España es una de las muchas acciones. Pero, como indican Elosua y Geisinger (2016), para que esta acción sea realmente útil, exige de un trabajo continuo de mejora, tanto procedimental, como formal y sustantiva. Y con ese objetivo de mejora se propone el CET-R. El objetivo último es claro: que los psicólogos tengan información contrastada y fiable que les permita hacer una mejor selección y uso de los tests disponibles. Todo ello repercutirá en una mejora de la práctica profesional y de su prestigio.

REFERENCIAS

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.



- Bartram, D. (1996). Test qualifications and test use in the UK: The competence approach. *European Journal of Psychological Assessment*, 12, 62-71.
- Bartram, D. (2002). *Review model for the description and evaluation of psychological tests*. Brussels: European Federation of Psychologists' Associations (EFPA).
- Bartram, D., Lindley, P. A. y Foster, J. M. (1990). *A review of psychometric tests for assessment in vocational training*. Sheffield, UK: The Training Agency.
- Bartram, D., Lindley, P.A. y Foster, J.M. (1992 a). *A review of psychometric tests for assessment in vocational training*. Leicester: BPS Books.
- Bartram, D., Lindley, P.A. y Marshall, L. (1992 b). *Update to the review of psychometric tests for assessment in vocational training*. Leicester: BPS Books.
- Bartram, D., Anderson, N., Kellett, D., Lindley, P.A. y Robertson, I. (Eds.). (1995). *Review of Personality Assessment Instruments (Level B) for use in occupational settings*. Leicester: BPS Books.
- Bartram, D., Burke, E., Kandola, R., Lindley, P., Marshall, L. y Rasch, P. (Eds.). (1997). *Review of Tests of Ability and Aptitude (Level A) for use in occupational settings*. Leicester: BPS Books.
- Buros, O. K. (1938). *The 1938 Mental Measurements Yearbook*. New Brunswick, NJ: Rutgers University Press.
- Carlson, J. F., Geisinger, K. F. y Jonson, J. L. (Eds.) (2014). *The nineteenth mental measurements yearbook*. Lincoln, NE: Buros Center for Testing.
- 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 (2nd ed.): Clinical and intellectual issues* (pp. 11-27). Washington, DC: American Psychological Association.
- De Boeck, P. y Elosua, P. (en prensa). Reliability and Validity: History, Notions, Methods, Discussion. En F. T. L. Leong, D. Bartram, F. Cheung, K. F. Geisinger, y D. Ilescu (Eds.), *The ITC international handbook of testing and assessment*. Oxford University Press.
- European Federation of Professional Psychologists' Associations (2005). *Meta-Code of ethics*. Brussels: Author (<http://www.efpa.eu>)
- Elosua, P. (2012). Tests publicados en España: Usos, costumbres y asignaturas pendientes. *Papeles del Psicólogo*, 33, 12-21.
- Elosua, P. & Geisinger, K (2016) Cuarta evaluación de tests editados en España: forma y fondo [Fourth Review of Tests Published in Spain: Form and Content]. *Papeles del Psicólogo*, 37, 82-88.
- Elosua, P. y Muñiz, J. (2013). Proyectos españoles para una mejora en el uso de los tests. *Revista Latinoamericana de Ciencia Psicológica*, 5, 139-143.
- Evers, A., Braak, M., Frima, R. y van Vliet-Mulder, J. C. (2009). *Documentatie van Tests en Testresearch in Nederland [Documentación de tests e investigación sobre tests en Holanda]*. Amsterdam: Boom test uitgevers.
- Evers, A., Muñiz, J., Hagemester, C., Hstmælingen, A., Lindley, P., Sjöberg, A. y Bartram, D. (2013). Assessing the quality of tests: Revision of the EFPA review model. *Psicothema*, 25, 283-291.
- Evers, A., Sijtsma, K., Lucassen, W. L y Meijer, R. R. (2010) The Dutch Review Process or evaluating the quality of psychological tests: History, procedure, and results. *International Journal of Testing*, 10, 295-317.
- Fernández-Ballesteros, R., De Bruyn, E., Godoy, A., Hornke, L., Ter Laak, J. y Vizcarro, C. et al. (2001). Guidelines for the assessment process (GAP): A proposal for discussion. *European Journal of Psychological Assessment*, 17, 187-200.
- Hambleton, R. K., Merenda, P. F. y Spielberger, C. D. (Eds.) (2005). *Adapting educational and psychological tests for cross-cultural assessment*. London: LEA.
- Hernández, A., Tomás, I., Ferreres, A. y Lloret, S. (2015). Tercera evaluación de test editados en España. *Papeles del Psicólogo*, 36, 1-8.
- ISO (2011). *Procedures and methods to assess people in work and organizational settings (part 1 and 2)*. Ginebra: ISO [versión española: Prestación de servicios de evaluación: procedimientos y métodos para la evaluación de personas en entornos laborales y organizacionales (partes 1 y 2). Madrid: AENOR, 2013].
- Lindley, P.A. (2009, Julio). Using EFPA criteria as a common standard to review tests and instruments in different countries. In D. Bartram (Chair), *National approaches to test quality assurance*. Symposium conducted at The 11th European Congress of Psychology, Oslo, Norway.
- Moosbrugger, H., Kelava, A., Hagemester, C., Kersting, M., Lang, F., Reimann, G., et al. (2009, Julio). The German Test Review System (TBS-TK) and first experiences. In D. Bartram (Chair), *National approaches to test quality assurance*. Symposium conducted at The 11th European Congress of Psychology, Oslo, Norway.
- Muñiz, J. (2012). Perspectivas actuales y retos futuros de la evaluación psicológica. En C. Zúñiga (ed.), *Psicología, sociedad y equidad*. Santiago de Chile: Universidad de Chile.
- Muñiz, J. y Bartram, D. (2007). Improving international tests and testing. *European Psychologist*, 12, 206-219.
- 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., Elosua, P., y Hambleton, R.K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25, 151-157.
- Muñiz, J., Fernández-Hermida, J. R., Fonseca-Pedrero, E., Campillo-Álvarez, A., & Peña-Suárez, E. (2011). Evaluación de tests editados en España. *Papeles del Psicólogo*, 32, 113-128.
- Nielsen, S. L. (2009, Julio). Test certification through DNV in Norway. In D. Bartram (Chair), *National approaches to test quality assurance*. In D. Bartram (Chair), *National approaches to test quality assurance*. Symposium conducted at The 11th European Congress of Psychology, Oslo, Norway.
- NIP (1969). *Documentatie van Tests en Testresearch in Nederland [Documentación de tests e investigación sobre tests en Holanda]*. Amsterdam: Nederlands Instituut van Psychologen.
- 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.

