

Impacto de la TIC en el entorno evaluativo. Innovaciones al servicio de la mejora continua

Impact of ICT on the assessment environment. Innovations for the purpose of continuous improvement

Paula Elosua
Universidad del País Vasco

La tecnología digital ya forma parte de las prácticas relacionadas con el uso de test, y de la evaluación psicológica y educativa; la forma en que se diseñan test, se recogen datos, o los modelos para su análisis han evolucionado. Como consecuencia, nos enfrentamos a retos como la formación del profesional, la colaboración con otras áreas, la expansión del concepto de test, y la deliberación sobre cuestiones éticas y legales. En este entorno dinámico, la experiencia y el conocimiento relacionado con la medición en psicología nos arrojan la capacidad y oportunidad de integrar y expandir los pilares en los que se asienta las buenas prácticas relacionadas con el uso de test: fiabilidad, validez y uso ético. En este trabajo, repasamos el desarrollo de las tecnologías de la información y comunicación, mostrando su impacto en la evaluación psicológica y educativa, con el objetivo de exponer el abanico de innovaciones que marcarán el desarrollo de la investigación y la práctica profesional los próximos años.

Palabras clave: TIC, Tecnología, Test, Evaluación.

Digital technology is already part of the practices related to test use as well as psychological and educational assessment; test design, data collection, and psychometric models have all evolved. As a consequence, we face a number of challenges such as professional training, collaboration with other areas, expansion of the testing concept, and deliberation on ethical and legal issues. In this dynamic environment, the experience and knowledge related to psychological measurement give us the capacity and opportunity to integrate and expand the pillars on which good test use practices are based: reliability, validity, and ethical use. In this paper, we review the development of the information and communication technologies, showing their impact on psychological and educational assessment, with the aim of exposing some innovations that will mark the development of research and professional practice in the coming years.

Key words: ICT, Technology, Test, Assessment.

El impacto de las tecnologías de la información y comunicación (TIC) sobre la construcción y el uso de test está transformando el entorno evaluativo. El test, como instrumento de medida al servicio del profesional, en cualquiera de los ámbitos de la psicología, educación y sanidad, y la evaluación, como actividad compleja que integra conocimiento, juicio clínico, información fiable y medidas psicométricas (American Psychological Association (APA, 2020)), se han abierto a las posibilidades ofrecidas por las TIC. Los escenarios virtuales interactivos, los nuevos modelos psicométricos, la utilización de dispositivos móviles, la recogida de datos de proceso (*log data*) o el elevado volumen de datos accesibles desde fuentes diversas (*big data*) han modificado el modelo estático clásico de la medición con test de lápiz/papel. La tecnología aporta la posibilidad de aplicar test en formatos digitales o de corregir ensayos de forma automatizada, pero además ha venido acompañada de nuevos constructos relacionados con las competencias en el manejo de las TIC o de nuevas formas de medir las habilidades blandas como liderazgo, comunicación, pensamiento crítico, aprendizaje colaborativo o trabajo en equipo, cada vez más relevantes en los medios educativo, laboral y social (Grundke et al., 2017; OECD, 2016).

Recibido: 1 noviembre 2021 - Aceptado: 21 diciembre 2021
Correspondencia: Paula Elosua. Universidad del País Vasco. Avda. de Tolosa, 70. 20018 San Sebastián. España.
E-mail: paula.elosua@ehu.es

Este trabajo ha sido parcialmente financiado por el Ministerio de Ciencia e Innovación (PID2019-103859RB-100).

El efecto del desarrollo tecnológico sobre la evaluación puede describirse como un proceso progresivo de asimilación que afecta a diferentes áreas relacionados con el uso de pruebas. Desde la utilización de un ordenador como mero soporte para la administración de un test de lápiz/papel (Elosua, 2021), hasta la integración de la tecnología en modelos sustantivos que facilitan la medición de las nuevas variables, o el nuevo paradigma de investigación guiado por datos (*data driven*), la tecnología digital ha impactado en todas las fases que acompañan la construcción y uso de test. En el ámbito de la evaluación educativa, por ejemplo, se han diferenciado tres estadios que marcan el impacto de las TIC (Bennet, 2015); en un primer momento se desarrolla la infraestructura y se hace uso de la tecnología para administrar los test tradicionales en un nuevo soporte; sobre esta infraestructura y aprovechando las ventajas de la utilización del ordenador, se construyen los test adaptativos (Wainer et al., 2000; Weiss, 1982), aparecen los nuevos formatos de ítem (multimedia, respuestas construidas cortas, tareas de rendimiento estático como ensayos) y la generación automática de ítems, o la automatización de la corrección de test; finalmente, en el estadio más avanzado la tecnología se integra de tal modo en el proceso evaluativo que forma parte del diseño de pruebas a partir de modelos cognitivos que haciendo uso de simulaciones y tareas interactivas reproducen entornos "reales" favorecedores de una evaluación auténtica.

En este trabajo repasamos las innovaciones relacionadas con las TIC que están teniendo un mayor impacto en el campo de la evaluación ligada al uso de test. El objetivo es apuntar aquellos aspectos que, desde nuestra perspectiva, están teniendo una mayor relevancia.

cia en la construcción y uso de los test; con esa finalidad y con el propósito de ofrecer un panorama general comenzamos con unas breves notas para definir y ubicar los elementos del campo de las TIC que más han podido influir en la situación actual.

DESARROLLO DE LA TECNOLOGÍA DIGITAL

Una de las características más salientes de la tercera revolución industrial, y que ha dado origen a la sociedad de la información (Rifkin, 2011) ha sido el desarrollo de la tecnología digital. Los mayores hitos están relacionados con la aparición y desarrollo de elementos como los ordenadores personales o internet, que han transformado la forma de comunicación humana.

Ordenadores personales

La idea del ordenador personal surgió hace aproximadamente 50 años; compañías como IBM, HP y Apple presentaron sus primeras máquinas en los años 80; el IBM *personal computer* apareció en 1981, HP comercializó su primer ordenador personal en el año 1980, y *The Apple Macintosh* se expuso en un anuncio durante la retransmisión de la *Super Bowl* el año 1984. La aparición del ordenador personal situó la capacidad de computación directamente en manos de millones de personas, y cambió con ello el paradigma anterior en el cual los privilegios de acceso de los usuarios individuales estaban estrictamente controlados por los administradores del sistema. La introducción posterior de la arquitectura cliente-servidor condujo además a la vinculación de los ordenadores personales (clientes) a computadoras (servidores) que permitían compartir grandes cantidades de datos (O'Regan, 2021).

Internet y WWW

El inicio formal de internet suele situarse en los años 60, en el momento en el que el departamento de defensa de Estados Unidos configuró la red ARPANET (*Advanced Research Project Agency*) que conectaba 4 centros de investigación norteamericanos. Esta primera red junto con la invención de la conmutación de paquetes (*packet switching*) que permitía aumentar la velocidad y efectividad del teleproceso, y la aparición del código ASCII creado por el Comité Estadounidense de Estándares en 1963 (*American Standard Code for Information Interchange*) dieron lugar a que en los años 70 se pudieran conectar alrededor de 30 instituciones.

Internet y *World Wide Web* (WWW) no son sinónimos. La aparición de la WWW, concebida por Tim Berners-Lee, se fija en el *Conseil Européen pour la Recherche Nucléaire* (CERN, Organización

Europea para la Investigación Nuclear) el año 1990. La WWW se fundamenta sobre avances anteriores relacionados con el desarrollo del hipertexto, del ratón y de internet. Berners-Lee creó un sistema que asignaba a cada página web una dirección estándar, la URL (*Universal Resource Locator*) a la cual se accedía por medio del protocolo HTTP (*Hypertext Transfer Protocol*), y programó un explorador que permitía solicitar, recuperar y visualizar páginas web en el PC local entre ordenadores conectados a través de internet. La invención de la *World Wide Web* fue un hito revolucionario. Transformó internet de un uso principalmente académico hasta donde ahora es una parte integral de la vida de las personas. Los usuarios podemos navegar por la web, hipervincular entre millones de ordenadores, y compartir información de forma rápida y sencilla. De acuerdo a *internetworld stats* (<https://www.internetworldstats.com>) se estima que la penetración de internet es a día de hoy del 65,6%; es decir, sobre una población mundial estimada en 7.8751765.587 personas, 5.1681780.607 utilizan internet.

Redes sociales

La ubicuidad de los teléfonos inteligentes o simplemente móviles, ha impulsado el crecimiento de las redes sociales, que a su vez han transformado las formas de comunicación humana. Las redes sociales son aplicaciones que permiten la creación e intercambio de contenido generado por el usuario; pueden ser horizontales o generalistas y verticales o especializadas en función de que tengan temáticas o actividades definidas; *LinkedIn* o *Researchgate* por ejemplo, serían redes verticales profesionales y de investigación.

Analizando el tipo y calidad de la información compartida en redes sociales, estas podrían clasificarse combinando modelos teóricos del campo de las redes con modelos sobre procesos sociales (Kaplan y Haenlein, 2010); un eje vendría definido por el nivel de presencia social y la riqueza del medio, es decir, por la cantidad de información que puede transmitirse en un intervalo de tiempo, y el segundo lo marcarían la auto-representación o deseo de control sobre las impresiones que se causan en el otro, y la revelación consciente o inconsciente de información personal (*self-disclosure*). (Tabla 2).

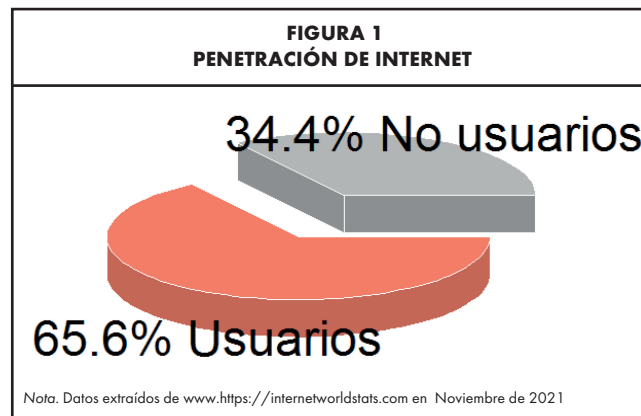
Inteligencia artificial

En el límite entre la tercera y la cuarta revolución industrial, definida como la revolución más rápida en el ámbito de la tecnología, y no comparable con nada en la historia de la humanidad en términos

**TABLA 1
TECNOLOGÍA DIGITAL**

| Año | Referencia |
|-----------|--|
| 1980 | Ordenadores personales |
| 1993 | <i>World Wide Web</i> (free) |
| 2003-2006 | <i>LinkedIn</i> , <i>Facebook</i> , <i>Twitter</i> |
| 2005 | <i>Big data</i> |
| 2013 | Internet de las cosas |
| 2020 | <i>Big data</i> + Inteligencia artificial |

**FIGURA 1
PENETRACIÓN DE INTERNET**



Nota. Datos extraídos de [www.internetworldstats.com](https://internetworldstats.com) en Noviembre de 2021

de impacto (Lee, 2018) la inteligencia artificial (IA) se erige en objeto de estudio y análisis desde ámbitos científicos, técnicos, políticos y sociales. La IA se ha nutrido de disciplinas tan variadas como la filosofía, las matemáticas, la economía, las neurociencias, la psicología, la ingeniería computacional, la cibernética y la lingüística. Hoy encontramos aplicaciones de IA en motores de búsqueda, diagnósticos médicos, reconocimiento de voz, control de robots, búsqueda web, publicidad e incluso juguetes.

Mencionar la inteligencia artificial lleva indudablemente a citar al matemático Alan Turing, que con sus trabajos a principios de la década de 1950 contribuyó al debate sobre máquinas pensantes, conciencia e inteligencia. Creó el famoso test de Turing para evaluar la consciencia de una máquina; en este experimento un juez evalúa una conversación en lenguaje natural entre dos partes, un humano y una máquina con el propósito de discernir cuál es la máquina. En esta línea de trabajo el alemán Joseph Weizenbaum programó el año 1966 en el *Massachusetts Institute of Technology* (MIT) el programa ELIZA; se trata de un *bot* que emulando a un psicoterapeuta de la escuela de Rogers interactúa con una persona sentada frente a una máquina de escribir (similar a un chat en línea). El software opera descomponiendo la entrada del usuario en sus partes constituyentes de voz y, a continuación, vuelve a escribirlas de un modo que aparenta un diálogo fluido. El autor se sorprendió al comprobar que muchos usuarios pensaban que el programa tenía una comprensión real. Esta observación le llevó a reflexionar sobre la ética y las implicaciones del campo de la inteligencia artificial (Weizenbaum, 1976), cuestión que impregna el desarrollo de la IA. El lector puede “conversar” con una versión en castellano de ELIZA en <http://deixi-labs.com/eliza.html>

El término inteligencia artificial aparece en un proyecto de investigación de verano de Dartmouth escrito por el informático John McCarthy el año 1955. Hoy, una de las definiciones más aceptadas de IA es la propuesta por Russell y Norvig (2021), según la cual la IA se centra en el estudio y construcción de agentes que hacen lo correcto, siendo lo correcto el objetivo marcado para el agente, y definiendo agente como algo que percibe su medio a través de sensores. En términos estadísticos simples lo correcto podría ser aquella decisión (estimación) que minimiza la función de pérdida. Esta definición ha sido aceptada por la Unión Europea que la reformula como: “El software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el Anexo I¹ y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa” (Comisión Europea, 2021).

El campo de la IA es amplio y encontramos áreas especializadas en computación, aprendizaje automático (*machine learning*), procesamiento de lenguaje natural, visión computacional y robótica. Además, y dadas las implicaciones de la IA, el capítulo dedicado a los aspectos éticos y legales relacionados con su implementación es una constante.

El dato

La disponibilidad de información proveniente de las redes sociales o de aparatos, sensores y servicios que capturan datos a su alrededor (internet de las cosas) ha causado un cambio de paradigma en la investigación, en el mercado, y en la industria. El origen del término *big data* se sitúa en el año 2005, y es atribuido a Roger Mouga-

TABLA 2
CLASIFICACIÓN DE LAS REDES SOCIALES

| Dimensión social | | | | | |
|--------------------------------------|----------------------|------|---|--|-----------------------------------|
| Presencia social / Riqueza del medio | | | | | |
| | | Baja | Media | Alta | |
| Dimensión de red | Auto-representación | Alta | Blogs Twitter | Redes sociales Facebook LinkedIn | Espacios virtuales Second life |
| | Información personal | Baja | Proyectos colaborativos Wikipedia BookCrossing Flick Slideshare | Contenidos Youtube | Juegos World of Warcraft |

Nota. Adaptado de Kaplan y Haenlein (2010)

¹ Anexo I. Estrategias de aprendizaje automático, incluidos el aprendizaje supervisado, el no supervisado y el realizado por refuerzo, que emplean una amplia variedad de métodos, entre ellos el aprendizaje profundo. Estrategias basadas en la lógica y el conocimiento, especialmente la representación del conocimiento, la programación (lógica) inductiva, las bases de conocimiento, los motores de inferencia y deducción, los sistemas expertos y de razonamiento (simbólico). Estrategias estadísticas, estimación bayesiana, métodos de búsqueda y optimización.

las. Hoy *big data* hace referencia a la información caracterizada por su **V**olumen, **V**elocidad y **V**ariación, y que requiere de métodos analíticos específicos para su tratamiento (De Mauro et al., 2016; Zicari, 2013). Las tres Vs características del *big data* han ido ampliándose con más sustantivos que intentan reflejar con mayor fidelidad su naturaleza; así encontramos determinantes como Valor, Veracidad, Visualización, Volatilidad, Validez y Viabilidad (Maté-Jimenez, 2014). El *big data* opera básicamente con modelos de aprendizaje automático (*machine learning*) que buscan patrones y relaciones para la clasificación y la predicción; es una aproximación a la investigación centrada en el dato, exploratoria, que aplica técnicas analíticas en la búsqueda de patrones en los datos. En contraposición, el acercamiento clásico, deductivo, centrado en la teoría, se asocia a la formulación y contraste de hipótesis sobre una muestra de datos. Entre ambas perspectivas que pueden definir un continuo, cada vez es mayor la postura que aboga por la complementariedad en pro del avance del conocimiento y de la productividad científica (Maass et al., 2018).

IMPACTO DE LAS TIC EN LOS TEST COMO INSTRUMENTOS DE EVALUACIÓN

Las innovaciones tecnológicas descritas en los puntos anteriores y algunas más que por limitaciones de espacio no podemos tratar (potencia computacional, robótica, desarrollo de software...) han impactado en todo el proceso relacionado con el uso de test; la construcción de ítems, el soporte y aplicación del test, la corrección de pruebas, la recogida de datos adicionales, los modelos psicométricos para tratarlos, el uso del dato como fuente de información... son áreas de trabajo e investigación que adecúan el test tradicional y las prácticas del siglo XX a las demandas sociales del siglo XXI.

Test adaptativos

El desarrollo de los test adaptativos informatizados (TAI) y la teoría psicométrica asociada a ellos construida sobre las bases de la teoría de respuesta al ítem, han facilitado la construcción de test personalizados y la comparabilidad de puntuaciones entre personas que reci-

ben ítems diferentes (Hambleton, 2006). El test adaptativo es un sistema compuesto por un banco de ítems con características psicométricas conocidas, y algoritmos de selección que en función del nivel de habilidad que se estima tras cada respuesta, eligen el ítem almacenado en el banco que por sus propiedades (parámetros) va a ofrecer la máxima información sobre el nivel de competencia de la persona evaluada (Olea et al., 2010; van der Linden y Glas, 2000).

Generación automatizada de ítems

La construcción automatizada de ítems o generación automatizada de ítems (GAI) se presenta como una solución a la creciente demanda de un mercado caracterizado por la implementación de test adaptativos computerizados, la aplicación de test por internet, la transparencia exigida a los proyectos evaluativos o el incremento de programas de evaluación educativa. La aplicación de los ítems, sobre todo en el ámbito educativo, ha pasado de considerarse una actividad ocasional a una actividad *ad hoc* que viene acompañada de la exigencia de disponer de un elevado número de ítems. Antes esta necesidad, la construcción artesanal que pone su foco atención en un único ítem que se construye, revisa, edita y calibra hasta que alcanza los estándares de calidad exigidos se presenta como un proceso poco eficiente.

La generación automática de ítems conjuga teorías cognitivas y psicométricas que permiten a partir de un modelo construir ítems con propiedades psicométricas preestablecidas. En este marco, la unidad de análisis es un modelo de ítem (Gierl et al., 2020) o a un modelo cognitivo (Embretson y Yang, 2006). La GAI puede representarse como un proceso que se desarrolla en 3 etapas: 1) se identifica el contenido que servirá de base para la generación de ítems; 2) se construye el modelo de ítem, y 3) se programan los algoritmos que a partir de las fases 1 y 2 generarán los ítems. Pueden consultarse varios modelos de generación de ítems en Bejar et. al. (2003), Case y Swanson (2002) o Gierl y Lai (2013).

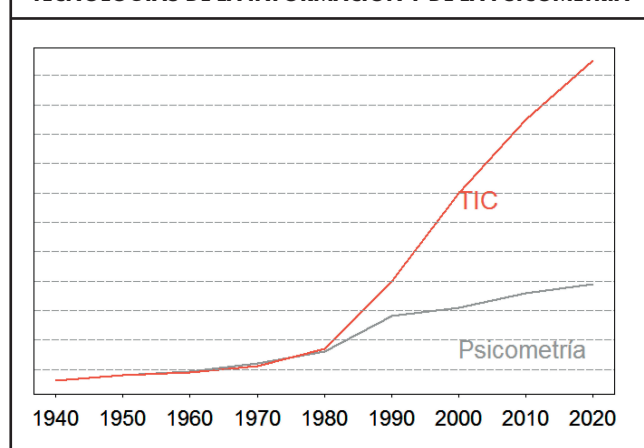
Aplicación de pruebas por ordenador

Una de las grandes ventajas de la aplicación de pruebas por ordenador es la posibilidad de recoger datos de proceso referidos a la interacción de la persona evaluada con el entorno evaluativo. Estos datos, *logdata*, pueden utilizarse para reconstruir conductas específicas, ahondar en teorías sobre el aprendizaje, estudiar diferencias entre grupos y, en definitiva, ampliar los estudios de validación. La información sobre el tiempo dedicado a cada elemento, puede ser útil por ejemplo, para analizar la motivación, la fatiga o la velocidad de ejecución; además, la aplicación computerizada permite recoger datos multimodales, datos de proceso multivariado como movimientos faciales, oculares, audio, resonancia magnética o tomografía computerizada que si bien todavía presentan dificultades de análisis serán objeto de investigación los próximos años (Guidry et al., 2013; Ramalingam y Adams, 2018; Scherer et al., 2015; en este monográfico Suarez et al., 2022).

Nuevos formatos de ítems

La mejora en las capacidades de los ordenadores personales junto al desarrollo de internet y la disponibilidad de dispositivo móviles permite aplicar test/ítems en formatos diferentes al clásico de lápiz/papel. Los nuevos formatos de ítems, que se enriquecen con la utilización del video o de la animación superan algunas de las des-

FIGURA 2
REPRESENTACIÓN FIGURATIVA DE LA EVOLUCIÓN DE LAS
TECNOLOGÍAS DE LA INFORMACIÓN Y DE LA PSICOMETRÍA



ventajas asociadas a los ítems de elección de respuesta y con ello, permiten la medición de aspectos difícilmente alcanzables con el formato de opción múltiple o con sus variantes.

Baste citar como ejemplo de la expansión y uso común de los nuevos formatos de ítems que la evaluación internacional PISA (*Programme for International Student Assessment*) gestionada por la OCDE introdujo la administración computerizada en el año 2006; a partir de la edición de 2015 PISA se diseña y aplica de forma digital (puede el lector ver ejemplos de ítems del programa PISA en <https://www.oecd.org/pisa/test/>).

La animación incluida en un ítem puede ser bidimensional, tridimensional simple y tridimensional fotorrealista o realidad virtual (Popp et al., 2016). La realidad virtual es una simulación, derivada de la industria del juego, que crea la sensación de presencia física (Linowes, 2015; Parisi, 2015). En principio, estos ítems prevén una validez aparente mayor, se acompañan de presentaciones dinámicas e interactivas que pueden generar una mejor aceptación e interés hacia la situación de test por parte de la persona evaluada, y favorecen la evaluación de tareas complejas (Bruk-Lee et al., 2013; Ryall et al., 2016). Sin embargo, los formatos multimedia también podrían introducir factores no deseados en el proceso evaluativo; por ejemplo, la caracterización sociodemográfica del avatar es un rasgo saliente que puede activar actitudes o experiencias afectivas que interactúan de forma no buscada ni deseada (Sadler et al., 2012). Son varias las críticas vertidas hacia este tipo de ítem por su falta de validez en el campo educativo (Young et al., 2012), si bien en medicina gozan de tradición y prestigio (McGaghie et al., 2010).

Corrección automatizada de ensayos

La aparición en los años 30 de los lectores ópticos incrementó la eficiencia en el proceso de corrección, y con ello el volumen de los test de selección de respuesta; sin embargo, la implementación y cada vez mayor aplicación de ítems de producción (ensayos, ítems abiertos) han impulsado la búsqueda de métodos de corrección eficientes para este formato de tarea. La corrección automatizada de ensayos se inició en los años 60 (Page, 1968), y con la sofisticación del procesamiento del lenguaje natural y el *machine learning*, la corrección automatizada de ensayos está hoy aceptada como parte de la práctica educativa. Son varios los estudios que la comparan con la evaluación mediatizada por jueces, y han mostrado su eficacia en la evaluación sobre diferentes criterios de corrección (Williamson et al., 2012). Las compañías más pujantes relacionadas con la evaluación educativa disponen de software diseñado con esta finalidad (*Pearson Test of English; ETS criterion; Accuplacer del College Board...*).

La idea que subyace a la corrección automatizada de ensayos es que el sistema es capaz de convertir la producción del estudiante (oral o escrita) en una puntuación (o varias puntuaciones) o *feedback*, que es preciso, fiable y está alineado con los constructos a evaluar. El proceso descansa en una muestra amplia de ensayos que representa el rango de posibles producciones y puntuaciones. Los ensayos se corrigen por expertos humanos que ofrecen al sistema la información necesaria para que este se entrene en la estimación de la puntuación verdadera. Con esa base, el software aprende a asociar las características distintivas de cada ensayo con las puntuaciones asignadas por

los expertos. Una vez construido el modelo el sistema es capaz de predecir la puntuación que los expertos asignarían a un nuevo ensayo.

Modelado psicométrico

Al mismo tiempo que la tecnología digital impregna la forma de construir, administrar y corregir test, los modelos psicométricos para el análisis de datos y estimación de las puntuaciones verdaderas y errores de medida se adaptan a las exigencias de los nuevos entornos. Los mayores retos a los que se enfrentan podrían ser la modelización de la multidimensionalidad, y la adaptación de la metodología construida para su uso en entornos controlados y estructurados para su uso en entornos dinámicos y con datos con un nivel de estructuración menor.

Modelos bifactoriales

Los modelos bifactoriales modelan la multidimensionalidad presente en muchos constructos analizados por la psicología, y se aplican cuando existe un factor general y factores específicos o factores de grupo. Aunque los modelos bifactoriales o modelos anidados fueron propuestos en 1937 (Holzinger y Swineford) su aplicación se ha extendido la última década (Rodríguez et al., 2016). Básicamente el modelo bifactorial y el modelo factorial de segundo orden podrían tener interpretaciones similares (Chen et al., 2006) pero los primeros cobran relevancia cuando el interés se centra en los factores de grupo, se desea analizar la relación entre éstos y los ítems que la conforman, o quiere profundizarse en la capacidad predictiva de las escalas parciales.

Modelos testlet

Los modelos TRI basados en *testlet* son formalmente modelos bifactoriales; la diferencia entre ambos proviene de la tradición en su uso que viene ligada a modelos factoriales y modelos TRI, o a campos de especialización psicológica o educativa. El *testlet* se construye por la necesidad de estimación en condiciones de violación de independencia local asociada a la aplicación de grupos de ítems dependientes de un contexto (Bradlow et al., 1999; Wainer et al., 2007). En un *testlet* cada ítem es un indicador de una dimensión general, y de una dimensión asociada a un grupo de ítems. La dimensión general representa la variable latente de interés central (por ejemplo, la competencia lectora) mientras que el resto se incorpora para tener en cuenta las dependencias adicionales entre los ítems que pertenecen al mismo subgrupo.

Modelado de ítems de elección forzosa

Frente al ítem comúnmente denominado Likert en el que una persona responde a una cuestión en una escala de respuesta ordenada, los ítems de elección forzosa o ítems ipsativos obligan a elegir entre dos o más enunciados que pueden ordenarse en función de las preferencias mostradas (en este monográfico Abad et al., 2022). Con ello se busca controlar el sesgo de respuesta (aquiescencia, deseabilidad social, tendencia central, severidad...) y mejorar el proceso evaluativo (Brown y Maydeu-Olivares, 2011, 2018; Chan, 2003).

Análisis de redes

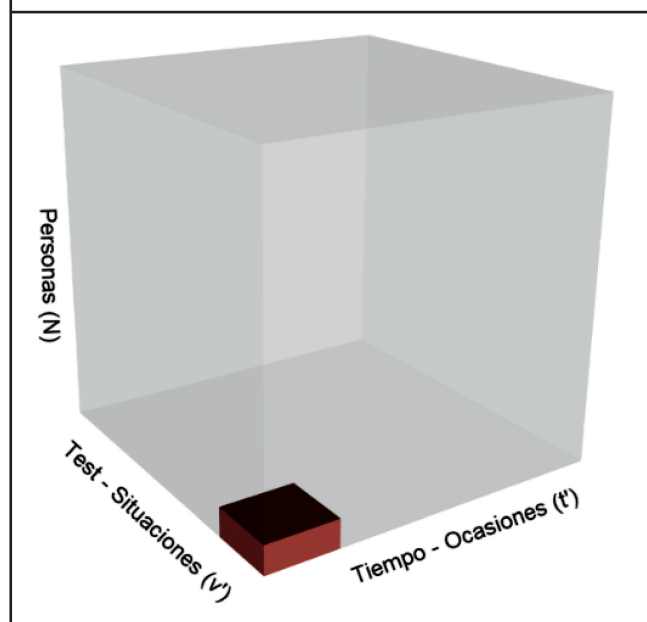
El análisis de redes en psicología es una alternativa a la visión psicométrica tradicional que relaciona un constructo con varios indica-

dores de los que es causa. En el modelado de redes los indicadores se perciben como *proxies* (aproximaciones) de variables que interactúan entre ellas; por ejemplo, desde la perspectiva tradicional síntomas como la falta de energía, problemas de sueño o baja autoestima estarían causados por la depresión mientras que en el modelo de redes esos síntomas constituyen una red de interacción mutua. Supone un acercamiento diferente a la modelización y estudio de los fenómenos psicológicos (Epskamp et al., 2018; Fonseca-Pedrero, 2018).

Big data. Redes sociales, wearables y dispositivos móviles

La tradición investigadora en ciencias sociales construida sobre la representación del cubo de datos de Cattell (1966) como un conjunto ordenado constituido por tres ejes (casos, variables y momentos temporales), ha dado paso al dato masivo, no estructurado y de alta dimensionalidad. Esta irrupción, en cierto modo cuestiona el concepto clásico de test como unidad básica para la recogida de información sobre comportamientos, actitudes o creencias. Las redes sociales como Facebook, LinkedIn o Twitter son fuentes continuas de datos que están dando lugar a una nueva línea de investigación en psicología (en este monográfico Andrés et al., 2022); la depresión, la ideación suicida, la personalidad o la selección de personal, por ejemplo, están siendo analizadas a través del análisis de la información desprendida de las redes sociales (Conway y O'Connor, 2016; Dwyer et al., 2018; Skaik y Inkpen, 2020; Woo et al., 2020). Pero además, el internet de la cosas o la utilización de dispositivos móviles abre la posibilidad de utilizar metodologías basada en el muestreo de experiencias (*Experience sampling methodology*, ESM; Myin-Germeys et al., 2018; Stieger et al., 2018) o evaluación ambulatoria (en este monográfico Fonseca-Pedrero et al., 2022) que aportan una perspectiva ecológica a la evaluación psicológica.

FIGURA 3
CUBO DE DATOS DE CATTELL Y BIG DATA



Aspectos éticos

De forma paralela al incremento de las aplicaciones de la IA, se desarrolla un debate sobre aspectos éticos, legales e implicaciones sociales. Son muchas las instituciones nacionales e internacionales que han creado comités de expertos *ad hoc* para la elaboración de documentos y directrices sobre la IA. En España en Julio del 2020 se constituyó el Consejo Asesor en inteligencia artificial, la Comunidad Europea publicó el año 2018 el plan coordinado sobre inteligencia artificial en el que se otorga el rol de vigilancia al *AI Watch*, y actualmente está trabajando en su reglamentación. La constitución de tales comités es un claro indicador del impacto tecnológico, económico, político y social de las nuevas tecnologías.

En una revisión sobre directrices y estándares en la que se estudian 84 documentos, se concluye que los problemas más tratados en los informes que analizan las cuestiones éticas implicadas con la IA están relacionadas con la transparencia, la justicia, la equidad y el sesgo (Jobin, Lenca y Vayena, 2019). Conceptos por otro lado que han sido, y son objeto de los estándares sobre construcción y uso de test, con los que estamos en permanente contacto (en este monográfico Hernández et al., 2022).

DISCUSIÓN

Los cimientos de la psicometría teórica actual se edificaron en la segunda mitad del siglo XX; el modelo de la teoría clásica de test, la formulación del modelo factorial o las primeras propuestas sobre la teoría de respuesta al ítem conocida también como "Nueva psicometría" corresponden a esa época (Lord y Novick, 1968; van der Linden y Hambleton, 1997). Desde entonces, la impulsión de la tecnología digital y la accesibilidad y potencia del software y hardware, han permitido la generalización del uso de esos modelos psicométricos acortando con ello la brecha entre psicometría teórica y aplicada (Elosua, 2012).

Pero la innovación no se centra únicamente en la socialización del modelo confirmatorio, importante por otro lado para el fortalecimiento de los estudios de validación. Junto a ella, se ha ampliado y enriquecido el concepto y uso de test; los entornos evaluativos son ahora diversos, complejos y dinámicos. Varias voces defienden que estamos en plena cuarta revolución industrial (Schwab, 2017); revolución caracterizada por el *big data*, la computación en la nube o el internet de las cosas. Si la tercera revolución industrial vino asociada a la explosión científica y a la tecnología de la información, la cuarta es un desarrollo de la anterior. La penetración de la tecnología digital en la evaluación nos ha acercado a un nuevo territorio en el cual las aportaciones de áreas de conocimiento como la ingeniería, lingüística computacional, informática y la inteligencia artificial abren nuevos campos de exploración, y cuestionan el concepto clásico de test; pero además, "el dato" se ha instalado en nuestras vidas. Este nuevo referente procedente de aplicaciones, dispositivos móviles y redes sociales está permitiendo analizar conductas, clasificar y predecir. Es cierto que los procedimientos de obtención de datos son diferentes, pero ambos acercamientos presentan afinidades en sus objetivos. El límite entre el test y el dato como instrumentos que a partir del análisis de información facilitan la toma de decisiones se diluye. La cuestión, o las cuestiones ante este horizonte, son varias ¿estamos los psicólogos preparados para acometer esta tarea? ¿cuentan los perfiles más tecnológicos (ingenieros, informáticos...)

con el suficiente *background* sustantivo para abordarla? La coordinación entre áreas, la formación del psicólogo en lenguajes como R o *Phyton* (Elosua, 2009, 2011), y el trabajo en equipo (Adjerid y Kelley, 2018; König et al., 2020; Oswald, 2020) pueden ser medios útiles que nos permitan afrontar la diversidad y el dinamismo de este tiempo. La experiencia adquirida por la psicometría en el campo de la medición, la comprensión de lo psicológico, y la clara consciencia sobre los problemas relacionados con la validez y el sesgo son valores que la psicología aporta, y de los que el *big data* puede claramente beneficiarse.

Pero no todo es *big data*; hemos presentado un abanico de líneas de trabajo actuales antes las cuales el profesional o académico puede tener varias reacciones; una de ellas es sentir las innovaciones presentadas como algo alejado de su quehacer habitual. El test de lápiz/papel, o las escalas Likert siguen estando omnipresentes en la psicología española; la construcción de test (Muñiz y Fonseca-Pedrero, 2019) y su adaptación (Muñiz et al., 2013) son territorios conocidos y fértiles para la academia, y siguen aportando conocimiento a la psicología. La tecnología digital, sin embargo, demanda algo más. Muchas de las innovaciones se estudian, analizan e implementan en compañías que se dedican al mercado relacionado con la construcción y uso de test, sobre todo, en los ámbitos educativo y organizacional. Los estudios sobre uso de test vienen señalando la diferencia entre sectores con respecto a las actitudes y utilización de la tecnología en la evaluación (Muñiz et al., 2020). Esta diferencia se agudiza cuando tratamos con tecnologías digitales (en este monográfico, Santamaría y Sánchez-Sánchez, 2022).

El test tradicional de lápiz/papel sigue vivo, pero hoy convive con desarrollos tecnológicos que han creado un entorno en el que lo virtual y digital ganan peso y atracción. El test surge como un instrumento de apoyo social, y en ello radica su pervivencia. Desde el punto de vista de la evaluación, la era digital crea un entorno compartido por la psicología, la educación, la ingeniería y la ciencia de datos (*data science*) en el que estamos abocados a participar de forma proactiva.

CONFLICTO DE INTERESES

No existe conflicto de intereses.

REFERENCIAS

- Abad, F. J., Schames, R., Sorrel, M., Nájera, P., García-Garzón, E., Garrido, L. E., y Jiménez, M. (2022). Construyendo tests adaptativos 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>
- Adjerid, I. y Kelley, K. (2018). Big data in psychology: A framework for research advancement. *The American Psychologist*, 73(7), 899-917. <https://doi.org/10.1037/amp0000190>
- American Psychological Association. (2020). *APA guidelines for psychological assessment and evaluation*. Descargado desde <https://www.apa.org/about/policy/guidelines-psychological-assessment-evaluation.pdf>
- Andrés, J. C., Aguado, D., y de Miguel, J. (2022). ¿Qué hay detrás de LinkedIn? Midiendo a través de rúbricas las LinkedIn Big Four Dimensions. *Papeles del Psicólogo*, 43(1), 12-20. <https://doi.org/10.23923/pap.psicol.2979>
- Bejar, I. I., Lawless, R., Morley, M. E., Wagner, M. E., Bennet, R. E., y Revuelta, J. (2003). A feasibility study of on-the-fly item generation in adaptive testing. *Journal of Technology, Learning and Assessment*, 2(3), 1-29.
- Bennett, R. E. (2015). The changing nature of educational assessment. *Review of Research in Education*, 39(1), 370-407. <https://doi.org/10.3102/0091732X14554179>
- Bradlow, E. T., Wainer, H., y Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, 64(2), 153-168. <https://doi.org/10.1007/bf02294533>
- Brown, A. y Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71, 460-502. <https://doi.org/10.1177/0013164410375112>
- Brown, A. y Maydeu-Olivares, A. (2018). Modelling Forced-Choice Response Formats. En P. Irving, T. Booth, y D. J. Hughes. (Eds.), *The Wiley Handbook of Psychometric Testing* (pp. 523-569). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118489772.ch18>
- Bruk-Lee, V., Drew, E. N., y Hawkes, B. (2013). Candidate reactions to simulations and media-rich assessments in personnel selection. En M. Fetzter y K. Tuzinski (Eds.), *Simulations for personnel selection* (pp. 43-60). Springer New York. https://doi.org/10.1007/978-1-4614-7681-8_3
- Case, S. M. y Swanson, D. B. (2002). *Constructing written test questions for the basic and clinical sciences*. National Board of Medical Examiners.
- Cattell, R. B. (1966). Data box: Its ordering of total resources in terms of possible relational systems. En R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 67-128). Rand-McNally.
- Chan, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika*, 30(1), 99-121. <https://doi.org/10.2333/bhmk.30.99>
- Chen, F. F., West, S., y Sousa, K. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41(2), 189-225. https://doi.org/10.1207/s15327906mbr4102_5
- Comisión Europea. (2021). *Propuesta de Reglamento del parlamento europeo y del consejo por el que se establecen normas armonizadas en materia de inteligencia artificial*.
- Conway, M. y O'Connor, D. (2016). *Social media, big data, and mental health: Current advances and ethical implications*. <https://doi.org/10.1016/j.copsyc.2016.01.004>
- De Mauro, A., Greco, M., y Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*, 65, 122-135.
- Dwyer, D. B., Falkai, P., y Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, 14(1), 91-118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>
- Elosua, P. (2009). ¿Existe vida más allá del SPSS? Descubre R. *Psicothema*, 21, 652-655.
- Elosua, P. (2011). *Introducción al entorno R*. Universidad del País Vasco. https://www.researchgate.net/publication/-264697893_Introduccion_al_entorno_R
- Elosua, P. (2012). Tests publicados en España: Usos, costumbres y asignaturas pendientes. *Papeles del Psicólogo*, 33, 12-21.
- Elosua, P. (2021). Aplicación remota de test: Riesgos y recomendaciones. *Papeles del Psicólogo*, 41(2), 33-39. <https://doi.org/10.23923/pap.psicol2021.2952>

- Embretson, S. y Yang, X. (2006). 23 Automatic item generation and cognitive psychology. En C. R. Rao y S. Sinharay (Eds.), *Handbook of statistics* (Vol. 26, pp. 747-768). Elsevier. [https://doi.org/10.1016/S0169-7161\(06\)26023-1](https://doi.org/10.1016/S0169-7161(06)26023-1)
- Epskamp, S., Maris, G., Waldorp, J., y Borsboom, D. (2018). Network psychometrics. En P. Irwing, T. Booth, y D. J. Hughes. (Eds.), *The Wiley Handbook of psychometric testing* (pp. 953-986). John Wiley & Sons. <https://doi.org/10.1002/9781118489772.ch30>
- Fonseca-Pedrero, E. (2018). Análisis de redes en psicología. *Papeles del Psicólogo*, 39, 1-12. <https://doi.org/10.23923/pap.psicol-12018.2852>
- Fonseca-Pedrero, E., Ródenas, Gabriel, Pérez-Albéniz, A., Al-Halabí, S., Pérez, M., y Muñoz, J. (2022). La hora de la evaluación ambulatoria. *Papeles del Psicólogo*, 43(1), 21-28 <https://doi.org/10.23923/pap.psicol.2983>
- Gierl, M. J. y Lai, H. (2013). Instructional topics in educational measurement (ITEMS) module: Using automated processes to generate test items. *Educational Measurement: Issues and Practice*, 32(3), 36-50. <https://doi.org/10.1111/emip.12018>
- Gierl, M. J., Lai, H., y Matovinovic, D. (2020). Augmented intelligence and the future of item development. En Hong Jiao y Robert W. Lissitz (Eds.), *Application of artificial intelligence to assessment* (pp. 1-25). Information Age Publishing-IAP.
- Grundke, R., Squicciarini, M., Jamet, S., y Kalamova, M. (2017). *Having the right mix: The role of skill bundles for comparative advantage and industry performance in GVCs*. OECD Science, Technology and Industry Working Papers, OECD Publishing.
- Guidry, B. W., Rupp, D. E., y Lanik, M. (2013). Tracing cognition with assessment center simulations: Using technology to see in the dark. En M. Feltzer y Tuzinski (Eds.), *Simulations for personal selection* (pp. 231-257). Springer Science+Business Media.
- Hambleton, R. K. (2006). Psychometric models, test designs and item types for the next generation of educational and psychological tests. En D. Bartram & R. K. Hambleton (Eds.), *Computer-based testing and the internet: Issues and avances* (pp. 77-90). John Wiley & Sons Ltd.
- Hernández, A., Elosua, P., Fernández-Hermida, J. R., y Muñoz, 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>
- Holzinger, K. J. y Swineford, S. (1937). The bi-factor method. *Psychometrika*, 47,41-54.
- Jobin, A., Ienca, M., y Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389-399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kaplan, A. M. y Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59-68. <https://doi.org/10.1016/j.bushor.2009.09.003>
- König, C., Demetriou, A., Glock, P., Hiemstra, A., Iliescu, D., Ionescu, C., Langer, M., Liem, C., Linnenbürger, A., Siegel, R., y Vartholomaios, I. (2020). Some advice for psychologists who want to work with computer scientists on big data. *Personnel Assessment and Decisions*, 17-23. <https://doi.org/10.25035/pad.2020.01.002>
- Lee, K. (2018). *AI superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin.
- Linowes, J. (2015). *Unity virtual reality: Explore the world of virtual reality by building immersive and fun VR projects using unity 3D*. Packt Publishing.
- Lord, F. M. y Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley
- Maass, W., Parsons, J., Purao, S., Storey, V. C., y Woo, C. (2018). Data-driven meets theory-driven research in the era of big data: Opportunities and challenges for information systems research. *Journal of the Association for Information Systems*, 19(12), <https://doi.org/10.17705/1jais.00526>.
- Maté-Jimenez, C. (2014). Big data. Un nuevo paradigma de análisis de datos. *Anales de mecánica y electricidad*, 5. https://revista-anales.icaei.es/web/n_29/seccion_1.html
- McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., y Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical Education*, 44, 50-63.
- Muñoz, J., Elosua, P., y Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: Segunda edición. *Psicothema*, 25, 2, 151-157. <https://doi.org/10.7334/psicothema2013.24>
- Muñoz, J. y Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31, 7-16. <https://doi.org/10.7334/psicothema2018.291>
- Muñoz, 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>
- Myin-Germeyns, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., y Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry*, 17(2), 123-132. <https://doi.org/10.1002/wps.20513>
- OECD. (2016). *Skills for a digital world*. OECD/Directorate for Science, Technology.
- Olea, J., Abad, F. J., y Barrada, J. R. (2010). Tests informatizados y otros nuevos tipos de tests. *Papeles del Psicólogo*, 31(1), 94-107.
- O'Regan, G. (2021). *A brief history of computing* (3ª edición). Springer.
- Oswald, F. L. (2020). Future research directions for big data in psychology. En S. E. Woo, L. Tay, y R. W. Proctor (Eds.), *Big data in psychological research*. (pp. 427-441). American Psychological Association. <https://doi.org/10.1037/0000193-020>
- Page, E. B. (1968). The use of the computer in analyzing student essays. *International Review of Education* 14, 210–225. <https://doi.org/10.1007/BF01419938>
- Parisi, T. (2015). *Learning virtual reality: Developing immersive experiences and applications for desktop, web, and mobile*. O'Reilly Media.
- Popp, E. C., Tuzinski, K., y Fetzer, M. (2016). Actor or avatar? Considerations in selecting appropriate formats for assessment content. En E. F. Drasgow (Ed.), *Improving educational and psychological measurement* (pp. 79-103). Routledge.
- Ramalingam, D. y Adams, R. J. (2018). How can the use of data from computer-delivered assessments improve the measurement of twenty-first century skills? En E. Care, P. Griffin, y M. Wilson

- (Eds.), *Assessment and teaching of 21st Century Skills: Research and Applications* (pp. 225-238). Springer International Publishing. https://doi.org/10.1007/978-3-319-65368-6_13
- Rifkin, J. (2011). *The third industrial revolution: how lateral power is transforming energy, the economy, and the world*. Palgrave macmillan.
- Rodriguez, A., Reise, S. P., y Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological methods*, 21(2), 137-150. <https://doi.org/10.1037/met0000045>
- Russell, S. y Norvig, P. (2021). Artificial intelligence: a modern approach. En *Series in Artificial Intelligence* (4ª edición). Prentice-Hall.
- Ryall, T., Judd, B. K., y Gordon, C. J. (2016). Simulation-based assessments in health professional education: A systematic review. *Journal of Multidisciplinary Healthcare*, 9, 69-82. <https://doi.org/10.2147/JMDH.S92695>
- Sadler, M. S., Correll, J., Park, B., y Judd, C. M. (2012). The world is not black and white: Racial bias in the decision to shoot in a multiethnic context. *Journal of Social Issues*, 68, 286-313. <https://doi.org/10.1111/j.1540-4560.2012.01749.x>
- Santamaría, P. y Sánchez-Sánchez, F. (2022). Cuestiones abiertas en el uso de las nuevas tecnologías en la evaluación psicológica. *Papeles del Psicólogo*, 43(1), 48-54. <https://doi.org/10.23923/pap.psicol.2984>
- Scherer, R., Greiff, S., y Hautamäki, J. (2015). Exploring the relation between time on task and ability in complex problem solving. *Intelligence*, 48, 37-50. <https://doi.org/10.1016/j.intell.2014.10.003>
- Schwab, K. (2017). *The Fourth Industrial Revolution*. Crown Publishing Group.
- Skaik, R. y Inkpen, D. (2020). Using social media for mental health surveillance: A review. *ACM Computing Surveys*, 53(6), 129:1-129:31. <https://doi.org/10.1145/3422824>
- Stieger, S., Lewetz, D., y Reips, U. (2018). Can smartphones be used to bring computer-based tasks from the lab to the field? A mobile experience-sampling method study about the pace of life. *Behavior Research Methods*, 50(6), 2267-2275. <https://doi.org/10.3758/s13428-017-0991-6>
- Suárez-Álvarez, J., Fernández-Alonso, R., García-Crespo, F. J., y Muñoz, J. (2022). El uso de las nuevas tecnologías en las evaluaciones educativas: La lectura en un mundo digital. *Papeles del Psicólogo*, 43(1), 36-47. <https://doi.org/10.23923/pap.psicol.2986>
- van der Linden, W. J. y Glas, G. A. W. (Eds.). (2000). *Computerized Adaptive Testing: Theory and Practice*. Kluwer Academic Publishers.
- van der Linden, W. J. y Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory*. Springer.
- Wainer, H., Bradlow, E. T., y Wang, X. (2007). *Testlet response theory and its applications*. Cambridge University Press.
- Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., y Mislevy, R. J. (Eds.). (2000). *Computerized adaptive testing: A primer* (2ª ed.). Routledge. <https://doi.org/10.4324/9781410605931>
- Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, 6, 473-492.
- Weizenbaum, J. (1976). *Power and human reason: From judgments to calculation*. W.H. Freeman y Co Ltd.
- Williamson, D. M., Xi, X., y Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice*, 31(1), 2-13. <https://doi.org/10.1111/j.1745-3992.2011.00223.x>
- Woo, S. E., Tay, L., y Proctor, R. W. (Eds.). (2020). *Big data in psychological research*. American Psychological Association.
- Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., y Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research*, 82(1), 61-89. <https://doi.org/10.3102/0034654312436980>
- Zicari, R. V. (2013). *Big data: Challenges and opportunities*. Chapman and Hall/CRC.