

DIFERENCIAS DE OPINIÓN SOBRE EL USO DE LOS TEST: SU RELACIÓN CON LA EDAD EL GÉNERO Y EL SECTOR LABORAL

DIFFERENCES IN OPINIONS REGARDING TESTING PRACTICES: THE ROLE OF AGE, GENDER AND JOB SECTOR

Ana Hernández¹, M^a Dolores Hidalgo², José Muñiz³ y José Ramón Fernández-Hermida⁴

¹Universidad de Valencia. ²Universidad de Murcia. ³Universidad Nebrija. ⁴Universidad de Oviedo

Los test son importantes herramientas que ayudan a los profesionales de la Psicología a tomar decisiones sobre las personas. Para plantear acciones encaminadas a mejorar su uso, es importante determinar qué factores afectan a las opiniones que los psicólogos tienen sobre la práctica de los test. El presente estudio analiza el rol de tres factores: el género, la edad y el sector laboral (público vs. privado). Asimismo, se contrasta la opinión de los profesionales colegiados con las de los profesores universitarios que imparten asignaturas relacionadas con los test y su uso (Psicometría y Evaluación Psicológica o Psico-educativa). Tras analizar una muestra de 1.248 colegiados y 95 profesores de distintas universidades, los resultados muestran algunas diferencias de opinión en función de todos los factores considerados. Se analizan los resultados y sus implicaciones de cara al futuro.

Palabras clave: Uso de test, Opiniones sobre la práctica de los test, Género, Edad, Sector laboral.

Tests are important tools that help psychologists to make decisions about people. In order to propose actions aimed at improving the use of tests, it is important to determine what factors relate to psychologists' opinions about testing practices. In this study we assess the role of three factors: gender, age, and job sector (public vs. private). In addition, we compare professionals' opinions with those reported by professors of test-related subjects (psychometrics, and psychological and psycho-educational assessment). After analyzing a sample of 1,248 members of the Spanish Psychological Association and 95 professors from different universities, results show some differences in opinions across all the factors analyzed. Results and future challenges are discussed.

Key words: Test use, Opinions about testing practices, Gender, Age, Job sector.

LOS TEST PSICOLÓGICOS Y SU USO

Los test permiten a los psicólogos obtener evidencias empíricas que les ayuden en el proceso de evaluación psicológica y en la toma de decisiones. Así lo vienen reconociendo durante años los profesionales de la Psicología cuando se les pregunta por el uso de los test en su práctica diaria (Muñiz y Fernández-Hermida, 2000, 2010; Muñiz et al., 2020). Pero para que los test sean realmente útiles se necesitan tres condiciones básicas: que las propiedades psicométricas de las puntuaciones obtenidas sean satisfactorias, que los profesionales tengan la formación adecuada, y que las pruebas se utilicen correctamente (Elosua y Muñiz, 2013; Elosua, 2017; Muñiz et al., 2020). Como señalan Hernández et al. (2020a) la primera condición se garantiza cuando, en la construcción y análisis de los test, se aplican los conocimientos psicométricos de forma rigurosa (American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME), 2018; Haladyna y Rodríguez, 2013; Irwing et al., 2018; Muñiz y Fonseca-Pedrero, 2019). Para velar por dicho rigor, dis-

tinias organizaciones nacionales e internacionales, tales como el Consejo General de Colegios Oficiales de Psicólogos (CGCOP), a través de su Comisión de Test, la Comisión Internacional de Test (ITC -International Test Commission), la Federación Europea de Asociaciones de Psicólogos (EFPA -European Federation of Psychological Associations), o las editoriales de test, realizan distintas acciones. Entre ellas destaca la publicación de directrices para la creación y análisis de los test (AERA, APA y NCME, 2018), y de directrices para la adaptación de los test de una cultura a otra (ITC, 2017; Muñiz et al., 2013; Hernández et al., 2020b). También destaca el desarrollo de modelos para evaluar la calidad de los test, como el de la EFPA (Evers et al., 2013) o el del CGCOP (Prieto y Muñiz (2000), revisado por Hernández et al. (2016)), así como la aplicación de dichos modelos a distintas pruebas para evaluar su calidad. Por ejemplo, en España, el modelo del CGCOP, o cuestionario de evaluación de test (CET) y su posterior revisión (CET-R), se ha aplicado a un total de 82 pruebas durante distintas rondas de evaluación (Muñiz et al., 2011; Ponsoda y Hontangas, 2013, Hernández et al., 2015; Elosua y Geisinger, 2016; Fonseca-Pedrero y Muñiz, 2017; Hidalgo y Hernández, 2019; Gómez-Sánchez, 2019; Viladrich et al., 2021), con nuevas evaluaciones en marcha. Estas revisiones (con informes cuantitativos y cualitativos) se publican en la página web del CGCOP (<https://www.cop.es/index.php?pa>

Recibido: 9 febrero 2020 - Aceptado: 16 marzo 2020

Correspondencia: Ana Hernández. IDOCAL. Universidad de Valencia, Av. Blasco Ibáñez, 21. 46010 Valencia. España.

E-mail: Ana.Hernandez@uv.es



ge=evaluacion-tests-editados-en-espana) y pueden ser consultados por cualquier profesional interesado. El objetivo es ayudar a estos profesionales a valorar de forma rigurosa los test que pueden utilizar para una determinada evaluación antes de realizar su elección.

Sin embargo, aunque estas acciones tienen como objetivo mejorar las pruebas y frenar el uso de los test con calidad cuestionable, ya Lilienfeld et al. (2006) advertían de que hay pruebas que mantienen su popularidad a pesar de ser problemáticas. Se aducen algunas razones, por ejemplo, el efecto Barnum, referido a la tendencia a aceptar interpretaciones generales o vagas ofrecidas por los test como descriptores válidos para prácticamente cualquier persona, lo que erróneamente incrementa la confianza en el test utilizado. También la fantasía del alquimista, relativa a la creencia de que los resultados de cualquier test, cuando se combinan con otro tipo de datos, ofrecen información clínica importante, de forma que los resultados proporcionados por un mal test, se toman como una fuente más de información válida, a pesar de que eso pueda mermar la calidad final de la evaluación. Finalmente, destaca la tradición clínica y la inercia institucional a usar determinadas pruebas, a pesar de que haya otras más actualizadas y mejores. Por ello, es necesario seguir realizando acciones que contribuyan a disminuir estos problemas.

La selección de pruebas adecuadas se relaciona estrechamente con las otras dos condiciones mencionadas para garantizar la utilidad de los test psicológicos que, a su vez, están íntimamente ligadas: que los profesionales tengan la formación adecuada, y que los test se utilicen correctamente. La formación en Psicometría y en el uso de los test como herramienta de evaluación son aspectos cruciales a la hora de hacer un buen uso de los test. Esta formación se da fundamentalmente en la universidad, en las asignaturas de Psicometría y de Evaluación Psicológica (Hidalgo y Hernández, 2019; Hernández et al., 2020a). A la hora de garantizar que los usuarios de test han recibido dicha formación, las editoriales de test (en España CEPE, GiuntiEOS Psychometrics, Pearson y TEA) juegan un papel crucial, ya que restringen la venta de los test psicológicos a quienes acrediten la cualificación requerida. Sin embargo, la formación universitaria no garantiza el buen uso de cualquier test, ni tampoco el buen uso de los test de por vida. Los problemas anteriormente mencionados para usar pruebas que no son idóneas (Lilienfeld et al., 2006), así como el desconocimiento sobre los nuevos avances psicométricos, el impacto de las nuevas tecnologías, o la falta de actualización sobre los test que se publican, pueden llevar a un mal uso de los mismos (Hernández et al., 2020a).

Por ello, además de la estrategia restrictiva que limita el uso de los test, es necesario realizar acciones informativas y formativas que promuevan el buen uso de los test. Entre las acciones informativas, tal y como señalan Muñiz et al., (2020), además de las directrices anteriormente menciona-

das (APA, AERA y NCME, 2018; ITC, 2017) y de la publicación de las evaluaciones de distintas pruebas del mercado siguiendo un modelo de calidad (Hernández et al., 2016) destaca también la publicación de códigos éticos, tales como el de la EFPA (2005), y de estándares sobre el proceso de evaluación, como las directrices de la asociación europea de evaluación psicológica (Fernández-Ballesteros et al., 2001) o la norma ISO10667 (AENOR, www.aenor.es), que regula la evaluación de personas en contextos laborales y organizacionales. Entre las acciones formativas destacan los cursos sobre construcción y evaluación de test ofrecidos por el CGCOP a través del Programa de Formación Continuada a Distancia (FOCAD) (Elosua, 2019; Muñiz y Fonseca-Pedrero, 2017), o los cursos que ofrecen las principales editoriales de test que operan en España.

Antes de diseñar nuevas acciones para mejorar el uso de los test psicológicos (ya sean restrictivas, formativas o informativas), es importante conocer las opiniones que los psicólogos tienen sobre los test y su uso. El CGCOP ha sido consciente de ello durante años. Por ello, a través de su comisión de test, ha impulsado la realización de encuestas entre sus colegiados para poder conocer y valorar cuáles son las necesidades, retos, y problemas a los que se enfrentan los profesionales de la Psicología al usar los test en su práctica profesional (Muñiz y Fernández-Hermida, 2000; 2010; Muñiz et al., 2020).

LA OPINIÓN DE LOS PSICÓLOGOS SOBRE LOS TEST

El último estudio (Muñiz et al., 2020), ofrece resultados interesantes, tanto en general como diferenciando por especialidad. En primer lugar, los psicólogos españoles (especialmente los del área de Psicología Educativa, seguidos de los del área de Psicología del Trabajo) muestran una actitud positiva hacia los test. Sin embargo, reconocen la necesidad de seguir formándose (e informándose) para usar los test adecuadamente. Asimismo, aunque en general están a favor de incrementar el control y la regulación de los test y su uso, hay diferencias por especialidad en algunas cuestiones, como la necesidad de introducir un sistema para acreditar la competencia de los usuarios de test o introducir más controles, siendo la especialidad clínica la más reacia a estas cuestiones. Por otra parte, los profesionales se muestran moderadamente escépticos a la incorporación de las nuevas tecnologías en la práctica de los test, sobre todo los clínicos. También reconocen que hay problemas en el uso de los test, si bien estos se dan con una frecuencia moderada. El problema más importante se refiere al precio de los test, que se considera demasiado elevado, sobre todo entre los especialistas de clínica-salud y educativa. Finalmente, aunque la evaluación de la calidad de los test que lleva a cabo la Comisión de Test del CGCOP es poco conocida (un 22,5% de los profesionales dicen conocerla), quienes la conocen la valoran muy positivamente.



OBJETIVO

Cabe esperar que las opiniones de los psicólogos sobre los test y su uso varíen dependiendo de otros factores, además de la especialidad. En el presente estudio se analiza el rol del género, la edad y el sector laboral (público vs. privado). Conocer estas opiniones ayudará a comprender mejor la realidad del uso de los test en los distintos grupos de profesionales y, a proponer, si fuese necesario, medidas dirigidas a colectivos específicos.

Por ejemplo, por lo que se refiere al género, existe cierta evidencia de que hombres y mujeres perciben de forma diferente el riesgo (que podría trasladarse en diferencias en la valoración de las consecuencias del uso de un test que no sea el más adecuado) y también de que las mujeres son más meticulosas que los hombres (Chan, 2011; Greenberg y Schneider, 1995; Gustafson, 1998). Estas diferencias podrían afectar a la manera de usar los test. Aunque son estereotipos que pueden no generalizarse en todos los casos, resulta interesante evaluar si se observan diferencias de opinión sobre el uso de los test entre hombres y mujeres. Por lo que se refiere al sector, los psicólogos que realizan su práctica en distintos sectores podrían disponer de distintos recursos, así como distintos niveles de autonomía para tomar decisiones, etc., lo que podría afectar a los test que se emplean, y a su uso (por ejemplo, haciendo algunas prácticas inadecuadas como fotocopiar hojas de respuesta). Finalmente, de especial interés resulta la edad, ya que suele acompañarse de una mayor experiencia profesional. Además, la educación universitaria ha cambiado mucho con la incorporación de España al Espacio Europeo de Educación Superior, cuyos planes de estudio se comenzaron a implementar entre 2009 y 2010. Ambos aspectos, formación y experiencia, junto con los cambios generacionales, pueden afectar a las opiniones que se tienen sobre los test. Finalmente, también contrastaremos la opinión de los profesionales colegiados con las de los profesores universitarios que imparten asignaturas relacionadas con los test y su uso (Psicometría y Evaluación Psicológica o Psico-educativa), ya que éstos juegan un papel crucial a la hora de formar a los (futuros) psicólogos en las competencias necesarias para el uso adecuado de los test.

PARTICIPANTES Y PROCEDIMIENTO

Los colegiados de los Colegios Oficiales de Psicólogos (COP) de nuestro país fueron invitados a participar en una encuesta a través de distintos medios de difusión, aportándose información sobre los objetivos y la importancia del estudio. La encuesta se aplicó *online*, a través de un enlace presentado en la página web del CGCOP. Al entrar a la encuesta, los participantes daban su consentimiento informado. Se garantizó el anonimato y el tratamiento confidencial de los datos. La encuesta fue respondida por 1.248 colegiados. El 73,7% fueron mujeres y la edad promedio fue de 46,3 años (DT=11,1). La mayoría (69,9%), trabajaba en el sector

privado y el 26,6% en el sector público, mientras que el 3,5% estaba en paro. Asimismo, el área de especialización mayoritaria fue Clínica-Salud (67,6%), seguida de Educativa (12,7%) y Trabajo (5,5%). Otras especialidades, como jurídica, deporte, servicios sociales, etc., conformaron el 14,1% restante. Una descripción más pormenorizada de la muestra puede consultarse en Muñiz et al. (2020).

Respecto a la muestra de profesores universitarios de Psicometría y de evaluación psicológica y psicoeducativa, estos fueron identificados a través de las páginas web de los grados de Psicología de las universidades españolas (en total, 316 profesores), y fueron invitados por email a participar en el estudio a través de una encuesta online, garantizándose la confidencialidad. Respondieron un total de 97 profesores (30,4%), aunque se eliminaron las respuestas de dos participantes que indicaron que no impartían ninguna de las asignaturas de interés. De los 95 profesores que se mantuvieron en la muestra, 73,7% de los participantes pertenecían a universidades públicas, el 51,6% eran varones y la mayoría (49,5%) eran profesores de Psicometría, seguidos del 35,8% de Evaluación Psicológica. Solo dos participantes (2,1%) indicaron ser profesores de evaluación psicoeducativa. El resto (12,6%), o bien impartían varias de las asignaturas mencionadas o combinaban alguna de estas asignaturas con otras. La antigüedad docente promedio fue de 17,5 años (DT=11,1).

LA ENCUESTA

Para recoger las opiniones de los psicólogos sobre los test y su uso empleé el cuestionario de 31 ítems descrito en detalle y publicado al completo en Muñiz et al. (2020), el cual se respondía mediante una escala graduada de 5 puntos (1-5) (ver ítems en Tabla 1). Considerando los resultados sobre la dimensionalidad del cuestionario, se diferenciaron 5 dimensiones: Actitudes hacia los test, Formación y conocimientos sobre los test, Necesidad de control y regulación de los test y su uso, Influencia de las nuevas tecnologías en el uso de los test, y Problemas en el uso de los test. Los coeficientes de fiabilidad reportados (alfa de Cronbach) oscilaron entre 0,62, para la dimensión de "Necesidad de control y regulación de los test y su uso", y 0,91 para la dimensión de "Problemas en el uso de los test". El coeficiente Omega, que no requiere que las medidas sean Tau-equivalentes (McDonald, 1999) presenta valores que oscilan entre 0,80 para la dimensión de "Necesidad de control y regulación de los test y su uso", y 0,94 para la dimensión de "Problemas en el uso de los test".

Aunque todas las dimensiones han sido comparadas a través de los distintos grupos de interés, en el presente estudio se ofrecen únicamente los análisis diferenciados por ítems que han ofrecido los resultados más relevantes. Asimismo, se analiza la pregunta adicional sobre si conocían las evaluaciones que realiza la comisión de test del CGCOP. A quienes sí la conocían, se les pidió que valoraran dichas evaluaciones (ver ítems en Tabla 3).



En la muestra de profesores, recogida con anterioridad (Hidalgo y Hernández, 2019), se empleó el cuestionario usado en el estudio sobre las opiniones de los psicólogos publicado en 2010 (Muñiz y Fernández-Hermida, 2010), que no incluía los ítems 25, 26 y 27 de la dimensión de Actitudes hacia los test del cuestionario de 2020. Además, se realizaron algunas modificaciones. Se eliminó el ítem 6, "Mis conocimientos actuales en relación con los test los recibí fundamentalmente durante la carrera de Psicología", por razones obvias. Asimismo, se incluyeron todos los problemas de la última versión del cuestionario excepto el último, referido al alto precio de los test. Finalmente, no se recogieron las opiniones sobre las evaluaciones de test que lleva a cabo la

comisión de test del CGCOP. En su lugar, se les preguntó si en las clases que impartían empleaban el modelo y las evaluaciones realizadas (estos últimos resultados pueden consultarse en Hidalgo y Hernández, 2019).

ANÁLISIS DE DATOS

Se obtuvieron los estadísticos descriptivos (medias y desviaciones típicas) de todos los ítems. Para comparar las puntuaciones se realizaron pruebas t de Student, para el caso del género, del tipo de muestra (profesionales vs. profesores) y del sector, para el que se diferenció entre sector privado y público, excluyéndose a las personas en paro. En el caso de la edad, se formaron 5 grupos de edad (ver Tabla 2) y se re-

TABLA 1
CUESTIONARIO UTILIZADO PARA RECOGER LAS OPINIONES SOBRE LOS TEST

Ítems

1. La formación recibida en la carrera de Psicología es suficiente para la utilización correcta de la mayoría de los test
2. La formación recibida en cursos y Máster es suficiente para el uso correcto de la mayoría de los test
3. La Federación Europea de Asociaciones de Psicólogos (EFPA) debería establecer un sistema para acreditar la competencia de los usuarios de test
4. Los profesionales disponen de suficiente información (revisiones independientes, investigaciones, documentación...) sobre la calidad de los test editados en nuestro país
5. En mi campo profesional los test informatizados están reemplazando progresivamente a los test de papel y lápiz
6. Mis conocimientos actuales en relación con los test los recibí fundamentalmente durante la carrera de Psicología
7. La aplicación de los test por Internet tiene muchas ventajas en comparación con la aplicación clásica de papel y lápiz
8. El uso de los test psicológicos debería restringirse a psicólogos cualificados
9. Aunque los no psicólogos podrían aplicar y puntuar los test, la interpretación e información sobre los resultados deberían estar restringidos a los psicólogos
10. Los informes generados automáticamente por ordenador no tienen ninguna validez
11. Los estándares y directrices que definen las cualidades técnicas mínimas de un test deberían ser obligatorios [por ejemplo los estándares de la Federación Europea de Asociaciones de Psicólogos (EFPA)] o los de la Asociación de Psicología
12. Se necesita legislación para controlar los abusos más serios con los test
13. La aplicación de los test por Internet pone en desventaja a algunas personas evaluadas
14. Todo aquel que sea capaz de demostrar su competencia en el uso de los test (sea psicólogo o no) debería ser autorizado para utilizarlos
15. Si se utiliza adecuadamente, Internet puede mejorar mucho la calidad de la aplicación de los test
16. Los controles sobre los test deberían ser mínimos, pues inhiben el desarrollo de nuevas ideas y nuevos procedimientos de evaluación
17. La aplicación de test por Internet no permite proteger la privacidad de los usuarios
18. Habría que permitir a los editores que vendiesen cualquier test que consideren adecuado
19. El Colegio Oficial de Psicólogos debería ejercer un papel más activo para regular y mejorar el uso que se hace de los test
20. La aplicación de test por Internet abre posibilidades de fraude
21. En el desempeño de mi profesión utilizo test habitualmente
22. Los test constituyen una excelente fuente de información si se combinan con otros datos psicológicos
23. Utilizados correctamente, los test son de gran ayuda para el psicólogo
24. Teniendo en cuenta todos los aspectos, creo que en la última década el uso de los test ha mejorado en mi país
25. La práctica profesional basada en evidencias potencia el uso de los test como fuentes de información
26. En la toma de decisiones profesionales doy un peso importante a los datos obtenidos mediante test
27. Antes de utilizar un test compruebo sus propiedades psicométricas
- 28-1. Hacer fotocopias de materiales sujetos a copyright
- 28-2. Hacer evaluaciones utilizando test inadecuados
- 28-3. No estar al día
- 28-4. No contrastar las interpretaciones con otros
- 28-5. No tener en cuenta los errores de medida de las puntuaciones
- 28-6. No restringir la aplicación de los test a personal cualificado
- 28-7. No tener en cuenta las condiciones locales (país, región) que pueden afectar a la validez
- 28-8. Hacer interpretaciones que van más allá de los límites del test
- 28-9. Utilizar test con baremos inadecuados
- 28-10. Utilizar algunos ítems sin los permisos correspondientes
- 28-11. Tener un precio muy elevado



alizaron ANOVAS. En caso de que estos ofrecieran resultados estadísticamente significativos, se realizaron comparaciones entre los grupos mediante la prueba de Tukey. Cuando no se cumplía el supuesto de homogeneidad de varianzas requerido en el ANOVA, se empleó la prueba robusta de Brown-Forsythe. Cuando los resultados fueron estadísticamente significativos, se empleó la prueba de Games-Howell para realizar comparaciones a posteriori. Todos los análisis se llevaron a cabo mediante el programa IBM-SPSS-26. Además, se obtuvieron indicadores del tamaño del efecto: η^2_{parcial} para los ANOVA y d de Cohen para las pruebas t (Cohen, 1992).

DIFERENCIAS DE OPINIONES SOBRE LOS TEST

Los resultados de las comparaciones para los ítems del cuestionario en función de los distintos grupos formados se presentan en la Tabla 2, indicando las comparaciones que resultaron estadísticamente significativas. Se comentan a continuación las diferencias más relevantes. Quienes estén interesados en los resultados completos, incluyendo los resultados por dimensiones, pueden dirigirse a la primera autora.

Edad

En primer lugar, por lo que se refiere a los ítems relativos a la dimensión de Formación y conocimientos sobre los test (ítems 1, 2, 4 y 6), solo hay diferencias estadísticamente significativas para el ítem 6, "Mis conocimientos actuales en relación con los test los recibí fundamentalmente durante la carrera de Psicología". Se observa que cuanto mayor es la edad, menor es la media en el ítem ($p = 0,012$; $\eta^2_{\text{parcial}} = 0,010$). Como es lógico, los profesionales siguen ganando conocimientos sobre los test a lo largo de su carrera profesional. Sin embargo, solo existen diferencias estadísticamente significativas entre el grupo más joven (entre 22 y 29 años) y el de más edad (60 años o más) ($p = 0,013$). Esta diferencia también sugiere que la formación actual puede ser más completa en materia de test que las que recibieron las primeras generaciones de psicólogos.

En segundo lugar, respecto a la regulación y el control en el uso de los test, los resultados indican que los psicólogos de más edad tienden a ser más favorables a introducir mecanismos de control, si bien la relación no siempre es lineal. Por ejemplo, hay diferencias de opinión respecto a que la EFPA establezca un sistema de acreditación de usuarios de test (ítem 3; $p = 0,017$; $\eta^2_{\text{parcial}} = 0,026$). Las diferencias alcanzan la significación estadística al comparar el grupo entre 30 y 39 años, menos favorable, y el grupo de 60 años o más, que es el más favorable ($p = 0,021$). También se observan diferencias estadísticamente significativas en el ítem 12 ($p = 0,001$; $\eta^2_{\text{parcial}} = 0,012$). Es también el grupo entre 30 y 39 años, junto con el grupo entre 40 y 49 años, quienes consideran menos necesaria la introducción de legislación que controle los abusos más serios, en comparación con el

grupo entre 50 y 59 años ($p = 0,006$ y $p = 0,008$, respectivamente), si bien los tres grupos se muestran favorables a dicha legislación, con medias cercanas a 4. También es entre los dos últimos grupos, donde se observan diferencias estadísticamente significativas sobre el papel que debería jugar el COP en la regulación y mejora del uso de los test (ítem 19; $p = 0,001$; $\eta^2_{\text{parcial}} = 0,014$). De nuevo es el grupo entre 50 y 59 años, en comparación con el grupo entre 40 y 49 quienes más importancia dan a que el COP juegue un papel más activo ($p < 0,001$).

En tercer lugar, por lo que se refiere a las actitudes hacia el uso de los test (ítems 21-26, junto con el ítem 27 que presentó una saturación factorial débil en la dimensión), se observan diferencias estadísticamente significativas en el ítem 25 ($p = 0,004$; $\eta^2_{\text{parcial}} = 0,012$). Destaca que son los jóvenes (los grupos entre 22 y 29, y entre 30 y 39 años) quienes más confían en la práctica profesional basada en la evidencia para potenciar el uso de los test en comparación con los profesionales de 60 años o más ($p = 0,031$ y $p = 0,004$, respectivamente).

En cuarto lugar, centrándonos en las nuevas tecnologías (informatización, automatización e internet) (ítems 5, 7, 10, 13, 15, 17 y 20), como podríamos anticipar, son los jóvenes quienes se muestran más favorables a la incorporación de dichas tecnologías. En concreto se observan diferencias estadísticamente significativas en el ítem 15 ($p = 0,007$; $\eta^2_{\text{parcial}} = 0,011$). Los dos grupos más jóvenes (entre 22 y 29, y entre 30 y 39 años) confían más en que Internet puede mejorar mucho la calidad de la aplicación de los test en comparación con el grupo de 60 años o más ($p = 0,023$ y $p = 0,008$, respectivamente). Por el contrario, son los grupos de más edad quienes más preocupación muestran por la protección de la privacidad de los usuarios cuando los test se aplican por internet (ítem 17; $p < 0,001$; $\eta^2_{\text{parcial}} = 0,0026$), dándose diferencias estadísticamente significativas entre los dos grupos más jóvenes y los dos de más edad y también entre el grupo entre 40 y 49 años y el grupo más edad (con valores p entre $p < 0,001$ y $p = 0,023$). Finalmente, también se observan diferencias estadísticamente significativas en el ítem 5 ($p = 0,006$; $\eta^2_{\text{parcial}} = 0,010$). Resulta curioso que el grupo más joven, entre 22 y 29 años, considera en menor medida que los test informatizados estén reemplazando a los de papel y lápiz en comparación con las opiniones de los profesionales más senior, entre 50 y 59 años y de 60 o más ($p = 0,032$ y $p = 0,010$, respectivamente). Probablemente los más jóvenes esperarían que dicho reemplazo se diera a un ritmo más rápido del que se está produciendo.

Finalmente, en cuanto a los problemas en el uso de los test (ítems 28.1 a 28.11), se observan diferencias estadísticamente significativas en los ítems 28.1 ($p = 0,005$; $\eta^2_{\text{parcial}} = 0,012$) y 28.11 ($p < 0,001$; $\eta^2_{\text{parcial}} = 0,020$). En concreto, se observa una tendencia a usar menos fotocopias de materiales sujetos a copyright (ítem 28.1) conforme aumenta la edad, si bien las únicas diferencias estadísticamente signifi-



cativas se dan entre el grupo de 30 a 39 años y el de 50 a 59 años ($p = 0,043$). En la misma línea, son los más jóvenes quienes más se quejan del alto precio de los test (ítem 28.11) con diferencias estadísticamente significativas entre los 3 grupos más jóvenes, y el grupo de 60 años o más ($p = 0,005$, $p < 0,001$ y $p = 0,008$, respectivamente).

Género

Tanto hombres como mujeres consideran que la formación universitaria en cursos y estudios de Máster y, especialmente

en licenciatura/grado, no son suficientes para el uso correcto de la mayoría de los test. Sin embargo, los hombres valoran dicha formación más positivamente que las mujeres, tanto por lo que se refiere a la formación durante la carrera (ítem 1; $p = 0,001$; $d = 0,22$) como a la formación en cursos y Máster (ítem 2; $p = 0,028$; $d = 0,14$). En cuanto a las prácticas en el uso de los test y los problemas que encuentran en su entorno profesional, mientras que los hombres indican que comprueban las propiedades psicométricas de los test que van a usar en mayor grado que las mujeres (ítem 27; $p =$

TABLA 2
COMPARACIONES DE LAS MEDIAS DE LOS ÍTEMS PARA LOS DISTINTOS GRUPOS ANALIZADOS

Nº de ítem y grupos con diferencias	EDAD										GÉNERO				SECTOR				PRFESIONALES vs. PROFESORES				
	22-29 años (N=79) A		30-39 años (N=300) B		40-49 años (N=342) C		50-59 años (N=362) D		60 o más (N=163) E		Hombre (N=328)		Mujer (N=920)		Público (N=332)		Privado (N=872)		Global Profesionales (N=1248)		Profesores (N=95)		
	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media
1. **Género, **Prof.	2,35	1,17	2,49	1,16	2,43	1,11	2,49	1,18	2,54	1,17	2,66	1,21	2,41	1,13	2,57	1,14	2,44	1,16	2,47	1,16	2,97	0,99	
2. *Género	3,08	1,07	3,09	1,09	2,99	1,05	3,04	1,00	3,07	1,03	3,16	1,04	3,01	1,05	3,08	0,97	3,03	1,07	3,05	1,05	3,20	0,95	
3. *Edad (B-E), *Sector, **Prof.	3,28	1,22	3,20	1,32	3,44	1,25	3,46	1,32	3,58	1,27	3,34	1,38	3,42	1,26	3,51	1,22	3,34	1,33	3,39	1,30	3,75	1,07	
4. **Prof.	2,90	0,96	3,12	1,12	3,00	1,10	3,06	1,08	3,00	1,07	3,09	1,09	3,03	1,08	3,00	1,05	3,06	1,10	3,04	1,09	2,77	0,88	
5. **Edad (A-D) (A-E), **Sector, *Prof.	3,00	1,24	3,30	1,31	3,30	1,29	3,46	1,27	3,58	1,28	3,36	1,22	3,36	1,31	3,20	1,30	3,44	1,29	3,36	1,29	3,08	1,20	
6. *Edad (A-E)	2,68	1,24	2,42	1,27	2,32	1,18	2,27	1,23	2,15	1,11	2,43	1,16	2,30	1,23	2,37	1,20	2,29	1,20	2,33	1,21	—	—	
7.	3,30	1,15	3,21	1,17	3,14	1,17	3,19	1,16	2,99	1,25	3,11	1,15	3,18	1,19	3,23	1,16	3,12	1,18	3,16	1,18	3,37	1,09	
8.	4,25	0,85	4,18	1,01	4,20	1,07	4,25	1,07	4,38	0,98	4,20	1,10	4,25	1,01	4,23	1,06	4,23	1,02	4,24	1,03	4,21	0,97	
9.	4,43	0,96	4,38	1,09	4,30	1,22	4,40	1,15	4,45	1,12	4,38	1,14	4,38	1,14	4,34	1,16	4,38	1,14	4,38	1,14	4,28	1,07	
10.	2,76	1,08	2,71	1,11	2,77	1,14	2,87	1,11	2,80	1,17	2,85	1,15	2,76	1,11	2,81	1,10	2,77	1,13	2,78	1,12	2,62	1,14	
11. **Edad (C-D), *Sector	4,04	0,84	4,18	0,79	4,10	0,92	4,32	0,82	4,10	0,99	4,17	0,90	4,18	0,86	4,27	0,79	4,15	0,89	4,18	0,87	4,34	0,78	
12. **Edad (B-D) (C-D), **Sector	3,96	0,94	3,90	1,04	3,92	1,00	4,17	0,99	4,13	1,02	4,10	1,09	3,99	1,02	4,12	0,89	3,96	1,06	4,02	1,01	4,01	0,97	
13.	3,51	1,07	3,45	1,08	3,38	1,14	3,38	1,07	3,40	1,13	3,32	1,16	3,44	1,07	3,34	1,06	3,42	1,11	3,41	1,10	3,60	0,98	
14. *Sector	2,16	1,19	2,14	1,25	2,11	1,28	1,93	1,24	2,15	1,42	2,11	1,29	2,07	1,28	1,95	1,19	2,11	1,31	2,08	1,28	2,34	1,30	
15. **Edad (A-E) (B-E), **Prof.	3,52	0,90	3,47	1,10	3,35	1,10	3,33	1,03	3,11	1,13	3,39	1,05	3,34	1,09	3,34	1,07	3,35	1,08	3,36	1,08	3,60	0,98	
16. **Prof.	1,95	0,93	1,78	0,87	1,85	0,91	1,77	1,00	1,93	1,11	1,76	0,93	1,85	0,97	1,87	0,98	1,80	0,95	1,83	0,96	1,53	0,73	
17. **Edad (A-D) (A-E) (B-D) (B-E) (C-E), *Prof.	2,10	1,08	2,19	1,11	2,37	1,13	2,57	1,21	2,70	1,30	2,42	1,25	2,41	1,16	2,37	1,14	2,42	1,20	2,41	1,18	2,14	1,02	
18.	1,66	0,86	1,82	1,07	1,80	1,06	1,70	0,95	1,74	1,04	1,84	1,06	1,73	1,00	1,75	1,01	1,76	1,03	1,76	1,02	1,94	0,91	
19. *Edad (C-D), *Prof.	4,14	0,80	4,11	0,99	3,92	1,05	4,22	0,92	4,13	1,03	4,09	1,05	4,09	0,96	4,17	0,91	4,05	1,02	4,09	0,99	3,91	0,83	
20.	3,59	1,13	3,52	1,12	3,42	1,13	3,54	1,18	3,68	1,15	3,49	1,16	3,53	1,14	3,48	1,08	3,53	1,18	3,52	1,14	3,52	0,97	
21. **Edad (B-E), *Sector	3,75	1,20	4,12	1,15	4,05	1,14	4,07	1,16	3,76	1,32	3,93	1,24	4,05	1,16	3,90	1,23	4,08	1,16	4,02	1,18	4,04	1,14	
22. **Prof.	4,53	0,77	4,49	0,83	4,47	0,86	4,46	0,86	4,44	0,89	4,47	0,89	4,48	0,83	4,41	0,86	4,49	0,85	4,47	0,85	4,74	0,55	
23. **Prof.	4,53	0,75	4,46	0,86	4,46	0,85	4,49	0,82	4,35	0,92	4,44	0,91	4,47	0,83	4,39	0,85	4,47	0,86	4,46	0,85	4,68	0,57	
24.	3,59	0,84	3,73	0,93	3,75	0,92	3,76	0,95	3,71	0,95	3,66	1,01	3,76	0,90	3,73	0,86	3,73	0,96	3,73	0,93	3,88	0,87	
25. **Edad (A-E) (B-E)	4,10	0,83	4,05	0,84	3,97	0,92	3,92	0,94	3,74	1,03	4,00	0,98	3,94	0,90	3,98	0,89	3,94	0,94	3,96	0,92	—	—	
26.	3,37	0,94	3,52	1,05	3,43	0,97	3,49	0,95	3,37	1,08	3,53	1,06	3,43	0,97	3,42	0,97	3,47	1,02	3,46	1,00	—	—	
27. **Género	3,71	1,16	3,88	1,11	3,80	1,18	3,97	1,16	4,01	1,08	4,06	1,10	3,83	1,16	3,98	1,09	3,84	1,18	3,89	1,15	—	—	
28-1. **Edad (B-D), **Prof.	3,39	1,44	3,24	1,40	3,04	1,46	2,93	1,41	2,88	1,36	3,05	1,40	3,05	1,43	3,05	1,43	3,03	1,42	3,05	1,42	3,48	1,31	
28-2. **Prof.	2,49	1,31	2,38	1,26	2,54	1,36	2,34	1,26	2,53	1,41	2,55	1,34	2,40	1,30	2,40	1,28	2,43	1,32	2,44	1,31	3,40	1,21	
28-3. **Prof.	3,20	1,16	3,10	1,30	3,16	1,28	3,02	1,21	3,00	1,30	3,10	1,26	3,08	1,26	3,05	1,25	3,07	1,27	3,08	1,26	3,62	1,07	
28-4. **Prof.	3,04	1,23	3,04	1,29	3,08	1,31	2,99	1,23	3,09	1,30	3,09	1,29	3,02	1,27	3,00	1,28	3,02	1,28	3,04	1,28	3,77	0,95	
28-5. **Prof.	3,04	1,26	3,09	1,25	3,06	1,26	2,90	1,24	2,98	1,26	3,10	1,30	2,97	1,24	2,92	1,26	3,02	1,25	3,01	1,25	3,93	1,17	
28-6. **Prof.	2,99	1,46	2,92	1,48	2,87	1,49	2,80	1,49	2,96	1,44	2,99	1,44	2,84	1,49	2,83	1,53	2,88	1,47	2,88	1,48	3,41	1,26	
28-7. **Prof.	2,97	1,40	2,85	1,33	2,94	1,35	2,88	1,32	2,89	1,32	2,95	1,36	2,87	1,32	2,82	1,31	2,89	1,34	2,89	1,33	3,40	1,19	
28-8. *Género, **Sector, **Prof.	3,05	1,34	2,90	1,34	2,90	1,37	2,75	1,37	2,94	1,33	3,03	1,39	2,81	1,34	2,69	1,33	2,92	1,36	2,87	1,36	3,69	1,12	
28-9. **Género, **Prof.	2,71	1,29	2,41	1,24	2,57	1,33	2,42	1,30	2,62	1,28	2,75	1,34	2,41	1,27	2,45	1,28	2,51	1,30	2,5	1,30	3,63	1,16	
28-10. **Prof.	2,32	1,35	2,18	1,20	2,35	1,36	2,19	1,24	2,31	1,31	2,36	1,32	2,21	1,26	2,19	1,29	2,26	1,28	2,25	1,28	3,29	1,26	
28-11. **Edad (A-E) (B-D) (B-E) (C-E), **género, **Sector	4,37	1,16	4,33	1,05	4,20	1,14	4,10	1,10	3,85	1,11	3,99	1,16	4,23	1,09	4,01	1,14	4,20	1,11	4,17	1,11	—	—	

NOTA: Edad, género y sector se refiere únicamente a la muestra de profesionales. Prof.: Profesionales vs. Profesores Ver enunciado de los ítems en Tabla 3.

* $p < 0,05$; ** $p < 0,01$; DT= Desviación Típica



0,002; $d = 0,20$), los hombres reportan que, en su entorno se hacen interpretaciones que van más allá de los límites del test (ítem 28.8; $p = 0,014$; $d = 0,16$) y se utilizan baremos inadecuados (ítem 28.9; $p < 0,001$; $d = 0,26$) con mayor frecuencia de lo que reportan las mujeres. Finalmente, las mujeres son las que más consideran que el precio de los test es muy elevado (ítem 28.11; $p = 0,001$; $d = 0,22$).

Sector

Los resultados muestran que los profesionales del sector público son más favorables al establecimiento de sistemas de regulación y control (con diferencias estadísticamente significativas en 4 de los 9 ítems de la dimensión: ítems 3, 11, 12 y 14). En particular, los colegiados del sector público están más a favor de la puesta en marcha un sistema de acreditación de la competencia de los usuarios de test por parte de la EFPA (ítem 3; $p = 0,038$; $d = 0,13$), en comparación con los del sector privado. Los colegiados del sector público también defienden en mayor medida que los estándares y directrices que marcan las cualidades mínimas de un test (como los de la EFPA o el CGCOP) sean obligatorias (ítem 11; $p = 0,038$; $d = 0,14$), dan más importancia a la necesidad de introducir legislación que controle los abusos más serios con los test (ítem 12; $p = 0,007$; $d = 0,16$) y, finalmente, se muestran más reticentes a que cualquier persona capaz demostrar su competencia en el uso de los test (sea psicólogo o no) sea autorizado para usarlos (ítem 14; $p = 0,046$; $d = 0,13$). Resulta interesante que es en el sector privado donde los test informatizados están remplazando a los test de papel y lápiz en mayor medida (ítem 5; $p = 0,005$; $d = 0,15$), siendo también los profesionales del sector privado quienes dicen usar los test de forma más habitual (ítem 21; $p = 0,026$; $d = 0,15$). Finalmente, por lo que se refiere a los problemas sobre el uso de los test, aunque las medias siempre están por debajo de 3, en el sector privado se reconoce que se hacen interpretaciones que van más allá de los límites de los test con mayor frecuencia que en el ámbito público (ítem 28.8; $p = 0,009$; $d = 0,17$) y se quejan en mayor medida del precio de los test (ítem 28.11; $p = 0,008$; $d = 0,17$).

Profesionales vs. Profesores

Los resultados indican que los profesores tienen un punto de vista más favorable sobre la formación que se recibe durante la carrera (ítem 1; $p < 0,001$; $d = 0,43$), si bien en ambos casos la puntuación no llega al punto medio de la escala 3. Curiosamente, sin embargo, son los profesores en comparación con los profesionales, quienes más ponen en duda que se disponga de suficiente información (revisiones independientes, investigaciones, documentación...) sobre la calidad de los test (ítem 4; $p = 0,005$; $d = 0,25$). En segundo lugar, considerando los ítems referidos a la regulación y el control sobre los test y su uso, los profesores son más favorables al establecimiento de un sistema de acreditación de usuarios de test (ítem 3; $p = 0,003$; $d = 0,28$), y consideran en menor

medida que los controles sobre el uso de los test debieran ser mínimos (ítem 16; $p < 0,001$; $d = 0,32$). Sin embargo, son los profesionales quienes están ligeramente más a favor de que el CGCOP juegue un papel más activo en la regulación y mejora del uso de los test (ítem 19; $p = 0,039$; $d = 0,18$). En tercer lugar, por lo que se refiere a las actitudes hacia los test, los profesores muestran opiniones aún más favorables que los profesionales en dos de los ítems; particularmente, indican que los test constituyen una excelente fuente de información combinados con otros datos (ítem 22; $p < 0,001$; $d = 0,32$), y que, usados correctamente, son de gran ayuda para el psicólogo (ítem 23; $p = 0,001$; $d = 0,26$). En cuarto lugar, respecto a los ítems referidos a las nuevas tecnologías, los profesores son más favorables al uso de internet en la aplicación de test. Consideran que si se usa adecuadamente puede mejorar la calidad de dicha aplicación (ítem 15; $p = 0,032$; $d = 0,22$) y cuestionan en menor medida que la privacidad quede expuesta al usar internet (ítem 17; $p = 0,013$; $d = 0,23$). En el ámbito académico parece que los test informatizados están remplazando a los de lápiz y papel a un ritmo más lento que en ámbito profesional (ítem 5; $p = 0,041$; $d = 0,22$). Finalmente, por lo que se refiere a los problemas, resulta curioso que todos los problemas sobre el uso de los test (ítems 28.1 a 28.10 en Tabla 1) son señalados con mayor frecuencia por los profesores que por los profesionales (Tabla 2) ($p = 0,004$ para el ítem 28.1, y $p < 0,001$, para los ítems 28.2 a 28.10, con tamaños del efecto que oscilan entre $d = 0,30$, para 28-1 y $d = 0,88$ para 28.9).

CONOCIMIENTO Y VALORACIÓN DE LAS EVALUACIONES DE TEST IMPULSADAS POR EL CGCOP

Por lo que se refiere al conocimiento de las evaluaciones de test llevadas a cabo por el CGCOP, conforme aumenta la edad de los grupos, más aumenta el porcentaje de profesionales que dice conocer las evaluaciones ($p < 0,001$; $\eta^2_{\text{parcial}} = 0,028$), con proporciones que varían entre el 0,05 para el grupo entre 22 y 29 años, y el 0,31 para el grupo de 60 años o más (Tabla 3). Entre quienes las conocen, no hay diferencias estadísticamente significativas en las valoraciones en función de la edad. En cuanto al género, la proporción de hombres que conocen dichas evaluaciones (0,34) difiere de la proporción de mujeres (0,18) ($p < 0,001$; $d = 0,39$), si bien, entre quienes conocen las evaluaciones no hay diferencias estadísticamente significativas a la hora de valorarlas. Por lo que se refiere al sector (público vs. privado), no hay diferencias estadísticamente significativas en el porcentaje de profesionales que conocen las evaluaciones en cada caso. Sin embargo, son los profesionales del sector público quienes consideran en mayor medida que estas evaluaciones son importantes para mejorar la calidad de los test ($p = 0,004$; $d = 0,32$). Finalmente, las diferencias entre las proporciones de profesores y profesionales que conocen las evaluaciones son muy marcadas: 0,71 para profesores y 0,23 para profesionales ($p < 0,001$; $d = 1,13$).



ALGUNAS CONCLUSIONES Y REFLEXIONES

Los resultados obtenidos ponen de manifiesto algunas diferencias de opinión de los profesionales de la Psicología sobre el uso de los test en función de la edad, el género y el sector laboral. Antes de analizar estas diferencias, cabe señalar que a) las opiniones sobre los test son positivas (Muñiz et al., 2020) y, cuando se observan diferencias estadísticamente significativas, éstas son de pequeña magnitud, ya que los tamaños del efecto son pequeños según los criterios tradicionales (Cohen, 1992); y b) las tendencias que señalamos son generales, ya que dentro de una misma categoría de análisis hay cierto grado de variabilidad en las opiniones y prácticas de uso de los test.

Los resultados del análisis en función de la edad indican, en primer lugar, que los profesionales, con los años, continúan ganando conocimientos sobre los test. Las estrategias formativas e informativas que impulsa el CGCOP deberían contribuir, al menos en parte, a ese desarrollo continuo a lo largo de la carrera profesional. Creemos que algunos de los temas que requieren especial atención por el impacto que está teniendo en los test y su uso son los modelos psicométricos de la Teoría de Respuesta a los Ítems (TRI), los test adaptativos informatizados (TAI), la tipificación continua o “continuous norming”, los modelos de diagnóstico clínico, los modelos de redes, o la evaluación ambulatoria.

Asimismo, se constata que los profesionales más jóvenes dan más importancia a la práctica basada en la evidencia para potenciar el uso de los test. Si esta tendencia se mantiene, los profesionales serán todavía más cuidadosos en la selección de test, basándose en las pruebas empíricas sobre su

calidad y utilidad. Con ello, cabe esperar que se minimicen problemas como el efecto Barnum, la fantasía del alquimista, o la tradición y la inercia institucional a usar determinadas pruebas (Lilienfeld et al., 2006). Se debe seguir trabajando en potenciar la práctica basada en la evidencia científica, y dirigir acciones específicas a los profesionales más veteranos. Los aspectos a considerar a la hora de seleccionar un test son bien conocidos independientemente del objetivo de evaluación (Canivez, 2019; Carretero-Dios y Pérez, 2007; Davis y Baillie, 2019), desde la delimitación conceptual del constructo a evaluar hasta la calidad y adecuación de las normas de interpretación. Todas estas cuestiones son consideradas en el modelo de evaluación de test (Hernández et al., 2016) y son recogidas en las evaluaciones que se publican en la web del CGCOP. Por ello hay que seguir difundiendo dichas evaluaciones, ya que solo un 22,5% de los colegiados dicen conocerlas. Fomentar la selección y uso de los test basándose en la evidencia científica contribuirá también a la mejora continua de la imagen y la confianza en la profesión, que no es cuestión baladí.

Con las nuevas generaciones se augura el remplazo de los test de papel y lápiz por tes informatizados y probablemente administrados online, ya que la gente más joven tiene una opinión más positiva hacia las nuevas tecnologías. Será cada vez más frecuente la obtención de puntuaciones y la generación de informes automatizados, la innovación en nuevos formatos de ítems/test, con contenidos multimedia y basados en juegos (la llamada gamificación), la evaluación on-line y la evaluación ambulatoria (Muñiz y Fonseca-Pedreiro, 2019; Parshall et al., 2010; Sanz et al., 2020; Seelow,

TABLA 3
CONOCIMIENTO Y VALORACIÓN DE LAS EVALUACIONES DE TEST LLEVADAS A CABO POR EL COP

Nº de ítem y grupos con diferencias	EDAD					GÉNERO				SECTOR				PRFESIONALES vs. PROFESORES								
	22-29 años (N=79) A		30-39 años (N=300) B		40-49 años (N=342) C		50-59 años (N=362) D		≥ 60 o más (N=163) E		Hombre (N=328)		Mujer (N=920)		Público (N=332)		Privado (N=872)		Profesionales (N=1248)		Profesores (N=95)	
	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT
29. Conoce la evaluación anual de test que realiza la Comisión de Test del COP (**Género, **Edad (A-C) (A-D) (A-E) (B-D) (B-E), **Prof.)	0,05	0,22	0,16	0,36	0,23	0,42	0,28	0,45	0,31	0,47	0,34	0,48	0,18	0,39	0,26	0,44	0,22	0,41	0,23	0,42	0,71	0,46
Si conocen la evaluación:																						
29-1. Me parece importante para mejorar la calidad de los test (**Sector)	4,50	1,00	4,32	0,66	4,41	0,80	4,61	0,60	4,37	0,82	4,48	0,66	4,45	0,76	4,62	0,51	4,39	0,79	4,46	0,72	—	—
29-2. Consulto las evaluaciones de los test en la página web del COP	3,75	1,89	3,72	1,28	3,73	1,19	3,89	1,14	3,84	0,90	3,94	1,05	3,72	1,20	3,94	1,03	3,75	1,18	3,81	1,15	—	—
29-3. Me ayuda en la elección de los test que utilizo	3,75	1,26	3,53	1,20	3,60	1,17	3,72	1,05	3,71	1,03	3,71	1,12	3,62	1,09	3,72	1,03	3,62	1,13	3,65	1,10	—	—
29-4. Es innecesaria, los psicólogos ya tienen suficiente información para saber qué test utilizar	1,50	1,50	1,79	1,16	1,63	0,98	1,58	1,03	1,53	0,95	1,60	1,15	1,63	0,92	1,56	1,03	1,65	1,02	1,62	1,01	—	—

* p < 0,05, ** p < 0,01 para las variables indicadas. Prof.: Profesionales vs. Profesores A los profesores no se les preguntaron las preguntas valorativas sobre la evaluación de test. La pregunta 29 se responde mediante la respuesta dicotómica "Sí" (1)/ "No" (0), por lo que la media representa la proporción de personas que conocen la evaluación. El resto de preguntas (29-1 a 29-4), se responde en una escala graduada de 5 alternativas (1-5). Edad, género y sector se refiere únicamente a la muestra de profesionales. DT= Desviación Típica



2019; Wan y Henly, 2012). Todas estas tendencias se enmarcan en la llamada Psicología 2.0 (Armayones et al., 2015), que está incidiendo en todos los aspectos de la evaluación psicológica, y que está siendo facilitada por el uso de teléfonos móviles y otros portables (Armayones et al., 2015; Chernyshenko y Stark, 2016; Muñiz y Fonseca-Pedrero, 2019). Incluso las redes sociales se han empleado en alguna ocasión para realizar evaluaciones basadas en “likes” (Kosinski et al., 2013). Para analizar este tipo de datos se necesitan nuevos modelos psicométricos, como los modelos de redes (Borsboom y Cramer, 2013; Fonseca-Pedrero, 2018), o los modelos de sistemas dinámicos (Nelson et al., 2017).

Creemos que estos avances tecnológicos abren grandísimas oportunidades en el mundo de la evaluación, pero también tienen sus riesgos. Por ejemplo, en un mercado dominado por el test presencial, la pandemia de la COVID-19 ha acelerado el proceso de tele-evaluación, ha hecho emerger la aplicación remota de test, y ha puesto de manifiesto la necesidad de contar con pruebas que puedan ser administradas con garantías en este contexto. A este respecto, Elosua (2021) presenta los principales riesgos de la administración remota de test (especialmente los que no se han creado específicamente para su administración on-line) y ofrece una serie de recomendaciones a seguir. Pero, además, Muñiz et al. (2020) advierten de la necesidad de ser precavidos y huir de algunas propuestas muy sofisticadas informáticamente para evaluar online constructos psicológicos y educativos que, a pesar de ser muy atractivas y estar impulsadas por fuertes campañas de marketing, no tienen detrás ninguna evidencia empírica que las avale. Las nuevas generaciones, más favorables a incorporar las nuevas tecnologías, deberán estar especialmente atentas a estos posibles fraudes.

Las diferencias de opinión en función del género y del sector (público o privado), son poco relevantes. El sector privado es más reacio en general a introducir mecanismos de control y regulación en el uso de los test (aun estando a favor), mientras que el sector público está siendo más lento en incorporar las nuevas tecnologías. Respecto al género, lo que más llama la atención es que las mujeres son las que consideran en mayor grado que el precio de los test es muy elevado. Esto sugiere una posible existencia de diferencias salariales en función del género, aunque sería necesario controlar muchas variables antes de afirmar que se da una brecha salarial. Algunos trabajos lo sugieren. Crothers et al. (2010) encuentran que las mujeres psicólogas norteamericanas cobran menos que sus colegas varones, controlando los años de experiencia y la formación, a pesar de ser una profesión en la que predominan las mujeres. Futuros estudios deberán indagar sobre este importante aspecto.

Cabe señalar que en este estudio hemos considerado la relación entre tres importantes variables sociodemográficas y laborales como la edad, el género y el sector (público o privado) y las opiniones sobre los test y su uso. Para futuros es-

tudios sería interesante recoger y analizar el papel de otras variables como la comunidad autónoma, la universidad de formación o el nivel educativo (grado/licenciatura, máster o doctorado), entre otras.

Finalmente, cabe destacar que las diferencias más marcadas, con mayores tamaños del efecto, se dan entre los profesionales y los profesores, especialmente a la hora de valorar la frecuencia con que se dan los distintos problemas relacionados con el uso de los test, con puntuaciones más altas en general en los profesores. Aunque las muestras difieren sustancialmente en tamaño, los resultados sugieren distintas opciones, que futuros estudios deberían explorar. Puede ser que se den más problemas en el ámbito académico, donde las consecuencias de un uso inadecuado de los test no tienen consecuencias directas tan importantes sobre las personas evaluadas, ya que en el mundo académico los test se emplean principalmente con fines de investigación (aunque obviamente hay unos criterios mínimos de calidad que se deben mantener). Otra posibilidad es que, por su condición de expertos en evaluación y psicometría, sean más exigentes en los estándares de calidad que se deberían cumplir y por eso valoran los problemas como más importantes. O finalmente, puede indicar que existe una brecha entre el mundo académico y el mundo profesional, tal y como ha sido puesto de manifiesto en algunas ocasiones por las editoriales de test (Hidalgo y Hernández, 2019).

Antes de concluir es necesario señalar que el presente estudio tiene algunas limitaciones. En primer lugar, aunque Muñiz et al. (2020) constataron que la muestra representaba razonablemente bien a la población de colegiados del COP en algunas de las principales variables, como la edad, el procedimiento de recogida de datos empleado no garantiza que la muestra sea representativa del total de psicólogas y psicólogos colegiados en nuestro país. Esta crítica es extensiva a la muestra de profesores empleada. Una segunda limitación, relacionada con la anterior, es el número reducido de participantes. El total de 1.248 participantes en la muestra de profesionales solo constituye el 1,7% de profesionales colegiados en el momento de la encuesta, si bien entre los profesores, la participación fue algo superior (30%). Finalmente, el tamaño de la muestra guarda relación con la última limitación. Algunos de los resultados presentados podrían deberse no únicamente a las variables analizadas sino a la interacción con otras de las variables (por ejemplo, que las diferencias observadas en función de la edad, dependan del tipo de sector). Aunque hubiera sido interesante analizar las interacciones, este análisis se descartó, ya que algunas de las categorías resultantes al cruzar las distintas variables, no quedaban suficientemente representadas en nuestra muestra (con casos extremos en los que no había ninguna observación). Para futuras encuestas de opinión sobre los test, sería conveniente reforzar la participación mediante distintas estrategias, con el fin de obtener muestras mayores que garanticen que todas las categorías de interés queden suficientemente representadas.



En suma, a pesar de las limitaciones mencionadas, los resultados del estudio permiten obtener conclusiones importantes. Aunque las medidas impulsadas por las distintas asociaciones e instituciones para mejorar los test y su uso parecen estar teniendo eco en los profesionales, todavía hay camino por hacer, especialmente si se consideran los cambios que las nuevas tecnologías están provocando en la forma de crear, administrar, puntuar e incluso interpretar los test. Por ello, desde la comisión de test del CGCOP es necesario seguir trabajando en mejorar las competencias y conocimientos de los usuarios de test, con acciones formativas e informativas, pero también en seguir mejorando la calidad de las pruebas para que los profesionales tengan a su abasto herramientas de calidad, sin olvidar la importancia del proceso evaluativo. En paralelo también se debe continuar impulsando la diseminación de las distintas acciones llevadas a cabo y de las herramientas que los psicólogos tienen a su alcance. En este sentido las universidades, a través de sus profesores, son agentes fundamentales para llegar a las generaciones más jóvenes.

CONFLICTO DE INTERESES

No existe conflicto de intereses

REFERENCIAS

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). *Estándares para pruebas educativas y psicológicas* (M. Lieve, Trans.). Washington, DC: American Educational Research Association. (Original work published 2014)
- 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, 153-160.
- Borsboom D. y Cramer, A.O.J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91-121. doi: 10.1146/annurev-clinpsy-050212-185608
- Canivez, G. L. (2019). Evidence-based assessment for school psychology: Research, training, and clinical practice. *Contemporary School Psychology*, 23, 194-200. doi: 10.1007/s40688-019-00238-z
- Carretero-Dios, H., y Pérez, C. (2007). Standards for the development and review of instrumental studies: Considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology*, 7, 863-882.
- Chan (2011) Feminising and masculinising primary teaching: a critical examination of the interpretive frameworks of male primary school principals in Hong Kong. *Gender and Education*, 23, 745-759. doi: 10.1080/09540253.2011.611041
- Chernyshenko, O. S. y Stark, S. (2016). Mobile psychological assessment. En F. Drasgow (Ed.), *Technology and testing* (pp. 206-216). Routledge. doi: 10.4324/9781315871493
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. doi: 10.1037/14805-018
- Crothers, L. M., Schmitt, A. J., Hughes, T. L., Lipinski, J., Theodore, L. A., Radliff, K., y Ward, S. (2010). Gender differences in salary in a female dominated profession. *Gender in Management: An International Journal*, 25, 605-626. doi: 10.1108/17542411011081392
- Davis, H. (H.) IV, y Baillie, P. (2019). Cognitive and personality assessment in applied sport psychology: Evidence-based test selection. En M. H. Anshel, T. A. Petrie, y J. A. Steinfeldt (Eds.), *APA handbook of sport and exercise psychology, Vol. 1. Sport psychology* (p. 99-110). American Psychological Association. doi: 10.1037/0000124-000
- Elosua, P. (2017). Avances, proyectos y retos internacionales ligados al uso de tests en psicología. *Estudios de Psicología*, 34, 201-220. doi: 10.1590/1982-02752017000200002
- Elosua, P. (2019). *Los test psicológicos en la práctica profesional*. FOCAD. Consejo General de Psicología de España.
- Elosua, P. (2021). Aplicación remota de test: riesgos y recomendaciones. *Papeles del Psicólogo*, 42, 33-37. doi: 10.23923/pap.psicol2021.2952
- Elosua, P. y Geisinger, K. F. (2016). Cuarta evaluación de test editados en España: forma y fondo. *Papeles del Psicólogo*, 37, 82-88.
- Elosua, P. y Muñiz, J. (2013). Proyectos españoles para una mejora en el uso de los tests. *Psiencia. Latin American Journal of Psychological Science*, 6, 5-9. doi: 10.5872/psiencia/5.2.75
- EFPA (2005). *Meta-Code of ethics*. (www.efpa.eu).
- Evers, A., Muñiz, J., Hagemester, C., Høstmaeligen, A., Lindley, P., Sjöberg, A., y Bartram, D. (2013). Assessing the quality of tests: Revision of the EFPA review model. *Psychothema*, 25, 283-291. doi: 10.7334/psychothema2013.97
- Fonseca-Pedrero, E. (2018). Análisis de redes en psicología. *Papeles del Psicólogo*, 39, 1-12. doi: 10.23923/pap.psicol2018.2852
- Fonseca-Pedrero, E. y Muñiz, J. (2017). Quinta evaluación de test editados en España: mirando hacia atrás, construyendo el futuro. *Papeles del Psicólogo*, 38, 161-166. doi: 10.23923/pap.psicol2017.2844
- Gómez-Sánchez, L. E. (2019). Séptima evaluación de test editados en España. *Papeles del Psicólogo*, 40, 205-210. doi: Séptima evaluación de test editados en España
- Greenberg, M. R., & Schneider, D. F. (1995). Gender differences in risk perception: Effects differ in stressed vs. non stressed environments. *Risk Analysis*, 15, 503-511. doi: 10.1111/j.1539-6924.1995.tb00343.x
- Gustafson, P. E. (1998). Gender Differences in risk perception: Theoretical and methodological perspectives. *Risk analysis*, 18, 805-811. doi: 10.1111/j.1539-6924.1998.tb01123.x
- Haladyna, T. M. y Rodríguez, M. C. (2013). *Developing*



- and validating test items. Routledge. doi: 10.4324/9780203850381
- Hernández, A., Elosua, P., Wechsler, S.M. y Muñiz, J. (2020a) El uso de los Test en España y Latinoamérica: Perspectivas actuales y retos futuros. *Anuario Internacional de Revisiones en Psicología*, 1, 67-94. doi: 10.14635/REVPSY.0.4.
- Hernández, A., Hidalgo, M. D., Hambleton, R. K., y Gomez-Benito, J. (2020b). International test commission guidelines for test adaptation: A criterion checklist. *Psicothema*, 32, 390-398. doi: 10.7334/psicothema2019.306
- Hernández, A., Ponsoda, V., Muñiz, J., Prieto, G. y Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los test utilizados en España. *Papeles del Psicólogo*, 37, 192-197.
- Hernández, A., Tomás, I., Ferreres, A. y Lloret, S. (2015). Tercera evaluación de test editados en España. *Papeles del Psicólogo*, 36, 1-8
- Hidalgo, M. D. y Hernández, A. (2019). Sexta evaluación de test editados en España: resultado e impacto del modelo en docentes y editoriales. *Papeles del Psicólogo*, 40, 21-30. doi: 10.23923/pap.psicol2019.2886
- Irwing, P., Booth, T. y Hughes, D. J. (Eds.) (2018). *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development*. John Wiley & Sons Ltd. doi: 10.1002/9781118489772
- ITC Guidelines for Translating and Adapting Tests (Second Edition) (2017). *International Journal of Testing*, 17, 1-34. (www.InTestCom.org)
- 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, 5802-5805. doi: 10.1073/pnas.1218772110
- Lilienfeld, S., Wood, J., y Garb, H. (2006). Why questionable psychological tests remain popular. *Scientific Review of Alternative Medicine*, 10, 6-15.
- McDonald R. P. (1999). *Test theory: A unified treatment*. Mahwah, Lawrence Erlbaum. doi: 10.4324/9781410601087
- Muñiz, J. y Fernández-Hermida, J.R. (2000). La utilización de los test en España. *Papeles del Psicólogo*, 76, 41-49.
- Muñiz, J. y Fernández-Hermida, J.R. (2010). La opinión de los psicólogos españoles sobre el uso de los test. *Papeles del Psicólogo*, 31, 108-121.
- Muñiz, J., y Fonseca, E. (2017). *Construcción de instrumentos de medida en psicología*. Madrid: FOCAD. Consejo General de Psicología de España
- Muñiz, J. y Fonseca, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31, 7-16. doi: 10.7334/psicothema2018.291
- Muñiz, J., Elosua, P. y Hambleton, R.K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25, 151-157. doi: 10.7334/psicothema2013.24
- Muñiz, J., Hernández, A., y Fernández-Hermida, J. R. (2020). La utilización de los test en España: el punto de vista de los psicólogos. *Papeles del Psicólogo*, 41, 1-15. doi: 10.23923/pap.psicol2020.2921
- Muñiz, J., Fernández-Hermida, J. R., Fonseca, E., Campillo, A. y Peña, E. (2011). Evaluación de test editados en España. *Papeles del Psicólogo*, 32, 113-128.
- Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T. y Hartmann, J. A. (2017). Moving from static to dynamic models of the onset of mental disorder. *JAMA Psychiatry*, 74, 528-534. doi: 10.1001/jamapsychiatry.2017.0001
- Parshall, C. G., Harnes, J. C., Davey, T., y Pashley, P. J. (2010). Innovative Items for Computerized Testing. En W. J. van der Linden, y C. A. W. Glas (Eds.), *Elements of Adaptive Testing, Statistics for Social and Behavioral Sciences* (pp. 215-230). Springer. doi: 10.1007/978-0-387-85461-8
- Prieto, G., y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-72.
- Sanz, S., Luzardo, M., García, C., y Abad, F. J. (2020). Detecting cheating methods on unproctored Internet tests. *Psicothema*, 32, 549-558. doi: 10.7334/psicothema2020.86
- Seelow, D. (2019). The art of assessment: Using game based assessments to disrupt, innovate, reform and transform testing. *Journal of Applied Testing Technology*, 20, 1-16.
- Viladrich, C., Doval, E., Penelo, E., Aliaga, J., Espelt, A., García-Rueda, R., & Angulo-Brunet, A. (2021). Octava evaluación de test editados en España: una experiencia participativa. *Papeles del Psicólogo*, 42, 1-9. doi: 10.23923/pap.psicol2020.2937
- Wan, L., y Henly, G. A. (2012). Measurement properties of two innovative item formats in a computer-based test. *Applied Measurement in Education*, 25, 58-78. doi: 10.1080/08957347.2012.635507