



# NUEVAS DIRECTRICES SOBRE EL USO DE LOS TESTS: INVESTIGACIÓN, CONTROL DE CALIDAD Y SEGURIDAD

## NEW GUIDELINES FOR TEST USE: RESEARCH, QUALITY CONTROL AND SECURITY OF TESTS

José Muñiz<sup>1</sup>, Ana Hernández<sup>2</sup> y Vicente Ponsoda<sup>3</sup>

<sup>1</sup>Universidad de Oviedo. <sup>2</sup>Universidad de Valencia. <sup>3</sup>Universidad Autónoma de Madrid

**Antecedentes.** Para llevar a cabo una evaluación psicológica rigurosa es necesario que los profesionales que la realizan tengan una preparación adecuada, que los tests utilizados muestren unas buenas propiedades psicométricas, y que se utilicen de forma correcta. El objetivo de este trabajo es presentar las directrices recientes de la Comisión Internacional de Tests sobre el uso de los tests en tres ámbitos: investigación, control de calidad y seguridad en el manejo de las pruebas. **Método.** Se revisarán y comentarán los directrices recientes desarrolladas por la Comisión Internacional de Tests. **Resultados.** Las nuevas directrices sobre el uso de los tests ofrecen todo un conjunto de recomendaciones teórico-prácticas para guiar la utilización adecuada de los tests en contextos de investigación, para desarrollar e implementar procesos de control de calidad efectivos, y para salvaguardar la seguridad de todos los datos implicados en un proceso evaluativo. **Conclusiones.** Las nuevas directrices desarrolladas por la Comisión Internacional de Tests contribuirán a una adecuada utilización de los tests en contextos de investigación, a una mejora en los procesos de control de calidad de las pruebas, y a garantizar la seguridad en los procesos evaluativos.

**Palabras clave:** Uso de tests, Comisión Internacional de Tests, Directrices, investigación, Seguridad, Control de calidad.

**Background.** In order to carry out a rigorous psychological evaluation three conditions must be met: the practitioners must have the appropriate qualifications, the tests must show good psychometric properties, and they must be used correctly. The aim of this paper is to present the recent guidelines developed by the International Test Commission on the use of tests in three areas: research, quality control, and security. **Method.** The guidelines developed by the International Test Commission will be analysed and discussed. **Results.** The new guidelines on the use of tests offer a whole range of theoretical and practical recommendations to guide the appropriate use of tests in research settings, in order to develop and implement effective quality control strategies, and to preserve the security of all of the data involved in the assessment process. **Conclusions.** The new guidelines developed by the International Test Commission will contribute to the correct use of tests in research settings, to an improvement in the quality control of testing, and to ensuring security in the assessment processes.

**Key words:** Test use, International Test Commission, Guidelines, Research, Quality control, Security.

Los tests constituyen unas herramientas fundamentales en la práctica profesional de los psicólogos, y como cualquier otra herramienta hay que utilizarla debidamente. La utilidad de los tests se asienta sobre tres pilares básicos: los profesionales deben tener una preparación pertinente, las pruebas unas propiedades psicométricas apropiadas, y el uso que se haga de ellas debe de ser el adecuado. Si se cumplen esos tres requisitos los tests serán de gran ayuda para los psicólogos a la hora de ejercer su profesión. Las universidades y otras instituciones educativas centran sus esfuerzos en la formación de los profesionales, los editores tratan de poner en el mercado las mejores pruebas posibles, y distintas organizaciones nacionales e internacionales se esfuerzan por mejorar el uso que se hace de los instrumentos de medida. Entre estas organizaciones merecen especial mención la Federación Europea de Asociaciones de Psicólogos (EFPA) y la

Comisión Internacional de Tests (ITC). A nivel nacional el Consejo General de Psicología de España (COP) colabora estrechamente con la EFPA y la ITC y es miembro de ambas. Estas organizaciones nacionales e internacionales llevan a cabo acciones y proyectos muy variados, los cuales pueden articularse en torno a dos grandes estrategias, que podemos denominar *restrictiva* e *informativa*. A continuación se describen brevemente ambas estrategias, siguiendo los trabajos previos de Muñiz y Bartram (2007), Muñiz y Fernández-Hermida (2010) y Muñiz (2012).

La estrategia *restrictiva* agrupa todo un conjunto de acciones llevadas a cabo para restringir el uso de los tests a aquellos profesionales que están realmente preparados para hacerlo. Los sistemas utilizados varían de unos países a otros (Bartram, 1996; Bartram y Coyne, 1998; Muñiz, Prieto, Almeida y Bartram, 1999), si bien uno de los más habituales en varios países, incluida España, es clasificar los tests siguiendo los criterios de la APA en tres categorías (A, B, C) de menos a más espe-



cializados, siendo exclusivo de los psicólogos el uso de los tests de las categorías B (tests colectivos de carácter cognoscitivo y Personalidad) y C (escalas individuales y tests proyectivos). Otra opción utilizada es que los profesionales obtengan una certificación específica en la que acrediten fehacientemente que conocen adecuadamente las pruebas. Aunque estas restricciones y otras son recomendables, no garantizan por sí solas un uso adecuado de los tests (Moreland, Eyde, Robertson, Primoff y Most, 1995; Simner, 1996), siendo necesario complementar esta estrategia con la difusión de información a todas las partes implicadas, tales como profesionales, usuarios, instituciones, y sociedad en general. Las acciones llevadas a cabo en el marco de la estrategia que hemos denominado *informativa* se refieren a todo tipo de iniciativas encaminadas a difundir información sobre la práctica de los tests. Se entiende que cuanto más información posean los profesionales, los usuarios, las familias, y en general todas las partes implicadas en el uso de los tests, menor será la probabilidad de que se haga un mal uso de las pruebas. En este sentido distintas organizaciones nacionales e internacionales han desarrollado códigos éticos y deontológicos, así como directrices varias para guiar el uso adecuado de los tests. Entre los primeros cabe destacar el meta-código ético de la EFPA (2005), el código desarrollado por el comité norteamericano para la buena práctica de los tests (2002), o las directrices de la asociación europea de evaluación psicológica (Fernández-Ballesteros et al., 2001). Véanse buenas revisiones en autores como Koocher y Keith-Spiegel (2007), Lindsay, Koene, Ovreeide y Lang (2008), o Leach y Oakland (2007), y sobre todo en el último número monográfico dedicado al tema por la revista *Papeles del Psicólogo* (2009). Aparte de estos códigos disponemos en la actualidad de un conjunto de directrices que marcan los pasos a seguir desde la propia construcción de la prueba, su aplicación, interpretación y aplicación de los resultados (Bartram, 1998; Brennan, 2006; Downing y Haladyna, 2006; Muñiz, 1997). Merecen mención especial los estándares técnicos desarrollados por la Asociación Americana de Psicología y otras dos organizaciones (AERA, APA y NCME, 2014), así como las directrices elaboradas por la Comisión Internacional de Tests (ITC) para la traducción y adaptación de los tests de unas culturas a otras (Hambleton, Merenda y Spielberger, 2005; Muñiz, Elosua y Hambleton, 2013). Para consultar otras directrices sobre el uso de los tests

en general, de los tests informatizados e Internet, o la utilización de los tests en el ámbito del trabajo y las organizaciones, véase, por ejemplo, el trabajo de Muñiz y Bartram (2007), o la página web de la ITC ([www.intestcom.org](http://www.intestcom.org)) y de la EFPA ([www.efpa.eu](http://www.efpa.eu)). También en la página web del Consejo General de Psicología de España, en el apartado de la Comisión de Tests, se puede consultar información de interés ([www.cop.es](http://www.cop.es)). Aparte de los códigos éticos y las directrices, hay dos medidas que merecen atención dentro de las acciones enmarcadas en la estrategia de la información; se trata por un lado de una nueva norma ISO-10667 que regula todo lo relativo a la evaluación de personas en contextos laborales, y por otro los modelos de evaluación de tests desarrollados en distintos países, entre ellos España (Prieto y Muñiz, 2000) que se aplican a los tests disponibles en el mercado con el fin de proporcionar información sobre sus puntos fuertes y aspectos a mejorar (Hernández, Tomás, Ferreres y Lloret, 2015; Muñiz, Fernández-Hermida, Fonseca-Pedrero, Campillo-Álvarez, y Peña-Suárez, 2011; Ponsoda y Hontangas, 2013).

Dentro de esta estrategia *informativa*, la ITC ha desarrollado recientemente tres documentos encaminados a mejorar el uso de los tests en tres ámbitos distintos. El primero es una declaración sobre el uso de los tests en investigación, el segundo son unas directrices sobre el control de calidad de los tests, y el tercero unas directrices sobre la seguridad de las pruebas. Estos tres documentos han sido recientemente traducidos al español y el objetivo central de este trabajo es presentarlos y subrayar sus aportaciones más relevantes. La versión completa en español de los tres documentos puede consultarse en la página web de la Comisión de Tests del Consejo General de Psicología de España ([www.cop.es](http://www.cop.es)), sección Comisión de Tests.

## EL USO DE LOS TESTS Y OTROS INSTRUMENTOS DE EVALUACIÓN EN INVESTIGACIÓN

Este documento fue iniciado por la profesora Fanny Cheung y preparado para la ITC por los profesores Dragos Iliescu y Dave Bartram, y fue traducido al español por José Muñiz.

Esta declaración de la ITC trata de aclarar todo lo relativo a la utilización de los tests con fines de investigación. Está dividido en siete partes: Permiso para usar los tests en investigación, Permiso de reimpresión, Modificación del test o de sus componentes, Uso ético de los tests, Documentación, Conflicto de intereses, y Utilización de



los tests de investigación en la práctica profesional. Se comentan a continuación algunos aspectos de cada apartado.

### PERMISO PARA USAR LOS TESTS EN INVESTIGACIÓN

En este apartado el aspecto clave que subraya la declaración de la ITC es que si los tests tienen *copyright* el investigador tiene que pedir permiso a quien sea el titular de ese derecho. Sólo las pruebas de libre acceso pueden ser utilizadas sin solicitar permiso, ahora bien, este tipo de situación es muy poco frecuente, aunque algunos investigadores no sean conscientes de ello. En este sentido, un error muy típico es asumir que si una prueba aparece publicada en una revista u otro documento esa prueba es de libre acceso. En la declaración de la ITC se recomienda que ante la más mínima duda debe de pedirse permiso a los autores para usar una prueba. Nótese lo que indica literalmente la declaración de la ITC: *Las versiones de los instrumentos para investigación se publican a menudo en revistas o en las páginas web de los autores. Cuando se publican en estos medios de libre acceso, los tests pueden parecer de uso público, sin embargo, por defecto, el autor es el titular del copyright hasta que ceda sus derechos a otra entidad, o dé un permiso explícito para su uso libre.*

### PERMISO DE REIMPRESIÓN

Aquí el mensaje sobre la reimpresión de los ítems de un test es muy claro, no puede hacerse sin el permiso correspondiente: *Los trabajos con copyright no deben de reproducirse, distribuirse ni exponerse públicamente; tampoco se deben llevar a cabo trabajos derivados de ellos sin el permiso del titular de los derechos de autor.* A la hora de presentar los resultados los investigadores pueden verse obligados a presentar algunos ítems para describir mejor sus datos, en ese caso o bien se pide permiso a los titulares de los derechos de autor, o como alternativa los investigadores pueden elaborar ítems similares a los del test original pero que no los reproducen, de modo que no se violan los derechos de autor.

### MODIFICACIÓN DEL TEST O DE SUS COMPONENTES

El mensaje de la declaración de la ITC sobre la modificación de un test es muy claro: *Cuando se utilicen tests sujetos a copyright, los investigadores no deben modificar ningún componente del test, pues se pondría en peligro la integridad de la prueba, se violaría los derechos de autor y se cometería una ilegalidad, salvo que los investigado-*

*res estén debidamente autorizados para ello.* Por supuesto, si el test es de uso público o se dispone del correspondiente permiso pueden llevarse a cabo modificaciones para ajustar la prueba a las condiciones de uso, *por ejemplo, para su utilización en otra cultura diferente a la que fue generado puede requerir su traducción a otro idioma, eliminar algunos aspectos, reformular otros, añadir ítems, o cambiar las instrucciones y la escala de los ítems.*

### USO ÉTICO DE LOS TESTS

En este sentido lo que se recoge en la declaración de la ITC está en línea con lo establecido en otras normas y códigos éticos en relación con el uso de los tests en el ámbito profesional: *El uso ético de los tests en investigación y en contextos profesionales es muy parecido. Quienes utilizan tests en investigación deben de actuar de forma ética y profesional, ser competentes en el uso de los tests, ser responsables de su uso, y garantizar la seguridad de los materiales utilizados y la confidencialidad de los resultados... La responsabilidad de un profesional cualificado abarca todo el proceso de evaluación, incluyendo la recogida de datos, codificación, análisis, realización de informes y la aplicación de los datos en sus distintas formas.*

### DOCUMENTACIÓN

La investigación requiere una documentación rigurosa que permita la replicación, así que cuando se utilicen tests hay que proporcionar una información completa sobre sus características y propiedades psicométricas, mereciendo especial mención el caso de pruebas de nueva creación, como bien establece la declaración de la ITC: *Los investigadores que construyen un nuevo instrumento de medida, como mínimo deben de proporcionar información sobre los fundamentos teóricos del test y su finalidad, el sistema seguido para la selección inicial de los ítems, cómo se analizaron posteriormente y los criterios de selección, el número de ítems de cada faceta, los métodos de escalamiento utilizados, además de información sobre las evidencias de validez y la precisión de las mediciones, por ejemplo fiabilidad u otros indicadores de la precisión de las mediciones, según el modelo de medida del test.*

### CONFLICTO DE INTERESES

Los investigadores tienen que hacer constar sus fuentes de financiación, por si ello pudiese afectar a su impar-



cialidad en el proceso investigador, como señala la declaración de la ITC: *Las investigaciones financiadas por agencias externas pueden tener un interés particular en los resultados, por ejemplo un departamento gubernamental que está tratando de implantar una determinada política, o una compañía editora de tests que es la titular del copyright.*

### UTILIZACIÓN DE LOS TESTS DE INVESTIGACIÓN EN LA PRÁCTICA PROFESIONAL

Cuando se utilizan los tests en contextos profesionales, clínico, educativo, trabajo, u otros, sus características y propiedades psicométricas tienen que ser muy exigentes, dado que se van a tomar decisiones importantes sobre las vidas de las personas. Estas condiciones pueden ser más relajadas cuando los tests se utilizan solamente para investigación, donde los datos suelen utilizarse de forma agregada, e incluso los propios tests están siendo contruidos y probados. Ahora bien, hay que ser muy cautelosos y advertir de las limitaciones cuando las versiones de investigación de los tests se utilizan en el campo profesional. Como bien señala la declaración de la ITC: *El autor del test tiene la responsabilidad de no contribuir a la utilización de los instrumentos de medida de investigación en la práctica profesional antes de que se haya publicado suficiente información sobre sus propiedades psicométricas.*

Confiamos en que estas recomendaciones de la ITC ayuden a los investigadores a mejorar el uso de los tests, lo cual contribuirá a su vez a una mayor calidad de las investigaciones, que sin duda repercutirá en la generación de unos instrumentos de medida de mayor calidad para la práctica profesional.

### DIRECTRICES PARA EL CONTROL DE CALIDAD DE LAS PUNTUACIONES DE LOS TESTS, SU ANÁLISIS Y LOS INFORMES SOBRE LAS PUNTUACIONES

Estas directrices fueron preparadas para el consejo de la ITC por Avi Allalouf y traducidas al español por Ana Hernández. Están motivadas por el hecho de que, como bien es sabido, errar es de humanos. Y a pesar de que la evaluación de personas mediante tests (entendidos estos en su acepción más amplia) es típicamente realizada por expertos profesionales, dicha evaluación no es una excepción y puede estar sujeta a errores.

Los errores cometidos pueden consistir, entre otros, en la aplicación de una plantilla de puntuación errónea, la conversión incorrecta de puntuaciones directas en esca-

las transformadas, la interpretación inadecuada de una puntuación (lo que podría depender de la transformación anterior), que se haya enviado un informe al cliente equivocado, o que se dé un retraso excesivo a la hora de informar de los resultados. Estos errores podrían tener consecuencias importantes para las personas evaluadas, la sociedad y la profesión. Por ejemplo, la aplicación de una plantilla errónea o la incorrecta transformación de una puntuación, podrían impedir que un candidato cualificado accediera a un determinado puesto de trabajo, o podría resultar en una intervención educativa inadecuada, o en conceder certificaciones académicas a personas que carecen de los conocimientos y habilidades requeridos. O por ejemplo un retraso a la hora de informar de los resultados podría causar problemas para quienes, por no recibir dicha información a tiempo, no pudieran acceder a un determinado puesto o institución. Pero además los errores también pueden afectar a las pruebas administradas y su reputación, con una reducción de la fiabilidad y validez de las puntuaciones obtenidas. Todo ello contribuiría a una pérdida de confianza en los tests educativos y psicológicos y a cuestionar los procesos de evaluación.

Por ello, los profesionales usuarios de tests deben ser capaces de anticipar los posibles errores, prevenirlos, y abordarlos, y este es el objetivo principal de las directrices. En concreto, las directrices de control de calidad se centran especialmente en los errores que pueden cometerse durante las fases de asignación de puntuaciones, análisis del test y elaboración de informes, aunque también puede resultar útil tener en cuenta estas directrices para las fases iniciales del proceso de evaluación: construcción o selección del test, y su administración. Las directrices están pensadas para situaciones de evaluación a gran escala (educativas o laborales), cuando el test constituye principalmente una medida de rendimiento, desempeño o habilidad (por contraposición a actitudes, preferencias, etc). Así pues, las directrices están principalmente dirigidas a los profesionales que participan en este tipo de evaluaciones: constructores de las pruebas, administradores, correctores, psicómetras, personas implicadas en el mantenimiento de la seguridad del test, psicólogos y educadores, o programadores informáticos, entre otros. Sin embargo, muchas de las propuestas incluidas en las directrices, también podrían aplicarse a evaluaciones a menor escala, con objetivos diferentes, o realizadas mediante otros tipos de pruebas (entrevistas, muestras de trabajo, etc).



Las directrices están estructuradas en dos partes. En la primera parte se presentan una serie de principios generales que sirven de guía. En la segunda se presentan las directrices detalladas paso a paso.

### PRINCIPIOS GENERALES

Estos se ocupan de siete cuestiones principales:

- 1) La necesidad de comprobar si existen procesos de control de calidad adecuados para aplicar a una situación evaluativa concreta, para, en caso de que no existan o no sean adecuados, desarrollar dichos procesos, adaptarlos, mejorarlos e implementarlos.
- 2) La necesidad de formular acuerdos sobre los principios básicos del proceso de evaluación entre los distintos profesionales implicados y, en ocasiones, entre las distintas partes interesadas. En este punto se incluyen cuestiones como el establecimiento de los objetivos de evaluación, el reparto de responsabilidades, la propuesta de un cronograma, la selección de las formas de puntuar más adecuadas, el mejor modo para transferir los datos, el tipo de informe a realizar y los destinatarios de dicha información, entre otras cuestiones.
- 3) La necesidad de asegurar que se dispone de los recursos necesarios (espaciales, financieros, materiales, temporales y de personal) para realizar todas las fases de evaluación según lo previsto, así como de prever recursos adicionales que pudieran necesitarse si surgiera un contratiempo.
- 4) La necesidad de, en ocasiones, realizar ajustes entre las necesidades y expectativas que tienen las partes interesadas en los resultados evaluativos (personas evaluadas, profesores, padres, etc), y las que tienen los responsables de la evaluación. Para ello es recomendable que haya una comunicación fluida entre las partes. En este punto se incluyen aspectos como el establecimiento de acuerdos, la responsabilidad en la toma de decisiones, y la posibilidad de que los examinados cuestionen o revisen los resultados y puedan hacer sugerencias.
- 5) La necesidad de asegurar que se cuenta con las personas y con el ambiente de trabajo adecuados para llevar a cabo el proceso de evaluación. En este punto se consideran los horarios y la forma de trabajo así como el apoyo al personal implicado en la evaluación (por ejemplo, mediante formación).
- 6) La necesidad de contar con uno o varios supervisores independientes que realicen un seguimiento de los pro-

cesos de control de calidad establecidos y que registren y notifiquen los errores que puedan observarse.

- 7) La necesidad de que todas las personas implicadas en el proceso de evaluación sigan los procedimientos acordados respecto a la documentación de actividades y el registro de errores mediante formularios estandarizados. Se deben establecer acuerdos sobre qué miembros del personal son responsables de cada fase del proceso y, cuando se observen fallos, informar con prontitud para prevenir que el error vuelva a darse en el futuro.

### DIRECTRICES DETALLADAS PASO A PASO

Éstas se presentan agrupadas en cinco apartados: la planificación y diseño del informe, la consideración de antecedentes y datos biográficos, las puntuaciones del test, su análisis, y la elaboración del informe. En todos estos apartados se sugieren una serie de acciones y pasos para asegurar la calidad del proceso. Se recomienda que estas acciones se lleven a cabo explícitamente en las evaluaciones a gran escala. Sin embargo, en evaluaciones a menor escala, aunque los principios de las directrices sigan siendo pertinentes, se podrían omitir o simplificar algunas de las fases. La razón radica en que algunos de los procedimientos propuestos requieren importantes recursos y están basados en modelos que requieren muestras grandes, por lo que podrían adaptarse para aplicarlos a muestras menores. A continuación se resumen algunas de las recomendaciones recogidas en los distintos apartados.

- 1) **Planificación y diseño del informe.** Puesto que todas las fases deben ir encaminadas a garantizar la calidad del producto de la evaluación, que es el informe, se recomienda que, antes de iniciar el proceso, se tomen decisiones sobre de qué informar, con qué tipo de puntuaciones, con cuánto detalle, a quién, cuándo, etc.
- 2) **Antecedentes y datos biográficos.** Se proponen recomendaciones relacionadas con este tipo de información, con el fin de verificar la identidad de las personas evaluadas manteniendo la confidencialidad de los datos, y de explicar resultados inesperados o inconsistentes con estudios previos.
- 3) **Puntuaciones.** Este apartado incluye varios subapartados: a) *La recogida y almacenamiento de las respuestas* de las personas evaluadas, que incluye recomendaciones sobre el almacenamiento de las hojas de respuesta y los datos electrónicos, el uso de códigos identificativos, la seguridad de datos, y la



necesidad de asegurar la corrección de los algoritmos y las tablas de conversión y baremos. b) La *obtención de las puntuaciones*, que incluye, entre otras cuestiones, el análisis de datos que asegure que las puntuaciones se encuentran dentro del rango esperado, la identificación y revisión de puntuaciones extremas, la identificación de personas con diferencias excesivas en las puntuaciones obtenidas en sub-tests correlacionados, y el análisis de las propiedades psicométricas de los ítems, que permitirá identificar errores en la plantilla de corrección, omisiones, etc. Y finalmente, c) para las *pruebas abiertas* de calificación del desempeño, muestras de trabajo, juegos de rol, entrevistas, etc., menos objetivas que los tests de elección múltiple, se mencionan una serie de precauciones adicionales. Se incluyen cuestiones relacionadas con la realización de cursos de formación para asignar puntuaciones, la elaboración de instrucciones de puntuación con ejemplos, o el número de evaluadores.

- 4) **Análisis del test.** Aquí las directrices se agrupan en 4 subapartados: a) En primer lugar se vuelve a hacer énfasis en el *análisis de ítems*, antes de obtener la puntuación total, con el fin de evaluar su calidad. b) *Equiparación/calibración de nuevas formas del test e ítems*, en caso de que la prueba se realice en varios momentos y/o con varias formas. En este punto se recomienda, entre otras cosas, desarrollar rutinas que aseguren que los procedimientos y diseños de equiparación especificados se han realizado correctamente y que se cumplen los supuestos en los que se basan, comprobar si diferentes procedimientos basados en diferentes supuestos ofrecen resultados similares, comparar las puntuaciones obtenidas con las que se anticiparon en función de los antecedentes y datos biográficos recogidos, o, cuando haya puntos de corte para diferenciar a los evaluados en función de su nivel, comprobar la similitud de las razones de aptos y no aptos en los distintos grupos evaluados. c) *Cálculo de puntuaciones estandarizadas*. En este punto se recomienda, entre otras cosas, comprobar la adecuación y precisión de la conversión realizada, convertir algunas puntuaciones manualmente y comparar los resultados con los generados por el ordenador, comparar los resultados obtenidos con diferentes programas, o comprobar la relación entre las puntuaciones directas y las estandarizadas usando gráficos de dispersión. Y finalmente, d) *Comprobaciones de seguridad de los tests*. Este apartado resume algunas de las

recomendaciones principales de las directrices de la ITC sobre la seguridad de los tests, de las que se ocupa también este artículo.

- 5) **Elaboración de informes.** Aquí se diferencian tres subapartados: a) *Informe sobre las puntuaciones*. Que incluye recomendaciones sobre el uso de grupos focales que permitan generar guías interpretativas, la especificación del nivel al que las puntuaciones pueden interpretarse de forma fiable, el uso de repositorios de datos para informar rápidamente de los resultados, o el asesoramiento de expertos relaciones públicas cuando los informes deban presentarse a políticos y medios de comunicación. b) *Medidas para mantener la seguridad de los informes*. En este apartado se recogen algunas recomendaciones sobre la rectificación de informes o la prevención de su falsificación. Y finalmente, c) *Documentación*. Se recomienda realizar un informe interno con información exhaustiva del proceso de obtención de puntuaciones y los estadísticos clave obtenidos, con el fin de avalar la precisión de todo el proceso. También se plantea la posibilidad de hacer públicos algunos estadísticos sobre resultados grupales (por ejemplo, por año, o por sexo) proporcionando una breve explicación sobre la interpretación de dichos estadísticos.

En definitiva, las directrices aglutinan una serie de sugerencias de control de calidad a lo largo de distintas fases del proceso de evaluación. El grado en que dichas recomendaciones se están siguiendo (o no) en una determinada situación de evaluación puede comprobarse siempre antes de pasar a una fase posterior, para verificar su cumplimiento o realizar las acciones correctivas pertinentes. Aunque algunas de las directrices sólo son aplicables a evaluaciones a gran escala, muchas de ellas son aplicables a cualquier situación de evaluación. Y aunque probablemente muchos profesionales implicados en la evaluación de personas (psicólogos y educadores) siguen muchas de las recomendaciones presentadas, la adopción sistemática de las directrices de control de calidad de la ITC contribuirá a prevenir los errores que pueden cometerse al realizar evaluaciones mediante tests.

## DIRECTRICES SOBRE LA SEGURIDAD DE LOS TESTS, EXÁMENES Y OTRAS EVALUACIONES

David Foster lideró la elaboración de estas directrices que han sido traducidas por Vicente Ponsoda. El Consejo de la ITC las aprobó en julio de 2014.

¿Por qué hacen falta ahora unas directrices sobre seguridad en los tests? La seguridad en las medidas tiene que



ver con la validez de las puntuaciones. Si hay problemas de seguridad dejamos de saber si la puntuación que estamos asignando al evaluado refleja realmente, como se supone, su nivel de conocimiento; pudiendo reflejar, en parte, su habilidad para hacer trampas al responder, el conocimiento previo de parte de las preguntas, que alguien le haya ayudado mientras respondía al test o después, etc. Incrementar la certeza de que las puntuaciones obtenidas en un test pueden interpretarse como el nivel del evaluado en el constructo que interesa evaluar es un asunto central, que facilita la justificación de las inferencias y decisiones basadas en ellas.

En las dos últimas décadas ha habido un considerable incremento de los problemas relacionados con la seguridad en las medidas en todo el mundo. Las dos principales razones son, por una parte, la cada vez más extendida evaluación informatizada, realizada muchas veces por Internet y muchas veces además sin vigilancia (evaluación conocida en inglés como *Unproctored Internet Testing*), y, por otra, la proliferación de tecnologías que permiten registrar, fotografiar, recibir información no autorizada, etc., con mucha facilidad y con instrumentos casi indetectables. A estas dos razones cabe añadir una tercera. Los problemas de seguridad son tanto más importantes cuanto más relevantes son las puntuaciones para el evaluado. Que la evaluación mediante tests sea cada vez más frecuente, en especial si son tests cuyos resultados tienen consecuencias importantes para los evaluados, es algo posiblemente también relacionado con los mayores problemas de seguridad en los tests observados en los últimos tiempos.

Las directrices se clasifican en tres apartados, según estén relacionadas con: a) el desarrollo del plan seguridad, b) la inclusión de las pautas sobre seguridad en el proceso de evaluación, y c) la respuesta a dar cuando se produce un fallo de seguridad.

Antes de resumir lo principal de las directrices de cada apartado, conviene hacer dos comentarios. En primer lugar, se debe prestar más y más atención a los posibles problemas de seguridad cuanto más importantes sean las consecuencias que se derivan para el evaluado, el profesor, el centro, etc. Cuando las consecuencias son poco o nada importantes, como, por ejemplo, cuando pedimos a nuestros estudiantes que respondan anónimamente a un test que queremos calibrar, no se esperan problemas de seguridad. En segundo lugar, como se ha indicado antes, la evaluación informatizada ha incrementado los problemas de seguridad, pero también en

tests de lápiz y papel pueden presentarse problemas de seguridad. Las directrices de hecho están pensadas para ambos tipos de tests e incluso pueden ser útiles cuando las evaluaciones sean no estandarizadas. Están pensadas, además, para un uso internacional; si bien hacen llamadas frecuentes a que se tengan en cuenta las normas y usos locales cuando vayan a ser aplicadas en cada caso concreto.

## DESARROLLO E IMPLEMENTACIÓN DE UN PLAN DE SEGURIDAD

Cuando se responde a un test se puede hacer trampas en las respuestas de muchas maneras. Algunas, conocidas, son las siguientes: el evaluado consigue parte o todo el test antes de responderlo, alguien ayuda al evaluado mientras responde al test, el evaluado utiliza medios no permitidos durante el examen (como calculadoras, teléfonos, etc.), otra persona hace el test en lugar del evaluado, alguien modifica las respuestas que el evaluado ha dado para incrementar su puntuación, el evaluado copia de otro evaluado sus respuestas, etc. Yee y MacKown (2009) listan treinta y siete maneras de hacer trampas en las evaluaciones educativas. La relación de métodos conocidos de robo del contenido del test es también considerable: robo de los cuadernillos o de los documentos digitales que contienen las preguntas; registro de las preguntas mediante, por ejemplo, una microcámara; la aplicación de procedimientos que permiten registrar toda la información que llega al ordenador en los tests informatizados; el evaluado memoriza unas cuantas preguntas, otros evaluados otras, y con este esfuerzo coordinado se consigue robar todo o casi todo el contenido del test; el contenido puede ser robado igualmente por alguna de las muchas personas que tienen acceso al test durante su elaboración; etc. Como se ha indicado, el objetivo de las directrices es reducir el riesgo de que surjan estos problemas en el programa de evaluación; y si se producen, dar pautas de cómo proceder.

Las trece directrices de este apartado indican que el elemento central para controlar la seguridad es el desarrollo de un *plan de seguridad* que indique a) quien es el responsable de cada una las fases del desarrollo del test, b) los derechos y deberes del evaluado y cómo dejar constancia de que el evaluado los conoce, c) qué hacer cuando se produce un fallo en la seguridad, y d) las exigencias que las tecnologías de la información y de la comunicación han de satisfacer para que la conservación y transmisión de los datos sean seguras.



En una última directriz de este apartado, se indica que las personas involucradas en el desarrollo y aplicación del test han de manifestar que conocen y aceptan las normas sobre no divulgación de sus contenidos, quien tiene el copyright del test, qué conductas se consideran fraudulentas y las posibles consecuencias de dichos actos.

### IMPLEMENTACIÓN DE UN SISTEMA DE SEGURIDAD EN EL PROCESO DE EVALUACIÓN CON TESTS

Las directrices de este segundo apartado son veinticinco e indican qué hacer para incrementar la seguridad en las distintas fases del desarrollo y aplicación del test. Las principales directrices se comentan brevemente a continuación.

Los evaluados han de registrarse formalmente y han de aplicarse estrictos procesos de autenticación, de forma que no haya dudas de que realiza la prueba quien se registró para hacerla.

Algunas de las características de la aplicación del test guardan una estrecha relación con la seguridad y por ello se recomienda evitar que los evaluados respondan al test más de una vez, que se estudie detenidamente el tiempo de administración del test para que no sobre tiempo, y que el evaluado no pueda volver a ver las preguntas que ya ha respondido. Se recomienda utilizar los formatos de ítems y tests que, manteniendo la calidad psicométrica, reducen la exposición de los ítems o cambian su orden de presentación, como ocurre con los tests adaptativos informatizados. Conviene controlar cuantas veces se ha administrado cada ítem; que el banco sea cuanto más grande, mejor; y se recomienda además la aplicación de tests de verificación cuando se apliquen los tests informatizados por Internet en administraciones no vigiladas.

Hay que proteger el contenido del test durante el proceso de elaboración y distribución, y cuando se está administrando. Como control de la seguridad durante el proceso de elaboración y distribución, las directrices proponen que solo las personas que han de trabajar con los ítems tengan acceso a ellos y por un tiempo limitado. A tal fin conviene establecer controles estrictos de acceso y que los autorizados firmen acuerdos de no divulgación del contenido. Se recomienda la encriptación como una medida de seguridad adicional. Como medidas para incrementar la seguridad durante la aplicación, se recomienda la participación de vigilantes motivados que preferiblemente no sean expertos en el contenido que se evalúa, y la instalación de cámaras que faciliten la vigi-

lancia a distancia que puedan registrar cualquier evento de interés ocurrido durante la aplicación del test. Si se detecta alguna anomalía se recomienda responder rápidamente interrumpiendo parcial o definitivamente la aplicación del test al evaluado, confiscando el equipo usado en el robo, si fuera el caso, y elaborando un informe sobre el incidente de seguridad. Otra directriz indica que los evaluados deben conocer las reglas de seguridad y las consecuencias de su violación antes de que se registren.

Conviene evaluar regularmente los resultados del test y de los ítems para comprobar si se ha hecho trampas y si los ítems son conocidos por los evaluados antes de la aplicación del test. Se puede comprobar si hay patrones de respuestas anormales (se aciertan los ítems difíciles y se fallan los fáciles), patrones de tiempo de respuesta anormales (tiempos muy cortos), un alto número de correcciones en la hoja de respuesta, similitud muy alta entre pares o grupos de evaluados (que puede indicar posible copia), mejoras fuertes cuando se repite el test (puede indicar que se ha hecho trampas), si hay cambios en la distribución de los parámetros de los ítems (puede indicar que se han filtrado), si funcionan como los ítems operativos del test otros tipos de ítems (por ejemplo, ítems nuevos que no han sido previamente expuestos), etc.

En lo relativo a la obtención y comunicación de las puntuaciones, se recomienda, especialmente en tests de lápiz y papel, informar de que la puntuación que se proporciona es "provisional" y que se comunicará la "definitiva" cuando se hayan considerado los informes de irregularidades que eventualmente hayan podido presentarse.

Conviene hacer un seguimiento de Internet para detectar si ha habido divulgación del contenido del test. Si fuera así, el programa de evaluación ha de ponerse en contacto con el responsable de la web y solicitar la eliminación de dichos contenidos, anunciar el inicio de acciones legales, etc.

### RESPUESTA ANTE UN FALLO EN EL SISTEMA DE SEGURIDAD

Diez son las directrices de esta tercera parte. Muestran qué hacer una vez que se ha producido un fallo en el sistema de seguridad.

Como se ha indicado más arriba, si los vigilantes o aplicadores del test ven que un evaluado está haciendo trampas o robando el contenido, se recomienda que la





aplicación al evaluado se suspenda provisional o definitivamente, y sean requisados, si la legislación lo permite, los instrumentos (cámara, teléfono, etc.) que haya utilizado, si fuera el caso. A continuación, el fallo en el sistema de seguridad debe ser investigado a fondo para determinar su extensión y la magnitud del daño, y revisar, en su caso, el plan de seguridad.

Se comentan a continuación algunas de las medidas concretas a tomar en relación al test, cuando se ha advertido un fallo de seguridad. El test que haya sido filtrado debe ser reemplazado cuando antes. Han de anularse las puntuaciones que se sepa que son incorrectas por haberse cometido algún tipo de fraude en el test y hay que recalificar al evaluado, bien a partir de sus respuestas a la parte del test no filtrada o pidiéndole que repita el test o responda a otro equivalente.

Una vez expuesto lo fundamental de las directrices sobre seguridad, conviene tener muy presente que la seguridad en los tests, como la seguridad en otros ámbitos, no es un asunto todo/nada, es más bien un continuo. El aumento en seguridad conlleva un coste y cada organización tiene que encontrar el punto de equilibrio. El objetivo de estas directrices no es, por tanto, indicar qué hacer para no tener problemas de seguridad, pues no existe una manera de evaluar que anule completamente la posibilidad de que alguien haga trampas o robe el contenido de la prueba. Como ocurre en otras áreas en las que se ha de cuidar la seguridad, es muy probable que las medidas de seguridad vayan siempre detrás de las nuevas conductas anómalas que vayan surgiendo (Foster, 2010). El objetivo de estas directrices es dar a conocer qué hacer y las mejores prácticas para que la evaluación con tests sea más segura y preservar su valor. La puesta en práctica de estas directrices ayudará a evitar que se produzcan fallos en el sistema de seguridad y a minimizar sus consecuencias si se produjeran.

En la última encuesta aplicada en España por el COP para evaluar la actitud ante los tests de los psicólogos españoles (Muñiz y Fernández-Hermida, 2010), la media de la pregunta "La aplicación de los tests por Internet abre posibilidades de fraude" fue 3.78, en la escala de 1 (total desacuerdo) a 5 (total acuerdo), por encima del punto neutro 3, parece indicar que los colegiados son conscientes de los riesgos de la administración por Internet de los tests. En otra encuesta (Ryan et al. 2015), aplicada a los responsables de selección de empresas americanas, europeas y asiáticas, se preguntó por las medidas de seguridad que venían aplicando en sus eva-

luaciones. En el caso de las evaluaciones por Internet no vigiladas, solo una medida es aplicada por más del 50% de los encuestados: el estricto control del tiempo de administración (59% de los encuestados). El uso de avisos acerca de que la evaluación incluye mecanismos para la detección de las conductas anómalas es señalada por el 40% y solo el 18% de los encuestados afirman aplicar los tests de verificación, a pesar de que una de las directrices los recomienda expresamente. En el caso de las evaluaciones informatizadas vigiladas, las medidas utilizadas por más del 50% de los encuestados son: estricto control de los tiempos de administración (66%), uso de contraseñas para acceder a los contenidos del test (58%), esmero en las tareas de vigilancia (56%) y no permitir el copie de los contenidos del test (55%). En el caso de los tests de lápiz y papel, las medidas que adoptan más del 50% de los encuestados coinciden con las cuatro acabadas de exponer y los porcentajes son muy similares. Los datos de esta encuesta, recogidos antes de la aprobación de las directrices, indican que la preocupación por la seguridad es real y se han venido tomando medidas en el pasado. Las nuevas directrices debieran facilitar a los profesionales la adopción de nuevas medidas y la elaboración de un sistema integrado para el control de la seguridad que resulte más eficaz que las medidas aisladas que se han venido aplicando.

### ALGUNAS REFLEXIONES FINALES

Se han presentado los aspectos fundamentales de la declaración de la ITC sobre el uso de los tests en investigación, así como las directrices sobre el Control de Calidad de los instrumentos de medida, y todo lo relativo a los distintos factores a tener en cuenta para garantizar la seguridad en el proceso de evaluación. Estas directrices constituyen una gran ayuda para ir mejorando el uso de los tests, pues como ya se ha señalado en la introducción, no es suficiente que un test tenga las propiedades psicométricas adecuadas y que los profesionales que lo utilizan tengan una buena cualificación, además hay que garantizar que se hace un uso adecuado de las pruebas, y a ello van encaminadas las directrices descritas. El uso correcto de los instrumentos de medida tiene importantes implicaciones a dos niveles; desde un punto de vista profesional es esencial que las personas sean evaluadas de forma rigurosa para que la toma de decisiones sobre distintos aspectos de su vida por parte de los psicólogos sea la adecuada, y se ajuste a la normativa deontológica correspondiente. Desde un punto de vis-



ta científico, la utilización de instrumentos de medida adecuados es la única forma de hacer avanzar la psicología científica e ir proporcionando nuevos instrumentos para la práctica profesional. Una ciencia psicológica sólida con resultados replicables sólo es posible si los instrumentos de medida utilizados tienen unas propiedades métricas adecuadas. Los problemas de replicabilidad de la investigación psicológica actual provienen de causas variadas, pero sin duda una de ellas está relacionada con los instrumentos de medida utilizados (Ioannidis et al., 2014; Koole y Lakens, 2012; Nosek y Lakens, 2014; Nosek et al., 2015). Confiamos en que las directrices presentadas contribuyan a ir mejorando la utilización de los tests tanto en contextos de investigación como aplicados.

Para facilitar una comprensión del encuadre de las directrices presentadas en el contexto de la evaluación psicológica actual, se comentan a continuación las perspectivas actuales y algunas de las vías de desarrollo futuro de la evaluación psicológica, siguiendo las líneas de lo expuesto en trabajos previos (Muñiz, 2012; Muñiz y Bartram, 2007; Muñiz, Elosua y Hambleton, 2013; Muñiz y Fernández-Hermida, 2010). La evaluación, al igual que la propia psicología, está en constante desarrollo, influenciada por distintos factores, pero sin duda alguna la fuerza más poderosa que impulsa los cambios son las nuevas tecnologías de la información, y en especial los avances informáticos, multimedia e Internet. Algunos expertos (Bennet, 1999, 2006; Breithaupt, Mills y Melican, 2006; Drasgow, Luecht y Bennet, 2006) consideran 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. Todo ello está haciendo cambiar el formato y contenido de las evaluaciones, surgiendo la duda razonable de si los tests de papel y lápiz tal como los conocemos ahora serán capaces de resistir este nuevo cambio tecnológico. En este contexto de cambio tecnológico emerge la Psicología 2.0 (Armayones et al., 2015), que pretende extender la psicología a través de las facilidades que ofrece Internet y las redes sociales. La evaluación no puede estar ajena a estas nuevas tendencias, emergiendo nuevos enfoques psicométricos conectados con el análisis de las grandes bases de datos (big data) de las que se dispone actualmente (Markovetz, Blaszkiewicz, Montag, Switala, y Schlaepfer, 2014). Por ejemplo, las ventajas potenciales de usar los teléfonos

móviles como terminales para la evaluación abren nuevas posibilidades para la psicometría del futuro (Armayones et al., 2015; Miller, 2012). Trabajos como el pionero de Kosinski, Stillwell y Graepel (2013) analizan con éxito la posibilidad de utilizar los “me gusta” de facebook como predictores de distintas características humanas, entre ellas los rasgos de la personalidad, lo que hace preguntarse si nuestros rastros en las redes sociales sustituirán algún día no muy lejano a los cuestionarios y tests tal como los conocemos ahora.

Según el profesor Hambleton (2004, 2006), seis grandes áreas atraerán la atención de investigadores y profesionales en los próximos años. La *primera* es el uso internacional de los tests, debido a la globalización creciente y a las facilidades de comunicación, lo cual plantea todo un conjunto de problemas de adaptación de los tests de unos países a otros (Byrne et al., 2009; Hambleton et al., 2005; Muñiz et al., 2013). Esta internacionalización ha puesto de manifiesto la necesidad de disponer de un marco general de evaluación que recoja las buenas prácticas de evaluación. Por ello el Instituto Internacional de Estandarización (ISO) ha desarrollado una nueva norma (ISO-10667), que recoge la normativa a seguir para una evaluación adecuada de las personas en entornos laborales y organizacionales (ISO, 2011). La *segunda* es el uso de nuevos modelos psicométricos y tecnologías para generar y analizar los tests. Cabe mencionar aquí toda la nueva psicometría derivada de los modelos de Teoría de Respuesta a los Ítems (TRI), los cuales vienen a solucionar algunos problemas que no encontraban buena solución dentro del marco clásico, pero como siempre ocurre a la vez que se solucionan unos problemas surgen otros nuevos que no estaban previstos (Abad, Olea, Ponsoda y García, 2011; De Ayala, 2009; Hambleton, Swaminathan y Rogers, 1991; Muñiz, 1997; Van der Linden y Hambleton, 1997). La *tercera* es la aparición de nuevos formatos de ítems derivados de los grandes avances informáticos y multimedia (Irvine y Kyllonen, 2002; Shermis y Burstein, 2003; Sireci y Zenisky, 2006; Zenisky y Sireci, 2002). Ahora bien, no se trata de innovar por innovar, antes de sustituir los viejos por los nuevos formatos hay que demostrar empíricamente que mejoran lo anterior, las propiedades psicométricas como la fiabilidad y la validez no son negociables. La *cuarta* área que reclamará gran atención es todo lo relacionado con los tests informatizados y sus relaciones con Internet. Mención especial merecen en este campo los Tests Adaptativos Informatizados



que permiten ajustar la prueba a las características de la persona evaluada, sin por ello perder objetividad o comparabilidad entre las personas, lo cual abre perspectivas muy prometedoras en la evaluación (Olea, Abad y Barrada, 2010). La evaluación a distancia o tele-evaluación es otra línea que se abre camino con rapidez, lo cual plantea, como hemos visto en el apartado de las directrices sobre seguridad, serios problemas de seguridad de los datos y de las personas, pues hay que comprobar que la persona que se está evaluando es la que realmente dice ser, sobre todo en contextos de selección de personal o de pruebas con importantes repercusiones para la vida futura de la persona evaluada (Bartram y Hambleton, 2006; Leeson, 2006; Mills et al., 2002; Parshall et al., 2002; Williamson et al., 2006; Wilson, 2005). Dentro de esta línea tecnológica también merecen especial mención los avances relativos a la corrección automatizada de ensayos, que plantea interesantes retos (Shermis y Burstein, 2003; Williamson, Xiaoming y Breyer, 2012). En quinto lugar cabe señalar un campo que puede parecer periférico pero que está cobrando gran importancia, se trata de los sistemas a utilizar para dar retroalimentación (feedback) de los resultados a los usuarios y partes legítimamente implicadas. Es fundamental que estos comprendan sin equívocos los resultados de las evaluaciones, y no es obvio cuál es la mejor manera de hacerlo, sobre todo si se tienen que enviar para su interpretación y explicación del profesional, como ocurre en numerosas situaciones de selección de personal, o en la evaluación educativa (Goodman y Hambleton, 2004). Finalmente, es muy probable que en un futuro haya una gran demanda de formación por parte de distintos profesionales relacionados con la evaluación, no necesariamente psicólogos, aunque también, tales como profesores, médicos, enfermeros, etc. No se trata de que estos profesionales puedan utilizar e interpretar los tests propiamente psicológicos, sino que demanden información para poder comprender y participar en los procesos evaluativos y de certificación que se desarrollan en su ámbito laboral.

Nuevas formas de evaluación emergen, pero los tests psicométricos seguramente seguirán siendo herramientas fundamentales, dada su objetividad y economía de medios y tiempo (Phelps, 2005, 2008). Estas son algunas líneas de futuro sobre las que muy probablemente girarán las actividades evaluadoras en un futuro no muy lejano, no se trata de hacer una relación exhaustiva ni mucho menos, sino indicar algunas pistas para orientarse en el

mundo rápidamente cambiante de la evaluación psicológica. Las directrices aquí presentadas tienen un carácter transversal, hallándose presentes en todas estas líneas de futuro esbozadas, pues en cualquiera de las circunstancias citadas siempre habrá pruebas que se utilicen en investigación, siempre será necesario llevar a cabo procesos rigurosos de control de calidad, y siempre habrá que asegurar la seguridad de todo el proceso evaluativo.

## REFERENCIAS

- Abad, F.J., Olea, J., Ponsoda, V., y García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.
- 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.
- Armayones, M., Boixadós, M., Gómez, B., Guillamón, N., Hernández, E., Nieto, R., Pousada, M., y Sara, B. (2015). Psicología 2.0: Oportunidades y retos para el profesional de la psicología en el ámbito de la e-salud. *Papeles del Psicólogo*, 36(2), 153-160.
- Bartram, D. (1996). Test qualifications and test use in the UK: The competence approach. *European Journal of Psychological Assessment*, 12, 62-71.
- Bartram, D. (1998). The need for international guidelines on standards for test use: A review of European and international initiatives. *European Psychologist*, 2, 155-163.
- Bartram, D. y Coyne, I. (1998). Variations in national patterns of testing and test use: The ITC/EFPPA international survey. *European Journal of Psychological Assessment*, 14, 249-260.
- Bartram, D. y Hambleton, R. K. (Eds.) (2006). *Computer-based testing and the internet: Issues and advances*. Chichester: Wiley.
- Bennett, R. E. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and practice*. 18(3), 5-12.
- Bennett, R. E. (2006). Inexorable and inevitable: The continuing story of technology and assessment. En D. Bartram and R. K. Hambleton (Eds.), *Computer-based testing and the internet: Issues and advances*. Chichester: Wiley.
- Breithaupt, K. J., Mills, C. N., y Melican, G. J. (2006). Facing the opportunities of the future. En D. Bartram y



- R. K. Hambleton (Eds.), *Computer-based testing and the Internet* (pp. 219-251). Chichester: John Wiley and Sons.
- Brennan, R. L. (Ed.) (2006). *Educational measurement*. Westport, CT: ACE/Praeger.
- Byrne, B. M., Leong, F. T., Hambleton, R. K., Oakland, T., van de Vijver, F. J., y Cheung, F. M. (2009). A critical analysis of cross-cultural research and testing practices: Implications for improved education and training in psychology. *Training and Education in Professional Psychology*, 3(2), 94-105.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.
- Downing, S. M. y Haladyna, T. M. (Eds.) (2006). *Handbook of test development*. Mahwah, NJ: LEA.
- Drasgow, F., Luecht, R. M. y Bennett, R. E. (2006). Technology and testing. En R. L. Brennan (Ed.), *Educational measurement*. Westport, CT: ACE/Praeger.
- European Federation of Professional Psychologists' Associations (2005). *Meta-Code of ethics*. Brussels: Author ([www.efpa.eu](http://www.efpa.eu)).
- 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.
- Foster, D. F. (2010). Worldwide Testing and Test Security Issues: Ethical Challenges and Solutions. *Ethics & behavior*, 20 (3-4), 207-228.
- Goodman, D.P. y Hambleton, R.K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*, 17, 145-220.
- Hambleton, R. K. (2004). Theory, methods, and practices in testing for the 21st century. *Psicothema*, 16 (4), 696-701.
- Hambleton, R. K. (2006). *Testing practices in the 21st century*. Key Note Address, University of Oviedo, Spain, March 8th.
- Hambleton, R. K., Merenda, P. F., y Spielberger, C. D. (Eds.) (2005). *Adapting educational and psychological tests for cross-cultural assessment*. London: LEA.
- Hambleton, R. K., Swaminathan, H., y Rogers, J. (1991). *Fundamentals of item response theory*. Beverly Hills, CA: Sage.
- Hernández, A., Tomás, I., Ferreres, A. y Lloret, S. (2015) Tercera evaluación de tests editados en España. *Papeles del Psicólogo*, 36, 1-8.
- Ioannidis, J. P. A., Munafo, M. R., Fusar-Poli, P., Nosek, B. A., & David, S. P. (2014). Publication and other reporting biases in cognitive sciences: Detection, prevalence, and prevention. *Trends in Cognitive Sciences*, 18, 235-241.
- Irvine, S. y Kyllonen, P. (Eds.) (2002). *Item generation for test development*. Mahwah, NJ: Lawrence Erlbaum.
- 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].
- Joint Committee on Testing Practices. (2002). *Ethical principles of psychologists and code of conduct*. Washington DC: Joint Committee on Testing Practices.
- Koocher, G. y Keith-Spiegel, P. (2007). *Ethics in psychology*. Nueva York: Oxford University Press.
- Koole, S. L. y Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science*, 7, 608-614.
- Kosinski, M., Stillwell, D., y Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behaviour. *Proceedings of the National Academy of Sciences (PNAS)*, 110(15), 5802-5805.
- Leach, M. y Oakland, T. (2007). Ethics standards impacting test development and use: A review of 31 ethics codes impacting practices in 35 countries. *International Journal of Testing*, 7, 71-88.
- Leeson, H. V. (2006). The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing*, 6, 1-24.
- Lindsay, G., Koene, C., Ovreide, H., y Lang, F. (2008). *Ethics for European psychologists*. Gottingen and Cambridge, MA: Hogrefe.
- Markovetz, A., Blaszkiewicz, K., Montag, C., Switala, C., y Schlaepfer, T. E. (2014). Psycho-Informatics: Big Data shaping modern psychometrics. *Medical Hypotheses*, 82(4), 405-411.
- Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, 7(3), 221-237.
- Mills, C.N., Potenza, M.T., Fremer, J.J., y Ward, W.C. (Eds.) (2002). *Computer-based testing: Building the foundation for future assessments*. Hillsdale, NJ: LEA.
- Moreland, K. L., Eyde, L. D., Robertson, G. J., Primoff, E.



- S., y Most, R. B. (1995). Assessment of test user qualifications. *American Psychologist*, 5, 1, 14-23.
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Pirámide.
- 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., Elosua, P., y Hambleton, R.K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25(2), 151-157.
- 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., Prieto, G., Almeida, L., y Bartram, D. (1999). Test use in Spain, Portugal and Latin American countries. *European Journal of Psychological Assessment*, 15(2), 151-157.
- Muñiz, J., Fernández-Hermida, J.R., Fonseca-Pedrero, E., Campillo-Álvarez, A. y Peña-Suárez, E. (2011). Evaluación de tests editados en España. *Papeles del Psicólogo*, 32, 113-128.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., D'Foe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T. A., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Levy Paluck, E., Simonsohn, U., Soderberg, C., Spellman, B. A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E. J., Wilson, R., y Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348, 1422-1425.
- Nosek, B. A. y Lakens, D. (2014). Registered reports. A method to increase the credibility of published results. *Social Psychology*, 45, 137-141.
- Olea, J., Abad, F., y Barrada, J. R. (2010). Tests informatizados y otros nuevos tipos de tests. *Papeles del Psicólogo*, 31(1), 94-107.
- Papeles del Psicólogo (2009). *Número monográfico sobre Ética Profesional y Deontología*. Vol. 30, 182-254.
- Parshall, C.G., Spray, J.A., Kalohn, J.C., y Davey, T. (2002). *Practical considerations in computer-based testing*. New York: Springer.
- Phelps, R. (Ed.) (2005). *Defending standardized testing*. Londres: LEA.
- Phelps, R. (Ed.) (2008). *Correcting fallacies about educational and psychological testing*. Washington: APA.
- Ponsoda, V. y Hontangas, P. (2013). Segunda evaluación de tests editados en España. *Papeles del Psicólogo*, 24, 82-90
- Prieto, G. y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-71.
- Ryan A.M., Inceoglu, I., Bartram, D., Golubovich, J., Grand, J., Reeder, M., Deros, E., Nikolaou, I. y Yao, X. (2015). Trends in testing: Highlights of a global survey. En Nikolaou, I. y J. Oostrom (Eds.). *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice*. Psychology Press-Taylor & Francis.
- Shermis, M. D. y Burstein, J. C. (Eds.) (2003). *Automated essay scoring*. London: LEA.
- Simner, M. L. (1996). Recommendations by the Canadian Psychological Association for improving the North American safeguards that help protect the public against test misuse. *European Journal of Psychological Assessment*, 12, 72-82.
- Sireci, S., y Zenisky, A. L. (2006). Innovative items format in computer-based testing: In pursuit of construct representation. En S. M. Downing y T. M. Haladyna (Eds.), *Handbook of test development*. Hillsdale, NJ: LEA.
- Van der Linden, W. J. y Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York, NY: Springer-Verlag.
- Williamson, D.M., Mislevy, R.J. y Bejar, I. (2006). *Automated scoring of complex tasks in computer based testing*. Mahwah, NJ: LEA.
- Williamson, D.M., Xi, X., y Breyer, J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice*, 31(1), 2-13.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: LEA.
- Yee, K., y MacKown, P. (2009). Detecting and preventing cheating during exams. In T. Twomey, H. White, & K. Sagendorf (Eds.), *Pedagogy not policing: Positive approaches to academic integrity at the University* (pp. 141 - 148). Syracuse: The Graduate School.
- Zenisky, A.L. y Sireci, S.G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education*, 15, 337-362.

