

# Avances, proyectos y retos internacionales ligados al uso de tests en Psicología

## *International advances, projects and challenges in psychological testing*

Paula **ELOSUA**<sup>1</sup>

### Resumen

El desarrollo de la psicometría y la normalización de la práctica profesional de la Psicología delimitan un espacio marcado por la importancia atribuida al uso correcto de los tests. En entornos sociales y científicos caracterizados por actitudes positivas hacia el uso de tests, disponibilidad de tests de calidad, regulación de la práctica profesional y colaboración internacional, se constituyen las asociaciones de tests nacionales e internacionales. Estas trabajan de forma conjunta en el establecimiento de directrices que, a través de una orientación pragmática, ofrecen guías para la buena práctica. En este trabajo describimos brevemente algunos de los últimos proyectos relacionados con el uso de tests, y apuntamos la importancia de las comisiones y las directrices para mejorar su uso.

**Palabras clave:** Administración de tests; Evaluación; Práctica profesional.

### Abstract

*The current development of psychometrics and the standardization of the psychological practice define a context marked by the importance given to the correct use of tests. National and international associations of tests are established in a social and scientific environment characterized by positive attitudes toward the use of tests, availability of quality tests, regulation of professional practice and international collaboration. These associations work together in establishing guidelines that offer guides to best practices by using a pragmatic orientation. In this paper, we briefly describe the latest developments in testing, and we point out the importance of the test commissions and guidelines to improve the use of tests.*

**Keywords:** Tests administration; Evaluation; Professional practice.

Los trabajos pioneros de Binet y Cattell sobre la medida de las aptitudes intelectuales a finales del siglo XIX marcaron en occidente el inicio del

uso de los tests psicológicos. Desde entonces, la historia de los tests (DuBois, 1970) ha descrito momentos de auge y declive. A una etapa de

▼ ▼ ▼ ▼ ▼

<sup>1</sup> Universidad del País Vasco, Facultad de Psicología. Av. Tolosa, 70, 20018, San Sebastián, España. E-mail: <paula.elosua@ehu.es>.

Agradecimientos: Trabajo financiado parcialmente por el Ministerio de Economía y Competitividad (PSI2014-54020), y por la Universidad del País Vasco (GIU15-24).

expansión y uso generalizado, prosiguió un periodo de degeneración en la década de los 70 caracterizado por un movimiento contrario al uso de los tests (Gould, 1981). Las críticas vertidas sirvieron como catalizador para el inicio de un proceso de revisión de prácticas y modelos, que dieron paso al momento actual en el que desde una perspectiva científica, profesional y ética se van delimitando las condiciones del uso correcto de los tests.

Las exigencias científicas, profesionales y éticas en el ámbito de los tests y en la práctica de la Psicología han evolucionado. Los avances en el campo de la investigación psicométrica y psicológica perfilan un área caracterizada por la disponibilidad de complejos modelos psicométricos, por el impacto de la tecnología en la construcción/administración de tests, y por una creciente e internacional conciencia sobre la importancia del uso correcto de los tests que subraya las consecuencias derivadas de un uso incorrecto. Esta terna delimita de alguna manera, la situación actual del uso de tests en la Psicología.

El objetivo de este trabajo es describir brevemente algunos de los aspectos destacados de cada uno de los elementos mencionados, dedicando especial atención a los proyectos de colaboración internacional cuyo objetivo es establecer normas universales que ayuden a mejorar la práctica profesional.

## Fiabilidad y validez

Los conceptos de fiabilidad y validez de los tests han evolucionado de la mano de la psicometría y de las exigencias sociales y profesionales. Junto a la evolución en las acepciones teóricas, se han desarrollado y ampliado los modelos psicométricos, y se han socializado los procedimientos operativos. Los modelos psicométricos se amplían para responder a nuevas demandas, y cada vez es más asequible y sencillo para el profesional aplicar las técnicas estadísticas más avanzadas. Las ediciones de los *Standards for Educational and Psychological Testing* (American Educational Research Association,

American Psychological Association, & National Council on Measurement in Education, 1966, 1974, 1999, 2014), a los que en adelante nos referiremos como estándares, son testigo de este proceso de mejora y adaptación continua que puede consultarse en De Boeck y Elosua (2016).

Introduciendo los términos brevemente, podemos decir que la fiabilidad se refiere a la reproducción o repetición de los resultados de una medición (consistencia). Si el mismo objeto se mide en diferentes ocasiones con el mismo procedimiento, pero los resultados de la medida varían, entonces el procedimiento no es perfectamente fiable. En el caso de la Psicología, la medición de atributos psicológicos siempre muestra variaciones (diferencias intraindividuales) que pueden asimilarse a la falta de fiabilidad del instrumento de medida (test) utilizado.

El concepto de fiabilidad como acercamiento al problema del error de medida se construyó inicialmente en Psicología a partir del estudio de las diferencias interindividuales (coeficiente de fiabilidad) y sobre él se erigió la Teoría Clásica de Tests (TCT); (Spearman, 1907). Ciertamente, la historia de la fiabilidad puede equipararse en gran medida a la historia de la TCT. Si bien siempre han existido iniciativas para la estimación del error de medida intraindividual (Lord, 1955; Mollenkopf, 1949; Thorndike, 1951) estas no terminaron de implantarse de forma generalizada hasta la aparición de la Teoría de Respuesta al Ítem (Hambleton & Swaminathan, 1985; Lord, 1980). La TRI ofrece un complemento a la TCT que asume que los errores de medida varían en función del nivel de competencia (habilidad) estimado.

El concepto de fiabilidad congrega dos aspectos, el error intrasujeto y el error intersujeto (Mellenbergh, 2011) relacionados respectivamente con la fiabilidad de las puntuaciones de una persona, y con la fiabilidad total del test. Esta doble naturaleza se recoge en los estándares, que hablan de "fiabilidad/precisión" y no únicamente de "fiabilidad". En contextos evaluativos, es decir, aquellos en que el objetivo es la interpretación de una puntuación, la relevancia práctica del Error Típico de Medida (ETM) es mayor que el de la

consistencia interna del test. El ETM aporta una vía para expresar la incertidumbre con relación a las puntuaciones que no ofrece, por ejemplo, el coeficiente alfa de Cronbach.

La validez es, de acuerdo a los estándares, el concepto más importante en el proceso de construcción/interpretación de un test. Esta idea básica, defendida en todas las ediciones de los estándares, sigue vigente. La definición de validez ha evolucionado desde la formulación inicial aportada por Kelley (1927). Según este autor la validez es una propiedad inherente a los ítems, según la cual un test es válido si mide lo que pretende medir. Esta forma de definir la validez progresó hacia una conceptualización construida sobre la evidencia empírica, que define las puntuaciones obtenidas por un test como indicadores de otras variables latentes; posteriormente se adoptó una definición de validez basada en la acumulación de pruebas empíricas para avalar la interpretación propuesta de las puntuaciones.

La evolución del concepto de validez ha sido descrita por varios autores (e.g., Angoff, 1988; Elosua, 2003; Kane, 2009; Newton & Shaw, 2014). Newton y Shaw (2014) especifican cinco períodos en la concepción de la validez: gestación, cristalización (1921-1951), fragmentación (1952 to 1974), unificación (1975-1999), y deconstrucción (2000-2012); esta última se caracteriza por un momento de discusión y debate sobre la definición y naturaleza de la validez de constructo, y sobre si la validez de constructo debería o no ser la base de toda validez. Otros autores (e.g., Elosua, 2003) han reducido la historia a tres estadios: una etapa operacional marcada por la diferenciación entre tres tipos de validez (validez de contenido, validez de criterio, y validez de constructo), un periodo teórico en el que domina la visión integradora del concepto de validez construida sobre la validez de constructo (Cronbach, 1984), y finalmente una fase contextual y argumental a partir de las propuestas de Messick (1989) y Kane (2001, 2006, 2013). Markus y Borsboom (2013) describen tres procesos interactivos recurrentes: expansión para abarcar nuevos problemas, unificación para la integración

de la diversidad y partición, relacionada con el interés en las diferencias. Entre ellas, los estándares actuales focalizan su atención sobre la integración.

Como consecuencia de la importancia otorgada a la interpretación de las puntuaciones, no es extraño que las consecuencias esperadas y no esperadas del uso de los tests hayan llegado a incluirse dentro del concepto de validez, y formen una etiqueta general conocida como validez consecuencial (Messick, 1995).

Independientemente de las disquisiciones teóricas sobre qué es y qué no es validez, la construcción de tests y su uso en la práctica de la Psicología como apoyo a la toma de decisiones es constante. Los tests se han extendido a todos los ámbitos de actuación (educativo, social, jurídico, organizacional, deportivo, clínico, investigador...) y sin un estudio empírico de validación de las puntuaciones que garantice en cada caso, su uso, las inferencias y las puntuaciones podrían carecer de sentido y ser inadecuadas.

## Desarrollo tecnológico

Si bien a lo largo del siglo XX los tests de "papel y lápiz" fueron claramente dominantes, durante las décadas de 1980 y 1990 asistimos a una etapa caracterizada por la aparición de los tests por computadora (CBT, *Computer Based Testing*). El CBT abrió expectativas prometedoras, que ciertamente no llegaron a materializarse de forma generalizada debido entre otros, a los costos de infraestructura y problemas logísticos asociados a la administración de tests por computadora.

La llegada de internet y de la *World Wide Web* transformó esta situación ofreciendo los medios para la administración de tests de forma generalizada y económicamente viable (Bartram & Hambleton, 2005). Baste citar como indicador del impacto de internet, que su penetración en la población mundial es del 48,7% y que el crecimiento en el número de usuarios entre 2000 y 2016 ha sido del 890,8% (De Argaez, 2016). Las diferencias logísticas más importantes entre los tests tradicionales de "papel y lápiz" y la administración

online son evidentes; los primeros necesitan ser impresos en papel, almacenados físicamente, y administrados/corregidos por personal cualificado. La administración *online* de tests centraliza el control de administración/corrección/almacenamiento en un solo punto.

La tendencia hacia los tests *online* además de suponer un incremento considerable del volumen de uso de tests, conlleva el diseño de nuevos tests en lugar de la adaptación de tests existentes. Si bien existen marcadas diferencias entre las áreas de actuación profesional de la Psicología con respecto a las necesidades y usos de los tests, y las características de una de ellas no se pueden generalizar al resto, la administración *online* se está erigiendo en el siglo XXI como el modo imperante en el área de la Psicología de las organizaciones (Bartram, 2016). Es menester recordar en este punto que los avances tecnológicos nunca pueden suplantar la validez y fiabilidad de las puntuaciones (Naglieri et al., 2004).

## Bases para el diseño de políticas activas

Los avances en los modelos psicométricos o la tecnología para la construcción/administración de tests deben de complementarse con un uso correcto. Las comunidades psicométrica y evaluativa, conscientes de la importancia de las buenas prácticas relacionadas con el uso de tests, han impulsado y amparado a través de organizaciones profesionales nacionales e internacionales, proyectos de mejora en este ámbito. El establecimiento de políticas activas para mejorar el uso de los tests por parte de las asociaciones nacionales se inició en la mayoría de los países en la década de los 80 (Poortinga & Klieme, 2016). El objetivo es tratar de ofrecer normas y herramientas que fomenten las buenas prácticas, adoptando para ello una perspectiva pragmática. Las guías ofrecidas o directrices, definen marcos de referencia que tutelan la práctica profesional; en ellas se recogen los avances metodológicos de mayor impacto y se recomiendan principios éticos que garantizan un uso correcto de los tests y en consecuencia una mejor evaluación psicológica.

Las directrices se construyen en un contexto marcado por actitudes positivas hacia los tests (Evers et al., 2012), disponibilidad de instrumentos de calidad, programas de formación y regulación de la práctica profesional (Elosua & Muñiz, 2013). Sin disponibilidad, formación y actitud sería difícil adoptar y observar cualquier directriz. Este marco contextual ha quedado reflejado en los dos estudios internacionales sobre las actitudes y uso de los tests entre psicólogos, que podemos considerar una pieza clave en el desarrollo de los proyectos formativos. La primera edición del proyecto se completó en el año 1999 y la segunda en el año 2009 (en el momento de publicación de este estudio se está desarrollando un estudio sobre actitudes/uso de tests en los países iberoamericanos). En la primera participaron 6 países europeos, y en la segunda el número de participantes se incrementó hasta 17. El estudio se apoya en un cuestionario de opinión que recoge información sobre 8 dominios; uso de los tests, actitudes de los psicólogos hacia los tests, necesidad de regulación del uso de los tests, uso de internet, formación y conocimientos sobre tests, permisividad en el uso de los tests, e información técnica sobre los tests de la que disponen los profesionales. Además se pide a los participantes que indiquen los tres tests que más utilizan en la práctica profesional. Los resultados más relevantes señalan que: (a) los psicólogos reconocen que la formación recibida en el grado de Psicología puede no ser suficiente para la correcta utilización de la mayoría de los tests, (b) la actitud de los psicólogos respecto al uso de los tests es positiva.

El punto relacionado con la disponibilidad de tests corre parejo con la adaptación de cuestionarios. Estudios recientes señalan que de los 10 tests más utilizados en Europa, 9 fueron construidos en lengua inglesa (Elosua & Iliescu, 2012), y de los 25 tests que más se utilizan en España, por citar un ejemplo, 17 son adaptaciones de versiones construidas en otro idioma (Elosua, 2012). El uso de versiones adaptadas del mismo test en diferentes países, o incluso el uso de un test dentro del mismo país con grupos minoritarios, exige la utilización de normas de estandarización

específicas. Sin embargo, los usuarios señalan que en muchas ocasiones no se dispone de normas distintas, y en estas circunstancias se asume que la distribución de las puntuaciones es (o debería ser) igual entre los diferentes grupos. Aunque es difícil encontrar evidencias al respecto, y sería necesario contrastar empíricamente esta aseveración, la impresión generalizada por parte de los usuarios de tests es que es mejor utilizar cuestionarios que no incorporan normas específicas a no utilizarlos (Poortinga & Klieme, 2016).

## Comisiones de tests y directrices internacionales

La regulación de la profesión de psicólogo, la formación y la disponibilidad de tests de calidad definen el marco científico/social que propicia la constitución de comisiones de tests nacionales e internacionales que trabajan en la observancia y apoyo a las buenas prácticas profesionales. La interacción entre las comisiones nacionales e internacionales es la base del trabajo de organizaciones como la Comisión Internacional de Tests (ITC, *International Test Commission*), la Federación Europea de Asociaciones de Psicólogos (EFPA, *European Federation of Psychologists' Associations*) o la Organización de Estándares Internacional (ISO, *International Standards Organization*) (Figura 1).

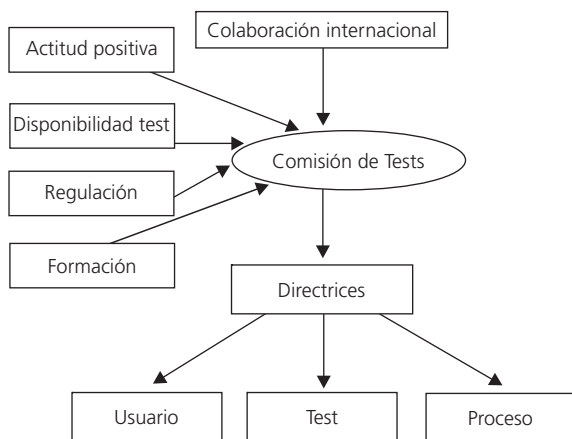


Figura 1. Directrices para mejorar el uso de tests.

La perspectiva adoptada por las comisiones nacionales o internacionales en cuanto a objetivos y ámbitos de actuación puede ser algo diferente, en tanto y en cuanto, las primeras han de atender a las especificaciones, usos, necesidades y marcos legislativos propios. Desde un enfoque internacional las comisiones intentan construir directrices que respetando las diferencias entre países, definan aquello que es constitutivo de la buena práctica profesional. Las directrices internacionales, por naturaleza, abordan problemas universales y no sugieren actuaciones que pudieran entrar en conflicto con leyes o regulaciones de carácter nacional. A pesar de las diferencias entre ambos acercamientos, existen estándares elaborados a nivel nacional que tienen un gran impacto internacional. Entre ellos los más relevantes son sin duda, los estándares de la American Educational Research Association, American Psychological Association, y National Council on Measurement in Education (2014) (Zumbo, 2014).

Con el objetivo de definir los aspectos abordados por las directrices relacionadas con el uso de los tests, Bartram y Hambleton (2016) diferencian tres facetas en la práctica profesional, que aunque actúan de forma interrelacionada merecerían tratamientos distintivos; hablan estos autores sobre: (a) el profesional que usa tests, (b) los tests que usa y (c) el proceso evaluativo en sí mismo. En este sentido, las directrices para un uso correcto de los tests deberían contemplar aspectos relacionados con: (a) la formación del usuario, (b) el cumplimiento por parte de los tests de requisitos técnicos mínimos, y (c) la observancia de buenas prácticas en el proceso evaluativo. No existe ninguna directriz que abarque en su totalidad los tres puntos; cada conjunto de directrices aborda con mayor o menor generalidad, puntos concretos.

Siendo reconocida la Comisión Internacional de Tests como una de las organizaciones de mayor impacto en el establecimiento de directrices prácticas relacionadas con el uso de tests, es menester mencionar el resultado de su trabajo. Hasta el momento la ITC ha publicado 6 conjuntos de directrices, a saber:

1. Directrices de la ITC para la traducción y adaptación de tests (International Test Commission, 2005);

2. Directrices de la ITC sobre el uso de tests (International Test Commission, 2001);

3. Directrices de la ITC sobre tests computarizados y tests administrados por internet (International Test Commission, 2006);

4. Directrices de la ITC sobre la seguridad de los tests, exámenes y otras evaluaciones (International Test Commission, 2016);

5. Directrices de la ITC para el control de calidad de las puntuaciones de los tests, su análisis y los informes sobre las puntuaciones (International Test Commission, 2014a);

6. Directrices de la ITC sobre el uso profesional de revisiones de tests, tests obsoletos y retirada de tests (International Test Commission, 2015).

Además de las directrices mencionadas, la ITC ha publicado una declaración sobre el uso de los tests y otros instrumentos de evaluación en investigación (International Test Commission, 2014b), y está trabajando en un nuevo conjunto de normas internacionales para la evaluación en contextos de diversidad lingüística (Elosua, 2016).

Todas las directrices adoptadas hasta ahora pueden descargarse desde la página de la Comisión Internacional de Tests ([www.intestcom.org](http://www.intestcom.org)); existen traducciones al español de la mayoría de ellas, a las que se puede acceder desde la página del Consejo General de la Psicología de España ([www.cop.es](http://www.cop.es)). Las primeras tres directrices están bien documentadas en trabajos y artículos desarrollados sobre ellas (e. g., Bartram, 2001; Coyne & Bartram, 2006; Muñiz, Elosua, & Hambleton, 2013) y del resto puede encontrarse una introducción general en Bartram y Hambleton (2016).

mejorar el uso de los tests, y con ello la práctica profesional. El objetivo es evaluar la calidad de los tests con el propósito de ofrecer al usuario información útil, objetiva e imparcial. Desde una perspectiva internacional relacionada con la revisión de la calidad de tests, la EFPA ha coordinado un proyecto en el cual se ha diseñado un cuestionario de evaluación que combina y amplía los modelos de evaluación de tests existentes en España, Holanda y Gran Bretaña (Evers et al., 2013). El cuestionario está dividido en tres secciones que contemplan: (a) descripción general; (b) valoración de las características y (c) evaluación global. La primera se corresponde con una ficha técnica de descriptores sobre el test que ofrece información sobre aspectos como el título, los objetivos, las puntuaciones o el tiempo de aplicación. El segundo apartado contiene ítems de respuesta graduada (inadecuada, adecuada con carencias, adecuada, buena, excelente) que permiten cuantificar la calidad de los estudios de fiabilidad, validación y baremación. Incluye también ítems abiertos sobre estos 3 últimos aspectos en los que se genera un juicio fundamentado. La última parte se destina a comentarios generales sobre el test, sus fortalezas/debilidades y una síntesis de la información más relevante para el uso correcto del test.

El procedimiento general de evaluación podría sintetizarse en 5 etapas: (a) elección de los tests a revisar; (b) identificación de revisores; (c) evaluación del test; (d) revisión y edición; y (e) comentarios de la editorial y memoria final. La versión española del cuestionario de evaluación de tests puede consultarse en Prieto y Muñiz (2000) y las fases más importantes del proceso se describen en Elosua y Geisinger (2016).

## Proceso evaluativo

Si bien en todas las directrices se aborda de algún modo el proceso evaluativo, la Asociación Europea de Evaluación Psicológica (EAPA; *European Association of Psychological Assessment*) creó un grupo de trabajo para el desarrollo de directrices específicas sobre los principios generales que guían

## Calidad de los tests

La evaluación de tests y divulgación de las consiguientes memorias se han implantado en el panorama internacional con el claro objetivo de

cada una de sus fases. La idea parte de que el proceso en sí es un elemento indispensable de cualquier aproximación a la evaluación psicológica, incluso desde orientaciones que no utilizan tests psicométricos. El resultado del trabajo se conoce como guías del proceso de evaluación (*Guidelines for the Assessment Process*; Fernández-Ballesteros et al., 2001).

Existe otro estándar internacional sobre evaluación que ha supuesto un paso importante en la estandarización y reconocimiento del proceso evaluativo; se trata de la norma ISO 10667 (International Organization for Standardization, 2011). La sigla ISO se refiere a la Organización Internacional para la Estandarización ([www.iso.org](http://www.iso.org)), que desarrolla normativas en todos los sectores industriales y de servicios. La EFPA inició un proceso para elaborar una norma ISO que regulase todo lo relativo a la evaluación de las personas en el ámbito laboral. Esta nueva norma es de gran interés para los psicólogos, dado su papel central en la evaluación de personas en contextos laborales. La norma no tiene rango legal en sentido estricto, pero constituye una importante norma reguladora del mercado, pues las empresas e instituciones pueden certificarse garantizando que cumplen el contenido incluido en ella. El objetivo es la regulación del proceso de evaluación de las personas en contextos laborales y organizacionales, cubriendo todo el proceso de evaluación, desde el establecimiento del contrato de evaluación hasta la utilización de los resultados, pasando por la metodología de la evaluación en sí misma. Es aplicable a los procedimientos y métodos utilizados a nivel individual (selección, consejo, formación...), grupal (clima y cohesión de equipos de trabajo) y organizacional (clima laboral, cultura de empresa, satisfacción...). En la norma se describen las competencias, obligaciones y responsabilidades de los clientes y de los proveedores del servicio de evaluación, antes, durante y después del proceso evaluativo.

## Conclusiones

Los tests son instrumentos de medida que acompañan al profesional de la Psicología en su

quehacer diario; en tanto que son herramientas técnicas, los tests han de cumplir ciertas condiciones psicométricas, sin las cuales no se puede reconocer su utilidad. Sin embargo, no es suficiente que un manual acredite que las puntuaciones que ofrece un test son fiables y válidas, ya que estas propiedades solo quedan garantizadas si el uso que se hace del test es correcto. Rigor técnico y uso apropiado han de fusionarse para que las puntuaciones alcancen su propósito. Se precisa para ello, competencia por parte del usuario, calidad por parte del test y observancia en el cumplimiento de buenas prácticas durante su uso.

Las organizaciones nacionales e internacionales relacionadas con el uso de los tests tratan de mejorar la práctica profesional utilizando como argumento principal la formación. Los proyectos relacionados con la mejora en el uso de los tests se construyen en entornos sociales y profesionales en los que se dispone de tests, existen posibilidades de formación, la práctica profesional está regulada y la cooperación internacional está asentada. A partir de la segunda década del siglo XX hemos asistido a avances en cada una de las áreas citadas, tanto en cantidad como en calidad: (a) la psicometría sigue estudiando y ofreciendo modelos formales que ayudan a medir variables psicológicas con mayor precisión y validez; (b) el trabajo de la comisión internacional de tests es aceptado como marco internacional dentro del cual se pueden acomodar usos y prácticas nacionales; (c) en las conferencias profesionales, independientemente del ámbito geográfico y especialidad, se organizan mesas de trabajo específicas sobre la práctica profesional; (d) la construcción y edición de tests sigue cada vez con mayor rigor estándares técnicos de calidad.

La colaboración internacional y los avances tecnológicos y científicos no cesan; asistimos a la llegada de nuevas prácticas en el uso de tests que serán objeto de atención por parte de los especialistas en las próximas décadas. La rápida evolución de internet y de los dispositivos móviles nos enfrentan a nuevas formas de construir/aplicar tests. Hoy se adquieren más dispositivos móviles que computadoras; la consecuencia directa es que es

posible acceder a un número cada vez mayor y más diverso de destinatarios. El conocido como *Movil Internet Testing* (King, Ryan, Kantrowitz, Grelle, & Dainis, 2015) está generando la atención de profesionales que ven en él un medio rápido, sencillo y económico de evaluar/diagnosticar al paciente, estudiante o candidato. Junto a ello la utilización de entornos de juego para la evaluación psicológica (*gamification*) que en principio incrementan la motivación en la tarea (CEB, 2014), genera oportunidades y nuevos problemas a los que habrá que prestar atención con el fin de conjugar los aspectos tecnológicos con los principios psicométricos y las normas éticas que rigen un buen uso.

Es cierto que el panorama descrito en este trabajo no es generalizable a todas las situaciones. Constatamos la presencia de grandes diferencias entre países con relación a la formación, regulación profesional y/o disponibilidad de tests, y carecemos de información sólida sobre situaciones particulares. Sería interesante diseñar estudios y ejecutar proyectos de colaboración internacional que permitan recoger datos fiables para llevar a cabo descripciones precisas del estado de la cuestión en cada país. Únicamente de ese modo podrán establecerse grupos de trabajo, y comisiones de tests que aboguen, velen, y sustenten la mejora del uso de tests en la práctica profesional de la Psicología.

## Referências

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1966). *Standards for educational and psychological tests*. Washington, D.C.: American Educational Research Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1974). *Standards for educational and psychological tests*. Washington, D.C.: American Educational Research Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological tests*. Washington, D.C.: American Educational Research Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological tests*. Washington, D.C.: American Educational Research Association.
- Angoff, W. H. (1988). Validity: An evolving concept. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp.19-32). Hillsdale: Laurence Erlbaum Associates.
- Bartram, D. (2001). The development of international guidelines on test use: The international test commission project. *International Journal of Testing*, 1(1), 33-53.
- Bartram, D. (2016). *The changing face of testing in work and organizational settings: A personal journey*. Paper presented at the 10th Conference of the International Test Commission, Vancouver, Canada.
- Bartram, D., & Hambleton, R. K. (2005). *Computer-based testing and the internet: Issues and advances*. New York: Wiley.
- Bartram, D., & Hambleton, R. K. (2016). The ITC Guidelines. International standards and guidelines relating tests and testing. In F. T. L. Leong, D. Bartram, F. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international handbook of testing and assessment*. New York: Oxford University Press.
- CEB. (2014). *Gamification in recruiting: Trends and best practices*. Retrieved September, 3, 2016, from <https://www.executiveboard.com/blogs/gamification-in-recruiting-trends-and-best-practices/>
- Coyne, I., & Bartram, D. (2006). Design and development of the ITC guidelines on computer-based and internet delivered testing. *International Journal of Testing*, 6(2), 133-142.
- Cronbach, L. J. (1984). *Essentials of psychological testing* (4rd ed.). New York: Harper.
- De Argaez, E. (2016). *Internet world stats*. Retrieved September 3, 2016, from <http://www.internetworldstats.com/>
- De Boeck, P., & Elosua, P. (2016). Reliability and validity: History, notions, methods, discussion. In F. T. L. Leong, D. Bartram, F. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international handbook of testing and assessment*. New York: Oxford University Press.
- DuBois, P. H. (1970). *A history of psychological testing*. Boston: Allyn & Bacon.
- Elosua, P. (2003). Sobre la validez de los tests. *Psicothema*, 15(2), 315-321.
- Elosua, P. (2012). Tests publicados en España: Usos, costumbres y asignaturas pendientes. *Papeles del Psicólogo*, 33(1), 12-21.



- Elosua, P. (2016). *Testing in linguistically diverse contexts*. Symposium presented at 10th Conference of the International Test Commission, Vancouver, Canada.
- Elosua, P., & Geisenger, K. (2016). Cuarta evaluación de tests editados en España: Forma y fondo. *Papeles del Psicólogo*, 37(2), 3-13.
- Elosua, P., & Iliescu, D. (2012). Tests in Europe. Where we are and where we should to go. *International Journal of Testing*, 12(2), 157-175.
- Elosua, P., & Muñoz, J. (2013). Proyectos españoles para una mejora en el uso de los Tests. *Psiencia. Latin American Journal of Psychological Science*, 5(2), 139-143.
- Evers, A., Muñoz, J., Bartram, D., Boben, D., Egeland, J., Fernández-Hermida, J.R., ... Urbánek, T. (2012). Testing practices in the 21st Century: Developments and European psychologists' opinions. *European Psychologist*, 17, 300-319.
- Evers, A., Muñoz, J., Hagemester, C., Høstmælingen, A., Lindley, P., Sjöberg, A., & Bartram, B. (2013). Assessing the quality of tests: Revision of the EFPA review model. *Psicothema*, 25(3), 283-291.
- Fernández-Ballesteros, R., De Bruyn, E. E. J., Godoy, A., Hornke, L., Ter Laak, J., Vizcarro, C., Westhoff, K., Westmeyer H., & Zacagnini, J. L. (2001). Guidelines for the Assessment Process (GAP): A proposal for discussion. *European Journal of Psychological Assessment*, 17(3), 187-200.
- Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.
- International Test Commission. (2001). International guidelines on test use. *International Journal of Testing*, 1(2), 95-114.
- International Test Commission. (2005). *ITC guidelines for translating and adapting tests*. Retrieved September 3, 2016, from [https://www.intestcom.org/files/guideline\\_test\\_adaptatiopn.pdf](https://www.intestcom.org/files/guideline_test_adaptatiopn.pdf)
- International Test Commission. (2006). International guidelines on computer-based and Internet-delivered testing. *International Journal of Testing*, 6(2), 143-172.
- International Test Commission. (2014a). ITC guidelines on quality control in scoring, test analysis, and reporting of test scores. *International Journal of Testing*, 14(3), 195-217.
- International Test Commission. (2014b). ITC statement on the use of tests and other assessment instruments for research purposes. Retrieved September 3, 2016, from [https://www.intestcom.org/files/statement\\_using\\_tests\\_for\\_research.pdf](https://www.intestcom.org/files/statement_using_tests_for_research.pdf)
- International Test Commission. (2015). ITC guidelines on practitioner use of tests revisions, obsolete tests, and test disposal. Retrieved September 3, 2016, from [https://www.intestcom.org/files/guideline\\_test\\_disposal.pdf](https://www.intestcom.org/files/guideline_test_disposal.pdf)
- International Test Commission. (2016). The ITC guidelines on the security of tests, examinations, and other assessments. *International Journal of Testing*, 16(3), 181-204.
- International Organization for Standardization. (2011). *Procedures and methods to assess people in work and organizational settings (part 1 and 2)*. Geneva: Author.
- Kane, M. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342.
- Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp.17-64), Westport: American Council on Education and Praeger.
- Kane, M. (2009). Validating the interpretations and uses of test scores. In R.W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp.39-64). Charlotte: Information Age Publishing.
- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Kelley, T. L. (1927). *Interpretation of educational measurements*. New York: MacMillan.
- King, D. D., Ryan, A. M., Kantrowitz, T., Grelle, D., & Dainis, A. (2015). Mobile internet testing: An analysis of equivalence, individual differences, and reactions. *International Journal of Selection and Assessment*, 23(4), 382-394.
- Lord, F. M. (1955). Estimating test reliability. *Educational and Psychological Measurement*, 1955(1), 325-336.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: Erlbaum.
- Markus, K. A., & Borsboom, K. A. (2013). *Frontiers of test validity theory*. New York: Routledge.
- Mellenbergh, G. J. (2011). *A conceptual introduction to psychometrics: Development, analysis, and application of psychological and educational tests*. The Hague: Eleven International Publishing.
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement*. Washington, D.C.: American Council on Education.
- Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, 50, 741-749.
- Mollenkopf, W. G. (1949). Variation of the standard error of measurement. *Psychometrika*, 14(3), 189-229.
- Muñoz, J., Elosua, P., & Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: Segunda edición. *Psicothema*, 25(2), 151-157.
- Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004).

- Psychological testing on the Internet: New problems, old issues. American Psychologist, 59(3), 150-162.*
- Newton, P. E., & Shaw, S. D. (2014). *Validity in educational & psychological assessment*. London: Thousand Oaks.
- Poortinga, Y. P., & Klieme, E. (2016). The history and status of testing across cultures and countries. In F. T. L. Leong, D. Bartram, F. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international handbook of testing and assessment*. New York: Oxford University Press.
- Prieto, G., & Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo, 77, 65-71.*
- Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology, 18(2), 161-169.*
- Thorndike, L. S. (1951). Reliability. In E. F. Lindquist (Ed.), *Educational measurement*. Washington, D.C.: American Council on Education.
- Zumbo, B. D. (2014). What role does, and should, the test standards play outside of the United States of America?. *Educational Measurement: Issues and Practice, 33(4), 31-33.*

Recibido: Septiembre 19, 2016  
Aprobado: Noviembre 4, 2016