

Cuestiones abiertas en el uso de las nuevas tecnologías en la evaluación psicológica

Open questions in the use of new technologies in psychological assessment

Pablo Santamaría y Fernando Sánchez-Sánchez

Departamento de I+D+i de Hogrefe TEA Ediciones, España

La aplicación de las nuevas tecnologías (NNTT) a la evaluación psicológica puede ampliar enormemente la capacidad de los profesionales para evaluar de forma más precisa y exhaustiva y agilizar los procesos de recogida y análisis de información. Para poder acceder a estas ventajas hay que considerar también los costes, desafíos y amenazas potenciales derivados de la implantación de estos nuevos sistemas de evaluación. Con este objetivo, en este trabajo se abordará en primer lugar en qué grado los profesionales de la psicología, los usuarios y las organizaciones están preparados para adoptar el uso de las NNTT. A continuación, se comentará el impacto que la aplicación de la tecnología está teniendo en los propios instrumentos y en el papel que los profesionales desempeñan. Tras este punto, se cuestionará en qué grado puede darse una cierta fascinación por las NNTT que ciega el análisis de los criterios psicométricos esenciales para cualquier medida. Finalmente, se apuntarán algunos desafíos que los creadores y editores de test deben afrontar durante el desarrollo de instrumentos basados en NNTT.

Palabras clave: TIC, Tecnología, Test, Evaluación.

The application of new technologies to the field of psychological assessment can greatly enhance the ability of professionals to carry out assessments more accurately and comprehensively and to streamline the processes of collecting and analyzing information. In order to access these benefits, the costs, challenges, and potential threats associated with the implementation of new assessment systems must also be considered. To this end, this paper will first address the extent to which psychology professionals, users, and organizations are prepared to adopt the use of new technologies. Then, we will discuss the impact that the application of technology is having on the instruments themselves and on the role that professionals play during the process. After this, we will question the extent to which a certain fascination with technology can blind the analysis of the essential psychometric criteria for any measurement. Finally, we will identify some of the challenges that test developers and publishers must face during the development of instruments based on new technologies.

Key words: ICT, Technology, Test, Assessment.

La implementación de las nuevas tecnologías (NNTT) en el ámbito de la evaluación psicológica abre un nuevo horizonte de posibilidades, recursos y planteamientos hasta hace poco tiempo inimaginables para la mayoría de los profesionales. El desarrollo de novedosas y sofisticadas técnicas de evaluación (p. ej., nuevos tipos de tareas, realidad virtual o interacciones con bots), la posibilidad de recoger grandes cantidades de datos y medidas (p. ej., registros continuos mediante dispositivos electrónicos personales o análisis de redes sociales) y la disponibilidad de nuevas técnicas para extraer, analizar y modelar grandes cantidades de información (big data, inteligencia artificial) han situado a los profesionales de la psicología y a los investigadores ante un nuevo y excitante escenario.

No obstante, cabe preguntarse si las NNTT emergentes acabarán constituyendo un nuevo mundo que rompa con el anterior (disrupción tecnológica) o si, por el contrario, serán solo juguetes, ilusiones o apariencias que acabarán desinflándose con el tiempo (Chamorro et al., 2016).

Tras una pandemia que ha acelerado la implantación y el uso de la tecnología, esta cuestión adquiere una vigencia mayor. La aplicación de las NNTT a la evaluación está llena de interrogantes, con luces prometedoras, pero también con sombras y aspectos inciertos.

En este monográfico se han abordado algunas de estas promesas y el presente artículo tratará de completar esa imagen mostrando, desde una perspectiva aplicada, algunos riesgos y efectos adversos potenciales que se derivan de su uso; el objetivo final sería ayudar a los profesionales a formarse una imagen crítica y compleja de cara a la adopción y uso de las NNTT en la evaluación psicológica.

PROFESIONALES, USUARIOS Y ORGANIZACIONES: ¿ESTAMOS PREPARADOS PARA EL USO DE LAS NNTT EN EVALUACIÓN?

La evaluación psicológica exige de los profesionales el dominio de un conjunto de técnicas específicas, una adecuada formación en el uso de los test y unos conocimientos suficientes sobre psicometría. La evaluación mediante NNTT, además de lo anterior, también requiere de un adecuado conocimiento y manejo de diferentes dispositivos electrónicos y de la tecnología en general (en este monográfico Elo-sua, 2022).

Siguiendo la clasificación de los usuarios de tecnologías en función de su grado de dominio (Rogers, 2003) (innovadores, primeros adoptantes, mayoría temprana, mayoría tardía y rezagados), encuestas sobre uso de NNTT dirigidas a profesionales de la psicología apuntan a que su posición predominante sería de mayoría tardía o de rezagados (Ramos-Álvarez y Rodríguez, 2021). Esto indicaría una relativamente baja adopción en el uso de las NNTT y que la competencia digital de los profesionales de la psicología podría ser un factor limitador para el desarrollo, implantación y buen uso de las nuevas herramientas de evaluación.

Recibido: 1 noviembre 2021 - Aceptado: 21 diciembre 2021

Correspondencia: Pablo Santamaría. Departamento de I+D+i de Hogrefe TEA Ediciones. C/ Cardenal Marcelo Spínola, 50. 28016 Madrid. España. E-mail: pablo.santamaria@teaediciones.com

A esta circunstancia se añade que: a) muchos de los entornos o contextos donde se ejerce la psicología no destacan por un alto nivel de adopción tecnológica, ni en términos de disponibilidad de equipamiento ni en términos de alfabetización digital de las personas a evaluar (p. ej., atención a las personas mayores, a familias en entornos desfavorecidos...); b) la literatura ha reportado la presencia de una relación entre una menor salud mental y un menor nivel socioeconómico (v.g. Reiss, 2013; Reiss et al., 2019) que plantea una potencial limitación al acceso y uso de NNNT; y c) la disponibilidad de equipos e instalaciones adecuadas para poder realizar las evaluaciones con garantías (zonas con cobertura WIFI o móvil, dispositivos actualizados...) supone otra potencial barrera, relacionada con los aspectos económicos y de dotación de los centros.

El coste para adquirir, mantener y actualizar tanto los dispositivos electrónicos (p. ej., tabletas, gafas de realidad virtual...) como las instalaciones (aulas, consultas o despachos con conexiones estables y rápidas) es elevado, y sin duda es mayor que las alternativas de evaluación más tradicionales basadas en papel y lápiz. Por ejemplo, podría mencionarse la desafortunada realidad del ejercicio de la psicología en parte del sector público (hospitales, juzgados, colegios...), un área en la que se realiza una parte importante de las evaluaciones, y que frecuentemente está infra-dotada e infrapresupuestada en material psicológico y tecnológico. A esta lista de potenciales barreras podrían añadirse las referidas a las restricciones de acceso a las redes impuestas por los protocolos de seguridad de los propios centros o el coste derivado del tratamiento informático de datos personales (necesidad de servidores seguros). El coste de adquisición y la rápida obsolescencia de los dispositivos y del software necesario suponen un reto mayúsculo a la hora de una gestión eficiente de los recursos para las instituciones y los profesionales. Este escenario de inversiones costosas con una vida útil corta influye en la implantación de las NNNT aplicadas a la evaluación.

Este conjunto de potenciales barreras afecta de forma relevante al uso que se hace de estas tecnologías. Baste citar como ejemplo que, en el año 2021, el 90% de las aplicaciones del SENA (Fernández-Pinto et al., 2015) o el 85% de las del PAI (Morey, 2013) fueron en formato de papel y lápiz, frente a un 10% y 15% en formato *online* respectivamente. Ambas pruebas son de las más utilizadas en la evaluación psicológica (Muñiz et al., 2020)

La tendencia es similar en el caso de las pruebas de rendimiento óptimo para la evaluación cognitiva basadas en tareas de ejecución. En el caso del Matrices (Matrices, Test de Inteligencia General; Sánchez-Sánchez et al., 2015), en 2021 un 87% de sus aplicaciones fueron realizadas en el formato de papel y lápiz y solo un 13% en el formato TAI (Test Adaptativo Informatizado), a pesar de las indudables ventajas de este último en cuanto a brevedad, precisión y calidad de la medida (Abad et al. 2020; Olea et al., 2010).

Un somero análisis de la relación entre la tecnología y los profesionales de la psicología evidencia un ritmo pausado de transición al uso de las NNNT en evaluación. Como se ha señalado, esto puede estar motivado, entre otras razones, por el nivel de adopción tecnológica tanto de los profesionales como de muchos de los destinatarios de las evaluaciones, así como a los condicionantes económicos y a las particularidades de los contextos donde se ejerce la práctica

profesional. Es previsible que las barreras relacionadas con la adopción tecnológica de los profesionales y usuarios se vayan reduciendo con las nuevas generaciones y que los dispositivos y conexiones sean cada vez más asequibles, con mayor capacidad y aplicaciones. Los profesionales y las instituciones interesadas en aprovechar las ventajas de las NNNT aplicadas a la evaluación deberían dirigir activamente una parte de sus esfuerzos a reducir estos condicionantes básicos de acceso, formación y disponibilidad, así como dar respuesta al desafío que supone a nivel de inversión, renovación y mantenimiento de los equipos y conexiones necesarias.

TEST PSICOLÓGICOS: ¿ESTÁN PREPARADOS PARA LAS NNNT Y PARA UN MUNDO SIN BARRERAS?

Al mismo tiempo que las NNNT facilitan el acceso a gran cantidad de información y servicios relacionados con la evaluación psicológica, plantean amenazas potenciales en relación con la vulnerabilidad de los test y al papel del profesional en el proceso evaluativo.

RESTRICCIÓN DEL ACCESO A LOS TEST PSICOLÓGICOS A PROFESIONALES

Cualquier test psicológico, adecuadamente construido y probado, es una herramienta valiosa en manos de los profesionales que lo sepan utilizar e interpretar adecuadamente. Una correcta evaluación requiere de un profesional cualificado que sea capaz de seleccionar los aspectos a evaluar, los test y técnicas que se deben aplicar y, posteriormente, integrar los resultados para poder llegar a un correcto diagnóstico, a una adecuada formulación del caso, a un informe de orientación o a una propuesta de contratación bien motivada.

Cuando se aplica un test se espera que la persona evaluada obtenga, directa o indirectamente algún beneficio de los resultados y, a la inversa, su uso inadecuado puede acarrear consecuencias negativas; ante esta circunstancia el profesional debe asegurarse de que el proceso de evaluación salvaguarda en todo momento los derechos e intereses de las personas implicadas (AERA et al., 2014).

Por este motivo, al igual que ocurre con la prescripción de fármacos y de pruebas radiológicas en el ámbito de la medicina (i.e., solo accesibles si las prescribe previamente un profesional acreditado), el acceso a los instrumentos de evaluación psicológica ha estado restringido tradicionalmente a los profesionales de la psicología o de disciplinas afines.

Una de las principales razones de la restricción de venta de los test psicológicos es evitar un mal uso que contribuya a la estigmatización o discriminación de las personas evaluadas (p. ej., aplicando etiquetas "diagnósticas" inadecuadas, excluyendo a una persona de un proceso de selección...). La restricción de venta de los test basada en el nivel de cualificación (a, b o c) que se requiere a los profesionales para poder usarlos (AERA et al., 2014) responde a un meditado sistema de control del buen uso de los test como herramientas profesionales que son. Desafortunadamente, muchos de los creadores de herramientas de evaluación basadas en NNNT resultan ajenos al mundo de la psicología y a las implicaciones éticas o deontológicas asociadas, ofreciéndose en muchos casos como una alternativa al profesional para poder obtener una evaluación.

Un buen ejemplo es la proliferación de las evaluaciones psicológicas ofrecidas directamente al usuario final (no al profesional) de for-

ma abierta en páginas webs o en las tiendas de aplicaciones (apps) para su uso tanto por parte del público general (madres, padres, la propia persona evaluada...), como por parte de profesionales de diversos sectores (profesores, educadores ...). Cualquier persona puede encontrar y adquirir herramientas para la evaluación de aspectos del desarrollo, aptitudes cognitivas o rasgos psicopatológicos –y que por tanto podrían considerarse clínicas–. Estas pruebas, promocionadas en ocasiones como “test diagnósticos”, son objeto de campañas de marketing *online* dirigidas específicamente a madres, padres o familiares, que pueden hacer uso de ellas *online* sin participación en el proceso de profesionales de la salud y quedando a su vez fuera de la supervisión de los colegios profesionales.

El problema de esta práctica, no es que se desarrollen y se ofrezcan herramientas diseñadas específicamente para familiares u otros profesionales fuera del ámbito de la psicología, algo que sería positivo y que implicaría un diseño *ex profeso* con ese objetivo, sino que se ofrezca la misma herramienta a todo el espectro posible de usuarios (familiares, profesionales de la psicología, profesionales de otras áreas...), poniendo una herramienta especializada en manos de personas sin la capacidad para interpretar y manejar las informaciones que proporciona.

Como se puede ver, la ausencia de barreras tecnológicas para el acceso a los instrumentos de evaluación por parte del público general no acarrea solo ventajas, también implica abrir la puerta a potenciales malas prácticas que quedan fuera de la supervisión y de la deontología profesional.

VULNERABILIDAD DE LOS TEST Y PROTECCIÓN DE LOS CONTENIDOS

Las restricciones de acceso a los test buscan también proteger las técnicas e instrumentos de evaluación para que puedan ser utilizados con eficacia, puesto que la utilidad de muchos test recae en la relativa novedad de la tarea o de los contenidos. Sirvan de ejemplo las pruebas de una batería de aptitudes intelectuales en un proceso de selección de personal o los estímulos de un test de memoria en un contexto clínico. Si las personas evaluadas tuvieran acceso previo a los mismos podrían entrenarse en las tareas de la batería o memorizar previamente los estímulos de la prueba de memoria, comprometiéndose la eficacia de las evaluaciones posteriores.

Cualquier tipo de exposición pública de los contenidos de un test afecta a su futuro uso, por lo que la necesidad de proteger los materiales y estímulos es un aspecto clave contemplado explícitamente en las normativas o códigos de buenas prácticas profesionales: “Todo tipo de material estrictamente psicológico tanto de evaluación como de intervención o tratamiento, queda reservado al uso de los/as Psicólogos/as, quienes por otra parte, se abstendrán de facilitarlos a otras personas no competentes” (artículo 19 del código deontológico de la psicología).

Las NNTT ofrecen soluciones eficaces –aunque costosas– para proteger los contenidos de los test, como el uso de test adaptativos informatizados basados en amplios bancos de ítems (en este monográfico Abad et al., 2022) o los sistemas de supervisión *online* de la persona que responde al test (*proctoring* en inglés). Pero frente a estas ventajas, también contribuyen a acrecentar los riesgos de exposición de una manera extrema; citamos como ejemplo las webs de

descargas de test o las plataformas de videos *online* en las que estudiantes o profesionales de la psicología suben videos mostrando los procedimientos de aplicación y corrección de numerosos test. Lamentablemente, a menudo en estos videos se muestran todos los ítems y estímulos, las respuestas correctas a cada uno de ellos y cómo obtener determinada puntuación. Si bien estos videos están realizados en su mayoría como recursos didácticos para otros profesionales de la psicología, la ausencia de restricciones en su visualización hace que su contenido esté virtualmente expuesto al mundo entero.

En esta misma línea cabría citarse la publicación de artículos en revistas de libre acceso o los repositorios *online* de tesis doctorales que contienen informaciones críticas de los test que se han utilizado en la investigación (ítems, la clave de corrección...) y acceder incluso a la información sobre las escalas de control y validez que utiliza el test para identificar disimulación, simulación u otros aspectos críticos.

No se ha de olvidar que los test son usados con frecuencia con fines muy sensibles, como en el contexto clínico o forense para tomar decisiones legales (incapacitación, imputabilidad...) o en contexto de selección de personal para elegir las personas idóneas para un cierto puesto. Son múltiples los intereses que pueden existir para acceder a sus ítems y a sus tareas con objetivos poco lícitos, lo que hace esencial una adecuada protección de sus estímulos.

Por lo tanto, resulta necesario resaltar la vulnerabilidad de los test a su exposición a no profesionales mediante las redes sociales e Internet y el negativo efecto que pueden tener a nivel aplicado. La sociedad de la información y el uso de las NNTT hace necesario acentuar la protección de los test.

DISPONIBILIDAD EN CUALQUIER MOMENTO Y EN CUALQUIER LUGAR

Las NNTT pueden contribuir a desdibujar la barrera geográfica de los test. Los servicios e instrumentos de evaluación *online* pueden ser utilizados fácilmente desde cualquier punto del planeta y estar disponibles las 24 horas del día, lo que está acercando la labor de los psicólogos a lugares y contextos antes difícilmente accesibles.

Esta circunstancia hace que pueda convertirse en una posible fuente de malas prácticas cuando determinados instrumentos desarrollados originalmente en una lengua y un país determinado se ofrecen y utilizan en otros países y lenguas diferentes sin seguir un adecuado proceso de adaptación (Muñiz et al., 2013). Es decir: a) mediante traducciones automáticas sin un adecuado proceso de adaptación lingüística y cultural, b) sin una baremación local adecuada, empleando los baremos originales del país donde se desarrolló o un agregado de datos de distintos países (denominados habitualmente como “baremos internacionales”), y c) sin aportar evidencias de validez obtenidas con las nuevas versiones, remitiéndose, a lo sumo, a las evidencias de la versión original del test que no han sido obtenidas ni en el país ni en la lengua con la que se evalúa.

Estas prácticas atentan de forma grave contra los estándares mínimos de calidad y uso de los test (AERA et al., 2014; ITC, 2018), que se deben aplicar tanto a las evaluaciones tradicionales (en papel y lápiz) como en las evaluaciones *online* o con nuevas tecnologías.

CALIDAD CIENTÍFICA DE LA EVALUACIÓN: ¿EVALUAMOS MEJOR MEDIANTE EL USO DE LAS NNTT?

Las cuestiones planteadas en los apartados anteriores tienen un importante impacto en la práctica aplicada de la evaluación psicológica mediante las NNTT y el uso que hacen de ella los profesionales. Pero, sin duda, la cuestión central debería ser la siguiente: ¿El uso de las NNTT aplicadas a la evaluación psicológica aumenta la calidad científica de la misma?

Varios artículos de este monográfico evidencian que claramente esto puede ser así, apuntando a un prometedor aumento de la calidad científica y de la utilidad práctica de las evaluaciones. No obstante, esto no siempre es así y se pueden encontrar múltiples ejemplos de plataformas de evaluación psicológica basadas en NNTT que descuidan de forma dramática los aspectos sustantivos a nivel psicológico y psicométrico. Desde una perspectiva legal, y dada la ausencia de un marco normativo restrictivo, poco se puede hacer para evitar estas deficitarias herramientas. Esta circunstancia hace que la responsabilidad recaiga en la habilidad del profesional para examinar críticamente y cribar la calidad de las herramientas con las que trabaja.

DIFERENCIANDO TEST Y PSEUDOTEST

Existen pocos estudios sobre la capacidad de los profesionales para poder realizar esta labor de criba y de valoración adecuada de la calidad de las herramientas de evaluación psicológica que utilizan NNTT. En un trabajo llevado a cabo con estudiantes graduados en psicología que asistían a diversos máster de especialización ($n = 232$) y con colegiados en ejercicio profesional que asistían a un curso de actualización sobre el buen uso de los test ($n = 73$) se pidió a los participantes que valorasen una plataforma *online* de evaluación psicológica con una excelente y atractiva presentación pero, a su vez, con graves problemas psicométricos (ausencia de baremos españoles, ausencia de evidencias de validez, presentación de datos de fiabilidad como si fueran de validez, traducción y no adaptación, baremos automáticos a partir de los propios casos clínicos evaluados...) (Santamaría, 2020). A los participantes se les pedía que estudiaran detenidamente la información disponible en la plataforma y que valoraran la herramienta como expertos asesores, tras lo cual deberían emitir un juicio sobre si recomendarían su uso o lo desaconsejarían. A pesar de las evidentes carencias, solo rechazaron el uso de la herramienta por su inadecuación un 29% de los graduados en psicología y un 42% de los profesionales colegiados; más de la mitad (55%) de los graduados y un tercio de los profesionales (33%) la valoraron positivamente, y recomendaron su uso, y un porcentaje elevado reconoció no sentirse competente para realizar esa valoración (16% de graduados y 25% de profesionales). Aunque preliminares y meramente exploratorios, estos datos alertan sobre la limitada capacidad de los graduados y de los profesionales para detectar herramientas claramente defectuosas y cómo el uso de la tecnología, de un atractivo diseño y de un adecuado marketing resulta persuasivo en estos casos, incluso sin una mínima calidad psicométrica subyacente.

Esta posible dificultad de los profesionales para valorar críticamente la calidad científica de las pruebas disponibles podría suplirse recurriendo a las valoraciones independientes realizadas por

instituciones profesionales (p. ej., la Comisión de test del COP o instituciones similares en otros países; en este monográfico Hernández et al., 2022). Desgraciadamente, muchas de las plataformas de evaluación basadas en NNTT quedan fuera del ámbito de actuación de estas instituciones –que en el caso de la Comisión de test del COP se ciñe a los test editados en España– y, por tanto, no son valoradas.

Estas limitaciones para poder valorar la calidad científica de algunas herramientas de evaluación que emplean NNTT se ve agravada por la dificultad para acceder a su información técnica. Como en cualquier test de evaluación psicológica, es esencial que los profesionales puedan disponer de la documentación en la que se detalle el proceso de construcción seguido, las evidencias psicométricas de fiabilidad y validez, las muestras de baremación utilizadas para el cálculo de las puntuaciones típicas, así como los procedimientos de aplicación y corrección (AERA et al., 2014). El hecho de que un test haya sido desarrollado con NNTT no lo exime del cumplimiento de los estándares requeridos a cualquier técnica de medición psicológica y que las evidencias disponibles sean expuestas de forma clara para su análisis por parte de los profesionales.

TECNOLOGÍA Y CONOCIMIENTO: UN EQUILIBRIO NECESARIO

La incorporación de las NNTT a la evaluación ha contribuido a la eliminación de muchos límites interprofesionales, haciendo necesario que en el desarrollo de las nuevas herramientas colaboren ingenieros, programadores, psicólogos clínicos, psicómetras y otros especialistas. Este trabajo multidisciplinar responde a la necesidad de asegurar que la medida contemple y recoja los aspectos sustantivos de lo que se desea evaluar y que los procedimientos técnicos para llevar a cabo la evaluación permiten hacerlo con garantías.

La colaboración multidisciplinar es necesaria para evitar caer en la confusión entre el medio (la tecnología) y el fin (la medida psicológica). Tal y como es definida por la RAE, la tecnología consiste en el “conjunto de teorías y de técnicas que permiten el aprovechamiento práctico del conocimiento científico” (RAE, 2014). No puede existir tecnología sin un conocimiento científico en el que basarse para extraer su aprovechamiento práctico. En este sentido, todas las técnicas de inteligencia artificial descansan en la calidad de los datos que se introduzcan, algo que dependerá del conocimiento científico para su correcta selección. Lamentablemente, a menudo el punto de partida es justo el contrario: dado que tenemos la técnica (programación e inteligencia artificial), se puede abordar cualquier asunto, sin necesitar el conocimiento científico que guíe, oriente o enmarque la tecnología; con ello, se sobrestima el valor aislado de la tecnología y se ignora en qué consiste realmente la evaluación psicológica.

La psicología tiene un importante desafío frente a una realidad en la que cada vez más profesionales ajenos a la psicología y vinculados a las NNTT desarrollan herramientas de evaluación psicológica con un limitado conocimiento científico de las variables que pretenden medir y de las cuestiones métricas básicas asociadas. Es esencial que la psicología a nivel social, institucional y profesional sepa reclamar su papel central en estos aspectos como garante de una adecuada calidad de las herramientas y de las evaluaciones, con el objetivo de asegurar una buena medición que evite consecuencias negativas para las personas evaluadas. Es necesaria e imprescindible

ble la colaboración entre el conocimiento científico acumulado en psicología sobre las diversas variables del comportamiento humano y la tecnología que permita un óptimo aprovechamiento del mismo. Sin esta interdisciplinariedad en grado de igualdad, ambas partes estarán perdidas.

LA TECNOLOGÍA POR LA TECNOLOGÍA: DISTINGUIR CUÁNDO Y CÓMO HACER USO DE LA TECNOLOGÍA

El conocimiento acumulado en psicometría sobre buenas prácticas a nivel de desarrollo de herramientas de medición es ingente. Este conocimiento debe ser aplicado también a las herramientas basadas en NNTT, desde la propia construcción del test, a su proceso de validación o al proceso de recogida de muestras de baremación o tipificación.

En la construcción de un test ha de hacerse un uso inteligente de las NNTT, integrando el conocimiento psicológico del área a evaluar. En este sentido, es importante considerar por un lado, su validez incremental, es decir, en qué medida la introducción de la tecnología mejora la calidad, la utilidad y el valor predictivo de la evaluación respecto a otras medidas ya disponibles; y por otro, el coste añadido, tanto de desarrollo como de uso. Desafortunadamente en muchas circunstancias no se lleva a cabo esta evaluación y es frecuente observar un uso de soluciones tecnológicas basadas en el mero hecho de su novedad o sofisticación: la tecnología por la tecnología.

Podemos citar como ejemplo la evaluación de la capacidad atencional en niños y adolescentes mediante tabletas y apps cuyo diseño y presentación resultan especialmente atractivos y estimulantes para esas edades. Paradójicamente, este aspecto lúdico y atractivo puede resultar contraproducente para evaluar adecuadamente casos con sospecha de presentar un trastorno por déficit de atención dado que el carácter atractivo de las tabletas y de los programas puede interferir precisamente en lo que sería el objeto de evaluación: la capacidad del niño/a de sostener su atención en tareas y entornos poco atractivos, poco estimulantes, monótonos y aburridos. Este ejemplo evidencia una clara divergencia entre los patrones que se suelen aplicar en las plataformas basadas en NNTT (entornos muy estimulantes, cambiantes y atractivos basados en la gamificación) y las características necesarias para una buena evaluación de la atención sostenida basados en el conocimiento del área (entornos poco estimulantes, poco cambiantes, poco atractivos y demandantes).

Si bien el futuro de la aplicación de las NNTT en la evaluación mediante el uso de tabletas y apps basadas en la gamificación es muy prometedor, la tecnología por la tecnología puede resultar descabzada si no se tienen en cuenta en profundidad los aspectos psicológicos que se están evaluando así como sus implicaciones.

MUESTRAS DE TIPIFICACIÓN Y TÉCNICAS DE BAREMACIÓN

Otro aspecto central en la construcción y el uso de test es el relativo a la obtención de las muestras de tipificación y a las técnicas de baremación empleadas, siendo esta fase de los estudios de desarrollo probablemente una de las más laboriosas y costosas. Las facilidades que ofrecen las NNTT en los procesos de recogida de muestras son enormes en comparación con los procedimientos más tradicionales en los que un profesional debe aplicar de forma pre-

sencial las pruebas tras haber seleccionado cuidadosamente a cada una de las personas que deben ser evaluadas para la muestra de tipificación. Existe un amplio corpus de conocimiento sobre los criterios a cumplir por las muestras de baremación en cuanto a representatividad. Estos criterios son trasladables y aplicables a los tests basados en NNTT, que por su propia naturaleza pueden mostrarse vulnerables a esta circunstancia. Por ejemplo, las recogidas de muestras a partir de reenvíos de mensajes por mensajería instantánea (v.g., *whatsapp*), correo electrónico o redes sociales, pueden adolecer de ausencia de control sobre las personas que responden y sobre las condiciones en que lo hacen (con qué grado de atención, en qué contexto...) pudiendo generar un marcado sesgo de selección. Desde otra perspectiva, las muestras de baremación recogidas en aplicaciones como la *App Store* de *Iphone* por ejemplo, podrían ser el origen de sesgo de selección debido al perfil sociodemográfico de las personas que tienen acceso a estos costosos dispositivos. Con relación a este punto, es importante tener en cuenta el sesgo asociado a las diferencias en el perfil de personas que tienen facilidad para el uso y acceso a la tecnología frente a aquellas que tienen dificultades, y que muy improbablemente participarán en este tipo de "estudios" (p. ej., mediante formularios on-line o test alojados en apps o webs). El hecho de poder recoger grandes cantidades de casos en poco tiempo no puede hacer olvidar la importancia de la calidad frente a la cantidad de las muestras.

Una cuestión más polémica sería el uso de empresas de encuestas on-line para recoger las muestras de tipificación de instrumentos de evaluación (p. ej., *Mturk*). Si bien su uso puede abaratar y agilizar enormemente los procesos de tipificación, no están exentas de críticas que cuestionan los posibles sesgos que pueden cometerse tanto por el perfil de las personas que habitualmente participan en estos procesos como por posibles alteraciones o falseamientos al mismo (v.g., uso de VPN para simular que la persona que está respondiendo se encuentra en España, cuando en realidad está respondiendo desde un país de Latinoamérica).

Si bien los problemas relacionados con el sesgo de selección no son nuevos para la psicometría, parece que no se le presta la debida atención cuando se trata de muestreos que utilizan NNTT, tal vez debido a: a) un desconocimiento por parte de los profesionales que las implementan (con escasa formación en psicometría), b) a cierto efecto halo de eficiencia y precisión con relación a las NNTT, c) el efecto "caja negra" de las NNTT que hace menos evidentes los procesos y menos obvios los posibles sesgos cometidos, o d) la tendencia a priorizar las muestras grandes (el tamaño) frente a su representatividad (su calidad), confiando en que los grandes números podrán compensar los sesgos que se cometan.

En síntesis, un adecuado uso de la tecnología en la evaluación requiere la integración de un profundo conocimiento de las NNTT –de sus ventajas, inconvenientes y posibilidades–, de los aspectos psicológicos sustantivos de las áreas que se pretenden medir y las cuestiones psicométricas asociadas. La ausencia de los aspectos sustantivos psicológicos y psicométricos frecuentemente provoca una tecnología de medición hueca.

Sea cual fuera la tecnología utilizada (clásicas o nuevas), las cuestiones esenciales y subyacentes que se requieren en una buena medida siguen siendo las mismas: una adecuada sistematización de la

evaluación, una muestra representativa para sus baremos y unas evidencias solventes en relación con su fiabilidad (consistencia interna, temporal e inter-jueces) y, por supuesto, con su validez. La validación de un instrumento –esto es, aportar evidencias teóricas y empíricas sobre qué se está midiendo– no es una cuestión accesoria, más bien es el aspecto central del test. Independientemente de que se esté hablando de un test de papel y lápiz o de un test basado en NNNT, la esencia del test y de su uso aplicado reside en el grado en que existen evidencias empíricas que respaldan las inferencias que se extraen de sus puntuaciones.

IMPLICACIONES DE LAS NNNT EN LA INDUSTRIA DE LOS TEST

Uno de los actores implicados de forma preminente en la aplicación de las NNNT a la evaluación psicológica son los creadores y editores de test, que se enfrentan a importantes desafíos en la implementación de las mismas en los instrumentos de evaluación. Uno de ellos viene definido por los diferentes horizontes temporales y ritmos de desarrollo de las NNNT y de la industria de los test.

En cierto sentido, ambos resultan contrapuestos; el horizonte temporal de las NNNT presenta un ciclo de vida rápido y efímero, que busca de manera persistente la mejora y la innovación como aspectos esenciales y definitorios (una mejor resolución, una conexión más rápida, una nueva característica que atraiga a los consumidores...). Por el contrario, el marco y el horizonte temporal de la creación de los instrumentos de evaluación psicológica busca la sistematización, la estandarización y la estabilidad en la medida. El objetivo es reducir el efecto de variables ajenas (modificaciones en las instrucciones, en la manera de responder...), de forma que el procedimiento esté cerrado y sea sistemático para que cualquier variación que se produzca tenga que ver con la persona evaluada (y su nivel en la variable medida); en este sentido los cambios en el formato pueden ser un problema para la medida.

Para ilustrar este punto, pensemos en las evaluaciones realizadas mediante ordenador. Desde las primeras propuestas, la evolución de los equipos informáticos ha permitido ofrecer a los usuarios nuevas formas de interactuar con los dispositivos (en este monográfico Elo-sua, 2022). Si bien originalmente las respuestas se indicaban con el teclado o con el ratón, pronto surgieron los *touchpad* en los portátiles, posteriormente las pantallas táctiles y actualmente el reconocimiento de voz. Desde el punto de vista de las NNNT el cambio es algo positivo y deseable y se busca constantemente la innovación y el desarrollo de formas novedosas de interactuar con los dispositivos (p. ej., el seguimiento de la mirada con las cámaras frontales de los ordenadores o teléfonos). La flexibilidad y el cambio son las señas de identidad de las NNNT. Ahora bien, la modificación de la conducta de respuesta puede ser un problema potencial para los test y su sistematización. Por ejemplo, los tiempos de respuesta no son los mismos con una pantalla táctil, con un ratón, con *touchpad* o con la voz. Y aplicar la prueba con uno u otro sistema (o con uno u otro tamaño de pantalla) afecta al propio proceso de evaluación (tamaño de estímulos, distractores, tiempos entre estímulos...) y como consecuencia a los baremos de referencia. Esto es lo que hace que en los test se requiera una uniformidad en los procedimientos de aplicación y que habitualmente se restrinja a una única manera de indicar las

respuestas para que puedan ser comparadas con los datos de baremación. En este sentido, la estandarización –entendida como la estabilidad, la uniformidad y la sistematicidad de las medidas– es seña de identidad de los test psicológicos.

Por otra parte, la construcción de test requiere una inversión económica de varios años para sistematizar la medida, estandarizarla, tipificarla y validarla. Crear un adecuado respaldo empírico a las inferencias que se realicen a partir de las mediciones requiere un extenso periodo de tiempo y recursos. Postular inferencias a partir de las puntuaciones de un test sobre los posibles rasgos de autismo de una persona, sobre su capacidad para portar armas, sobre su grado de psicopatía o sobre su competencia como madre o padre son cuestiones complejas que requieren un extenso y complejo proceso de validación de varios años. Esta circunstancia se acentúa en el caso de nuevas medidas basadas en NNNT dado que, precisamente por su carácter novedoso e innovador, requieren de más investigación y evidencias que las respalden ante la relativa ausencia de literatura previa al respecto.

La industria de los test busca, en cierta medida, lo contrario a la industria de las NNNT: ciclos relativamente largos de vida de los test que permitan la acumulación de evidencias acerca de la validez de los usos de la prueba en diferentes poblaciones y con diferentes propósitos.

Este contraste entre los ciclos de vida de ambos, NNNT y test, supone un desafío continuo. Así, por ejemplo, el plazo que requiere el desarrollo, tipificación y validación de un test suele oscilar entre 3 y 5 años. En ese lapso, seguramente se producirán cambios en los sistemas operativos, en los lenguajes de programación y aparecerán o desaparecerán determinados programas o dispositivos (véanse los casos relativos a la desaparición de la tecnología Flash, a la consola Kinect o a las gafas de realidad virtual...). A menudo, estos cambios durante el desarrollo de test basados en NNNT provocan que, una vez ya finalizado, haya que reiniciar el proyecto para adaptarlo a la nueva tecnología y evitar problemas de compatibilidad o que el diseño quede obsoleto. Un problema aún más grave es la dependencia existente con respecto a los proveedores de la tecnología que se emplee (ya sea un dispositivo, un lenguaje de programación...) y que pueden variar sus condiciones a lo largo del tiempo.

Todo esto supone un reto muy importante a nivel económico que afecta al desarrollo de los test, ya que requiere no solo una inversión adicional durante la propia construcción del test, sino también en su mantenimiento, y afecta a su potencial vida útil, resultando en algunos casos incierta.

A todo lo anterior se suma que el número de usuarios de los test es infinitamente menor que el de usuarios de NNNT en general, lo que complica aún más la ecuación para la creación de test basados en NNNT: un elevado coste de desarrollo para una vida del test potencialmente más corta (por la rápida obsolescencia de las tecnologías), con un mercado reducido de potenciales compradores (a menudo auto restringido a profesionales como se ha comentado) y con un coste de mantenimiento y actualización muy superior a las versiones en papel y lápiz.

Sin duda, esta combinación de factores hacen complicada la tarea de desarrollo de test con NNNT, por lo que a veces buscan vías de solución poco adecuadas como: a) tratar de aumentar indiscriminadamente el número de potenciales usuarios (ampliándolo al público

general o a todo tipo de profesionales), b) intentar ahorrar los costes de los estudios controlados de tipificación y validación (indispensables para realizar los adecuados análisis psicométricos y para la obtención de muestras representativas); o c) en una vía intermedia aumentar los potenciales clientes “exportando” los test a otros países sin realizar los adecuados estudios de validación o tipificación para esas nuevas lenguas. Estas prácticas suponen una de las principales amenazas al buen uso de las NNTT en la evaluación.

A MODO DE RESUMEN

Las NNTT revolucionarán la forma en que concebimos la evaluación psicológica en los próximos años. Diversos aspectos tratados en otros artículos de este monográfico ilustran cómo supondrán nuevas vías de información al mundo psicológico que serán extremadamente valiosas.

No obstante, el ritmo de esta transición hacia el uso de las NNTT en evaluación parece aún pausado, probablemente condicionado, entre otras razones, por el nivel de adopción tecnológica tanto de los profesionales y los destinatarios de las evaluaciones como por los condicionantes económicos y por las particularidades de los contextos donde se ejerce la práctica profesional (colegios, hospitales, juzgados, centros de salud mental, gabinetes privados...).

Esta revolución requiere de una adecuada integración de tecnología y psicología, algo que, a nuestro juicio, todavía está lejos de ser una realidad. Por una parte, la tecnología sin un adecuado conocimiento científico en el que apoyarse resulta descabezada y desaprovecha el gran bagaje de la psicología y la psicometría a la hora de medir aspectos cognitivos, emocionales o conductuales, omitiendo cuestiones relevantes e introduciendo sesgos evitables en la obtención de los datos y en la generación de resultados, llevando a un mal uso con consecuencias negativas poco deseables.

La psicología y la psicometría corren el riesgo de quedarse relegadas de un futuro donde las NNTT abarcarán todos los aspectos de la vida del ser humano; para ello, han de acelerar el ritmo de adopción tecnológica, reivindicar su papel central como expertos en evaluación psicológica y alertar de las malas prácticas que suponen omitir el cumplimiento de los estándares psicométricos en cuanto a lo que significa una adecuada adaptación, baremación o validación que fundamente sólidamente las inferencias a extraer. Todo ello supone responder a muy diferentes desafíos, tal y como se han planteado en el artículo, tanto a nivel de formación como de competencia tecnológica, encontrando el equilibrio entre transparencia y protección, entre estabilidad y cambio, entre sistematización e innovación y entre los distintos tiempos y horizontes de tecnología y psicología en el buen uso de las herramientas de evaluación.

Del satisfactorio encuentro de ambas, tecnología y psicología, depende el buen curso en la evaluación psicológica en los próximos años.

CONFLICTO DE INTERESES

No existe conflicto de intereses.

REFERENCIAS

Abad, F. J., Sánchez-Sánchez, F., y Santamaría, P. (2020). *MATRI-CES-TAI. Test Adaptativo de Inteligencia General*. TEA Ediciones.

Abad, F. J., Schames, R., Sorrel, M., Nájera, P., García-Garzón, E., Garrido, L. E., y Jiménez, M. (2022). Construyendo tests adapta-

tivos de elección forzosa “on the fly” para la medición de la personalidad. *Papeles del Psicólogo*, 43(1), 29-35. <https://doi.org/10.23923/pap.psicol.2982>

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Psychological Association.

Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., y Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? *Industrial and Organizational Psychology*, 9(3), 621-640. <https://doi.org/10.1017/iop.2016.6>

Elosua, P. (2022). Impacto de la TIC en el entorno evaluativo. Innovaciones al servicio de la mejora continua. *Papeles del Psicólogo*, 43(1), 3-11. <https://doi.org/10.23923/pap.psicol.2985>

Fernández-Pinto, I., Santamaría, P., Sánchez-Sánchez, F., Carrasco, M. A., y Del Barrio, V. (2015). *SENA. Sistema de Evaluación de Niños y Adolescentes*. TEA Ediciones.

Hernández, A., Elosua, P., Fernández-Hermida, J. R., y Muñiz, J. (2022). Comisión de Test: Veinticinco años velando por la calidad de los test. *Papeles del Psicólogo*, 43(1), 55-62. <https://doi.org/10.23923/pap.psicol.2978>

International Test Commission (2018). Guidelines for translating and adapting tests (Second edition). *International Journal of Testing*, 18(2), 101-134. DOI:10.1080/15305058.2017.1398166

Morey, L. C. (2013). *PAI. Inventario de Evaluación de la Personalidad* (Ortiz-Tallo, M. Santamaría, P., Cardenal, V y Sánchez, M. P., adaptadores). TEA Ediciones.

Muñiz, J., Elosua, P. y Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: Segunda edición [International Test Commission Guidelines for test translation and adaptation: Second edition]. *Psicothema*, 25(2), 151-157.

Muñiz, J., Hernández, A., y Fernández-Hermida, J. R. (2020). Utilización de los test en España: El punto de vista de los psicólogos. *Papeles del Psicólogo*, 41(1), 1-15. <https://doi.org/10.23923/pap.psicol2020.2921>

Olea, J., Abad, F. J., y Barrada, J. R. (2010). Tests informatizados y otros nuevos tipos de tests. *Papeles del Psicólogo*, 31(1), 97-107.

Ramos-Álvarez, R. y Rodríguez, M. (9-11 de julio de 2021). *Evolución del uso de las tecnologías en psicología en España: 2013-2020*. V Congreso Nacional de Psicología. España.

Real Academia Española (2014). *Diccionario de la lengua española* (Vol. 23). Autor.

Reiss, F. (2013). Socioeconomic inequalities and mental health problems in children and adolescents: A systematic review. *Social science & Medicine*, 90, 24-31. <https://doi.org/10.1016/j.socscimed.2013.04.026>

Reiss, F., Meyrose, A. K., Otto, C., Lampert, T., Klasen, F., y Ravens-Sieberer, U. (2019). Socioeconomic status, stressful life situations and mental health problems in children and adolescents: Results of the German BELLA cohort-study. *PLoS one*, 14(3), e0213700. DOI: 10.1371/journal.pone.0213700

Rogers, E. (2003). *Diffusion of innovations* (6ª ed.). Free Press.

Sánchez-Sánchez, F., Santamaría, P., y Abad, F. J. (2015). *MATRI-CES. Test de Inteligencia General*. TEA Ediciones.

Santamaría, P. (12 de mayo de 2020). ¿Es oro todo lo que reluce en evaluación psicológica con nuevas tecnologías? I Jornada de Actualización en Evaluación Psicológica. España.