

# QUINTA EVALUACIÓN DE TESTS EDITADOS EN ESPAÑA: MIRANDO HACIA ATRÁS, CONSTRUYENDO EL FUTURO

Fonseca-Pedrero<sup>1</sup>, Eduardo

QUINTA EVALUACIÓN DE TESTS EDITADOS EN ESPAÑA: MIRANDO HACIA ATRÁS, CONSTRUYENDO EL FUTURO

Papeles del Psicólogo, vol. 38, núm. 3, 2017

Consejo General de Colegios Oficiales de Psicólogos, España

**Disponible en:** <https://www.redalyc.org/articulo.oa?id=77853188001>

## QUINTA EVALUACIÓN DE TESTS EDITADOS EN ESPAÑA: MIRANDO HACIA ATRÁS, CONSTRUYENDO EL FUTURO

*Eduardo Fonseca-Pedrero1*  
*Universidad de La Rioja, España*  
*eduardo.fonseca@unirioja.es*

Redalyc: <https://www.redalyc.org/articulo.oa?id=77853188001>

### RESUMEN:

La psicología como ciencia necesita de una continua y rigurosa evaluación tanto de sus métodos y procedimientos como de sus técnicas e instrumentos. Los tests, una de las principales herramientas que utilizan los psicólogos en su quehacer diario, no pueden ser una excepción. Se hace necesario someter a escrutinio los instrumentos de evaluación psicológica mediante un proceso de revisión sistemático, riguroso e independiente que garantice su calidad y utilidad. Con este fin, desde el año 2010 la Comisión Nacional de Tests del Colegio Oficial de Psicólogos viene revisando anualmente algunos de los tests más utilizados por los psicólogos españoles. Tal proceso de revisión se basa en un modelo europeo de evaluación de la calidad de los tests adaptado al español y que se operativiza mediante el Cuestionario de Evaluación de Tests-Revisado. En este trabajo se presenta la quinta evaluación de tests editados en España. Se evaluaron un total de once tests, participaron un total de cuatro casas editoriales y 24 expertos en un proceso de revisión por pares. En líneas generales puede afirmarse que la calidad de los tests evaluados fue adecuada, si bien se han encontrado algunas fortalezas y debilidades. Finalmente se comentan las principales conclusiones y se analizan las posibles líneas de futuro en la evaluación de los tests en España.

**PALABRAS CLAVE:** Tests, Uso de los tests, Evaluación de tests, Psicometría, Calidad de los tests, Medición.

### ABSTRACT:

Psychology, as a science, needs a continuous and rigorous evaluation of both its methods and procedures as well as its techniques and instruments. The test is one of the main tools used by psychologists in their daily work, and cannot be an exception. The measuring instruments of psychological evaluation must be analyzed through a systematic, rigorous and independent process that guarantees their quality and usefulness. To this end, since 2010 the National Test Commission of the Spanish Psychological Association has been reviewing annually some of the tests most used by Spanish psychologists. This review process is based on the European test review model adapted to Spanish and it is performed through the Tests Review Questionnaire-Revised. In the present work the fifth evaluation of tests published in Spain is presented. A total of eleven tests were evaluated, with the participation of four publishing houses, and 24 experts in a peer review process. In general terms, the quality of the measurement instruments evaluated in this fifth revision was adequate, although there are a number of strengths and weaknesses. Finally the main conclusions are discussed and the possible future research areas in the evaluation of the tests in Spain are analyzed.

**KEYWORDS:** Tests, Use of tests, Evaluation of tests, Psychometrics, Quality of the tests, Measurement.

## QUINTA EVALUACIÓN DE TESTS EDITADOS EN ESPAÑA: MIRANDO HACIA ATRÁS, CONSTRUYENDO EL FUTURO

La psicología como ciencia necesita de una continua y rigurosa evaluación de sus métodos, procedimientos, técnicas e instrumentos. Los tests, una de las principales herramientas que utilizan los psicólogos en su práctica profesional, no pueden ser una excepción. A partir de ellos se toman decisiones que afectan a la vida de las personas y a las organizaciones teniendo un claro impacto en la sociedad. Por ende, los tests se deben someter a escrutinio mediante un proceso de revisión sistemático, riguroso e independiente que garantice su calidad y utilidad. Como señalan Muñiz et al. (2011), la utilización correcta de los tests requiere, por

---

### NOTAS DE AUTOR

*eduardo.fonseca@unirioja.es*

un lado, de instrumentos con adecuadas propiedades psicométricas (p.ej., fiabilidad de las puntuaciones, evidencias de validez, etc.), y por otro, de una acertada preparación técnica para su correcta utilización por parte de los profesionales. Los tests deben tener una calidad y rigor empíricamente demostrados. Además, los psicológicos deben ser competentes y tener información contrastada que les permita seleccionar aquellos tests más adecuados para su propósito (Hernández, Ponsoda, Muñiz, Prieto, y Elosua, 2016). Todo ello queda recogido en el Código Deontológico del Psicólogo y en las distintas directrices desarrolladas a tal efecto tanto a nivel nacional como internacional (Muñiz, 1997; Muñiz y Bartram, 2007). Por ejemplo, el artículo 17º del Código Deontológico menciona que el psicólogo tiene que estar profesionalmente preparado y especializado en la utilización de métodos, instrumentos, técnicas y procedimientos que adopte en su trabajo y debe reconocer los límites de su competencia y las limitaciones de sus técnicas.

El Consejo General de la Psicología de España (COP) desde el año 2010 viene revisando, de forma anual, los tests editados en España (Elosua y Geisinger, 2016; Hernández, Tomás, Ferreres, y Lloret, 2015; Muñiz et al., 2011; Ponsoda y Hontangas, 2013). Hernández et al. (2016) indican que el objetivo último de la evaluación de la calidad de los tests utilizados en España es proporcionar a los usuarios de instrumentos de y evaluación psicológica información precisa y accesible sobre la calidad de los tests disponibles. Actualmente se acaba de finalizar la quinta edición. El principal objetivo de este artículo es presentar los principales resultados de esta última evaluación de los tests editados en España. El primer lugar se realiza un breve recorrido histórico donde se examinan los principales hitos del proceso de evaluación de tests editados en nuestro país. Se pretende ofrecer con ello una panorámica del estado actual de la cuestión. A continuación se detalla de forma pormenorizada el proceso de revisión y los principales resultados de esta quinta evaluación de tests. Finalmente, en tercer lugar, se comentan las principales conclusiones así como futuras líneas de trabajo.

#### EVALUACIÓN DE TESTS EN ESPAÑA: EL CAMINO RECORRIDO

Cuando se les pregunta a los profesionales de la psicología cuales son las principales demandas sobre el uso de tests refieren, entre otras, la necesidad de disponer información técnica y psicométrica sobre los mismos que les ayude a tomar decisiones fundamentadas (Muñiz y Fernández-Hermida, 2010; Evers et al., 2017). Para dar respuesta a estas y otras cuestiones la Comisión de Tests de la European Federation of Psychologists' Associations (EFPA, <http://www.efpa.eu/>) desarrolló un modelo de evaluación tests que fue adaptado al español por Prieto y Muñiz (2000). En el año 2010 la Comisión Nacional de Tests del COP decidió poner en marcha la primera revisión de tests cuyos resultados fueron publicados en el 2011 (Muñiz et al., 2011). Posteriormente, casi de forma anual, se ha venido realizando una nueva revisión de tests. La evolución histórica del proceso de evaluación de tests editados en España así como la información relevante de cada evaluación anual se recoge en la Figura 1. Todos los tests evaluados en las cinco revisiones se presentan en la Figura 2. Asimismo, los resultados de todas y cada una de las evaluaciones realizadas hasta la fecha se encuentran disponibles en la web del COP: <https://www.cop.es/index.php?page=evaluacion-tests-editados-en-espana> Se invita a los profesionales a visitar y analizar los tests evaluados. Como se puede observar en las Figuras 1 y 2 hasta el momento se han revisado un total de 55 tests procedentes de diferentes casas editoriales, han participado más de 100 revisores y cinco coordinadores. Es digno de mención que el número de expertos que revisan las pruebas no es la suma aritmética de las cinco evaluaciones, pues algunos de ellos han participado en varias ediciones. También se quiere mencionar que los profesionales utilizan estos informes disponibles en la página web del COP. Un indicador de ello puede ser el número de descargas de estos informes. Con este fin se ha realizado un análisis del volumen de descargas en el periodo comprendido entre los años 2012 y 2017. De años anteriores no se dispone de esta información. Los resultados del número de descargas es el siguiente: Sistema para la Evaluación de la Conducta Adaptativa n=105, Inventario Breve de 18 Síntomas n=101, Escala de Memoria de Wechsler-IV n=65, Test Boehm de Conceptos Básicos-3 Preescolar n=101, Batería de Evaluación Cognitiva de la Lectura y Escritura n=97, Cuestionario Oviedo para la Evaluación de la Esquizotipia n=94, Inventario de Ansiedad de Beck n=111, Inventario de Depresión de Beck-II n=104, Cuestionario para la Detección de los Trastornos del Comportamiento en Niños y

Adolescentes n=105, Inventario Multifásico de la Personalidad de Minnesota-2 Reestructurado n= 164, Escala de inteligencia de Wechsler para niños-IV n=134 y 16 Factores de Personalidad, quinta edición n=119.

El modelo desarrollado por la EFPA para la evaluación de la calidad de los tests se operativiza en un instrumento de medida denominado Cuestionario de Evaluación de Tests (CET) (Muñiz et al., 2011). El CET permite una evaluación cualitativa y cuantitativa del test que es objeto de revisión, esto es, trata de examinar su calidad y utilidad. El CET se puede consultar en el trabajo de Muñiz et al. (2011). Posteriormente, en el año 2013, la EFPA revisó este modelo de evaluación (Evers et al., 2013). Los cambios introducidos en el modelo unido a modificaciones sugeridas en las diferentes evaluaciones anuales por los coordinadores y revisores así como a las nuevas directrices de los Standards (American Educational Research Association, American Psychological Association, y National Council on Measurement in Education, 2014) (p. ej., las referidas a la obtención de evidencias de validez y su taxonomía) llevaron a la modificación del CET dando lugar al CET-Revisado (CET-R). Estos cambios se han introducido en la presente quinta evaluación de tests. Para un análisis más detallado de los cambios introducidos se puede consultar el excelente trabajo de Hernández et al. (2016). Además, el CET-R se puede descargar en el siguiente enlace: <http://www.cop.es/uploads/pdf/CET-R.pdf> El CET-R, al igual que su versión predecesora, incorpora tres grandes apartados: a) descripción general del test; b) valoración de las características del test (p.ej., fiabilidad, evidencias de validez, etc.); y c) valoración global del test. Es una herramienta interesante a utilizar por todos los profesionales de la psicología no solo con fines prácticos o profesionales sino también con fines formativos. Más aún, el CET-R puede ser utilizado a modo de guía o estándar, contribuyendo a mejorar el proceso de construcción y edición de tests publicados en España, aspecto que redundará a su vez en la capacidad de toma de decisiones por parte de los profesionales de la psicología.

El proceso de revisión de tests comienza por la designación por parte de la Comisión Nacional de Tests de COP de un coordinador para la evaluación de ese año. Esta comisión en colaboración con los editores también designa los tests que serán revisados en esa anualidad. Una vez seleccionados los tests, entre 10 o 12 según el año, el coordinador selecciona un conjunto de expertos en la temática que revisan las pruebas mediante un proceso de evaluación por pares (similar al que se utiliza en la revisión de artículos científicos). La selección de estos revisores se lleva a cabo siempre de acuerdo a unos criterios y estándares científicos, que con puntuales modificaciones, se mantienen invariantes a través de las sucesivas evaluaciones. Se seleccionan dos revisores, siempre que las circunstancias lo permitan, uno de ellos con un perfil más técnico-psicométrico, y otro con un perfil más centrado en aspectos sustantivos. Se garantiza su anonimato referido a la prueba que valora, se solicita su participación y la ausencia de conflicto de intereses. En la gran mayoría de los casos, y salvo causas de fuerza mayor, los revisores aceptan gratamente la tarea encomendada. A todos los revisores se les entrega una copia del test que van a evaluar (cedido de forma gratuita por los editores) así como una gratificación económica simbólica de 50 euros. Cada revisor realiza un informe independiente del test que envía al coordinador en un plazo de tres meses. Obviamente durante todo el proceso el coordinador está abierto a responder a cualquier duda o sugerencia tanto de los revisores como de los editores. A continuación el coordinador analiza los comentarios y puntuaciones procedentes de los dos revisores y confecciona un primer informe final. Este informe es enviado a las respectivas casas editoriales que, en un plazo de aproximadamente de un mes, responden de forma razonada, si así fuera el caso, a aquellos comentarios y/o puntuaciones que estimen oportuno. De esta forma todas las partes implicadas forman parte del proceso. Finalmente, con dichos comentarios el coordinador elabora un informe final de cada test que es enviado a la Comisión Nacional de Tests y que posteriormente es subido a la página web para su consulta (véase el link arriba). Cabe subrayar dos aspectos importantes, en primer lugar, el proceso de evaluación de las pruebas es totalmente independiente tanto de las editoriales y autores de los tests, como del COP, solo depende de los evaluadores y del coordinador de la evaluación. Como todo proceso de evaluación por pares no es perfecto, nada humano lo es, pero nos parece el mejor de los posibles. En segundo lugar, la evaluación de los tests

no constituye una certificación de las pruebas por parte de la Comisión de Tests del COP, se trata de una evaluación en la que se ponen sobre la mesa los puntos fuertes y débiles de las pruebas tal como los estiman los evaluadores en un momento dado, siguiendo el modelo de evaluación de tests CET-R. La filosofía que guía las evaluaciones de los tests no es la censura de las pruebas, se trata de ayudar a los editores y autores a mejorar sus tests, aportando las evidencias de validez que garanticen las inferencias hechas a partir de las puntuaciones. Como bien señalara el egregio físico Lord Kelvin, aquello que no se evalúa no se puede mejorar.

#### QUINTA EVALUACIÓN DE TESTS

En este apartado se comentan el proceso de evaluación, los tests revisados, los revisores participantes y los principales resultados de la quinta evaluación de tests editados en España.

### Proceso de evaluación

El proceso de revisión general que se ha seguido en la presente evaluación se ha mencionado anteriormente, si bien aquí se quieren comentar de forma breve aquellos aspectos específicos concernientes a la quinta evaluación de tests. Una vez designado los tests a valorar el coordinador (primer autor de este trabajo) realizó una selección de expertos para la revisión de los tests, a los cuales se les envió una invitación formal por email (enero 2016). De los seleccionados solo dos de ellos no pudieron participar, por motivos totalmente comprensibles, por lo que se seleccionaron otros dos nuevos revisores. A continuación se procedió al envío del material que incluía el modelo de evaluación (CET-R), las instrucciones de cumplimentación y el test a evaluar (documentación, CD, hojas de corrección, etc.) (marzo/abril 2016). Durante todo este proceso se solucionaron todas aquellas posibles dificultades y cuestiones que fueron surgiendo. Por ejemplo, a modo anecdótico una de las pruebas fue enviada a EEUU. Cuando revisor recibió el test este llegó abierto ya que había sido inspeccionado en aduanas. Entre los meses de julio y septiembre de 2016 los revisores fueron enviando las evaluaciones al coordinador. A continuación el informe preliminar se remitió a los editores que respondieron de forma eficaz entre los meses de noviembre y diciembre de 2016. Finalmente entre diciembre de 2016 y enero del 2017 se confeccionaron los informes finales que fueron enviados al COP y fueron colgados en la página web anteriormente citada.

### Tests y revisores seleccionados

Los tests seleccionados en la quinta evaluación se recogen en la Tabla 1. Como se puede observar fueron un total de 11 procedentes de cuatro casas editoriales (CEPE, EOS, PEARSON, TEA). La lista original de tests seleccionados por la Comisión Nacional de Tests para esta quinta evaluación sufrió dos ligeras modificaciones que se comentan a continuación. Primera, originalmente fueron 12 las pruebas elegidas. Se había incluido el Personality and Preference Inventory (PAPI), comercializado en España por FACTHUM y cuyo propietario internacional es CUBIKS limited, no obstante el director de la casa editorial CUBIKS limited explicó en una carta debidamente razonada al COP que el PAPI se encontraba en pleno proceso de actualización, por lo que sería necesario posponer su evaluación para futuras ediciones. Segundo, también se seleccionó el Test de Inteligencia Verbal-E2 (INVE-E2) de CEPE, que por estar también en proceso de actualización, se acordó por unanimidad sustituirlo por la batería PAIB 2 y 3. Todos estos cambios fueron consensuados tanto por la Comisión Nacional de Tests del COP como por los respectivos editores.

En la Tabla 2 se recogen los 24 revisores que participaron en esta quinta evaluación de tests. Como se comentó anteriormente se seleccionaron dos para cada prueba, si bien en dos casos concretos donde se encontraron claras divergencias entre revisores se seleccionó un tercero que ayudó al coordinador a confeccionar el informe final del respectivo test. La selección de los revisores se llevó a cabo considerando, en la medida de lo posible y al igual que en evaluaciones previas, los siguientes criterios: a) ausencia de



conflicto de intereses u otras cuestiones éticas; b) ampliar las instituciones participantes en la evaluación nacional de tests (p. ej., consejerías de sanidad y educación, psicólogos clínicos, entidades de evaluación y medición extranjeras, etc.); c) no haber participado en evaluaciones previas; y d) jóvenes investigadores o profesionales de la psicología. Se pretendió abrir el abanico de revisores y expertos fuera de la academia y más allá del contexto español, aspecto que pretendía, en esencia, dar a conocer y extender el uso del modelo español de evaluación de tests entre los profesionales de la psicología. Queremos mostrar nuestro más sincero agradecimiento al buen hacer de todos los revisores y editores que han participado en esta quinta evaluación, este proceso no hubiese llegado a buen puerto sin su estimable ayuda y colaboración.

## Resultados

En la Tabla 3 se recogen los principales resultados derivados de la quinta evaluación de tests en España. Como se puede observar las puntuaciones fueron, en la mayoría de los casos, superiores a 3,5 (calificación de buena a excelente). Del total de valoraciones únicamente en cuatro casos las puntuaciones fueron inferiores a 3. Como fortalezas se debe mencionar que los tests revisados presentan, en términos generales, una excelente calidad de sus materiales y documentación, adecuados niveles de fiabilidad y excelentes baremos además de suficientes evidencias de validez. Como limitaciones o debilidades se podrían mencionar la falta de estudios sobre funcionamiento diferencial de los ítems en algunos de los tests analizados o la total ausencia de análisis psicométricos bajo el prisma de la Teoría de Respuesta a los Ítems (TRI).

Hay dos aspectos que requieren un comentario: a) la dificultad a la hora de interpretar, analizar y puntuar determinados ítems del CET-R como, por ejemplo, los referidos al análisis de ítems o la validez criterial; y b) en dos casos concretos se hallaron diferencias sustanciales entre las puntuaciones y comentarios enviados al coordinador por los dos revisores, por lo que hubo que recurrir a una tercera opinión. Referente al primer punto resultados similares se han encontrado en evaluaciones previas (Elosua y Geisinger, 2016; Hernández Baeza et al., 2015; Muñiz et al., 2011; Ponsoda y Hontangas, 2013). En este sentido algunos de los ítems del CET-R parecen ser difíciles de valorar por los expertos. Ya en evaluaciones previas, y con la finalidad de subsanar esta limitación, se recurrió a incorporar unas pautas generales para la utilización del CET/CET-R así como una mayor explicación de algunos de los ítems. No obstante, parece ser que esta dificultad continúa en la quinta evaluación. Referente al segundo punto (parcialmente relacionado con el anterior) no cabe duda que el proceso de revisión por pares es uno de los muchos posibles para la evaluación de la calidad de tests, y tiene pros y contras. En el mundo de revisión por pares, bien sea de artículos científicos, proyectos, contratos, o en nuestro caso tests, es frecuente encontrar discrepancias entre los revisores, por lo que hay que recurrir a un tercer evaluador e incluso a varias rondas de evaluación. Es algo consustancial al proceso de revisión y no necesariamente se debe interpretar como algo negativo. En futuras ediciones tal vez sería necesario seguir mejorando la comprensión de algunos de los ítems del CET-R, aspecto que se podría solventar, por ejemplo, desarrollando video-tutoriales, ejemplos concretos o incorporando más información técnica-psicométrica en el CET-R.

## ALGUNAS CONCLUSIONES

Tras un lustro de evaluaciones de tests editados en España se puede decir que el proceso se está consolidando y goza de buena salud, no obstante todavía nos encontramos en los estadios iniciales, más aún si nos comparamos con otros países, por ejemplo Holanda, que ha evaluado todos los tests editados en ese país (Evers, 2012), o Estados Unidos, donde la organización especializada BUROS ya ha evaluado más de 10.500 tests (Elosua y Geisinger, 2016). Hasta el momento en España se han revisado un total de 55 tests procedentes de diferentes casas editoriales, han participado más de 100 revisores y cinco coordinadores. Actualmente ya

se ha comenzado la sexta revisión, coordinada por la Catedrática de la Universidad de Murcia María Dolores Hidalgo Montesinos. En términos generales, la evaluación de la calidad de los tests mediante el CET/CET-R en las últimas cinco evaluaciones arrojan unos resultados que demuestran que nuestros tests tienen una calidad razonablemente buena, con claras fortalezas pero también aspectos a mejorar.

La incorporación del CET-R en esta quinta evaluación ha sido totalmente adecuada. Además, los resultados encontrados tanto a nivel cuantitativo como cualitativo han sido plenamente satisfactorios, si bien es cierto que se tienen que incorporar nuevos datos sobre el estudio del funcionamiento diferencial de los ítems o invarianza de medición, así como nuevos estudios bajo la óptica de la TRI que añadan información, entre otros aspectos, de las curvas características de los ítems o de las funciones de información de los tests. Se debe recordar que el estudio del funcionamiento diferencial de los ítems, permite garantizar, en cierta medida, la equidad en el proceso de medición. La ausencia de funcionamiento diferencial en un ítem supone que la probabilidad de respuesta correcta depende únicamente del nivel del participante en la variable objeto de medición, y no está condicionada por la pertenencia a un grupo o característica (p. ej., género, cultura, etnia, etc.). En suma, hay que garantizar que la evaluación se lleva a cabo con equidad y de forma justa. Los autores del presente trabajo son plenamente conocedores de que tanto los investigadores y los profesionales de la psicología como los editores están trabajando intensamente en estas y otras líneas de interés.

El CET-R parece ser una herramienta interesante para mejorar el uso de los tests y la práctica profesional de los psicólogos. Como bien señalan Elosua y Geisinger (2016) la publicación de revisiones independientes sujetas a procedimientos estandarizados y criterios científicos suponen un impulso en el proceso continuo de la mejora de la construcción/edición/uso de tests. Además, el CET-R es una herramienta que se podría utilizar a modo de gold standard, obviamente reconociendo sus limitaciones, para ayudar a la formación tanto de los futuros profesionales de la psicología como en los cursos de reciclaje además de la construcción/edición de tests. La formación y concienciación por parte de los profesionales y diferentes organizaciones es necesaria ya que el uso (in)correcto de los tests tiene claras repercusiones en la vida de las personas u organizaciones que son objeto de evaluación y medición. Una correcta utilización de los tests, en cualquiera de innumerables ámbitos donde se usan, mejorará sin lugar a dudas la práctica profesional, y como no, la visión de la psicología como ciencia y profesión.

El proceso de revisión de tests es laborioso y complejo, además de enriquecedor. Este proceso de evaluación de tests editados en España se encuadra dentro de un conjunto más amplio de organizaciones y de directrices internacionales para el buen uso de los tests. Por ejemplo, como bien señala Elosua (2017), la Comisión Internacional de Tests (ITC, International Test Commission) ha desarrollado seis conjuntos de directrices: a) traducción y adaptación de tests; b) uso de los tests; c) tests computerizados y tests administrados por internet; d) seguridad de los tests, exámenes y otras evaluaciones; e) control de calidad de puntuaciones de los tests, su análisis y los informes sobre las puntuaciones; y f) uso profesional de revisiones de los tests, tests obsoletos y retiradas de tests. Muchas de ellas han sido traducidas al español por los miembros de la Comisión Nacional de Tests del COP (Hernández et al., 2016; Muñiz, Elosua, y Hambleton, 2013). Para una información más detallada invitamos al lector a visitar la página web de la ITC en la sección guidelines (<https://www.intestcom.org/>) así como excelentes trabajos previos (Leong, Bartram, Cheung, Geisinger, y Iliescu, 2016; Wells y Faulkner-Bond, 2016).

### CONSTRUYENDO EL FUTURO

En los últimos años los avances en evaluación y medición psicológica y educativa han sido evidentes (Fonseca-Pedrero y Muñiz, 2016), y todo hace pensar que el futuro nos deparará muchos más. A continuación se comentan algunas líneas de investigación futuras concernientes a la evaluación de la calidad de los tests. Estas líneas se insertan, como no puede ser de otro modo, dentro del marco más general de la evaluación psicológica y educativa.

Primero, habrá que evaluar al evaluador, es decir, por un lado tendremos que estimar las propiedades psicométricas del modelo de evaluación CET-R, y por otro la fiabilidad interjueces de los revisores de los tests.

Son tareas pendientes. Además, sería muy conveniente desarrollar una versión web del CET-R que pudiera ser utilizada por los profesionales.

Segundo, y dado que la psicología como ciencia y profesión se encuentra en continua evolución, es sumamente importante que el modelo de evaluación de calidad de tests (CET-R) se adapte rápidamente a aquellos hechos más destacables que acontezcan en el área de la evaluación y medición psicológica y educativa. Por ejemplo, son interesantes las nuevas formas de evaluación que nos llegan de la mano de las nuevas tecnologías. El gran cambio en el ámbito de la tecnología utilizada en la evaluación es el paso progresivo de las pruebas de papel y lápiz clásicas a la utilización de la tecnología informática. Autores como Drasgow (2016) o Sireci y Faulkner-Bond (2016) indican que las nuevas tecnologías están influyendo sobre todos los aspectos de la evaluación psicológica, tales como el diseño de los tests, la construcción de los ítems, la presentación de los ítems, la puntuación de los tests y la evaluación a distancia, por citar algunas. Citaremos solo dos formas de evaluación emergentes: los tests adaptativos informatizados (TAIs) y la evaluación ambulatoria.

Los TAI, aunque se llevan utilizando bastantes años, todavía no se ha extendido su uso, por lo que siguen siendo una interesante línea futura de trabajo (Olea, Abad, y Barrada, 2010; Zenisky y Luecht, 2016). En España actualmente se está trabajando en varios proyectos, por ejemplo, la evaluación de la inteligencia desde el modelo de matrices de RAVEN, la evaluación del clima organizacional en contextos laborales o la evaluación de la personalidad emprendedora en jóvenes.

La evaluación ambulatoria es un tema de estudio clásico en psicología, si bien está resurgiendo con nuevos bríos en la actualidad (Trull y Ebner-Priemer, 2009, 2013; van Os, Delespaul, Wigman, Myin-Germeys, y Wichers, 2013). La evaluación ambulatoria abarca una amplia gama de métodos de evaluación que tratan de estudiar las experiencias de las personas en su entorno natural y en la vida diaria. Debido a ciertas limitaciones de los tests tipo autoinforme (p. ej., sesgo retrospectivo, contexto de evaluación artificial, etc.) recientes investigaciones proponen desplazarse, o al menos incorporar, nuevas formas de evaluación, que permitan comprender determinadas variables y constructos psicológicos desde una perspectiva más dinámica, personalizada, contextual y ecológica. Se pretende analizar utilizando registros estructurados de preguntas las conductas, los sentimientos, las cogniciones, las emociones y/o los síntomas de las personas mediante dispositivos móviles en su contexto real diario. Para ello habitualmente se realizan evaluaciones varias veces al día durante un período temporal (típicamente una semana) que permita captar la variabilidad de los fenómenos. Las preguntas se activan mediante un “beep” en un marco temporal fijado por el investigador, por ejemplo, entre las 10 am y las 10 pm. Además, estos “beeps” pueden presentarse de forma aleatoria y/o en intervalos de tiempo predeterminados, por ejemplo, cada 90 minutos. A lo largo de cada día se recogen diferentes muestras de comportamiento, aproximadamente 6-8 por día durante 7 días. Todos estos datos se vuelcan a una plataforma para su análisis posterior. Se trata, pues, de un abordaje complementario a los procedimientos tradicionales de evaluación psicométrica basada en el lápiz y papel en contextos más o menos artificiales y de corte más bien transversal y retrospectivo.

Tercero, se tiene que contemplar la posibilidad de incorporar al modelo de evaluación de tests (CET-R) nuevos modelos psicométricos como el análisis de redes (Borsboom y Cramer, 2013; Fonseca-Pedrero, 2017), los modelos multidimensionales dentro de la TRI (Reckase, 2009) u otros modelos matemáticos procedentes de la teoría de los sistemas dinámicos o la teoría del caos (Nelson, McGorry, Wichers, Wigman, y Hartmann, 2017), por citar algunos.

Finalmente, la continua formación de los profesionales de la psicología es la piedra angular para los años venideros, pues cada vez será más frecuente e ineludible poseer un conocimiento en técnicas y procedimientos de evaluación que vayan más allá de la mera administración de pruebas tipo tests o entrevistas. El desarrollo de las neurociencias y de la robótica, así como los enfoques traslacionales y multidisciplinarios en ciencia son un claro ejemplo.



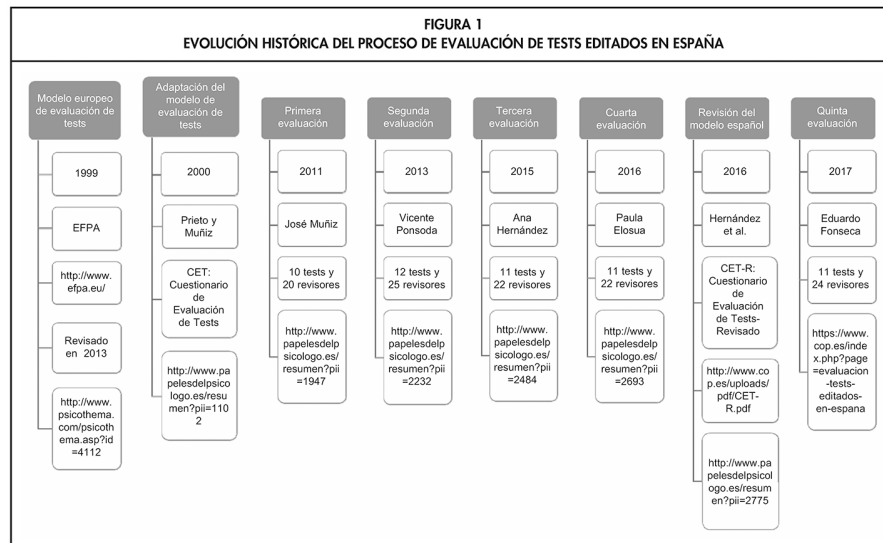
Se percibe una atmósfera de cambio entre los profesionales de la psicología. Con total seguridad, mucho de las nuevas líneas de investigación comentadas aquí se irán consolidando en los años venideros. El camino que queda por recorrer es, sin duda alguna, fascinante.

### AGRADECIMIENTOS

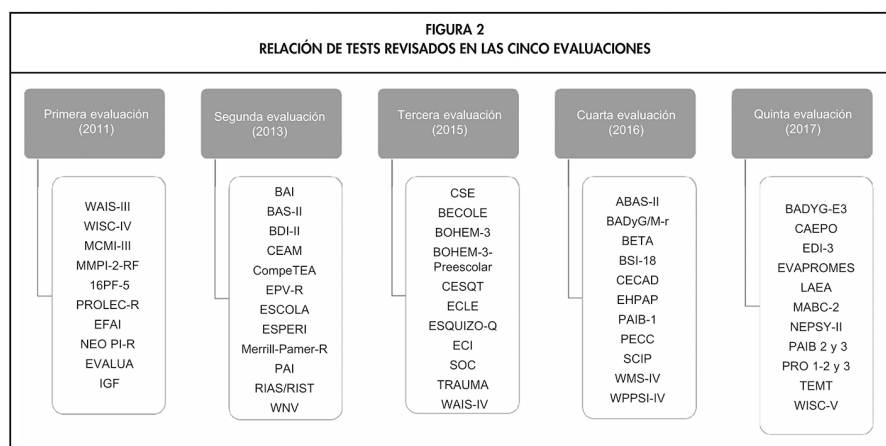
Deseamos expresar nuestro más sincero agradecimiento a los miembros de la Comisión Nacional de Tests y al personal administrativo del COP, así como a los revisores y a las distintas casas editoriales, sin cuya ayuda y colaboración este trabajo no se hubiese podido realizar.

### Conflicto de intereses

No existe conflicto de intereses



**FIGURA 1**  
**EVOLUCIÓN HISTÓRICA DEL PROCESO DE EVALUACIÓN DE TESTS EDITADOS EN ESPAÑA**



**FIGURA 2**  
**RELACIÓN DE TESTS REVISADOS EN LAS CINCO EVALUACIONES**

<b>TABLA 1</b> <b>RELACIÓN DE INSTRUMENTOS DE MEDIDA ANALIZADOS</b> <b>EN LA QUINTA EVALUACIÓN DE TESTS</b>		
<b>Acrónimo</b>	<b>Test</b>	<b>Editorial</b>
NEPSY-II	NEPSY-II Batería Neuropsicológica Infantil	Pearson Education
WISC-V	Escala de Inteligencia de Wechsler para Niños-V	Pearson Education
MABC-2	Evaluación de la Batería de Evaluación del Movimiento para niños-2	Pearson Education
TEMT	Test de Evaluación Matemática Temprana	EOS
EVAPROMES	Evaluación de los Procesos Metacognitivos en Escritura	EOS
LAEA	Listado de Adjetivos para la Evaluación del Autoconcepto	TEA ediciones
EDI-3	Inventario de Trastornos de la Conducta Alimentaria-3	TEA ediciones
CAEPO	Cuestionario de Afrontamiento al Estrés para Pacientes Oncológicos	TEA ediciones
BADYG-E3	Batería de Aptitudes Diferenciales y Generales Renovado E3	CEPE, S.L.
PRO 1-2 y 3	Prueba de Rendimiento Ortográfico Primaria (1-2) y Secundaria (3)	CEPE, S.L.
PAIB 2 y 3	Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas (2 y 3)	CEPE, S.L.

**TABLA 1**  
**RELACIÓN DE INSTRUMENTOS DE MEDIDA**  
**ANALIZADOS EN LA QUINTA EVALUACIÓN DE TESTS**

**TABLA 2**  
**REVISORES QUE HAN PARTICIPADO EN LA**  
**QUINTA EVALUACIÓN DE TESTS**

Nombre	Filiación
Carmen García García	Universidad Autónoma de Madrid
Javier Suárez Álvarez	Universidad de Oviedo
Beatriz Lucas Molina	Universidad de Valencia
Alicia Pérez de Albéniz	Universidad de La Rioja
Javier Ortuño-Sierra	Universidad Loyola Sevilla
Félix Inchausti Gómez	Consejería de Salud de Navarra
Jorge López Puga	Universidad Católica de Murcia
Ana Vanesa Valero	Universidad de La Rioja
Irene Solbes Canales	Universidad Complutense de Madrid
Joan Guàrdia Olmos	Universidad de Barcelona
David Álvarez García	Universidad de Oviedo
Ricardo Olmos Albacete	Universidad Autónoma de Madrid
Isabel Benítez Baena	Universidad de Loyola Sevilla
Susana Al-Halabí Díaz	Universidad de Oviedo
Gloria García Fernández	Universidad Complutense de Madrid
Susana Sierra Baigrie	Universidad de Oviedo
Cristina Anguiano Carrasco	Educational Testing Service (ETS)
Mercedes Páino Piñero	Universidad de Oviedo
Roberto Lozano Herce	Consejería de Educación de La Rioja
Silvia Sastre i Riba	Universidad de La Rioja
Juana Gómez Benito	Universidad de Barcelona
Joan Pere Ferrando	Universidad Rovira i Virgili
Nekane Balluerka Laso	Universidad del País Vasco
Maite Garaigordobil Landazabal	Universidad del País Vasco

**TABLA 2**  
**REVISORES QUE HAN PARTICIPADO EN LA QUINTA EVALUACIÓN DE TESTS**

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

Características	Tests										
	NEPSY-II	WISC-V	MABC-2	EVAPROMES	TEMT	LAEA	EDI-3	CAEPO	BADYG-E3	PRO 1-2 y 3	PAIB 2 y 3
Materiales y documentación	5	5	5	4	5	4	5	3,5	4,5	4	4
Fundamentación teórica	5	5	5	4,5	5	5	5	3,5	5	5	4
Adaptación	4	4,5	5	—	5	—	5	—	—	—	—
Análisis de los ítems	—	4,5	4	4	4,5	4	4	—	5	4	4
Validez: contenido	4	5	4	4	4	4	4,5	—	5	3,5	4
Validez: relación con otras variables	3,5	4,5	4	4,5	3,5	4	4	2,5	3,5	2,5	3
Validez: estructura interna	—	4,5	4,5	4	3,5	4	5	3	4	3	2
Validez: análisis del DIF	—	—	3	5	—	—	4	—	—	—	—
Fiabilidad: equivalencia	—	—	—	—	3	—	—	—	—	—	—
Fiabilidad: consistencia interna	4	5	4	5	4,5	4,5	4	4,5	5	4	4
Fiabilidad: estabilidad	3,5	4	3	—	3	5	3,5	—	—	—	—
Fiabilidad: TRI	—	—	—	—	—	—	—	—	—	—	—
Fiabilidad: inter-jueces	5	5	5	—	—	—	—	—	—	—	—
Boremos e interpretación de puntuaciones	4	5	4,5	4	4,5	5	4,5	2,5	5	3,5	3,5

Nota. Las puntuaciones de la tabla están hechas en una escala de 1 a 5, y corresponden a las siguientes valoraciones: 1 = inadecuada; 2 = adecuada pero con carencias; a partir de 2,5 = adecuada; a partir de 3,5 = buena; a partir de 4,5 = excelente. Cuando aparece el símbolo (—) significa que no se aporta información o no procede.

## RESUMEN DE LAS CALIFICACIONES DE LOS TESTS ANALIZADOS EN LA QUINTA EVALUACIÓN

## REFERENCIAS

American Educational Research Association, American Psychological Association, y National Council on Measurement in Education. (2014). Standards for Educational and Psychological Testing. Washington, D.C.: Author.

- Borsboom, D., y Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91–121.
- Dragow, F. (Ed.). (2016). *Technology and testing*. New York: Routledge.
- Elosua, P. (2017). Avances, proyectos y retos internacionales ligados al uso de tests en Psicología. *Estudios de Psicología*, 34, 201–210.
- Elosua, P., y Geisinger, K. F. (2016). Cuarta evaluación de tests editados en España: Forma y fondo. *Papeles del Psicólogo*, 37, 82–88.
- Evers, A. (2012). The internationalization of test reviewing: Trends, differences, and results. *International Journal of Testing*, 12, 136–156.
- Evers, A., McCormick, C., Hawley, L., Muñoz, J. et al. (2017). Testing practices and attitudes toward tests and testing: An international survey. *International Journal of Testing*, 17, 158–190.
- Evers, A., Muñoz, J., Hagemister, C., Høtmølingen, A., Lindley, P., Sjöberg, A., y Bartram, D. (2013). Assessing the quality of tests: revision of the EFPA review model. *Psicothema*, 25, 283–91.
- Fonseca-Pedrero, E. (2017). Análisis de redes: ¿una nueva forma de comprender la psicopatología? *Revista de Psiquiatría y Salud Mental*, <https://doi.org/10.1016/j.rpsm.2017.06.004>.
- Fonseca-Pedrero, E., y Muñoz, J. (2016). Avances en evaluación psicológica. *Papeles del Psicólogo*, 37, 2-2.
- Hernández, A., Ponsoda, V., Muñoz, J., Prieto, G., y Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 37, 161–168.
- Hernández Baeza, A., Tomás, I., Ferreres, A., y Lloret, S. (2015). Tercera evaluación de tests editados en España. *Papeles del Psicólogo*, 36, 1–8.
- Leong, F. T. L., Bartram, D., Cheung, F. M., Geisinger, K. ., y Iliescu, C. (2016). *The ITC International Handbook of Testing and Assessment*. New York: Oxford University Press.
- Muñoz, J. (1997). Aspectos éticos y deontológicos de la evaluación psicológica. En A. Cordero (Coord.), *La Evaluación psicológica en el año 2000*. Madrid: TEA Ediciones. (págs. 307-345).
- Muñoz, J. y Bartram, D. (2007). Improving international tests and testing. *European Psychologist*, 12(3), 206-219.
- Muñoz, J., Elosua, P., y Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25, 151–157.
- Muñoz, 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.
- Muñoz, J., Fernández-Hermida, J. R., Fonseca-Pedero, E., Campillo-Álvarez, Á., y Peña-Suárez, E. (2011). Evaluación de los tests editados en España. *Papeles del Psicólogo*, 32, 113–128.
- Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T. W., y Hartmann, J. A. (2017). Moving from static to dynamic models of the onset of mental disorder. *JAMA Psychiatry*, 74, 528-534.
- Olea, J., Abad, F., y Barrada, J. R. (2010). Tests informatizados y otros nuevos tipos de tests. *Papeles del Psicólogo*, 31, 94-107.
- Ponsoda, V., y Hontangas, P. (2013). Segunda evaluación de tests editados en España. *Papeles del Psicólogo*, 34, 82–90.
- Prieto, G., y Muñoz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-77.
- Reckase, M. D. (2009). *Multidimensional Item Response Theory*. New York, NY: Springer-Verlag New York.
- Trull, T. J., y Ebner-Priemer, U. W. (2009). Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: Introduction to the special section. *Psychological Assessment*, 21, 457–462.
- Trull, T. J., y Ebner-Priemer, U. W. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology*, 9, 151–176.
- van Os, J., Delespaul, P., Wigman, J., Myin-Germeys, I., y Wichers, M. (2013). Beyond DSM and ICD: introducing “precision diagnosis” for psychiatry using momentary assessment technology. *World Psychiatry*, 12, 113–117.

- Wells, C. S., y Faulkner-Bond, M. (2016). Educational Measurement. From Foundations to Future. New York, NY: The Guilford Press.
- Zenisky, A. L., y Luecht, R. M. (2016). The future of computer-based testing: Some new paradigms. En C. S. Wells and M. Faulkner-Bond (Eds.), Educational measurement. From foundations to future (págs. 221-238). New York: Guilford Press.