

PAPELES DEL PSICÓLOGO

PSYCHOLOGIST PAPERS

SECCIÓN MONOGRÁFICA:
TECNOLOGÍA DIGITAL, TEST Y EVALUACIÓN



LA DIMENSIÓN DE CRECIMIENTO PERSONAL - PSICOTERAPIA
ANALÍTICO FUNCIONAL EN VÍCTIMAS DE VIOLENCIA DE PAREJA

Ámbito: Papeles del Psicólogo / Psychologist Papers es una revista científico-profesional, cuyo objetivo es publicar revisiones, meta-análisis, soluciones, descubrimientos, guías, experiencias y métodos de utilidad para abordar problemas y cuestiones que surgen en la práctica profesional de cualquier área de la psicología. Se ofrece también como foro para contrastar opiniones y fomentar el debate sobre enfoques o cuestiones que suscitan controversia. Los autores pueden ser académicos o profesionales, y se incluyen tanto trabajos por invitación o recibidos de manera tradicional. Todas las decisiones se toman mediante un proceso de revisión anónimo y riguroso, con el fin de asegurar que los trabajos reflejan los planteamientos y las aplicaciones prácticas más novedosas.

Scope: Papeles del Psicólogo / Psychologist Papers is a scientist-practitioner journal, whose goal is to offer reviews, meta-analyses, solutions, insights, guidelines, lessons learned, and methods for addressing the problems and issues that arise for practitioners of every area of psychology. It also offers a forum to provide contrasting opinions and to foster thoughtful debate about controversial approaches and issues. Authors are academics or practitioners, and we include invited as well as traditional submissions. All decisions are made via anonymous and rigorous peer review process, to ensure that all material reflects state-of-the-art thinking and practices.

Sumario

Contents

Sección monográfica

1. Presentación
Paula Elosua
3. Impacto de la TIC en el entorno evaluativo. Innovaciones al servicio de la mejora continua
Paula Elosua
12. ¿Qué hay detrás de LinkedIn? Midiendo a través de rúbricas las LinkedIn Big Four Dimensions
José Carlos Andrés, David Aguado y Jesús de Miguel
21. La hora de la evaluación ambulatoria
Eduardo Fonseca-Pedrero, Gabriel Ródenas-Perea, Alicia Pérez-Albéniz, Susana Al-Halabi, Marino Pérez y José Muñiz
29. Construyendo test adaptativos de elección forzosa "On The Fly" para la medición de la personalidad
Abad, Francisco J.; Kreitchmann, Rodrigo S.; Sorrel, Miguel A.; Nájera, Pablo; García-Garzón, Eduardo; Garrido, Luis Eduardo y Jiménez, Marcos
36. El uso de las nuevas tecnologías en las evaluaciones educativas: La Lectura en un mundo digital
Javier Suárez-Álvarez, Rubén Fernández-Alonso, Francisco J. García-Crespo y José Muñiz
48. Cuestiones abiertas en el uso de las nuevas tecnologías en la evaluación psicológica
Pablo Santamaría y Fernando Sánchez-Sánchez
55. Comisión de test: Veinticinco años velando por la calidad de los test
Ana Hernández, Paula Elosua, José Ramón Fernández-Hermida y José Muñiz

Artículos

63. La consecución del funcionamiento psicológico pleno: La dimensión de crecimiento personal
Carlos Freire, María del Mar Ferradás, José Carlos Núñez y Antonio Valle
74. Uso de estrategias de la psicoterapia analítica funcional para apoyar a mujeres latinas víctimas de violencia de pareja
Amanda M. Muñoz-Martínez y Renzo Aguilar-Cacho

Special section

1. Introduction
Paula Elosua
3. Impact of ICT on the assessment environment. Innovations for the purpose of continuous improvement
Paula Elosua
12. What's behind LinkedIn? Measuring the LinkedIn Big Four Dimensions through rubrics
José Carlos Andrés, David Aguado and Jesús de Miguel
21. The time of ambulatory assessment
Eduardo Fonseca-Pedrero, Gabriel Ródenas-Perea, Alicia Pérez-Albéniz, Susana Al-Halabi, Marino Pérez and José Muñiz
29. Building adaptive forced choice tests "On The Fly" for personality measurement
Abad, Francisco J.; Kreitchmann, Rodrigo S.; Sorrel, Miguel A.; Nájera, Pablo; García-Garzón, Eduardo; Garrido, Luis Eduardo and Jiménez, Marcos
36. The use of new technologies in educational assessments: Reading in a digital world
Javier Suárez-Álvarez, Rubén Fernández-Alonso, Francisco J. García-Crespo and José Muñiz
48. Open questions in the use of new technologies in psychological assessment
Pablo Santamaría and Fernando Sánchez-Sánchez
55. Test commission: Twenty five years working on test quality
Ana Hernández, Paula Elosua, José Ramón Fernández-Hermida and José Muñiz

Articles

63. Achieving full psychological functioning: The dimension of personal growth
Carlos Freire, María del Mar Ferradás, José Carlos Núñez and Antonio Valle
74. Using functional analytic psychotherapy strategies for supporting Latinas victims of interpersonal partner violence
Amanda M. Muñoz-Martínez and Renzo Aguilar-Cacho

Edita / Publisher
Consejo General de la Psicología de España

Director / Editor
Serafin Lemos Giráldez (Univ. de Oviedo)

Directores asociados / Associated Editors
José Ramón Fernández Hermida (Univ. de Oviedo), José Carlos Núñez Pérez (Univ. de Oviedo), José María Peiró Silla (Univ. de Valencia) y Eduardo Fonseca Pedrero (Univ. de la Rioja)

Consejo Editorial / Editorial Board
Mario Alvarez Jiménez (Univ. de Melbourne, Australia); Imanol Amayra Caro (Univ. de Deusto); Antonio Andrés Pueyo (Univ. de Barcelona); Neus Barrantes Vidal (Univ. Autónoma de Barcelona); Adalgisa Battistelli (Univ. de Bordeaux, Francia); Elisardo Becoña (Univ. de Santiago de Compostela); Amalio Blanco (Univ. Autónoma de Madrid); Carmen Bragado (Univ. Complutense de Madrid); Gualberto Buela (Univ. de Granada); Esther Calvete (Univ. de Deusto); Antonio Cano (Univ. Complutense de Madrid); Enrique Cantón (Univ. de Valencia); Pilar Carrera (Univ. Autónoma de Madrid); Juan Luis Castejón (Univ. de Alicante); Alex Cohen (Louisiana State University, USA); María Crespo (Univ. Complutense de Madrid); Martin Debbané (Université de Genève, Suiza); Paula Elosua (Univ. del País Vasco); José Pedro Espada (Univ. Miguel Hernández); Lourdes Ezpeleta (Univ. Autónoma de Barcelona); Jorge Fernández del Valle (Univ. de Oviedo); Raquel Fidalgo (Univ. de León); Franco Fraccaroli (Univ. de Trento, Italia); Maite Garaigordobil (Univ. del País Vasco); José Manuel García Montes (Univ. de Almería); César González-Blanch Bosch (Hospital Universitario 'Marqués de Valdecilla', Santander); Joan Guardia Olmos (Univ. de Barcelona); José Gutiérrez Maldonado (Univ. de Barcelona); Juan Herrero Olaizola (Univ. de Oviedo); M^a Dolores Hidalgo (Univ. de Murcia); Cándido J. Inglés Saura (Univ. Miguel Hernández); Juan E. Jiménez (Univ. de La Laguna); Barbara Kozusznik (Univ. de Silesia, Polonia); Francisco Labrador (Univ. Complutense de Madrid); Concha López Soler (Univ. de Murcia); Nigel V. Marsh (James Cook University, Singapore); Emiliano Martín (Dept^o. de Familia. Ayuntamiento de Madrid); Vicente Martínez Tur (Univ. de Valencia); Carlos Montes Piñeiro (Univ. de Santiago); Luis

Montoro (Univ. de Valencia); José Muñiz (Univ. de Oviedo); Marino Pérez (Univ. de Oviedo); Salvador Perona (Univ. de Sevilla); Ismael Quintanilla (Univ. de Valencia); José Ramos (Univ. de Valencia); Georgios Sideridis (Harvard Medical School, USA); Ana Sornoza (Univ. de Valencia); M^a Carmen Tabernero (Univ. de Salamanca); Antonio Valle Arias (Univ. de A Coruña); Miguel Angel Vallejo (UNED); Oscar Vallina (Hospital Sierrallana de Torrelavega); Carmelo Vázquez (Univ. Complutense de Madrid); Antonio Verdejo (Monash University, Australia); Miguel Angel Verdugo (Univ. de Salamanca)

Diseño y maquetación / Design and layout
Cristina García y Juan Antonio Pez

Administración y publicidad / Management and advertising
Arantxa Sánchez y Silvia Berdullas

Consejo General de la Psicología de España
C/ Conde de Peñalver, 45-3^a planta
28006 Madrid - España
Tels.: 91 444 90 20 - Fax: 91 309 56 15
Web: <http://www.papelesdelpsicologo.es>
E-mail: papeles@cop.es

Depósito Legal
M-27453-1981 / ISSN 0214-7823

Los editores no se hacen responsables de las opiniones vertidas en los artículos publicados.

Papeles del Psicólogo / Psychologist Papers está incluida en las bases de datos PsycINFO, Clarivate Analytics (Emerging Sources Citation Index), Psycodoc y del ISOC (Psedisc), del DOAJ (Directory of Open Access Journals), Elsevier Bibliographic Database: SCOPUS, Redalyc, IBECS, EBSCO, Dialnet e In-RECS; y también se puede consultar en la página WEB del Consejo General de la Psicología de España:
<https://www.cop.es>



En conjunto, creo que *Papeles del Psicólogo* nos brinda la oportunidad de disfrutar de lecturas amenas, variadas y científicamente sólidas que muestran parte del trabajo que estamos llevando a cabo en España en un área relevante para el desarrollo de la psicología. Nos gustaría que la formación del profesional pudiera abarcar algunos de los puntos que se tratan en el monográfico porque en un entorno dinámico y a veces algo desestructurado, las competencias relacionadas con la medición y uso de test, en su sentido más amplio, son herramientas para el desarrollo de una perspectiva crítica y sólida con la que el psicólogo puede posicionarse ante perfiles ajenos a la disciplina. La era digital ha creado un hábitat compartido por la psicología, la educación, la ingeniería y la ciencia de datos en el que tenemos asignado un papel preferente; sigamos haciendo un buen uso de él.

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>

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>

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>

Fonseca-Pedrero, E., Ródenas, Gabriel, Pérez-Albéniz, A., Al-Halabí, S., Pérez, M., y Muñiz, J. (2022). La hora de la evaluación ambulatoria. *Papeles del Psicólogo*, 43(1), 21-28 <https://doi.org/10.23923/pap.psicol.2983>

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>

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>

Suárez-Álvarez, J., Fernández-Alonso, R., García-Crespo, F. J., y Muñiz, 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>

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, España

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).

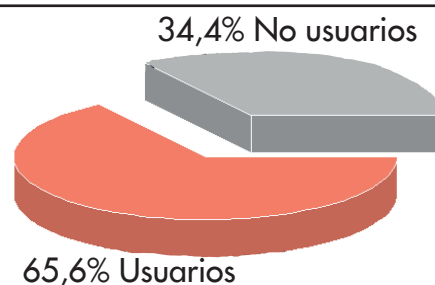
Inteligencia artificial

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	World Wide Web (free)
2003-2006	LinkedIn, Facebook, Twitter
2005	Big data
2013	Internet de las cosas
2020	Big data + Inteligencia artificial

FIGURA 1
PENETRACIÓN DE INTERNET



Nota. Datos extraídos de <https://www.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**ariiedad, 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 psicómétricos 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).

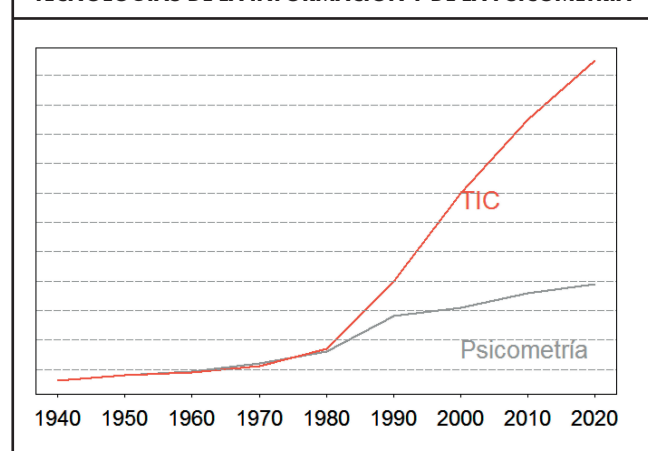
Aplicación de pruebas por ordenador

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

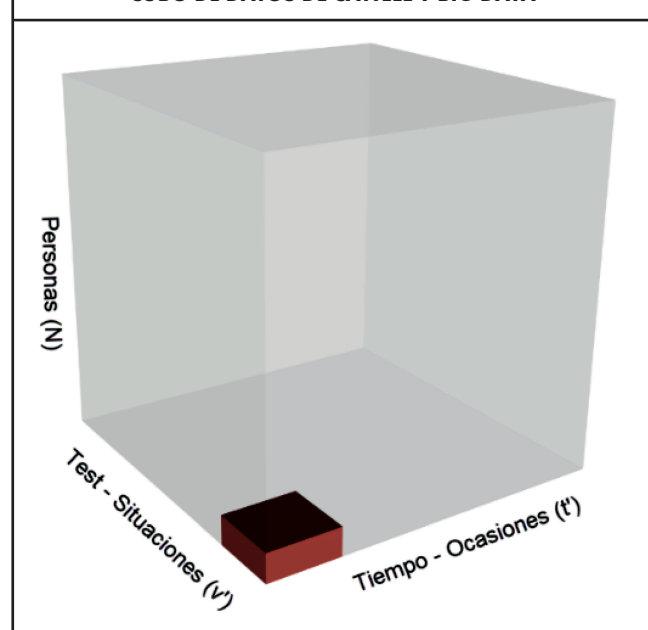
El análisis de redes en psicología es una alternativa a la visión psicométrica tradicional que relaciona un constructo con varios indica-

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, Ienca 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. Irwing, 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>
- Bruck-Lee, V., Drew, E. N., y Hawkes, B. (2013). Candidate reactions to simulations and media-rich assessments in personnel selection. En M. Fetzner 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.psicol.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.psico-12018.2852>
- Fonseca-Pedrero, E., Ródenas, Gabriel, Pérez-Albéniz, A., Al-Halabí, S., Pérez, M., y Muñiz, 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. Fetzter 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 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>
- 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.ica.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.
- Muñiz, J., Elosua, P., y Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: Segunda edición. *Psicothema*, 25.2, 151-157. <https://doi.org/10.7334/psicothema2013.24>
- Muñiz, 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>
- 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>
- 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 Fetzter, 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.

¿Qué hay detrás de LinkedIn? Midiendo a través de rúbricas las LinkedIn Big Four Dimensions

What's behind LinkedIn? Measuring the LinkedIn Big Four Dimensions through rubrics

José Carlos Andrés¹, David Aguado^{1,2} y Jesús de Miguel^{1,3}

¹Universidad Autónoma de Madrid, España. ²Instituto de Ingeniería del Conocimiento, España. ³Centro de Investigación para la Efectividad Organizacional, España

LinkedIn es la red social más utilizada en el ámbito profesional para funciones de reclutamiento y selección de personal. Aunque los profesionales de Recursos Humanos realizan inferencias sobre los candidatos en base a la información contenida en el perfil de LinkedIn, estas inferencias no se realizan de forma sistematizada y objetiva. En esta investigación se proponen rúbricas, basadas en el modelo LinkedIn Big Four, como herramienta para hacer más sistemáticas y objetivas las inferencias de los profesionales de selección. Para el análisis se utilizó una muestra de profesionales del sector IT (n = 105). Los resultados muestran unos resultados de fiabilidad adecuados a los estándares al uso y una evidencia inicial de validez que relaciona las puntuaciones en las rúbricas con las competencias genéricas de los profesionales.

Palabras clave: LinkedIn Big Four, rúbricas, reclutamiento online, Selección de personal online.

LinkedIn is the most used social network in the professional field for recruitment and selection. Although human resources professionals make inferences about candidates based on the information contained in the LinkedIn profile, these inferences are not made in a systematized or objective way. In this research, rubrics, based on the LinkedIn Big Four model, are proposed as a tool to make the inferences of selection professionals more systematic and objective. A sample of professionals from the IT sector (n = 105) was used. Results show adequate reliability and initial evidence of validity by means of correlation between the rubrics and generic competencies scores.

Key Words: LinkedIn big four, Rubrics, online recruitment, online selection.

INTRODUCCIÓN

Las organizaciones, cada vez más, conceden al proceso de reclutamiento y selección un papel fundamental en el éxito y la competitividad organizacional (Deros y De Fruyt, 2016). En este contexto de "la guerra" por el talento (Frasca y Edwards, 2017) la tecnología se ha erigido como un elemento clave (Deros y De Fruyt, 2016; García-Izquierdo et al., 2019; Ryan y Deros, 2019) para la identificación y selección de profesionales con un adecuado ajuste al puesto y a la organización (García-Izquierdo et al., 2015; Sylva y Mol, 2009).

Entre estos usos tecnológicos, junto con la utilización de los test informatizados, la gamificación, los recursos audiovisuales, y las páginas web corporativas entre otros (Elosúa, 2022), las Social Networks Websites (SNW) se han convertido en una de las principales herramientas utilizadas tanto por los profesionales de ámbitos y ocupaciones diversas, como por los profesionales del reclutamiento y la selección (Chapman y Mayers, 2015; Nikolaou, 2014; Woods et al, 2020). Para los primeros, la búsqueda de un puesto de trabajo es una de las razones más importantes para unirse a una red social profesional como LinkedIn (Frasca y Edwards, 2017) ante la que muestran una actitud positiva (Aguado et al., 2016). Para los segundos, las SNW son una herramienta para, en el reclutamiento online, realizar el cribado de candidatos y configurar un banco sobre el que desarrollar los procesos evaluativos asociados al proceso de selección (Black y Johnson, 2012; Chiang y Suen, 2015).

Ahora bien, el uso de las SNW para reclutamiento y selección conlleva algunas dificultades. Por ejemplo, la tradicional división entre el reclutamiento (atracción de un número suficiente de candidatos adecuados) y la selección (evaluación de su ajuste a los requerimientos de la posición) puede difuminarse (Aguado et al., 2016). Así, los profesionales de recursos humanos infieren, a partir de la información que el candidato aporta en su perfil, características sobre su personalidad, sus competencias, e incluso predicen la calidad de su desempeño futuro en la empresa (Van Iddekinge et al., 2016), y en base a ello evalúan la idoneidad del candidato en cuanto a su ajuste al puesto y a la organización (Back et al., 2010; Chamorro-Premuzic y Steinmetz, 2013; Kluemper et al., 2012; Roulin y Bangerter, 2013). Este proceso de toma de decisiones asociado al análisis de la información pública del candidato y no a un proceso de medición estandarizado, puede acompañarse de déficits en la fiabilidad de las medidas, y de falta de validez por la escasez de evidencias asociadas. Limitaciones que, además, pueden producir sesgos al utilizar información que no está relacionada con el desempeño en el trabajo (García-Izquierdo et al., 2015; Seiter y Hatch, 2005; Shannon y Stark, 2003; Villeda et al., 2019).

En este contexto, diferentes autores reclaman tanto mayor investigación sobre el uso de las SNW para el reclutamiento y la selección (e.g. Ryan y Deros, 2016), como el desarrollo de herramientas que permitan una utilización de éstas acorde a los criterios psicométricos establecidos (Ötting y Maier, 2018).

En esta línea, el objetivo de nuestro trabajo es construir cuatro rúbricas específicas para valorar la información presente en los perfiles de LinkedIn, tomando como modelo sustantivo el LinkedIn Big Four (Aguado et al., 2019), y obtener evidencia inicial acerca de su fiabilidad y validez. Para ello se explora la fiabilidad de las rúbricas

Recibido: 1 noviembre 2021 - Aceptado: 10 diciembre 2021

Correspondencia: David Aguado. Universidad Autónoma de Madrid e Instituto de Ingeniería del Conocimiento. España.

E-mail: david.aguado@uam.es

a través de su consistencia interna, fiabilidad temporal y fiabilidad inter-jueces; y su validez predictiva mediante el análisis de su relación con las competencias genéricas establecidas en el Great Eight Competency Model (Bartram, 2005).

FUNDAMENTACIÓN TEÓRICA

Las SNW en el contexto del Reclutamiento y la Selección

Una SNW se caracteriza por permitir al usuario: (a) definir un perfil dentro de un sistema acotado; (b) articular una lista de contactos con los que compartir la información; y (c) ver y recorrer la lista propia y las de los demás usuarios para identificar oportunidades de conexión y contacto (Boyd y Ellison, 2007). Además, las redes sociales profesionales como LinkedIn, permiten poner a disposición de un gran número de usuarios de Internet información relativa a los proyectos profesionales, los puestos desempeñados y la contribución realizada a organizaciones profesionales (Black y Johnson, 2012; Nikolaou, 2014).

Entre las ventajas de la utilización de las SNW como medio de reclutamiento online podemos destacar: (a) la rapidez en la que los candidatos pueden aplicar a las vacantes actuales; (b) la calidad de la información respecto de las competencias, habilidades y experiencia específicas del candidato; (c) el ahorro tanto en coste como en tiempo invertido; (d) la imagen de marca que la organización ofrece a los candidatos; (e) el incremento en la capacidad para alcanzar un número mayor y más variado de candidatos; y (f) la facilidad de aproximación a los candidatos pasivos, que no están en búsqueda activa de empleo (Ghazzawi y Accoume, 2014; Kaur 2015; Lievens y Harris, 2003).

Por otro lado, desde la perspectiva del candidato, las SNW también tienen importantes ventajas. Entre ellas se valora especialmente la inmediatez del proceso (Nikolaou, 2014), la posibilidad de búsqueda y comparación entre las distintas ofertas de empleo y la facilidad de aplicación a las mismas (Sylva y Mol, 2009), la rapidez para establecer contacto con más oportunidades de trabajo (Lievens y Harris, 2003), y el acceso a un mayor flujo de información sobre

el puesto y la organización, que permite tomar mejores y más rápidas decisiones sobre a qué vacantes aplicar (Galanaki, 2002; Kashi y Zheng, 2013). Además, al igual que para las organizaciones, también supone un ahorro de costes para los candidatos (Sylva y Mol, 2009). Adicionalmente, las SNW permiten a los usuarios controlar la forma en la que se presentan a sí mismos, pudiendo escoger qué información, comentarios, fotografías, etc., ofrecer en su perfil para crearse una determinada imagen online (Roulin, 2014).

Una mirada estructural sobre LinkedIn: LinkedIn Big Four

Entre las SNW profesionales, LinkedIn es probablemente la más utilizada por los profesionales del reclutamiento y la selección (Aguado et al., 2016; Chiang y Suen, 2015). LinkedIn permite a los usuarios incluir información equivalente a su currículum, combinando información personal y profesional (Chiang y Suen, 2015). Así, a partir de ella, los profesionales del reclutamiento y la selección son capaces de extraer información representativa sobre las habilidades específicas que muestran los candidatos. Por ejemplo, utilizando la experiencia profesional, la trayectoria académica o las tareas de voluntariado, los profesionales infieren información sobre las capacidades de liderazgo; mediante la observación de la estructura del perfil se valoran las habilidades de planificación; y, a través de las descripciones que el usuario hace en su perfil sobre su experiencia profesional, se valoran sus habilidades de comunicación (Roulin y Levashina, 2019). Por otro lado, la información contenida en el perfil, que es gestionada por el propio usuario, es también relevante para su posicionamiento en los procesos de reclutamiento. Por ejemplo, los reclutadores prestan más atención a los perfiles con más información y realizan peores valoraciones de aquellos en la que ésta es incompleta (McCabe, 2017; Shahani-Denning et al., 2017).

En la Tabla 1 se observa la organización del perfil de LinkedIn en sus 8 aspectos diferentes.

Sobre esta estructura de LinkedIn, Aguado et al., (2019) proponen el LinkedIn Big Four (LKBF). Mediante el LKBF la información existente en los perfiles de LinkedIn es organizada en cuatro dimensiones:

TABLA 1
PRINCIPALES ELEMENTOS EN EL PERFIL LINKEDIN

Aspecto	Descripción
Datos básicos	Fotografía personal; nombre; número de contactos; titular profesional; ubicación; información de contacto.
Extracto	Resumen con palabras clave (hasta 2.000 caracteres)
Actividad	Publicaciones, imágenes, comentarios sobre publicaciones de otros usuarios.
Experiencia	Descripción de la experiencia profesional y de los logros conseguidos
Estudios	Titulaciones acreditadas
Certificaciones y cursos	Certificaciones logradas y cursos realizados
Experiencia como voluntario	Experiencias y causas de voluntariado
Aptitudes y validaciones	Aptitudes del usuario y validaciones realizadas de cada aptitud por parte de otros usuarios
Recomendaciones	Recomendaciones realizadas y recibidas
Logros	Publicaciones; idiomas; proyectos; premios
Intereses	Seguimiento de: Personalidades relevantes; empresas; grupos; y universidades.

1) amplitud de la experiencia profesional; 2) capital social; 3) interés por mantener la formación actualizada; y 4) amplitud de la información no profesional (ver Tabla 2). Estas dimensiones son generalizables a través de cualquier perfil profesional y han resultado ser eficaces tanto en el análisis de las diferencias entre profesionales TIC en función de su experiencia, como en la predicción de su rendimiento profesional (Aguado et al., 2019).

Midiendo las dimensiones LKBF mediante rúbricas

La utilización de rúbricas para la evaluación en contextos educativos y organizacionales tiene una amplia tradición (e.g., Hung et al., 2013; Nadelson et al., 2015; Riebe y Jackson, 2014; Vaughan et al., 2019). En esencia, una rúbrica es una guía que permite atribuir una puntuación a una determinada dimensión a partir del cumplimiento de un conjunto de criterios (Wenzlaff et al., 1999). Así, una rúbrica consiste en un conjunto de elementos que describen a través de diferentes niveles escalados el rendimiento o ejecución en una determinada tarea o dimensión (Unal et al., 2012).

En nuestro trabajo, para la medida de las cuatro dimensiones especificadas en el LKBF se desarrollaron cuatro rúbricas diferentes (una para cada dimensión, ver Apéndice I). En su elaboración se siguieron los estándares al uso (Wenzlaff et al., 1999) y se especificaron tanto los aspectos del perfil LinkedIn a evaluar, como los elementos a utilizar para ello. Para el establecimiento de las diferentes categorías de respuesta, para cada elemento se analizaron las distribuciones de frecuencias encontradas por Aguado et al., (2019) en su estudio original realizado sobre 618 perfiles de LinkedIn en el sector TIC. Como resultado de este ejercicio, los diferentes elementos contemplados en las rúbricas son valorados en escalas de 10 categorías, cinco categorías o dos categorías (presencia-ausencia). Debido a estas diferencias en las escalas de respuesta, siguiendo las recomendaciones de Meade (2010), la puntuación total en cada rúbrica se obtiene mediante la adición de las puntuaciones estandarizadas de cada elemento.

MÉTODO

Participantes

La muestra está compuesta por 105 candidatos a cubrir vacantes del sector TIC: El 79% de los participantes son hombres con edades comprendidas entre los 22 y los 57 años (edad media 32.27). Todos los participantes fueron informados de los objetivos de la investigación y se requirió su consentimiento informado. Se aclaró que la no participación en la investigación no tenía ninguna consecuencia en el proceso selectivo. Ningún participante declinó participar en el estudio.

Instrumentos

Perfil LinkedIn de los participantes

Con el objetivo de mantener estable la información de los perfiles a analizar, estos fueron impresos en formato pdf. Para ello se utilizó una licencia *recruiter* de LinkedIn.

Competencias Genéricas

La valoración de las competencias de los participantes se realizó con el test PRISMA 4D (Instituto de Ingeniería del Conocimiento, 2015). A través de esta prueba se exploran las ocho competencias definidas en el *Great Eight Competency Model* (Bartram, 2005) mediante el análisis de (a) los comportamientos del evaluado, a los que se denomina destrezas; (b) la experiencia profesional en puestos que favorecen el desarrollo de cada competencia, a lo que se denomina experiencia; (c) el desarrollo de acciones formativas que permitan al evaluado adquirir habilidades relacionadas con cada competencia, a lo que se denomina formación; y (d) las actitudes del evaluado hacia cada competencia, lo que se denomina preferencias. Por tanto, para cada competencia, mediante PRISMA 4D se obtienen cuatro puntuaciones diferentes.

Rúbricas para la valoración de las cuatro dimensiones del perfil LinkedIn

Se utilizaron las cuatro rúbricas descritas anteriormente (ver Apén-

TABLA 2
LINKEDIN BIG FOUR DIMENSIONS (AGUADO ET AL., 2019)

Dimensiones y Elementos	Descripción
Amplitud de la Experiencia Profesional. Número de experiencias distintas reflejadas en el perfil; número de roles distintos desempeñados en la actividad profesional; número de empresas en las que ha desempeñado actividad profesional; número de líneas dedicadas a desarrollar la experiencia profesional; número de meses que suman la experiencia laboral del candidato.	Esta dimensión expresa la amplitud y profundidad de la experiencia profesional. Permite acercarse a las diferentes experiencias laborales del individuo, así como a su duración y a los roles desempeñados.
Capital Social. Número de contactos que tiene el candidato en su perfil; número de validaciones de aptitudes; recomendaciones recibidas por el candidato; número de empresas a las que sigue el candidato; número de grupos a los que sigue el candidato.	Refleja la intensidad de la interacción del participante con la comunidad presente en la red social, así como el grado de dinamismo de dicha actividad.
Interés por mantener conocimientos actualizados. Número de titulaciones universitarias (grado o posgrado); número de cursos adicionales a la formación universitaria; número de universidades a las que sigue el candidato.	Integra la información sobre la trayectoria académica reglada y no reglada. Refleja el interés por mantenerse actualizado en los contenidos relevantes para la actividad profesional.
Amplitud de la Información no profesional. Número de categorías rellenadas en el perfil que muestran información sobre el candidato; número de idiomas que refiere conocer el candidato (excepto el español); intereses; causas benéficas; existencia de la sección "Acerca de" o "Extracto".	Señala el interés del usuario por ofrecer un perfil LinkedIn completo, de forma que otros usuarios puedan conocer sus intereses más allá del ámbito estrictamente profesional.

dice 1): amplitud de la experiencia profesional (LKBF1), capital social (LKBF2), interés por mantener la formación actualizada (LKBF3), y amplitud de la información no profesional (LKBF4).

Con el objetivo de realizar una valoración inicial de la calidad del contenido de las rúbricas, siguiendo las recomendaciones al uso (Curran et al., 2011; Roblyer & Wiencke, 2003), trabajamos en un Grupo Focal formado por 12 expertos. Estos fueron elegidos tomando como criterio su experiencia en la utilización de LinkedIn para tareas de reclutamiento y selección de personal. El trabajo del Grupo Focal se encaminó a valorar el grado en el que las rúbricas eran fáciles de utilizar, amigables, y entendibles, tanto en los aspectos que contienen como en la valoración que se solicita al evaluador. Como resultado de este proceso se realizaron modificaciones en las nomenclaturas de las secciones de LinkedIn y en la redacción de algunos enunciados.

Procedimiento

Los participantes cumplimentaron el cuestionario PRISMA 4D en el marco de los diferentes procesos selectivos en los que participaban (los procesos se desarrollaron a lo largo del año 2019 y principios de 2020). Tras ello, el primer autor del trabajo recolectó los perfiles LinkedIn de los participantes y los guardó como documentos *Printable Document Format* (pdf).

La fiabilidad de las rúbricas se exploró a través de (a) la fiabilidad inter-jueces (los 12 expertos que formaron parte del grupo focal); (b) la fiabilidad test-retest (a través de las valoraciones de 6 de los 12 expertos, transcurrido un lapso de tiempo de 6 semanas); y (c) la consistencia interna (a través de las puntuaciones que los 12 expertos otorgaron a los diferentes elementos de las rúbricas en la primera valoración de los perfiles). La validez de las puntuaciones obtenidas con las rúbricas se examinó a partir de las correlaciones entre dichas puntuaciones y las obtenidas en las destrezas, experiencia, formación y preferencias, relativas a las 8 competencias genéricas.

RESULTADOS

Evidencias de Fiabilidad

En la tabla 3 pueden observarse los resultados obtenidos en relación a la fiabilidad de las medidas. La fiabilidad inter-jueces de las cuatro medidas resultó ser superior a los estándares recomendados (Stellmack et al., 2009) en todas las dimensiones excepto en la tercera (LKBF3). El índice Kappa promedio para los 12 reclutadores resultó ser .86, .94, .65, y .83 respectivamente para cada una de las cuatro medidas. Los resultados referidos a la fiabilidad temporal

también resultaron ser adecuados para todas las medidas. Los índices de correlación de Pearson obtenidos para cada una de las cuatro dimensiones valoradas fueron 1, 1, .97, y .99 respectivamente. Finalmente, respecto de la consistencia interna, se obtuvieron valores por encima del estándar habitual ($> .70$) para LKBF1 (.89), LKBF2 (.72), y LKBF4 (.70), pero para la dimensión LKBF3 (.40) no se alcanzó dicho estándar.

Evidencias de Validez

En la Tabla 4 se muestran las correlaciones entre las puntuaciones obtenidas en las rúbricas y las obtenidas en los cuatro elementos de cada competencia: destrezas, formación, experiencia y preferencias.

Un primer análisis de los resultados muestra que las puntuaciones obtenidas en las rúbricas parecen ser útiles para inferir tanto el grado en el que los profesionales han desarrollado sus competencias a partir de la experiencia profesional, como el grado en el que son efectivos en las competencias a través de su comportamiento. La puntuación en la rúbrica LKBF1 (amplitud de la experiencia profesional) correlaciona significativamente con la experiencia en las competencias Liderar-Decidir ($r = .30$; $p < .05$) y Analizar-Usar conocimiento ($r = .30$, $p = .05$); la puntuación en la rúbrica LKBF2 (capital social) correlaciona significativamente con la experiencia en las 8 competencias: Liderar-Decidir ($r = .41$, $p < .05$), Cooperar-Respeto ($r = .27$, $p < .05$), Comunicar-Relacionar ($r = .46$, $p < .001$), Analizar-Usar conocimiento ($r = .46$, $p < .001$), Aprendizaje e Innovación ($r = .40$, $p < .05$), Planificación y Organización ($r = .30$, $p < .05$), Adaptación y Afrontamiento ($r = .30$, $p < .05$) y Logro y Emprendimiento ($r = .41$, $p < .05$). La puntuación en LKBF3 (interés por mantener la formación actualizada), correlaciona significativamente con la experiencia en la competencia Comunicar-Relacionar ($r = .37$, $p < .05$). Finalmente, la puntuación en la rúbrica LKBF4 (amplitud de la información no profesional) muestra correlaciones significativas en la componente experiencia con 6 de las 8 competencias: Cooperar-Respeto ($r = .31$, $p < .05$), Comunicar-Relacionar ($r = .39$, $p < .05$), Analizar-Usar conocimiento ($r = .29$, $p < .05$), Planificación y Organización ($r = .39$, $p < .05$), Adaptación y Afrontamiento ($r = .32$, $p < .05$) y Logro y Emprendimiento ($r = .32$, $p < .05$).

Al observar las correlaciones con las destrezas observamos que las puntuaciones obtenidas en LKBF1 y en LKBF3 no correlacionan significativamente con el comportamiento relacionado con las competencias. Por otro lado, la puntuación en la LKBF2 muestra correlaciones significativas con Liderar-Decidir ($r = .32$, $p < .05$),

TABLA 3
ESTADÍSTICOS DESCRIPTIVOS E INDICADORES DE FIABILIDAD DE LAS RÚBRICAS

	Media	DT	Fiabilidad Inter-jueces	Fiabilidad Temporal	Consistencia Interna
LKBF1: Amplitud de la Experiencia Profesional	-.34	4.50	.86 (.95)	1	.89
LKBF2: Capital Social	-.15	3.40	.94 (.99)	1	.72
LKBF3: Interés Formación Actualizada	-.23	1.69	.65 (.84)	.97	.40
LKBF4: Amplitud Información no Profesional	-.21	3	.83 (.93)	.99	.70

Nota: los valores de la media están normalizados; DT, desviación típica; Fiabilidad Inter-jueces (Cohen Kappa, entre paréntesis weighted kappa). Fiabilidad Temporal (Correlación de Pearson); Consistencia Interna (Alpha de Cronbach)

Comunicar Relacionar ($r = .40, p < .05$), Aprendizaje e Innovación ($r = .40, p < .05$) y Logro y Emprendimiento ($r = .40, p < .05$). Finalmente, la puntuación en LKBF4 correlaciona significativamente en el nivel de las destrezas con Aprendizaje e Innovación ($r = .29, p < .05$) y con Logro y Emprendimiento ($r = .33, p < .05$).

Por otro lado, los resultados nos muestran que no existe ninguna correlación significativa entre las puntuaciones en las rúbricas y el aspecto de "formación" de las competencias. De manera similar, las puntuaciones en las rúbricas solo muestran dos relaciones significativas con las "preferencias" de los participantes respecto de las competencias: para las actitudes relacionadas con el logro y el emprendimiento, y para las relacionadas con la comunicación y la relación, en ambos casos con la puntuación obtenida en LKBF2 (Capital Social) ($r = .29; p < .05$) y ($r = .32; p < .05$) respectivamente.

DISCUSIÓN

En este trabajo se han mostrado las propiedades psicométricas iniciales de cuatro rúbricas diseñadas para valorar los perfiles de LinkedIn a la luz de las dimensiones propuestas por Aguado et al., (2019) en el modelo LinkedIn Big Four. Los resultados muestran una buena fiabilidad de las medidas obtenidas por las rúbricas tanto en términos de fiabilidad inter-jueces, como de fiabilidad temporal. La fiabilidad expresada a partir del coeficiente Alfa de Cronbach resultó ser adecuada para tres de las medidas obtenidas con las rúbricas, pero no para la rúbrica diseñada para la medición del LKBF3 (interés por mantener la formación actualizada). Así mismo, nuestros resultados aportan evidencia inicial de la validez de las rúbricas para informar acerca de las competencias genéricas de los candidatos. La medida del LinkedIn Big Four que se realiza con las rúbricas

TABLA 4
CORRELACIONES ENTRE LAS PUNTUACIONES OBTENIDAS EN LAS RÚBRICAS Y LAS PUNTUACIONES EN LAS COMPETENCIAS

	LKBF1: Amplitud de la Experiencia Profesional	LKBF2: Capital Social	LKBF3: Interés Formación Actualizada	LKBF4: Amplitud Información no Profesional
Liderar-Decidir				
Formación	.06	.23	.02	.14
Actitudes	.18	.18	-.07	.12
Experiencia	.30*	.41*	.17	.23
Destrezas	.13	.32*	-.06	.19
Cooperar-Respeto				
Formación	-.01	.15	.03	.18
Actitudes	.07	.13	-.06	.09
Experiencia	.02	.27*	.18	.31*
Destrezas	.14	.19	.03	.25
Comunicar Relacionar				
Formación	.07	.16	.13	.17
Actitudes	.03	.32*	.10	.16
Experiencia	.25	.46**	.37*	.39*
Destrezas	.18	.40*	.16	.24
Analizar-Usar conocimiento				
Formación	.08	.15	.03	.07
Actitudes	.09	.14	.01	.13
Experiencia	.30*	.46**	.23	.29*
Destrezas	.19	.22	.08	.25
Aprendizaje e Innovación				
Formación	.06	.22	.05	.07
Actitudes	.06	.14	.03	.04
Experiencia	.22	.40*	.11	.23
Destrezas	.24	.40*	.03	.29*
Planificación y Organización				
Formación	.17	.25	.12	.14
Actitudes	-.02	.07	.01	.04
Experiencia	.27*	.30*	.14	.39*
Destrezas	.20	.20	-.07	.17
Adaptación y Afrontamiento				
Formación	.12	.23	-.08	.00
Actitudes	-.03	.15	-.03	.06
Experiencia	.23	.30*	.07	.32*
Destrezas	.12	.24	-.03	.22
Logro y emprendimiento				
Formación	.03	.16	.01	.03
Actitudes	-.09	.29*	-.01	.01
Experiencia	.24	.41*	.19	.32*
Destrezas	.16	.40*	.08	.33*

Nota: * $p < .05$; ** $p < .01$

muestra una fuerte relación con los componentes experiencial y comportamental de las competencias, pero no con los componentes formativos ni con los actitudinales.

Más específicamente, la puntuación en capital social (LKBF2) muestra relación significativa y positiva con los componentes de experiencia en todas las competencias, denotando que el desarrollo por parte de los participantes de posiciones laborales que les permiten desarrollar sus competencias corre parejo con el desarrollo de su capital social expresado en LinkedIn. De igual forma, la relación entre LKBF2 y los comportamientos desarrollados por los participantes es fuerte y positiva para determinadas competencias. Así, se muestra cómo la mayor amplitud del capital social de los participantes parece indicar también un mayor desarrollo de sus comportamientos de Liderazgo y Toma de Decisiones, Cooperación y Relación con Otros, Análisis y Utilización de conocimiento, Aprendizaje e Innovación, y Logro y Emprendimiento. Por otro lado, la dimensión de amplitud de la experiencia profesional (LKBF1) refleja bien la experiencia adquirida por los participantes en posiciones laborales que les han permitido desarrollar las competencias de Liderazgo y Toma de Decisiones, Análisis y Utilización de conocimiento y Planificación y Organización. Por otro lado, la dimensión relativa a la amplitud de la información no profesional representada en LinkedIn (LKBF4) parece reflejar la experiencia adquirida por los participantes en posiciones laborales que favorecen el desarrollo de la Cooperación y el Respeto, la Comunicación y la Relación con otros, el Análisis y la Utilización de conocimientos, la Planificación y Organización, la Adaptación y el Logro y el Emprendimiento.

Estos hallazgos parecen estar en línea con los resultados de estudios anteriores donde se ofrece evidencia sobre la relación entre la experiencia laboral y el capital social y las competencias personales relacionadas con el pensamiento estratégico, el pensamiento emprendedor, la dirección de personas y el liderazgo, la toma de decisiones y la resolución de problemas (Dragoni et al., 2011; Gabriellson & Politis, 2012; Krueger, 2007; Lord & Hall, 2005; Mumford et al., 2000).

Por último, la medida realizada de la dimensión LKBF3 (interés por mantener los conocimientos actualizados) no tiene la misma capacidad que las otras tres para reflejar los aspectos competenciales evaluados. En resumen, los resultados parecen mostrar que las rúbricas presentadas pueden ser un instrumento útil para explorar el desarrollo de competencias que los usuarios de los perfiles de LinkedIn han obtenido a través de la experiencia.

Lo anterior tiene varias implicaciones prácticas para el profesional del reclutamiento y la selección. La primera de ellas es que puede utilizar una herramienta estándar para realizar su aproximación a los perfiles de LinkedIn de los candidatos. Esta estandarización permite a los profesionales realizar comparaciones entre los diferentes perfiles profesionales valorados. La segunda implicación tiene que ver con la posibilidad de que los profesionales, utilizando las medidas proporcionadas por las rúbricas, puedan realizar inferencias sobre las competencias de los candidatos apoyadas en los resultados presentados.

Nuestro estudio no está exento de algunas limitaciones. En concreto, solo se contempla la información de carácter estructural presente en el perfil y, en ningún caso, el contenido específico presentado en el perfil. Contemplar dicha información proporcionaría información adicional que también debiera conectarse con las competencias ad-

quiridas por los usuarios de los perfiles. En este sentido, estudios posteriores podrían indagar en la relación existente entre ese contenido específico informado en LinkedIn y las competencias de los usuarios.

CONCLUSIONES

Los resultados presentados en este estudio suponen evidencia inicial de la calidad de las rúbricas desarrolladas para la medición de las cuatro dimensiones del modelo LinkedIn Big Four. Profesionales y académicos tienen en ellas una herramienta estándar para explorar estas dimensiones y utilizar las medidas con ellas obtenidas para los diferentes procesos de toma de decisiones que se producen en el ámbito aplicado de la evaluación de personas en los contextos organizacionales.

Además, parece evidente que las SNW han venido para quedarse y que los profesionales de la psicología en el desarrollo de sus tareas de carácter evaluativo y/o diagnóstico, tienen en ellas una fuente importante de información acerca del comportamiento de los individuos. Nuestro estudio pone de manifiesto que la forma en la que los usuarios de LinkedIn trabajan con su perfil puede ser entendida como una expresión de sus competencias individuales. El modo de acercamiento al perfil de LinkedIn a través de un conjunto de medidas estandarizadas a través de las rúbricas, bien pudiera trasladarse a la valoración de la información presente en otras redes sociales, con otros objetivos evaluativos diferentes de la selección de personal. El comportamiento de los individuos en las redes sociales es, en definitiva, una expresión particular de las diferencias individuales que caracterizan a cada persona. Y, desde este punto de vista, el psicólogo tiene en ellas un recurso de gran valor. Su utilización, en cualquier caso, debe estar sujeta al menos a dos principios generales: (a) utilización acorde a las convenciones sociales y legales relacionadas con la protección de la privacidad; y (b) exploración a partir de instrumentos que permitan obtener una información válida y fiable respecto de los objetivos del profesional.

CONFLICTO DE INTERESES

No existe conflicto de intereses

REFERENCIAS

- Aguado, D., Andrés, J. C., García-Izquierdo, A. L., & Rodríguez, J. (2019). LinkedIn "Big Four": Job Performance Validation in the ICT Sector. *Journal of Work and Organizational Psychology*, 35(2), 53-64. <https://doi.org/10.5093/jwop2019a7>
- Aguado, D., Rico, R., Rubio, V. J., & Fernández, L. (2016). Applicant reactions to social network web use in personnel selection and assessment. *Journal of Work and Organizational Psychology*, 32(3), 183-190. <https://doi.org/10.1016/j.rp-to.2016.09.001>
- Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, 21(3), 372-374. <https://doi.org/10.1177/0956797609360756>
- Bartram, D. (2005). The Great Eight competencies: a criterion-centric approach to validation. *Journal of Applied Psychology*, 90(6), 1185-1203. <https://doi.org/10.1037/0021-9010.90.6.1185>
- Black, S. L., & Johnson, A. F. (2012). Employers' use of social net-

- working sites in the selection process. *The Journal of Social Media in Society*, 1(1), 7-28.
- Boyd, D., & Ellison N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Chamorro-Premuzic, T., & Steinmetz, C. (2013). Technology and psychology are reshaping the search for the best employees. *Scientific American Mind*, 1(1), 43-47. <https://fermion.com.au/wp-content/uploads/The-Perfect-Hire-1.pdf>
- Chapman, D. S., & Mayers, D. (2015). Recruitment processes and organizational attraction. In *Employee Recruitment, Selection, and Assessment* (2nd ed., pp. 39-54). Psychology Press. <https://doi.org/10.4324/9781315742175>
- Chiang, J. K. H., & Suen, H. Y. (2015). Self-presentation and hiring recommendations in online communities: Lessons from LinkedIn. *Computers in Human Behavior*, 48, 516-524. <http://dx.doi.org/10.1016/j.chb.2015.02.017>
- Curran, V., Hollett, A., Casimiro, L. M., McCarthy, P., Banfield, V., Hall, P., Lackie, K., Oandasan, I., Simmons, B. & Wagner, S. (2011). Development and validation of the interprofessional collaborator assessment rubric (ICAR). *Journal of Interprofessional Care*, 25(5), 339-344. <http://dx.doi.org/10.3109/13561820.2011.589542>
- Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347-360. <http://dx.doi.org/10.1080/02602938.2015.1111294>
- Derous, E., & De Fruyt, F. (2016). Developments in recruitment and selection research. *International Journal of Selection and Assessment*, 24(1), 1-3. <https://doi.org/10.1111/ijsa.12123>
- Dragoni, L., Oh, I. S., Vankatwyk, P., & Tesluk, P. E. (2011). Developing executive leaders: The relative contribution of cognitive ability, personality, and the accumulation of work experience in predicting strategic thinking competency. *Personnel Psychology*, 64(4), 829-864. <http://dx.doi.org/10.1111/j.1744-6570.2011.01229.x>
- 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>
- Frasca, K. J., & Edwards, M. R. (2017). Web-based corporate, social and video recruitment media: Effects of media richness and source credibility on organizational attraction. *International Journal of Selection and Assessment*, 25(2), 125-137. <http://dx.doi.org/10.1111/ijsa.12165>
- Gabrielsson, J., & Politis, D. (2012). Work experience and the generation of new business ideas among entrepreneurs: An integrated learning framework. *International Journal of Entrepreneurial Behavior & Research* 18(1), 48. <http://dx.doi.org/10.1108/13552551211201376>
- Galanaki, E. (2002). The decision to recruit online: A descriptive study. *Career Development International*, 7(4), 243-251. <https://doi.org/10.1108/13620430210431325>
- García-Izquierdo, A. L., Ramos-Villagrasa, P. J., & Castaño, A. M. (2015). e-Recruitment, gender discrimination, and organizational results of listed companies on the Spanish Stock Exchange. *Revista de Psicología del Trabajo y de las Organizaciones*, 31(3), 155-164. <https://dx.doi.org/10.1016/j.rpto.2015.06.003>
- García-Izquierdo, A. L., Aguado, D., & Ponsoda-Gil, V. (2019). New insights on technology and assessment: introduction to JWOP special issue. *Revista de Psicología del Trabajo y de las Organizaciones*, 35(2), 49-52. <https://doi.org/10.5093/jwop2019a6>
- Ghazzawi, K., & Accoume, A. (2014). Critical success factors of the e-recruitment system. *Journal of Human Resources Management and Labor Studies*, 2(2), 159-170.
- Hung, H. T., Chiu, Y. C. J., & Yeh, H. C. (2013). Multimodal assessment of and for learning: A theory-driven design rubric. *British Journal of Educational Technology*, 44(3), 400-409. <http://dx.doi.org/10.1111/j.1467-8535.2012.01337.x>
- Instituto de Ingeniería del Conocimiento (2015). PRISMA 4D. Retrieved from <https://www.iic.uam.es/soluciones/recursos-humanos/evaluacion-por-perfiles-profesionales/>
- Kashi, K., & Zheng, C. (2013). Extending Technology Acceptance Model to the E-recruitment Context in Iran. *International Journal of Selection and Assessment*, 21(1), 121-129. <https://doi.org/10.1111/ijsa.12022>
- Kaur, P. (2015). E-recruitment: A conceptual study. *International Journal of Applied Research*, 1(8), 78-82.
- Kluemper, D. H., Rosen, P. A., & Mossholder, K. W. (2012). Social networking websites, personality ratings, and the organizational context: More than meets the eye? *Journal of Applied Social Psychology*, 42(1), 1143-1172. <https://doi.org/10.1111/j.1559-1816.2011.00881.x>
- Krueger Jr, N. F. (2007). What lies beneath? The experiential essence of entrepreneurial thinking. *Entrepreneurship Theory and Practice*, 31(1), 123-138. <http://dx.doi.org/10.1111/j.1540-6520.2007.00166.x>
- Lievens, F., & Harris, M. M. (2003). Research on Internet recruiting and testing: Current status and future directions. In C. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (1st ed., 131-166). Wiley Online Library. <https://doi.org/10.1002/0470013346.ch4>
- Lord, R. G., & Hall, R. J. (2005). Identity, deep structure and the development of leadership skill. *The Leadership Quarterly*, 16(4), 591-615. <http://dx.doi.org/10.1016/j.leaqua.2005.06.003>
- McCabe, M. B. (2017). Social media marketing strategies for career advancement: An analysis of LinkedIn. *Journal of Business and Behavioral Sciences*, 29(1), 85-149.
- Meade, A. W. (2010). A taxonomy of effect size measures for the differential functioning of items and scales. *Journal of Applied Psychology*, 95(4), 728. <http://dx.doi.org/10.1037/a0018966>
- Mumford, M. D., Marks, M. A., Connelly, M. S., Zaccaro, S. J., & Reiter-Palmon, R. (2000). Development of leadership skills: Experience and timing. *The Leadership Quarterly*, 11(1), 87-114. [https://doi.org/10.1016/S1048-9843\(99\)00044-2](https://doi.org/10.1016/S1048-9843(99)00044-2)
- Nadelson, L. S., Pfister, J., Callahan, J., & Pyke, P. (2015). Who is doing the engineering, the student or the teacher? The development and use of a rubric to categorize level of design for the elementary classroom. *Journal of Technology Education*, 26(2), 22-45. <http://dx.doi.org/10.21061/jte.v26i2.a2>
- Nikolaou, I. (2014). Social networking web sites in job search and employee recruitment. *International Journal of Selection and Assessment*

- essment, 22(2), 179-189. <https://doi.org/10.1111/ij-sa.12067>
- Ötting, S. K., & Maier, G. W. (2018). The importance of procedural justice in human-machine interactions: Intelligent systems as new decision agents in organizations. *Computers in Human Behavior*, 89, 27-39. <https://doi.org/10.1016/j.chb.2018.07.022>
- Riebe, L., & Jackson, D. (2014). The use of rubrics in benchmarking and assessing employability skills. *Journal of Management Education*, 38(3), 319-344. <http://dx.doi.org/10.1177/1052562913511437>
- Roblyer, M. D., & Wiencke, W. R. (2003). Design and use of a rubric to assess and encourage interactive qualities in distance courses. *The American Journal of Distance Education*, 17(2), 77-98. http://dx.doi.org/10.1207/S15389286AJDE1702_2
- Roulin, N. (2014). The influence of employers' use of social networking websites in selection, online self-promotion, and personality on the likelihood of faux pas postings. *International Journal of Selection and Assessment*, 22(1), 80-87. <https://doi.org/10.1111/ij-sa.12058>
- Roulin, N., & Bangerter, A. (2013). Social networking websites in personnel selection. *Journal of Personnel Psychology* 12(1), 143-151. <https://doi.org/10.1027/1866-5888/a000094>
- Roulin, N., & Levashina, J. (2019). LinkedIn as a new selection method: Psychometric properties and assessment approach. *Personnel Psychology*, 72(2), 187-211. <https://doi.org/10.1111/peps.12296>
- Ryan, A. M., & Derous, E. (2016). Highlighting tensions in recruitment and selection research and practice. *International Journal of Selection and Assessment*, 24(1), 54-62.
- Ryan, A. M., & Derous, E. (2019). The unrealized potential of technology in selection assessment. *Journal of Work and Organizational Psychology*, 35(2), 85-92. <https://doi.org/10.5093/jwop2019a10>
- Seiter, J. S., & Hatch, S. (2005). Effect of tattoos on perceptions of credibility and attractiveness. *Psychological Reports*, 96(3_suppl), 1113-1120. <http://dx.doi.org/10.2466/pr0.96.3c.1113-1120>
- Shahani-Denning, C., Patel, V., & Zide, J. (2017). Recruiter and applicant use of LinkedIn: A spotlight on India. *The Psychologist-Manager Journal*, 20(2), 90-105. <https://doi.org/10.1037/mgr0000052>
- Shannon, M. L., & Stark, C. P. (2003). The influence of physical appearance on personnel selection. *Social Behavior and Personality: An International Journal*, 31(6), 613-623. <https://doi.org/10.2224/sbp.2003.31.6.613>
- Stellmack, M. A., Konheim-Kalkstein, Y. L., Manor, J. E., Massey, A. R., & Schmitz, J. A. P. (2009). An assessment of reliability and validity of a rubric for grading APA-style introductions. *Teaching of Psychology*, 36(2), 102-107. <http://dx.doi.org/10.1080/00986280902739776>
- Sylva, H., & Mol, S. T. (2009). E-Recruitment: A study into applicant perceptions of an online application system. *International Journal of Selection and Assessment*, 17(3), 311-323. <https://doi.org/10.1111/j.1468-2389.2009.00473.x>
- Unal, Z., Bodur, Y., & Unal, A. (2012). A standardized rubric for evaluating Webquest design: reliability analysis of ZUNAL Webquest design rubric. *Journal of Information Technology Education: Research*, 11(1), 169-183. <http://dx.doi.org/10.28945/1688>
- Van Iddekinge, C. H., Lanivich, S. E., Roth, P. L., & Junco, E. (2016). Social media for selection? Validity and adverse impact potential of a Facebook-based assessment. *Journal of Management*, 42(7), 1811-1835. <http://dx.doi.org/10.1177/0149206313515524>
- Vaughan, B., Yoxall, J., & Grace, S. (2019). Peer assessment of teamwork in group projects: Evaluation of a rubric. *Issues in Educational Research*, 29(3), 961-978.
- Villeda, M., McCamey, R., Essien, E., & Amadi, C. (2019). Use of social networking sites for recruiting and selecting in the hiring process. *International Business Research*, 12(3), 66-78. <http://dx.doi.org/10.5539/ibr.v12n3p66>
- Wenzlaff, T. L., Fager, J. J., & Coleman, M. J. (1999). What is a rubric? Do practitioners and the literature agree?. *Contemporary Education*, 70(4), 41.
- Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, 29(1), 64-77. <https://doi.org/10.1080/1359432X.2019.1681401>

Apéndice I

Rúbricas para la recogida de la información contenida en los perfiles de LinkedIn

Rúbrica 1: Amplitud de la Experiencia Profesional

Elemento 1: número de experiencias distintas reflejadas en el perfil (contar las distintas experiencias reflejadas por el candidato en su perfil). 1 (puntuación = 1); 2 (puntuación = 2); 3 (puntuación = 3); de 4 a 5 (puntuación = 4); > 5 (puntuación = 5)

Elemento 2: número roles distintos desempeñados en la actividad profesional (contar los roles distintos en cuanto a puesto, categoría, etc.). 1 (puntuación = 1); 2 (puntuación = 2); 3 (puntuación = 3); de 4 a 5 (puntuación = 4); > 5 (puntuación = 5)

Elemento 3: número de empresas en las que ha desempeñado actividad profesional (contar las empresas en las que ha tenido actividad profesional). 1 (puntuación = 1); 2 (puntuación = 2); 3 (puntuación = 3); de 4 a 5 (puntuación = 4); > 5 (puntuación = 5)

Elemento 4: número de líneas dedicadas a describir la experiencia profesional (contar las líneas en las que se extiende el apartado de experiencia profesional (se cuentan las líneas en el formato que aparece en el perfil). De 0 a 3 (puntuación = 1); de 4 a 7 (puntuación = 2); de 8 a 10 (puntuación = 3); de 11 a 13 (puntuación = 4); de 14 a 18 (puntuación = 5); de 19 a 24 (puntuación = 6); de 25 a 32 (puntuación = 7); de 33 a 42 (puntuación = 8); de 43 a 62 (puntuación = 9); y > 62 (puntuación = 10)

Apéndice I

Rúbricas para la recogida de la información contenida en los perfiles de LinkedIn (Continuación)

Elemento 5: número de meses que suman la experiencia laboral del candidato (sumar en meses la duración de las distintas actividades profesionales señaladas en el perfil). De 0 a 49 (puntuación = 1); de 50 a 79 (puntuación = 2); de 80 a 102 (puntuación = 3); de 103 a 116 (puntuación = 4); de 117 a 131 (puntuación = 5); de 132 a 148 (puntuación = 6); de 149 a 179 (puntuación = 7); de 180 a 208 (puntuación = 8); de 209 a 250 (puntuación = 9); y > 250 (puntuación = 10)

Rúbrica 2: Capital Social

Elemento 1: número de contactos que tiene el candidato en su perfil. De 0 a 23 (puntuación = 1); de 24 a 49 (puntuación = 2); de 50 a 74 (puntuación = 3); de 75 a 91 (puntuación = 4); de 92 a 112 (puntuación = 5); de 113 a 138 (puntuación = 6); de 139 a 169 (puntuación = 7); de 170 a 213 (puntuación = 8); de 214 a 311 (puntuación = 9); > 312 (puntuación = 10)

Elemento 2: número de validaciones de aptitudes (sumar el número de aptitudes validadas en total). 0 (puntuación = 1); de 1 a 17 (puntuación = 2); de 18 a 44 (puntuación = 3); de 45 a 89 (puntuación = 4); > 89 (puntuación = 5)

Elemento 3: recomendaciones recibidas por el candidato (referir si el candidato ha recibido alguna recomendación). No ha recibido recomendaciones (puntuación = 0); sí ha recibido recomendaciones (puntuación = 1)

Elemento 4: número de empresas a las que sigue el candidato (contar las empresas a las que sigue el candidato en el apartado intereses). De 0 a 2 (puntuación = 1); 3 (puntuación = 2); de 4 a 5 (puntuación = 3); de 6 a 12 (puntuación = 4); y > 12 (puntuación = 5)

Elemento 5: número de grupos a los que sigue el candidato (contar los grupos a los que sigue el candidato en el apartado intereses). 0 (puntuación = 1); 1 (puntuación = 2); de 2 a 3 (puntuación = 3); de 4 a 7 (puntuación = 4); y > 7 (puntuación = 5)

Elemento 6: número de noticias que sigue el candidato (referir si el candidato sigue alguna noticia o no en el apartado intereses). No sigue noticia alguna (puntuación = 0); sigue alguna noticia (puntuación = 1)

Rúbrica 3: Interés en mantener la Formación Actualizada

Elemento 1: número de titulaciones universitarias (grado o posgrado) referidos por el candidato (contar las distintas formaciones universitarias reflejadas por el candidato en su perfil). 0 (puntuación = 1); 1 (puntuación = 2); 2 (puntuación = 3); 3 (puntuación = 4); > 3 (puntuación = 5)

Elemento 2: número de cursos adicionales a la formación universitaria referidos por el candidato (referir si el candidato ha realizado algún curso adicional a la formación universitaria). No refiere ningún curso adicional (puntuación = 0); refiere algún curso adicional (puntuación = 1)

Elemento 3: número de universidades a las que sigue el candidato (referir si el candidato sigue a alguna universidad). No sigue a ninguna universidad (puntuación = 0); sigue a alguna universidad (puntuación = 1)

Rúbrica 4: Amplitud de la Información No Profesional

Elemento 1: número de categorías rellenas en el perfil que muestran información sobre el candidato (contar las categorías que el candidato ha relleno). De 0 a 4 (puntuación = 1); de 5 a 6 (puntuación = 2); 7 (puntuación = 3); 8 (puntuación = 4); y > 8 (puntuación = 5)

Elemento 2: número de idiomas que refiere conocer el candidato, excepto el español (contar los idiomas referidos, excepto el español). 0 (puntuación = 1); 1 (puntuación = 2); 2 (puntuación = 3); 3 (puntuación = 4); y > 3 (puntuación = 5)

Elemento 3: intereses referidos por el candidato (verificar si el candidato ha referido intereses o no). No refiere ningún interés en su perfil (puntuación = 0); refiere algún interés en su perfil (puntuación = 1)

Elemento 4: causas benéficas referidas por el candidato (verificar si el candidato ha referido causas benéficas o no). No refiere ninguna causa benéfica en su perfil (puntuación = 0); refiere alguna causa benéfica en su perfil (puntuación = 1)

Elemento 5: existencia de la sección "Acerca de" o "Extracto" (verificar si el candidato ha relleno la sección "Acerca de" o "Extracto"). No ha relleno la sección "Acerca de" o "Extracto" (puntuación = 0); ha relleno la sección "Acerca de" o "Extracto" (puntuación = 1)

La hora de la evaluación ambulatoria

The time of ambulatory assessment

Eduardo Fonseca-Pedrero¹, Gabriel Ródenas-Perea², Alicia Pérez-Albéniz¹, Susana Al-Halabi³, Marino Pérez³ y José Muñiz⁴

¹Universidad de La Rioja, España. ²Hospital Virgen del Rocío, España. ³Universidad de Oviedo, España. ⁴Universidad Nebrija, España

La evaluación ambulatoria aglutina un conjunto de métodos que permiten evaluar mediante dispositivos móviles, y en múltiples momentos temporales, el comportamiento de las personas en su entorno natural y contexto diario. Permite una evaluación más precisa, dinámica, contextual e ideográfica que los métodos clásicos, abriendo nuevos horizontes con claras implicaciones para el diagnóstico y la intervención psicológica. El objetivo de este trabajo es realizar una introducción a la evaluación ambulatoria. En primer lugar, se realiza una delimitación conceptual y se comentan las cuestiones que viene a solucionar y sus posibles beneficios. En segundo lugar, se exponen aspectos relacionados con la metodología, abordando los diseños, los tipos de datos y un protocolo general de evaluación. En tercer lugar, se comentan algunas de las principales limitaciones, y se exponen las aplicaciones más relevantes. Finalmente, se comentan algunas recomendaciones para la aplicación de este tipo de metodología, y se analizan los retos y perspectivas futuras.

Palabras clave: Evaluación ambulatoria, Muestreo de experiencias, Evaluación ecológica momentánea, Modelo contextual, Dispositivos electrónicos.

Ambulatory assessment brings together a set of methods that make it possible to evaluate, through mobile devices, and at multiple moments in time, the behavior of people in their natural environment and daily context. It allows a more precise, dynamic, contextual, and ideographic evaluation than the classical approaches, opening new horizons with clear implications for psychological intervention. The main goal of this paper is to provide an introduction to ambulatory assessment. First, a conceptual delimitation is made and the issues to be solved are discussed, as well as their possible benefits. Second, aspects related to the methodology are exposed, addressing the designs, data types, and a general evaluation protocol. Third, some of the main limitations are discussed, and the most relevant applications are presented. Finally, some recommendations for the application of this type of methodology are discussed, and some challenges and future perspectives are analyzed.

Key words: Ambulatory assessment, Experience sampling methodology, Ecological momentary assessment, Contextual model, Electronics devices.

La Psicología, como ciencia y como profesión, se encuentra en continua mudanza, evolución y metamorfosis (Pérez-Álvarez, 2018). La incorporación de nuevos modelos psicológicos, el rescate de algunos injustamente olvidados, las tecnologías de la información, o los diferentes avances científicos permiten abrir prometedores horizontes. Como ocurre en otras ciencias y profesiones, hay numerosas cuestiones que aún continúan sin una respuesta definitiva, y que suponen un reto o desafío para el progreso de la Psicología, y, entre ellas, se encuentra el problema de la evaluación que se aborda en el presente trabajo. El tipo de evaluación que se lleva a cabo, además de los problemas tecnológicos inherentes a su aplicación, tiene repercusiones directas sobre el tipo de modelos que se utilicen para entender el comportamiento humano. El tipo de evaluación no es neutra, condiciona, incluso puede determinar, la teorización psicológica. Como se verá más adelante, los avances que supone la evaluación ambulatoria van a permitir formular modelos más realistas y ajustados del comportamiento.

Actualmente, la mayoría de los procedimientos de evaluación en psicología se basan en el análisis descriptivo de conductas, tales como afectos, estados, rasgos, síntomas, etc., y se llevan a cabo mediante el uso de test de lápiz y papel o entrevistas. Además, se

administran en contextos artificiales, como consultas, centros de salud, laboratorios, aulas, etc. Se trata generalmente de evaluaciones estáticas, llevadas a cabo en amplios intervalos temporales y con poca frecuencia, lo que conduce a una colección de indicadores únicos e instantáneos del comportamiento. Sin embargo, los grandes avances actuales de las tecnologías de la información permiten llevar a cabo evaluaciones con mayor validez ecológica, imprescindibles para realizar diagnósticos precisos, seguimientos personalizados, o formular intervenciones eficaces, eficientes y efectivas (Muñiz y Fonseca-Pedrero, 2019). Al lado de una metodología evaluativa estática, anclada en el pasado, y la utilización del modelo biomédico en psicología, puede conducir, en algunos campos como la psicopatología, a la reificación, la cosificación y a razonamientos tautológicos (Pérez-Álvarez, 2012). Un caso paradigmático sería la psicosis, que tras décadas de estudio bajo la óptica del modelo biomédico no se ha hallado la supuesta alteración biológica que la genera (p. ej., Fonseca-Pedrero, 2019; McCutcheon et al., 2021; Pérez-Álvarez, 2012, Radua et al., 2018). Y eso es así porque los signos y síntomas de los problemas psicológicos, lejos de constituir indicadores patognómicos de una supuesta enfermedad mental, como postula el modelo biomédico, forman parte de la diversidad humana, y deben comprenderse como formas de responder y afrontar las diferentes vicisitudes de la vida (Pérez-Álvarez, 2019). La conducta humana se explica en función de su biografía particular, en un momento temporal y contexto socio-cultural determinado, por lo que hay que analizarla en el contexto de la persona y sus circunstancias, comprender las razones, más

Recibido: 1 noviembre 2021 - Aceptado: 16 diciembre 2021

Correspondencia: Eduardo Fonseca-Pedrero. Universidad de La Rioja. C/ Luis de Ulloa, s/n, Edificio VIVES. 26002 Logroño. La Rioja. España. E-mail: eduardo.fonseca@unirioja.es

que las causas (Fonseca-Pedrero, 2021a, 2021b). Por ello, se necesitan modelos alternativos que recojan la naturaleza dinámica y contextual del comportamiento humano, en la vida real y cotidiana, considerando al individuo concreto (Borsboom, 2017; Fonseca-Pedrero, 2018; Myin-Germeys et al., 2016; Pérez-Álvarez, 2018, 2020; Pérez-Álvarez y García-Montes, 2018; Trull y Ebner-Priemer, 2020, van Os et al., 2013). Estos nuevos planteamientos conceptuales exigen también un cambio de metodología evaluativa, que vaya más allá de las evaluaciones tradicionales, y que reciba el nombre genérico de evaluación ambulatoria (EA). La idea de la EA no es nueva conceptualmente, pero ha recibido un gran impulso con el desarrollo de nuevas tecnologías como los *smartphone*, los dispositivos móviles y las aplicaciones de diverso tipo (*apps*), que facilitan la evaluación a distancia en contextos naturales. El uso de esta tecnología está permitiendo nuevas formas de evaluación e intervención de los fenómenos psicológicos, potenciando la incorporación de modelos, herramientas y procedimientos que permitan implementar enfoques psicológicos contextuales, dinámicos, idiográficos, adaptativos, fenomenológicos, funcionales y multicausales.

En el presente trabajo se pretende hacer una introducción comprensiva de la EA. En primer lugar, se realiza una delimitación conceptual y se comentan las cuestiones que viene a solucionar y los posibles beneficios. En segundo lugar, se exponen aspectos relacionados con la metodología, como los diseños, los tipos de datos y un protocolo general de evaluación. En tercer lugar, se comentan algunas de las principales limitaciones, luego se explican algunas aplicaciones al campo de la Psicología, y se introduce una reciente plataforma desarrollada para su uso en el contexto español. A continuación, se ofrecen algunas recomendaciones prácticas sobre cómo implementar este tipo de metodología, y se finaliza con un apartado de recapitulación, donde se analizan algunos retos y perspectivas futuras. No se pretende en este breve trabajo llevar a cabo una exposición exhaustiva, remitiendo para ello al lector a trabajos más especializados (Carpenter et al., 2016; De Moor et al., 2009; Hektner et al., 2007; Hirschtritt y Insel, 2018; Mehl, 2012; Myin-Germeys et al., 2018; Myin-Germeys y Kuppens, 2021; Russell y Gajos, 2020; Stone y Shiffman, 1994; Torous et al., 2021; Trull y Ebner-Priemer, 2013, 2020).

DELIMITACIÓN CONCEPTUAL

Se puede concebir la EA como un procedimiento sistemático y estructurado de observación del comportamiento (afectos, cognición, experiencias, procesos, estados mentales, etc.) de una persona en el contexto natural y real, mediante un dispositivo, generalmente electrónico, como PDA, *smartphone*, etc. La EA abarca una amplia gama de métodos de evaluación que tratan de estudiar las experiencias de las personas en su entorno natural, en su vida diaria. Según Trull y Ebner (2013) se utiliza el término EA para representar un paraguas conceptual y metodológico que incluye el método de muestreo de experiencias (*Experience Sampling Methodology*, ESM) (Csikszentmihalyi y Larson, 1987), la evaluación ecológica momentánea (*Ecological Momentary Assessment*, EMA), o simplemente evaluación momentánea (Stone y Shiffman, 1994), y los registros psicofisiológicos, biológicos y de comportamiento. Históricamente, la ESM se ha relacionado con el uso de diarios de papel y lápiz, la EMA se imple-

menta generalmente mediante diarios electrónicos o teléfonos móviles, y los registros psicofisiológicos se obtienen usando sensores o actígrafos. Algunos autores procedentes del área de la salud mental han denominado este campo como *fenotipo digital* (Insel, 2017; Torous et al., 2017).

De forma sintética, puede decirse que la EA se caracteriza básicamente por:

- Constituir un enfoque idiográfico que permite el examen de múltiples procesos individuales (p. ej., emocionales, conductuales, psicofisiológicos, estados mentales).
- Recoger información de la persona en entornos del mundo real, en la vida cotidiana de las personas, en su hábitat natural.
- Evaluar los comportamientos (estados, experiencias, procesos psicológicos, etc.) actuales (o muy recientes) o en el tiempo real (en el momento en el que ocurren, de momento a momento) de los individuos.
- Recoger información mediante evaluaciones múltiples (de forma intensiva) de cada individuo en el tiempo (típicamente varias veces al día, varias veces a la semana).

Como ya se ha señalado, la EA trata de dar respuesta a ciertas limitaciones de la evaluación psicológica clásica, incorporando la naturaleza ideográfica del comportamiento, el contexto cotidiano donde ocurre la conducta de las personas, la naturaleza contextual del comportamiento (ambiente físico, ambiente social, cultura), y la naturaleza dinámica y los mecanismos etiológicos del comportamiento humano.

Por lo tanto, la EA pretende evaluar:

- A las personas teniendo en cuenta sus circunstancias biográficas y contextuales, se trata de un enfoque personalizado.
- A la persona en relación e interdependencia con el contexto de su vida cotidiana conforme ocurren sus comportamientos en el flujo diario. La posibilidad de evaluar factores relevantes del contexto donde acontecen los fenómenos psicológicos de interés, así como las respuestas de la persona a los mismos son de sumo interés. Se entiende que, para explicar la conducta o las experiencias de las personas, es necesario analizarlas en interacción con el contexto.
- El comportamiento de las personas en toda su amplitud, complementando los autoinformes con heteroinformes y registros psicofisiológicos, biológicos y de comportamiento. Y no solo desde una perspectiva cuantitativa, sino también cualitativa o mixta (p. ej., tipo de pensamientos, forma y contenido de las alucinaciones o delirios, naturaleza y gravedad de la ideación suicida).
- La evolución y dinámicas de cambio de la conducta en el transcurso del tiempo. Se trata de captar la variabilidad temporal mediante la valoración momentánea de las fluctuaciones de momento a momento de los comportamientos (estados mentales, experiencias, etc.), es decir, los cambios y la evolución a nivel micro.

Se trata, en suma, de una evaluación de carácter intensivo y micro-longitudinal que a) minimice el sesgo asociado a la evaluación retrospectiva clásica, b) permita analizar los patrones personales de variación (intra, inter, y contextuales) y comprender cómo se desarrollan y despliegan (o no) en el transcurso del tiempo, c) mejore aspectos relacionados con la precisión (fiabilidad) y validez de las puntuaciones, y d) permita analizar los posibles mecanismos de tipo

conductual, aspecto que contribuye a una comprensión más profunda y minuciosa (radical) del comportamiento humano.

En síntesis, la EA permite una evaluación más precisa, dinámica, contextual e ideográfica, capturando de forma más exacta y rigurosa la compleja naturaleza del comportamiento humano. Su objetivo sería la comprensión de los procesos psicológicos dinámicos a medida que se desarrollan en la vida cotidiana y en función de las circunstancias (necesariamente contextuales) de las personas (véase Figura 1). Igualmente, permite indagar sobre los posibles mecanismos de cambio, más allá de la mera descripción clínica o topográfica. Estos aspectos permitirían, en el campo de la Psicología clínica, por ejemplo, predecir la recurrencia o aparición de experiencias de malestar en el futuro, monitorizar la intervención, controlar los efectos secundarios del tratamiento, predecir el éxito del tratamiento, prevenir la posible recaída o brindar un tratamiento individualizado donde y cuando sea necesario.

DISEÑO Y PROCEDIMIENTO

La EA consiste en un procedimiento sistemático y estructurado de observación del comportamiento humano. La variabilidad en cuanto a tipos de diseños de investigación, tipo de datos que se obtiene, y posibles maneras de formular los ítems y las preguntas es enorme. La utilización de uno u otro vendrá determinada por el objetivo y el fenómeno estudiado, así como por posibles variables relacionadas (disponibilidad de respuesta del o de la participante, acceso a internet, etc.). Por ejemplo, si el interés es evaluar la ideación suicida en adolescentes que acuden al centro escolar, tal vez carecería de sentido realizar la EA mientras se encuentran en la jornada escolar o haciendo una actividad extraescolar, y dejarla para otros contextos como el de la familia o el ocio.

En cuanto al diseño de investigación, se pueden encontrar diseños evento-contingente, los diseños temporales y diseños mixtos. En los diseños evento-contingente la "activación" de la respuesta a las preguntas viene determinada por la ocurrencia de un evento (p. ej., la persona comienza a tener pensamientos obsesivos y esto hace que responda a las preguntas). En los diseños temporales, la "activación" se realiza en función del tiempo, esto es, se utiliza el tiempo como la unidad de muestreo. En este caso, la persona responde cada cierto momento a lo largo del día (p. ej. de forma aleatoria, cada 90 minutos, se activan las preguntas en el teléfono de la persona). Los diseños mixtos consisten en la combinación de los diseños evento-contingente y temporal. En cuanto a los datos, pueden ser activos/pasivos, así como objetivos/subjetivos. Entre los datos denominados activos se podría encontrar el autoinforme o las respuestas a un test, mientras que entre los pasivos estarían la información proveniente del GPS, las horas sueño, la tasa cardiaca, etc. En cuanto a los datos objetivos/subjetivos, serían datos objetivos aquellos como la tasa cardiaca, número de pasos, etc., mientras que los datos subjetivos, como su nombre indica, serían aquellos recogidos mediante autoinforme. Otras clasificaciones consideran si los datos son de carácter autoinformado, observacional, psicofisiológico (p. ej., actividad electrodérmica de la piel), biológico (p. ej., cortisol) y conductual (p. ej., motor).

La heterogeneidad es la regla a la hora de diseñar un estudio que implemente la EA (Smyth et al., 2021). Aquí únicamente nos vamos

a referir a un ejemplo prototípico. En un estudio se pueden realizar evaluaciones varias veces al día durante un periodo temporal determinado (aproximadamente 6-8 veces por día durante 7 días). Las preguntas se activan fijando un intervalo temporal concreto (p. ej., entre las 10:00h y las 22:00h) y se presentan de forma aleatoria en intervalos de tiempo predeterminados, por ejemplo, cada 90 minutos. Este conjunto de preguntas se suele completar, aproximadamente, en 1-2 minutos. Las preguntas se establecen en función del objeto de estudio, siendo la escala tipo *Likert* el formato más utilizado. En la Figura 2 se representa un ejemplo del método de muestreo de experiencias. En la Figura 3 se recogen dos ítems a modo de ejemplo. Con frecuencia los ítems que se formulan suelen comentar o incluir

FIGURA 1
EJEMPLO DE EVALUACIÓN AMBULATORIA
(TOMADO DE FONSECA-PEDRERO, 2019)

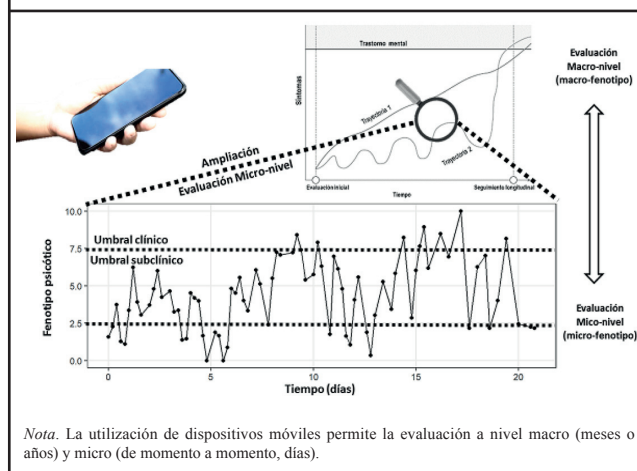
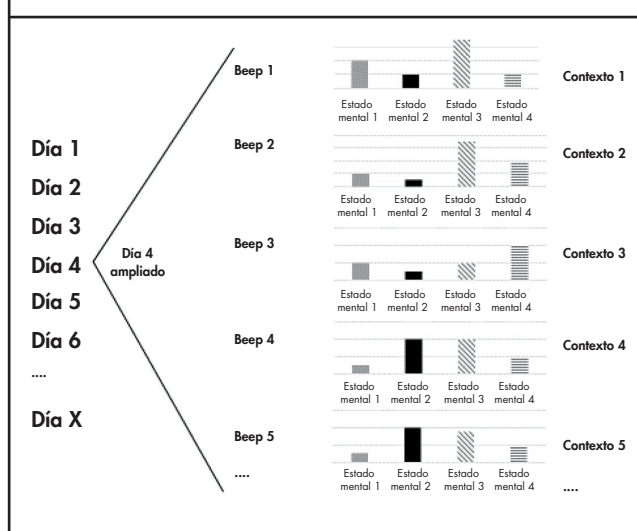


FIGURA 2
REPRESENTACIÓN DEL MÉTODO DE MUESTREO DE EXPERIENCIAS
(TOMADO DE FONSECA-PEDRERO, 2019)



las expresiones “En este momento...”, “En este instante...”, “Ahora mismo...”, “Desde el último beep...” (nótese que las posibilidades en la construcción de ítems son casi ilimitadas). En cualquier caso, dado que, al final, lo que se está construyendo es un test, se deberían seguir las directrices internacionales para la construcción de ítems, así como las cuestiones referentes a la calidad y uso de los test (Haladyna y Rodríguez, 2013; Hernández et al., 2016; Lane et al., 2016; Muñiz et al., 2020; Muñiz y Fonseca-Pedero, 2019). Tanto las cuestiones referidas a la construcción de ítems como de test, no es una cuestión baladí y debe realizarse con rigor. Igualmente, esta forma de evaluar permite ir más allá de una perspectiva cuantitativa e incorporar una perspectiva claramente cualitativa y fenomenológica. Se podrían formular preguntas específicas para cada individuo de cara a recabar información concreta y particular sobre sus experiencias, esto es, que las personas escriban lo que piensan, sienten o están haciendo. Esta cuestión se puede aplicar, por ejemplo, a la recogida de información sobre el tipo de experiencias psicóticas que están sucediendo en el momento (Raugh et al., 2021), el tipo de pensamientos suicidas (Rath et al., 2019), o la naturaleza de los pensamientos frente al consumo de sustancias (Burgess-Hull y Epstein, 2021), por citar solo tres casos. Como se puede observar, se trata de una información valiosa, detallada y profunda procedente de múltiples niveles, y que permite una mejor comprensión del fenómeno de estudio, captar patrones de comportamiento que no son fácilmente visibles, y diseñar y monitorizar el tratamiento.

Aunque no es un tema menor, por falta de espacio no se abordan aquí las cuestiones relativas a los tipos de enfoques estadísticos y análisis de datos asociados a la EA. Nótese que el volumen de datos que se genera es ingente, aspecto que entronca con el *big data*, siendo el tipo de diseño más utilizado el multinivel o de efectos mixtos. Tal vez, el avance más interesante es que la EA encaja perfectamente con los modelos de sistemas complejos di-

námicos, y el modelo de redes (Borsboom, 2017; Fonseca-Pedero, 2017).

CUESTIONES DE FONDO

Como se ha visto, la EA viene a responder a cuestiones y limitaciones no solo prácticas sino epistemológicas de gran calado, como sería el uso extensivo de los autoinformes, los sesgos de respuesta, las limitaciones del modelo biomédico, etc. No obstante, la EA no está exenta de problemas, en la Tabla 1 se recogen las principales ventajas y limitaciones en el uso de estas metodologías. La necesidad de motivar a los participantes para que respondan de forma adecuada y con la frecuencia deseada, disponer de un dispositivo móvil y tal vez conexión a Internet, la complejidad de la tecnología percibida por algunos participantes, la confidencialidad, la ley de protección de datos, etc., son algunas posibles desventajas (Palmier-Claus et al., 2011). Al igual que ocurre en la tele-evaluación o la administración remota de test (Elosua, 2021), las cuestiones relativas a la privacidad y la confidencialidad son las más importantes. Además, no se debe perder de vista a las personas que por las características de sus problemas (p. ej., ideación delirante) u otros factores, rechazan el uso de esta tecnología. Aunque en los últimos años se han realizado avances, no obstante, aún estamos asistiendo al desarrollo temprano de esta metodología y es necesario continuar investigando, pues aún hay muchas cuestiones que analizar y resolver.

TABLA 1
VENTAJAS Y LIMITACIONES EN EL USO DE LA EVALUACIÓN AMBULATORIA

Ventajas	Limitaciones
✓ Método integrado en la vida cotidiana de las personas. ✓ Evaluación dinámica ideográfica in vivo en el contexto natural de la persona. ✓ Recoger información mediante evaluaciones múltiples (de forma intensiva), de momento a momento. ✓ Obtención de información más válida (ecológica), solucionando los sesgos de memoria y pérdida de información. ✓ Monitorización de variables psicológicas de interés (niveles micro y macro). ✓ Utilización de la información recogida para realizar predicciones conductuales o la evolución y dinámicas de cambio de las personas. ✓ Posibilidad de intervenir inmediatamente en el contexto natural de la persona. ✓ Incrementar el acceso de la población a intervenciones psicológicas basadas en la evidencia.	✓ Riesgo de ausencia de marco teórico detrás de los procedimientos. ✓ Falta de homogeneidad en los diseños. Problemas de validez. ✓ Dificultades de adaptación de instrumentos clásicos de lápiz y papel que mantengan sus propiedades psicométricas. ✓ Posible superioridad de otros métodos de evaluación (e. g. entrevista) ✓ Dificultades de adherencia a este método de algunas poblaciones (p. ej., personas con psicosis). ✓ Alta variabilidad de fuentes de información. Sesgos en la obtención muestras. ✓ Privacidad y acceso a los datos. ✓ Saturación y cansancio en el uso de los dispositivos. ✓ Posibles problemas asociados a las tecnologías de la información, software, acceso y conexión a internet o apps (p. ej., notificaciones).

FIGURA 3
EJEMPLO DE ÍTEMS ADMINISTRADOS EN EVALUACIÓN AMBULATORIA (REALIZADO CON: [HTTPS://WWW.EVALUACIONAMBULATORIA.COM/](https://www.evaluacionambulatoria.com/))

APLICACIONES PRÁCTICAS

La EA ha crecido de forma exponencial en los últimos años, ganando importancia y popularidad en las distintas áreas de la Psicología, sobre todo, en el ámbito clínico. Ello va asociado a la disponibilidad e incremento del uso de las tecnologías de la información en un conjunto importante de la sociedad. Se comentan a continuación aplicaciones prácticas a diferentes campos de la Psicología, en particular al ámbito de la salud mental: ansiedad y depresión (Hall et al., 2021), psicosis (Bell et al., 2017), conducta suicida (Sedano-Capdevila et al., 2021) y trastornos de conducta alimentaria (Smith y Juarascio, 2019), entre otros.

Se han llevado a cabo intervenciones con *smartphone* para la autogestión de sintomatología ansiosa y depresiva, tanto en poblaciones clínicas como en no clínicas (Hall et al., 2021; Torous et al., 2021). También se han encontrado los posibles beneficios para el trastorno bipolar, tanto para el manejo de síntomas depresivos como (hipo) maníacos (Torous et al., 2021). A este respecto, las aplicaciones (*apps*) para teléfonos inteligentes podrían proporcionar un mecanismo accesible, escalable y de bajo costo para brindar intervenciones efectivas de autocuidado para los síntomas de depresión y ansiedad, particularmente para las poblaciones no clínicas y aquellos que no pueden acceder a los servicios en persona.

La EA ha sido utilizada también ampliamente en el campo de la psicosis (Myin-Germeys et al., 2009; Thomas et al., 2019). A pesar de lo que podamos pensar inicialmente, las personas con psicosis toleran y aceptan con frecuencia este tipo de procedimientos (entre un 69% y 88% completan las evaluaciones). Algunas personas si informaron de efectos negativos, por ejemplo, tomar mayor conciencia de sus síntomas o cierta reactividad al método en perfiles concretos de personas (p.ej., con sintomatología paranoide). Entre sus aplicaciones hemos encontrado su uso enfocado a la monitorización de síntomas y el automanejo en el funcionamiento diario (Bell et al., 2017). Por ejemplo, en un estudio con 17 pacientes con psicosis y utilizando una aplicación de monitorización pasiva instalada en el *smartphone* durante tres meses, pudieron detectar desviaciones en los patrones de movilidad y comportamiento social en las dos semanas previas a una recaída (Barnett et al., 2018).

También podría resultar especialmente útil aplicar estos procedimientos para el estudio de la conducta suicida, dada su naturaleza dinámica y fluctuante (Rath et al., 2019; Sedano-Capdevila et al., 2021). Se está comenzando a utilizar para evaluar el riesgo suicida en distintos momentos y estudiar posibles variables predictoras que permitan la anticipación e intervención de forma precisa. Por ejemplo, el afecto negativo y la alteración del sueño, entre otros, se han sugerido como predictores a corto plazo de pensamientos y comportamientos suicidas (Sedano-Capdevila et al., 2021).

En suma, la EA es una herramienta potencialmente útil en la práctica clínica y en la investigación. Los primeros resultados indican, entre otros, su buena aceptabilidad y una mejoría en los resultados clínicos de las intervenciones.

UNA PLATAFORMA PÚBLICA EN LA PSICOLOGÍA ESPAÑOLA

Dentro del marco de la ciencia abierta y con la finalidad de hacer accesible esta metodología a los profesionales de la Psicología y a la so-

ciudad es su conjunto, nuestro grupo de trabajo ha desarrollado una nueva plataforma (www.evaluacionambulatoria.com) y una *app* asociada, denominada EVAMBU. La *app* permite responder a las preguntas sobre salud mental y bienestar emocional al usuario (p. ej., un joven con TDAH). Por su parte, la web está exclusivamente diseñada para el uso por parte de profesionales de la salud mental. Esta web permite a los profesionales confeccionar los test, escalas y encuestas que posteriormente responderán los usuarios. La plataforma admite cualquier tipo de test (auditivo, visual, etc.) y preguntas diversas (formato de respuesta Likert, abierto, respuestas múltiples, etc.). Además, permite adaptar las preguntas a las personas con algún tipo de discapacidad. Los datos que van contestando las personas usuarias se vuelcan en un archivo Excel para su posterior procesamiento por parte del profesional. Es imprescindible que el usuario dé su consentimiento informado. La *app* y la web garantizan la ley de protección de datos.

DIRECTRICES GENERALES SOBRE MEJORES PRÁCTICAS

Dada la novedad de la EA y de sus metodologías asociadas como la EMA o la ESM, es importante mencionar, al menos, unas directrices generales sobre las mejores prácticas disponibles (Heron et al., 2017; Trull y Ebner-Priemer, 2020), que también tienen relación con el fenotipado digital (Hirschtritt y Insel, 2018; Torous et al., 2021) y la incorporación de las tecnologías de la información al campo de la Psicología. Estas directrices generales pueden venir a solucionar posibles limitaciones metodológicas la forma de informar sobre este tipo de estudios. Por ejemplo, según una reciente revisión sistemática solo el 2% de estudios justificaba el tamaño muestral, el 17% explicaron el proceso de selección de la muestra y el diseño del programa de evaluación, únicamente el 30% informó de las propiedades psicométricas de los ítems en el estudio y el 46% describió el conjunto final de los datos (número de participantes, duración estudio y tasas de adherencia y pérdida de participantes o respuestas) (Trull y Ebner-Priemer, 2020). Siguiendo el trabajo de Trull y Ebner-Priemer (2020) y su aplicación a la psicopatología, se pueden establecer una serie de buenas prácticas de cara a desarrollar e informar sobre este tipo de trabajos:

- Selección y tamaño de muestra:** es importante indicar el tamaño muestral, cómo ha sido seleccionado, y el análisis de la potencia estadística.
- Tipo de diseño de investigación:** se debe explicar la justificación del diseño muestral (p. ej., evento-contingente, temporal) y el fundamento de la densidad de muestreo (p. ej., evaluaciones por día) y programación (es decir, cuándo están programadas las evaluaciones). Asimismo, se deben proporcionar detalles técnicos del muestreo (p. ej., capacidad para suspender o retrasar respuestas; tiempo de respuesta en cada sesión, detalles de ramificación, número de beeps por intervalo, etc.).
- Selección e información sobre las medidas:** se debe informar del texto completo de los ítems administrados, el marco temporal de respuesta, las opciones de respuesta o escala. Igualmente, se debe informar sobre las propiedades psicométricas de los ítems del estudio actual (intra e inter-persona), así como de dónde proceden los ítems o si son de reciente construcción.
- Devices y software utilizados:** describir completamente el *hardware* y el *software* utilizados en el estudio.

- e) *Cumplimiento de respuesta*: definir los datos válidos y valores perdidos (tanto para la muestra general como para cada uno de los y las participantes del estudio). Informar igualmente de los análisis descriptivos de los datos válidos (p. ej., media por persona, rango, porcentaje de participantes por encima y por debajo del umbral del 80%).
- f) *Formación de los participantes, seguimiento y remuneración*: describir qué procedimientos se han seguido para formar a los y las participantes del estudio (qué, cómo, cuándo, etc.), los procedimientos utilizados para mejorar el cumplimiento y la participación (p. ej., gratificación, llamadas, SMS, privacidad, etc.). Una cuestión relevante tanto en este apartado como en los anteriores se refiere a la privacidad y seguridad. Sin un enfoque basado en la privacidad y la protección de los datos de las personas usuarias, los medios electrónicos carecerán de la confianza necesaria para su aceptación.
- g) *Gestión y análisis de datos*: describir el conjunto de datos final referido al número de informes (total; promedio de personas; promedio de grupo), días en estudio y tasas de retención, y tasas de respuesta demorada o suspendida (si corresponde). Concerniente a la preparación para análisis de datos, se debe describir las variables predictoras y en qué nivel; e informar sobre las covariables incluidas en los modelos. Respecto al análisis de datos, se deben describir los niveles de análisis (momentáneo, día, persona); explicar cómo se tiene en cuenta el tiempo en los análisis; especificar y justificar opciones de efectos aleatorios versus efectos fijos en los modelos; y describir el análisis de datos, así como el software estadístico utilizado.

Un punto relevante en el estudio de la EA es la referida a la tasa de cumplimiento de respuesta (Rintala et al., 2019, 2020; Wen et al., 2017; Williams et al., 2021). En un reciente estudio se analizó el grado de cumplimiento de respuesta con un protocolo ESM (Rintala et al., 2019). Los resultados indicaron un cumplimiento aceptable general con una tasa de respuesta promedio del 78% (IC del 95% [0,74, 0,82]) ($N = 1717$ participantes, 92.394 evaluaciones). Sin embargo, el cumplimiento disminuyó a lo largo de los días ($p < .001$), alcanzando un mínimo en el quinto día con un 73% (IC del 95% [0,68, 0,77]). El cumplimiento también varió significativamente entre las evaluaciones dependiendo del tiempo dentro de un día ($p < .001$), con el mayor cumplimiento entre las 12h p.m. y 13.30h p.m. (83%; IC del 95% [0,80, 0,86]) y cumplimiento más bajo entre las 7:30h a.m. y las 9h a.m. (56%; IC del 95% [0,50, 0,62]). Las personas con experiencias psicóticas mostraron un menor grado de cumplimiento que los participantes control (70% frente a 83%, respectivamente; $p < 0,001$). Los hallazgos sugieren un cumplimiento aceptable en un protocolo de ESM de 4 a 6 días de estudio con una alta frecuencia de 10 evaluaciones por día a pesar de las fluctuaciones entre y dentro de los días de estudio. En otro trabajo (Rintala et al., 2020), los investigadores encontraron que los participantes que informaron de un mayor afecto positivo en general mostraron mejor adherencia al procedimiento. Sentirse perturbado por un beep (pitido/sonido del móvil que es indicador del inicio del momento de evaluación), estar fuera de casa, el uso de medicamentos o un intervalo entre beeps más prolongado disminuyó las posibilidades de cumplimiento del beep posterior. Si bien las personas participantes

con depresión tendían a presentar mayor adherencia al procedimiento de evaluación, las posibilidades de cumplir disminuían por las noches y durante el transcurso de los días de estudio. Los resultados sugieren que la alteración del beep, estar fuera del hogar, el uso de medicamentos y el intervalo entre beeps podrían disminuir las posibilidades de cumplimiento del beep posterior.

OTEANDO EL FUTURO

El empleo de la EA tiene claras implicaciones, tanto para el diagnóstico (más preciso, contextual, dinámico, e individualizado), como para el diseño de tratamientos y la búsqueda de mecanismos explicativos del comportamiento. Implica un profundo cambio en la forma de comprender, analizar, evaluar e intervenir, y está en consonancia con la necesidad de que la Psicología vaya evolucionando hacia una evaluación personalizada, dinámica, intensiva, ecológica, etiológica, contextual y colaborativa. Este nuevo enfoque evaluativo trata de superar algunos de los problemas asociados a las evaluaciones psicométricas tradicionales, como son la falta de validez ecológica o los sesgos asociados a las evaluaciones retrospectivas. La EA también permite recoger datos que van más allá de un plano psicológico, como la actividad motora, los patrones de sueño, o la frecuencia cardíaca. A raíz de las crecientes evidencias sobre su utilidad para comprender la fenomenología y los mecanismos de acción dinámicos, y aprovechando el potencial que ofrecen las nuevas tecnologías en el ámbito de la salud mental, se están impulsando también las intervenciones dispensadas en dichos contextos reales. Esta estrategia es la denominada *Ecological Momentary Interventions* (Balaskas et al., 2021; Heron y Smyth, 2010; Myin-Germeys et al., 2016). Las intervenciones ecológicas momentáneas son tratamientos que se brindan a las personas durante su vida cotidiana (es decir, en tiempo real) y en entornos naturales (es decir, el mundo real). El tratamiento se administra mediante tecnología móvil y se puede implementar de forma independiente o como complemento del tratamiento existente (Heron y Smyth, 2010). Las intervenciones ecológicas momentáneas presentan una amplia gama de características que permiten, entre otros, el autocontrol de los síntomas mediante el uso de recordatorios momentáneos o instrucciones (incluida la adherencia a los medicamentos), el manejo de los síntomas, el desarrollo de las habilidades de la vida diaria, o el logro de metas. Por ejemplo, están apareciendo aplicaciones para intervenir sobre síntomas concretos, por ejemplo, sobre alucinaciones verbales auditivas que generan malestar (Bell et al., 2018).

Estos nuevos acercamientos metodológicos ambulatorios de carácter poliédrico constituyen una herramienta fundamental para una comprensión más integral del comportamiento humano, facilitando la evaluación centrada en la persona, el diagnóstico preciso y la intervención eficaz. Todo lo cual contribuirá a mejorar la calidad de vida de las personas y de la sociedad en su conjunto, fin último de los investigadores y profesionales de la Psicología.

CONFLICTO DE INTERESES

No existe conflicto de intereses.

REFERENCIAS

Balaskas, A., Schueller, S. M., Cox, A. L., y Doherty, G. (2021). Ecological momentary interventions for mental health: A scoping review.

- PLOS ONE, 16(3), e0248152. <https://doi.org/10.1371/JOURNAL.PONE.0248152>
- Bell, I. H., Lim, M. H., Rossell, S. L., y Thomas, N. (2017). Ecological momentary assessment and intervention in the treatment of psychotic disorders: A systematic review. *Psychiatric Services*, 68(11), 1172–1181. <https://doi.org/10.1176/appi.ps.201600523>
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16, 5–13. <https://doi.org/10.1002/wps.20375>
- Burgess-Hull, A., y Epstein, D. H. (2021). Ambulatory assessment methods to examine momentary state-based predictors of opioid use behaviors. *Current Addiction Reports*, 8(1), 122–135. <https://doi.org/10.1007/S40429-020-00351-7>
- Carpenter, R. W., Wycoff, A. M., y Trull, T. J. (2016). Ambulatory assessment: New adventures in characterizing dynamic processes. *Assessment*, 23(4), 414–424. <https://doi.org/10.1177/1073191116632341>
- Csikszentmihalyi, M., y Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease*, 175(9), 526–536. <https://doi.org/10.1097/00005053-198709000-00004>
- De Moor, M. H. M., Distel, M. A., Trull, T. J., y Boomsma, D. I. (2009). Assessment of borderline personality features in population samples: Is the Personality Assessment Inventory-Borderline Features Scale measurement invariant across sex and age? *Psychological Assessment*, 21, 125–130.
- Elosua, P. (2021). Aplicación remota de test: Riesgos y recomendaciones. *Papeles del Psicólogo*, 42, 33–37. <https://doi.org/10.23923/pap.psicol2021.2952>
- Fonseca-Pedrero, E. (2017). Network analysis: A new way of understanding psychopathology? *Revista de Psiquiatría y Salud Mental*, 10, 206–215. <https://doi.org/10.1016/j.rpsm.2017.06.004>
- Fonseca-Pedrero, E. (2018). Análisis de redes en psicología. *Papeles del Psicólogo*, 39, 1–12. <https://doi.org/10.23923/pap.psico-12018.2852>
- Fonseca-Pedrero, E. (coord.) (2019). *Tratamientos psicológicos para la psicosis*. Pirámide.
- Fonseca-Pedrero, E. (coord.) (2021a). *Manual de tratamientos psicológicos. Adultos*. Pirámide.
- Fonseca-Pedrero, E. (coord.) (2021b). *Manual de tratamientos psicológicos. Infancia y adolescencia*. Pirámide.
- Haladyna, T. M., y Rodríguez, M. C. (2013). *Developing and validating test items*. Routledge.
- Hall, M., Scherner, P. V., Kreidel, Y., y Rubel, J. A. (2021). A systematic review of momentary assessment designs for mood and anxiety symptoms. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/FPSYG.2021.642044>
- Hektner, J., Schmidt, J., y Csikszentmihalyi, M. (2007). *Experience sampling method: measuring the quality of everyday life*. SAGE Publications.
- Heron, K. E., Everhart, R. S., McHale, S. M., y Smyth, J. M. (2017). Using mobile-technology-based ecological momentary assessment (ema) methods with youth: A systematic review and recommendations. *Journal of Pediatric Psychology*, 42(10), 1087–1107. <https://doi.org/10.1093/JPEPSY/JSX078>
- Heron, K. E., y Smyth, J. M. (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology*, 15, 1–39. <https://doi.org/10.1348/135910709X466063>
- Hirschtritt, M. E., y Insel, T. R. (2018). Digital technologies in Psychiatry: Present and future. *Focus*, 16(3), 251–258. <https://doi.org/10.1176/APPI.FOCUS.20180001>
- Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA*, 318(13), 1215–1216. <https://doi.org/10.1001/jama.2017.11295>
- Lane, S., Raymond, M. R., y Haladyna, T. M. (2016). *Handbook of test development (2nd edition)*. Routledge.
- McCutcheon, R. A., Merritt, K., y Howes, O. D. (2021). Dopamine and glutamate in individuals at high risk for psychosis: A meta-analysis of in vivo imaging findings and their variability compared to controls. *World Psychiatry*, 20(3), 405–416. <https://doi.org/10.1002/WPS.20893>
- Mehl, M. (2012). *Handbook of research methods for studying daily life*. Guilford Press.
- 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, 1–15. <https://doi.org/10.23923/pap.psico-12020.2921>
- Myin-Germeys, I., Kananova, 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>
- Myin-Germeys, I., Klippel, A., Steinhart, H., y Reininghaus, U. (2016). Ecological momentary interventions in psychiatry. *Current Opinion in Psychiatry*, 29, 258–263. <https://doi.org/10.1097/YCO.0000000000000255>
- Myin-Germeys, I., y Kuppens, P. (2021). *The Open Handbook of Experience Sampling Methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies*. The center for Research on Experience sampling and Ambulatory methods Leuven (REAL) - Belgium.
- Palmier-Claus, J. E., Myin-Germeys, I., Barkus, E., Bentley, L., Udachina, A., Delespaul, P. A., Lewis, S. W., y Dunn, G. (2011). Experience sampling research in individuals with mental illness: Reflections and guidance. *Acta psychiatrica Scandinavica*, 123(1), 12–20. <https://doi.org/10.1111/j.1600-0447.2010.01596.x>
- Pérez-Álvarez, M. (2020). El embrollo científico de la psicoterapia: cómo salir. *Papeles del Psicólogo*, 41, 174–183. <https://doi.org/10.23923/pap.psicol2020.2944>
- Pérez-Álvarez, M. (2018). Para pensar la psicología más allá de la mente y el cerebro: un enfoque transteórico. *Papeles del Psicólogo*, 39, 161–173. <https://doi.org/10.23923/pap.psicol2018.2875>
- Pérez-Álvarez, M. (2019). La psicoterapia como ciencia humana, más que tecnológica. *Papeles del Psicólogo*, 40, 1–14. <https://doi.org/10.23923/pap.psico19.2877>
- Pérez Álvarez, M. (2012). *Las raíces de la psicopatología moderna: La melancolía y la esquizofrenia*. Pirámide.

- Pérez Álvarez, M., y García-Montes, J. M. (2018). Evaluación fenomenológica más allá de los síntomas. En Fonseca-Pedrero, E. (coord), *Evaluación de los trastornos del espectro psicótico* (pp. 331–363). Pirámide.
- Radua, J., Ramella-Cravaro, V., Ioannidis, J. P. A., Reichenberg, A., Phipphathsanee, N., Amir, T., Yenn Thoo, H., Oliver, D., Davies, C., Morgan, C., McGuire, P., Murray, R. M., y Fusar-Poli, P. (2018). What causes psychosis? An umbrella review of risk and protective factors. *World Psychiatry*, 17, 49–66. <https://doi.org/10.1002/wps.20490>
- Rath, D., de Beurs, D., Hallensleben, N., Spangenberg, L., Glaesmer, H., y Forkmann, T. (2019). Modelling suicide ideation from beep to beep: Application of network analysis to ecological momentary assessment data. *Internet Interventions*, 18. <https://doi.org/10.1016/J.INVENT.2019.100292>
- Raugh, I. M., James, S. H., Gonzalez, C. M., Chapman, H. C., Cohen, A. S., Kirkpatrick, B., y Strauss, G. P. (2021). Digital phenotyping adherence, feasibility, and tolerability in outpatients with schizophrenia. *Journal of Psychiatric Research*, 138, 436–443. <https://doi.org/10.1016/J.JPSYCHIRES.2021.04.022>
- Rintala, A., Wampers, M., Myin-Germeys, I., y Viechtbauer, W. (2019). Response compliance and predictors thereof in studies using the experience sampling method. *Psychological Assessment*, 31(2), 226–235. <https://doi.org/10.1037/PAS0000662>
- Rintala, A., Wampers, M., Myin-Germeys, I., y Viechtbauer, W. (2020). Momentary predictors of compliance in studies using the experience sampling method. *Psychiatry Research*, 286. <https://doi.org/10.1016/J.PSYCHRES.2020.112896>
- Russell, M. A., y Gajos, J. M. (2020). Annual Research Review: Ecological momentary assessment studies in child psychology and psychiatry. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 61(3), 376–394. <https://doi.org/10.1111/JCPP.13204>
- Sedano-Capdevila, A., Porras-Segovia, A., Bello, H. J., Baca-García, E., y Barrigon, M. L. (2021). Use of ecological momentary assessment to study suicidal thoughts and behavior: A systematic review. *Current Psychiatry Reports*, 23(7). <https://doi.org/10.1007/s11920-021-01255-7>
- Smith, K. E., y Juarascio, A. (2019). From ecological momentary assessment (ema) to ecological momentary intervention (emi): Past and future directions for ambulatory assessment and interventions in eating disorders. *Current Psychiatry Reports*, 21(7). <https://doi.org/10.1007/s11920-019-1046-8>
- Smyth, J. M., Jones, D. R., Wen, C. K. F., Matera, F. T., Schneider, S., y Stone, A. (2021). Influence of ecological momentary assessment study design features on reported willingness to participate and perceptions of potential research studies: an experimental study. *BMJ Open*, 11(7). <https://doi.org/10.1136/BMJOPEN-2021-049154>
- Stone, A. A., y Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, 16, 199–202.
- Torous, J., Onnela, J., y Keshavan, M. S. (2017). New dimensions and new tools to realize the potential of RDoC: Digital phenotyping via smartphones and connected devices. *Translational Psychiatry*, 7, e1053. <https://doi.org/10.1038/tp.2017.25>
- Torous, John, Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., y Firth, J. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*, 20(3), 318–335. <https://doi.org/10.1002/wps.20883>
- Trull, T. J., y Ebner-Priemer, U. W. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology*, 9, 151–176. <https://doi.org/10.1146/annurev-clinpsy-050212-185510>
- Trull, T. J., y Ebner-Priemer, U. W. (2020). Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology*, 129(1), 56–63. <https://doi.org/10.1037/ABN0000473>
- van Os, J., Delespaul, P., Wigman, J., Myin-Germeys, I., y Wichers, M. (2013). Psychiatry beyond labels: Introducing contextual precision diagnosis across stages of psychopathology. *Psychological Medicine*, 43(07), 1563–1567. <https://doi.org/10.1017/S0033291713000937>
- Wen, C. K. F., Schneider, S., Stone, A. A., y Spruijt-Metz, D. (2017). Compliance with mobile ecological momentary assessment protocols in children and adolescents: A systematic review and meta-analysis. *Journal of Medical Internet Research*, 19(4). <https://doi.org/10.2196/JMIR.6641>
- Williams, M. T., Lewthwaite, H., Frayse, F., Gajewska, A., Ignatavicius, J., y Ferrar, K. (2021). Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: systematic review and meta-analysis. *Journal of Medical Internet Research*, 23(3). <https://doi.org/10.2196/17023>

Construyendo Test adaptativos de elección forzosa "On The Fly" para la medición de la personalidad

Building adaptive forced choice tests "On The Fly" for personality measurement

Abad, Francisco J.¹; Kreitchmann, Rodrigo S.¹; Sorrel, Miguel A.¹; Nájera, Pablo¹; García-Garzón, Eduardo²; Garrido, Luis Eduardo³ y Jiménez, Marcos¹

¹Universidad Autónoma de Madrid, España. ²Universidad Camilo José Cela, España. ³Pontificia Universidad Católica Madre y Maestra, República Dominicana

Los nuevos desarrollos metodológicos y tecnológicos de la última década permiten resolver, o al menos atenuar, los problemas psicométricos de los test de elección forzosa (EF) para la medición de la personalidad. En estas pruebas, a la persona evaluada se le muestran bloques de dos o más frases de parecida deseabilidad social, entre las que debe elegir aquella que le represente mejor. De esta manera, los test de EF buscan reducir los sesgos de respuesta en pruebas de autoinforme. No obstante, su uso no está exento de riesgos y complicaciones si no se elaboran adecuadamente. Afortunadamente, los nuevos modelos psicométricos permiten modelar las respuestas en este tipo de test, así como optimizar su construcción. Más aún, permiten la construcción de Test Adaptativos Informatizados de EF (TAI-EF) "on-the-fly", en los que cada bloque se construye en el mismo momento de aplicación, emparejando óptimamente las frases de un banco previamente calibrado.

Palabras clave: Personalidad, Elección forzosa, Test adaptativo informatizado, On-the-fly.

The new methodological and technological developments of the last decade make it possible to resolve or, at least, attenuate the psychometric problems of forced-choice (FC) tests for the measurement of personality. In these tests, the person being tested is shown blocks of two or more sentences of similar social desirability, from which he or she must choose which one best represents him or her. Thus, FC tests aim to reduce response bias in self-report questionnaires. However, their use is not without risks and complications if they are not created properly. Fortunately, new psychometric models make it possible to model responses in this type of test and to optimize their construction. Moreover, they allow the construction of "on the fly" computerized adaptive FC tests (CAT-FC), in which each item is constructed on the spot, optimally matching sentences from a previously calibrated bank.

Key words: Personality, Forced-choice, Computerized adaptive tests, On-the-fly.

Diversos estudios de meta-análisis avalan el papel predictivo de las variables de personalidad, medidas por autoinforme, en contextos organizacionales, educativos y de la salud (p.ej., Barrick & Mount, 1991; Judge et al., 2013; Otero et al., 2020). Sabemos, por ejemplo, que la Responsabilidad y la Estabilidad Emocional tienen una validez predictiva generalizada a través de distintas ocupaciones y criterios, mientras que otras dimensiones, como la Apertura a la experiencia, la Afabilidad o la Extraversión, también resultan relevantes en contextos particulares y en la predicción de criterios específicos. En el ámbito educativo, la Responsabilidad y la Estabilidad emocional juegan un papel importante en la predicción del rendimiento académico (Poropat, 2009; Richardson et al., 2012), mientras que la responsabilidad y la afabilidad predicen conductas indebidas (p.ej., copiar en exámenes; Cuadrado et al., 2021). A pesar de estos resultados, las dudas sobre la sensibilidad del autoinforme a los efectos de la deseabilidad social y del falseamiento (*faking*) han acompañado a estas pruebas desde su

creación. En concreto, en contextos de selección se espera de los candidatos cierta distorsión de las respuestas para dar una imagen más positiva de sí mismos, bien por autoengaño, bien deliberadamente para ser elegidos. Estas distorsiones producen fuertes incrementos en las medias de las puntuaciones en la dirección percibida como deseable y reducen la fiabilidad y la variabilidad de las puntuaciones (Viswesvaran & Ones, 1999; Salgado, 2016). Estos efectos se encuentran tanto en estudios experimentales, cuando se comparan respuestas honestas y con deshonestidad inducida, como, aunque más atenuados, en contextos aplicados, cuando se comparan muestras de solicitantes a puestos laborales y de personas que ya ocupan puestos laborales (Salgado, 2005). Por tanto, sin menoscabo de que en contextos reales se encuentren resultados positivos sobre la validez predictiva de las puntuaciones de personalidad, la evidencia previa sugiere que el ordenamiento de los candidatos podría ser distinto según falseen sus respuestas o no. Salgado (2005) describe algunas estrategias para reducir los efectos del falseamiento, tales como informar a los evaluados de que existe la posibilidad de ser penalizados si distorsionan sus respuestas o utilizar baremos específicos, confeccionados con muestras de solicitantes.

Otra posibilidad es utilizar formatos de respuesta más robustos al falseamiento. En el ámbito de las pruebas de autoinforme, puede distinguirse entre el formato tradicional de escala Likert (i.e., ítems o enunciados para los que el evaluado tiene que indicar su grado de acuerdo) y el formato de elección forzosa (i.e., bloques de uno o más enunciados, entre los que el evaluado debe hacer una elección

Recibido: 1 noviembre 2021 - Aceptado: 16 diciembre 2021

Correspondencia: Abad, Francisco J.

E-Mail: fjose.abad@uam.es.

Este trabajo fue realizado con la ayuda financiera del Ministerio de Ciencia, Innovación y Universidades (Proyecto PSI2017-85022-P) y por la Cátedra de Modelos y Aplicaciones Psicométricas (Instituto de Ingeniería del Conocimiento y Universidad Autónoma de Madrid).

-p.ej., indicar el que mejor le representa- o un ordenamiento -p.ej., ordenarlos parcial o totalmente en función del grado en que le describen). En la Tabla 1 se muestran ejemplos de ítems de elección forzosa con distintos formatos.

El formato Likert es susceptible no sólo a los efectos del falseamiento o la deseabilidad social, sino también a la presencia de otros sesgos de respuesta tales como los sesgos de aquiescencia, negatividad, de tendencia central o respuesta extrema, etc. La existencia de sesgos de respuesta puede distorsionar la estructura factorial de la escala y derivar en desajuste (p.ej., Abad et al., 2018), además de producir una sobreestimación de la fiabilidad y alterar las estimaciones de validez convergente. Por el contrario, estos sesgos no son aplicables al formato de elección forzosa. Especialmente, se espera que, si los bloques se forman con ítems igualados en deseabilidad social, la susceptibilidad al falseamiento se reduzca.

PROBLEMAS DE LAS PUNTUACIONES OBTENIDAS EN PRUEBAS DE ELECCIÓN FORZOSA

A pesar de lo anterior, el uso del formato de elección forzosa no ha estado libre de controversia. En primer lugar, se ha puesto en duda su mayor resistencia al falseamiento (p.ej., Heggstad et al., 2006). No obstante, los estudios de meta-análisis más recientes sugieren que el efecto del falseamiento es menor en pruebas de elección forzosa (Cao & Drasgow, 2019; Martínez & Salgado, 2021). En segundo lugar, los test de elección forzosa pueden derivar en puntuaciones con propiedades *ipsativas*, en las que la interpretación de una puntuación es relativa al resto de puntuaciones del mismo sujeto. Por ejemplo, una persona muy organizada y muy sociable puede coincidir en su respuesta con otra poco organizada y poco sociable, por considerarse ambas más organizadas que sociables. Si las puntuaciones son completamente *ipsativas*, la suma de las puntuaciones de cada sujeto dará

lugar a un mismo valor constante y las interpretaciones *normativas* (p.ej., concluir que la primera persona es más organizada que la segunda) serían arriesgadas. En estos casos, la aplicación de las técnicas tradicionales de análisis psicométrico derivará en artefactos metodológicos (Hicks, 1970). Por ejemplo, la correlación esperada promedio entre las dimensiones tenderá a ser negativa. De la misma manera, las correlaciones entre las puntuaciones en esas dimensiones y cualquier criterio externo será cero. Estos resultados proceden de las covarianzas negativas que se producen al forzar a una persona a elegir un enunciado frente a otro. Por ejemplo, suponga el caso extremo en el que un test incluya 20 bloques de dos ítems, uno puntuando positivamente en Extroversión y otro puntuando positivamente en Responsabilidad. Si sumamos +1 en Extroversión por cada ítem de Extroversión elegido y +1 en Responsabilidad por cada ítem de Responsabilidad elegido, la correlación entre ambas escalas será -1 y la suma de las puntuaciones en ambas escalas será 20, independientemente de las elecciones de los evaluados.

La ipsatividad no es una cuestión de todo o nada, ni va asociada al formato en sí, sino que depende del diseño del test y de los bloques (p.ej., número de ítems por bloque, naturaleza unidimensional/multidimensional de los bloques, polaridad directa/inversa de los ítems que forman los bloques, número de dimensiones evaluadas, correlación entre las dimensiones medidas o modo de puntuación). En este sentido, las puntuaciones en una prueba de elección forzosa pueden ser (Hicks, 1970): (a) totalmente *ipsativas*; (b) *cuasi-ipsativas* o parcialmente *ipsativas*; y (c) *normativas*. Las puntuaciones *normativas* pueden obtenerse, por ejemplo, si los ítems de un mismo bloque pertenecen a la misma dimensión. Las puntuaciones pueden hacerse parcialmente *ipsativas* si, por ejemplo, los evaluados ordenan parcialmente -más que completamente- las alternativas, las escalas difieren en el número de ítems, o una de las dimensiones no se

TABLA 1
EJEMPLOS DE ÍTEM CON FORMATO DE ELECCIÓN FORZOSA

Tipo de formato (entre paréntesis, para cada frase, dimensión medida y polaridad)	Elección/ puntuación	Elección/ puntuación
PICK.PAIR. <i>Elige la frase que mejor te representa:</i> A. Creo que los demás tienen buenas intenciones (Af+) B. Hago listas de cosas a hacer (Re+)	A / Af: +1 Re: +0	B/ Af: +0 Re: +1
PICK. <i>Elige la frase que mejor te representa:</i> A. Me siento relajado la mayor parte del tiempo (Es+) B. Creo que los demás tienen buenas intenciones (Af+) C. Hago listas de cosas a hacer (Re+)	A/ Es: +1 Af: +0 Re: +0	B/ Es: +0 Af: +1 Re: +0
MOLE (ej., Heggstad et al., 2006). <i>Indica las frases que mejor (↑) y peor (↓) te representan:</i> A. Evito el material de lectura difícil (Ap-) B. Solo me siento cómodo con amigos (Ex-) C. Creo que los demás tienen buenas intenciones (Af+) D. Hago listas de cosas a hacer (Re+)	C↑; B↓ / Ap: +0 Ex: -(-1) Af: (+1) Re: +0	B↑; C↓ / Ap: +0 Ex: -(-1) Af: (-1) Re: +0
RANK. <i>Ordena las frases según el grado en que te representan, desde "más parecido a ti" hasta "menos parecido a ti":</i> A. Me siento relajado la mayor parte del tiempo (Es+) B. Creo que los demás tienen buenas intenciones (Af+) C. Hago listas de cosas a hacer (Re+) D. Me gusta aprender cosas nuevas (Ap+)	A>B>C>D / Es: +4 Af: +3 Re: +2 Ap: +1	D>B>C>A / Es: +1 Af: +3 Re: +2 Ap: +4

Nota. Af: Afabilidad; Ap: Apertura; Es: Estabilidad emocional; Ex: Extroversión; Re: Responsabilidad; +: ítem directo o polaridad positiva; -: ítem inverso o polaridad negativa.

puntuación. Las escalas cuasi-ipsativas dan lugar a puntuaciones que no suman una constante para todos los individuos, pero pueden mantener cierta interdependencia; esto es, el problema se reduce, pero puede no eliminarse (Brown & Maydeu-Olivares, 2018). Algunos meta-análisis muestran que las pruebas cuasi-ipsativas tienen mayor validez predictiva (Salgado et al., 2015; Salgado & Táuriz, 2014) y son más robustas al falseamiento (Martínez & Salgado, 2021).

Los cuatro formatos más frecuentes de las pruebas de elección forzosa son (ver Tabla 1): (a) elegir el ítem que mejor te describe de entre dos enunciados (PICK-PAIR), (b) elegir el ítem que mejor te describe de entre más de dos enunciados (PICK), (c) elegir el ítem que más te describe y el que menos (MOLE, de "Most and Least"), y (d) ordenar las alternativas según el grado en el que te describen (RANK). En cuanto a la puntuación tradicional, en los formatos PICK y PICK-PAIR se puede puntuar +1 en la dimensión si la polaridad del ítem elegido es positiva (ver Tabla 1) o -1, si fuera negativa (i.e., ítem inverso). En el formato RANK se pueden asignar valores entre 1 y K, siendo K el número de frases a ordenar, mientras que en el formato MOLE se pueden asignar puntuaciones -1, 0 o 1, dependiendo de la elección concreta y la polaridad de los ítem seleccionados (ver dos ejemplos en Tabla 1). Hontangas et al. (2015, 2016) encontraron, por simulación, que el formato MOLE proporcionaba resultados similares al RANK y, en ambos casos, superiores al PICK. Sin embargo, Cao y Drasgow (2019) encuentran que el formato PICK es más resistente al falseamiento que el formato MOLE, indicando que este último, además, implica una mayor carga cognitiva para responder.

PUNTUANDO BLOQUES DE ELECCIÓN FORZOSA DESDE LA TRI

En los últimos años, se ha sugerido que muchos de los problemas de las puntuaciones en pruebas de elección forzosa pueden deberse al propio procedimiento clásico de puntuación, pudiendo superarse mediante el modelado de las respuestas desde la Teoría de la Respuesta al Ítem (TRI; p.ej., Brown & Maydeu-Olivares, 2011; Hontangas et al., 2015, 2016; Morillo et al., 2016). La TRI permite modelar las probabilidades de respuesta a los bloques en función de los niveles de rasgo, lo que posibilita alcanzar, bajo ciertas condiciones, una interpretación normativa de las puntuaciones (i.e., posibilita comparaciones entre individuos). El uso de modelos de TRI viene acompañado de diversas ventajas (Olea et al., 2010): (a) permite evaluar la precisión para cada nivel de rasgo, en vez de asumir que todas las personas son evaluadas con la misma fiabilidad; (b) permite obtener puntuaciones en la misma escala, aun cuando se apliquen distintos ítems; y (c) permite el desarrollo de aplicaciones avanzadas, como los Test Adaptativos Informatizados (TAIs). La principal característica de los TAIs es que los ítems administrados se ajustan al nivel de rasgo que va manifestando el evaluado según sus respuestas a los ítems previos. El uso de un TAI permite obtener medidas más eficientes (igual precisión en menor tiempo), así como medidas con un nivel de precisión más homogéneo a través del nivel de rasgo.

Se han propuesto diversos tipos de modelos de TRI para describir el proceso de comparación de ítems dentro de un bloque, de entre los que destacan el MUPP (Multi-unidimensional Pairwise Preference) y el TIRT (Thurstonian Item Response Theory).

El modelo MUPP fue desarrollado por Stark et al. (2005) para bloques de dos ítems, cada uno midiendo una dimensión distinta. En pri-

mer lugar, se define un modelo para la probabilidad de que una persona esté de acuerdo con el contenido de un ítem. Puede asumirse que esta probabilidad sigue un modelo de dominancia (Morillo et al., 2016) o un modelo de punto ideal (Stark et al., 2005). En un modelo de dominancia la probabilidad de acuerdo con un ítem (p.ej., "Creo que los demás tienen buenas intenciones") aumenta en función del nivel de rasgo (p.ej., la Afabilidad). En un modelo de punto ideal la función de probabilidad de respuesta es unimodal; esto es, la probabilidad de acuerdo aumenta en función del nivel de rasgo hasta alcanzar un máximo y luego se reduce. Por ejemplo, la probabilidad de acuerdo con el ítem "A veces puedo persuadir a mis amigos de que hagan las cosas a mi manera" puede ser máxima para personas que tienen una cierta capacidad de persuasión, pero menor para personas que nunca persuaden a sus amigos o para personas que siempre les persuaden. En segundo lugar, a partir de estas probabilidades de acuerdo con los ítems, puede obtenerse la probabilidad de preferir un ítem sobre otro dentro de un bloque (ver, por ejemplo, Morillo et al., 2016).

El modelo TIRT (Brown & Maydeu-Olivares, 2011) se basa en la ley del juicio comparativo de Thurstone y asume un modelo de dominancia. En este modelo se desglosa la respuesta a cada bloque en un conjunto de comparaciones binarias. Por ejemplo, supongamos que alguien establece que, en un bloque de tres ítems, la frase que más le representa es la B y la que menos la A. Ese ordenamiento ($B > C > A$) se podría representar con tres variables, una por comparación binaria: $X_{AB} = 0$ (i.e., prefiere el ítem B al A), $X_{BC} = 1$ (i.e., prefiere el ítem B al C) y $X_{AC} = 0$ (i.e., prefiere el ítem C al A). Una vez creadas estas variables pueden estimarse los modelos de TRI mediante análisis factorial (Brown & Maydeu-Olivares, 2012). En el caso de bloques de dos enunciados, Morillo et al. (2016) muestran que, cuando se asume el modelo de dominancia, el MUPP es equivalente al TIRT.

¿Qué modelo es mejor? Los modelos MUPP de punto ideal son más flexibles, pero quizás innecesariamente complejos. La cuestión decisiva se encontraría en aceptar o no la necesidad de usar ítems con funciones de probabilidad unimodales. Ítems como "A veces puedo persuadir a mis amigos de que hagan las cosas a mi manera" suelen desecharse en el análisis psicométrico previo y con frecuencia son ambiguos e incluso frustrantes para los respondientes (Brown & Maydeu-Olivares, 2010). No obstante, algunos de los mayores éxitos en la aplicación de la TRI a ítems de elección forzosa se han conseguido con modelos de punto ideal.

En cualquier caso, la ventaja de usar modelos de TRI radica en obtener puntuaciones con menor, o incluso nula ipsatividad, aunque el grado en que esto se consigue dependerá del diseño del test, como se describe a continuación.

FACTORES GENERALES QUE AFECTAN A LA EFICACIA DE LOS TEST DE ELECCIÓN FORZOSA

Como se ha mencionado, los modelos de TRI no dan necesariamente lugar a puntuaciones con propiedades normativas. Frick et al. (2021) señalan algunos factores que afectan a la eficacia en la construcción de ítems de elección forzosa en el caso de modelos de dominancia, aunque sus conclusiones no pueden considerarse definitivas. En primer lugar, su recomendación más importante es la de utilizar tanto bloques homopolares positivos, formados por ítems que miden las dimensiones en la misma dirección (p.ej., A. Creo que es

emocionante hablar con muchas personas diferentes {Ex+}; B. Me siento cómodo conmigo mismo {Es+}), como bloques heteropolares, formados por ítems que miden las dimensiones en direcciones contrarias (p.ej., A. Me gusta hablar con extraños {Ex+}; B. Me preocupo por las cosas {Es-}). Por ejemplo, si se utilizan únicamente bloques del primer tipo puede ser más difícil saber si alguien escoge el ítem A por tener alta extroversión o por tener baja estabilidad, mientras que incluir bloques del segundo tipo ayudará a distinguir esos dos perfiles. A pesar de esto, la necesidad de usar bloques heteropolares es discutible, ya que puede ser más difícil igualar los ítems en deseabilidad social (Bürkner et al., 2019; Lee & Joo, 2021), facilitando el falseamiento que se pretende prevenir en este formato. Por otro lado, Morillo et al. (2016) y Kreitchmann et al. (2021) han mostrado la posibilidad de estimar con precisión sin bloques heteropolares siempre que se haga un ensamblaje óptimo y haya un rango suficiente en los pesos de los ítems (algo que también apuntan Frick et al., 2021).

Otro factor importante es el tamaño de los bloques. Incrementar su tamaño (p.ej., usando tripletas) puede reducir la ipsatividad, pero incrementa la carga cognitiva al requerir más comparaciones por bloque (Sass et al., 2020). De hecho, Frick et al. (2021) encuentran fiabilidades similares cuando se comparan pares y tripletas si se mantiene constante el número de comparaciones binarias. Otro problema de los bloques de más de dos ítems es que, al aplicar el TIRT en ausencia de bloques heteropolares, se tiende a sobrestimar la fiabilidad.

Otros factores relevantes para la presencia de ipsatividad son las correlaciones entre dimensiones y el número de estas. Cuanto menor sea el número de dimensiones medidas o mayor la correlación positiva, mayor será la ipsatividad. Por ejemplo, a partir de los resultados en sus estudios de simulación, Bürkner et al. (2019) sugieren que con cinco o menos dimensiones y bloques homopolares no se pueden alcanzar mediciones precisas, mientras que con 30 sí obtienen una buena recuperación (no consideraron casos intermedios, entre 6 y 29 factores). Fisher et al., (2019) también son pesimistas en cuanto al uso del TIRT, encontrando peor validez referida a criterio en contextos de selección. En realidad, son varios los estudios empíricos que encuentran que las fiabilidades de las pruebas pueden ser bajas (p.ej., Kreitchmann et al., 2019) o que las correlaciones entre rasgos pueden verse distorsionadas (p.ej., Morillo et al., 2016). Probablemente, estas inconsistencias entre estudios se deben a la dificultad para construir buenos bloques homopolares de elección forzosa.

LA CONSTRUCCIÓN DE BLOQUES DE ELECCIÓN FORZOSA

La clave para el éxito en la construcción de una prueba de elección forzosa es el emparejamiento de los ítems en deseabilidad social, atendiendo a la información que proporciona el bloque en su conjunto. Respecto al emparejamiento en deseabilidad, suele recurrirse a valoraciones de expertos (o de muestras similares a la que es objeto de evaluación) para puntuar la deseabilidad social de los ítems. En este punto, Pavlov et al., (2021) destacan la importancia de emparejar los ítems no solo por deseabilidad social, sino teniendo en cuenta el consenso de los jueces en la valoración.

Respecto a la formación de bloques informativos, la utilización de un modelo de TRI permite anticipar cuánta información proporcionará el bloque cuando se aplique (esto es, el grado en que reducirá la varianza

error de los niveles de rasgo estimados) y ensamblar ítems en bloques para maximizar la información. No obstante, un problema suele ser el tamaño del universo potencial de bloques. Por ejemplo, ensamblar 60 ítems en 30 bloques de 2 deriva, aproximadamente, en 2.92×10^{40} cuestionarios posibles (Kreitchmann et al., 2021). Para resolver este problema, Kreitchmann et al. (2021) adaptan el algoritmo genético NHBSA (node histogram-based sampling algorithm; Tsutsui, 2006) al problema de ensamblar ítems en bloques y proporcionan una implementación amigable en Shiny que permite diseñar una prueba de elección forzosa (<https://psychometricmodelling.shinyapps.io/FCoptimization/>). Kreitchmann et al. (2021) encontraron que el algoritmo propuesto era más eficiente que los métodos ya existentes (p.ej., al azar con restricciones de contenido o fuerza bruta). En resumen, la calidad de una prueba de elección forzosa dependería, como en test tradicionales, de la calidad psicométrica de sus componentes: los bloques que lo forman.

TEST ADAPTATIVOS INFORMATIZADOS EN PERSONALIDAD CON FORMATO LIKERT

En el ámbito de la personalidad pueden encontrarse algunos ejemplos de TAI para medir los Big Five con escalas Likert. Destaca el trabajo pionero de Reise y Henson (2000) para el NEO-PI-R en el que encontraron que un TAI-unidimensional de tan solo cuatro ítems por faceta (esto es, reduciendo la longitud a la mitad) proporcionaba una buena recuperación de los niveles de rasgo. También se han desarrollado TAI basados en modelos multidimensionales asumiendo factores correlacionados (p.ej., Makransky et al., 2013; Nieto et al., 2018) y basados en el modelo bifactor (Nieto et al., 2018), aplicados dentro de cada dominio de personalidad (p.ej., Extraversión). Los TAI multidimensionales muestran cierta ventaja al tener en cuenta las correlaciones entre las distintas facetas (p.ej., en el estudio de Makransky et al., 2013, se obtuvo una correlación promedio elevada de 0,7 para las facetas del dominio de estabilidad emocional). En los estudios de Nieto et al. (2017; 2018) se investigaron las correlaciones entre las puntuaciones obtenidas en los TAI con las obtenidas en el banco completo. Para los dominios, con 12 ítems por dominio, se alcanzaban correlaciones promedio de 0,89 para el TAI unidimensional (y para las escalas cortas), y de 0,94 para los TAI multidimensionales (Nieto et al., 2018). Estos últimos, además, proporcionaban un mejor balance en la proporción de ítems aplicados en cada faceta (i.e., mayor validez de contenido). Para las facetas, los TAI multidimensionales alcanzaban una correlación promedio más baja que los unidimensionales (0,87 vs. 0,95), pero con la mitad de los ítems.

CONSTRUYENDO TAI DE ELECCIÓN FORZOSA ADAPTATIVOS

Las ventajas de un TAI pueden hacerse especialmente importantes en ítems de pocas categorías de respuesta, como en el formato PICK-PAIR, ya que en esos casos el rango de niveles de rasgo para los que el ítem es preciso es estrecho. Existen múltiples TAI de elección forzosa (TAI-EF), siendo el más famoso el TAPAS (p.ej., Stark et al., 2014), que mide 22 dimensiones de personalidad y que se entiende como un "test a la carta", pudiéndose elegir, por ejemplo, las dimensiones a evaluar, el tipo de test (adaptativo o fijo) y el formato (p.ej., binario, politómico, de elección forzosa unidimensional o de elección forzosa multidimensional) en función del contexto de aplicación (p.ej., mayor o menor previsión de deseabilidad social). Las

versiones adaptativas permiten reducir la longitud a la mitad (Drasgow et al., 2012).

La efectividad de un TAI-EF multidimensional depende de: (a) el banco de bloques ensamblado y (b) la regla de selección. Respecto al primer punto, lo mencionado en apartados anteriores para la construcción de test fijos es aplicable. Los bloques pueden emparejarse, por ejemplo, atendiendo a un algoritmo genético, produciendo bancos óptimos de los que seleccionar bloques adaptativamente. En cuanto a la regla de selección, existen distintas variantes. En modelos unidimensionales, la varianza error es inversamente proporcional a la información del test, que es la suma de las funciones de información de los ítems. Igualmente, la matriz de varianzas-covarianzas error en modelos multidimensionales es la inversa de la matriz de información. A pesar de la aparente similitud, esta diferencia implica que distintas reglas (p.ej., regla-T: maximizar la información de cada dimensión al añadir el ítem; regla-A: minimizar la varianza error de cada dimensión al añadir el ítem) dan distintos resultados.

Por ejemplo, Kreitchmann et al. (enviado) parten de un banco de 240 ítems (48 ítems por dimensión) que ensamblan en un banco de 120 bloques. El número de bloques posibles, excluyendo los unidimensionales, era de 23.040. En este caso, se compararon los resultados para TAIs de distinta longitud (i.e., 30 y 60 bloques) y regla de selección (p.ej., regla-T y regla-A), partiendo de un banco construido según el algoritmo genético o de un banco conformado por bloques al azar. Para la mejor regla de selección se encontró que, en promedio, el uso de un banco óptimo frente a un banco aleatorio podía incrementar el coeficiente de fiabilidad 0,05 puntos (de 0,80 a 0,85) y, más importante, reducir la ipsatividad de las puntuaciones (el sesgo negativo de las correlaciones entre dimensiones y en la relación con el criterio se reducía en 0,04 puntos). En cuanto a la regla de selección, Kreitchmann et al. (enviado) encontraron que, en consonancia con investigaciones previas (p.ej., Mulder & van der Linden, 2009), la regla-A, al minimizar directamente las varianzas de error, proporcionaba mejores resultados. Este resultado es importante, puesto que algunos investigadores utilizan la regla-T por eficiencia computacional (Chen et al., 2020).

CONSTRUYENDO TAIS DE ELECCIÓN FORZOSA ADAPTATIVOS ON-THE-FLY

Como se ha mencionado, el uso de un algoritmo genético permite la optimización de un test fijo o de un banco. El siguiente paso natural es el de construir bloques on-the-fly; esto es, ensamblar “al vuelo” los enunciados en bloques en el momento de aplicar el TAI. Esta idea constituye el corazón del TAPAS, que parte de un conjunto amplio de ítems (enunciados) calibrados de los que se derivan, mediante el modelo MUPP, un conjunto gigante de bloques potenciales de los que se selecciona en cada momento el más informativo. Este proceder no está exento de supuestos, ya que asume la veracidad del modelo MUPP y la ausencia de efectos del contexto (i.e., el funcionamiento de cada ítem no depende del ítem con el que se empareja). Aunque puede haber efectos del contexto, este supuesto de invarianza se puede sostener razonablemente en la práctica (Lin & Brown, 2017; Morillo et al., 2019). Lin y Brown (2017) sugieren que los efectos de contexto pueden reducirse si se emparejan los ítems en deseabilidad social (de lo contrario, el ítem claramente más deseable será más elegido por ser percibido como la “respuesta co-

recta”) e indican que debe evitarse incluir dentro del mismo bloque ítems similares en contenido (p.ej., *Soy una persona animada en la conversación; Evito hablar de mis éxitos*), ya que esto puede modificar el significado de los ítems (en el ejemplo, *Evito hablar de mis éxitos* dejaría de ser un marcador de modestia para ser un marcador de extroversión). En todo caso, los resultados de validez predictiva del TAPAS, cuya prueba adaptativa se basa en esa invarianza, son positivos (p.ej., Trent et al., 2020).

Kreitchmann et al. (enviado) encuentran que un TAI-EF on-the-fly, bajo ese supuesto de invarianza y un procedimiento de selección óptimo, muestra una pequeña mejora respecto a un TAI-EF basado en un banco óptimo (p.ej., 0,01 en el coeficiente de fiabilidad), pero grandes mejoras en el control de la exposición, ya que al incrementarse el número de bloques posibles es más difícil que dos evaluados reciban exactamente los mismos bloques.

DISCUSIÓN

El avance de la tecnología y el desarrollo de modelos psicométricos en las dos últimas décadas está permitiendo dar respuesta a un problema clásico: la medición de la personalidad en contextos de selección en los que la deseabilidad social puede ser alta. En 2005, Salgado incluía las pruebas de elección forzosa como una solución no recomendable, en parte por la dificultad de análisis que sufre este tipo de pruebas. La evidencia sobre su mayor robustez al falseamiento y mayor validez predictiva parece inclinar la balanza hacia una visión más positiva, siempre que se resuelvan los problemas de ipsatividad de las puntuaciones. Desde la TRI, se han propuesto distintos modelos (Brown & Maydeu-Olivares, 2011; Morillo et al., 2016; Stark et al., 2005) que ayudan a resolver el problema de la ipsatividad. No obstante, las demandas de evaluación, el ensamblaje de los bloques y el tamaño de los bloques pueden tener gran importancia. En general, el comportamiento de las pruebas será mejor cuantas más dimensiones se midan y menos correlacionadas estén. En el ensamblaje de bloques es relevante no sólo el emparejamiento por deseabilidad social, sino también su contribución a reducir la varianza error. A este respecto, la necesidad de usar bloques heteropolares para eliminar la ipsatividad de las puntuaciones (Frick et al., 2021) o no (Morillo, 2018; Kreitchmann et al., 2021) es un tema de debate. Para Bürkner et al. (2019), los bloques heteropolares podrían ser contraproducente en contextos aplicados. Nuestra recomendación es el uso de algoritmos de optimización para conformar los bloques homopolares óptimamente. Esto puede resultar costoso, ya que requiere recoger información sobre la deseabilidad de los ítems, así como su calibración previa. No obstante, no debemos olvidar que: (a) el análisis exploratorio de la estructura puede resultar más complejo *a posteriori*, en bloques bidimensionales; y (b) el ensamblaje no óptimo dará lugar a problemas de ipsatividad de las puntuaciones. Por último, el tamaño de los bloques introduce una complejidad añadida en la creación de bloques óptimos puesto que, a medida que se incrementa el tamaño de los bloques, el número de bloques posibles entre los que elegir incrementa exponencialmente, haciendo menos factible explorar el universo de posibilidades.

En definitiva, aunque las pruebas de elección forzosa existen desde mucho tiempo atrás su uso se ha visto reducido por las limitaciones que se les atribuían. Sin embargo, cada vez se entiende mejor que este tipo de pruebas no constituye una categoría homogénea,

siendo importante comprender cómo el diseño de la prueba y el modo de puntuación influye en su robustez al falseamiento, en la resolución del problema de la ipsatividad y, en definitiva, en su validez predictiva. Los meta-análisis más recientes muestran que, en contextos aplicados, el uso de pruebas de elección forzosa cuasi-ipsativas constituye una estrategia prometedora para obtener una mayor validez predictiva en el ámbito de la personalidad, con mayor resistencia al falseamiento que otros formatos (Martínez y Salgado, 2021). Finalmente, el avance de las nuevas tecnologías y el desarrollo de nuevos modelos psicométricos se presentan como dos potentes aliados que posibilitan la construcción de pruebas adaptadas al vuelo, optimizando el diseño de las pruebas y la puntuación de estas.

CONFLICTO DE INTERESES

No existe conflicto de intereses

REFERENCIAS

- Abad, F. J., Sorrel, M. A., García, L. F., & Aluja, A. (2018). Modeling general, specific, and method variance in personality measures: Results for ZKA-PQ and NEO-PI-R. *Assessment*, 25(8), 959-977. <https://doi.org/10.1177/1073191116667547>
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A Meta-Analysis. *Personnel Psychology*, 44(1), 1-26. <https://doi.org/10.1111/j.1744-6570.1991.tb00688.x>
- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71(3), 460-502. <https://doi.org/10.1177%2F0013164410375112>
- Brown, A., & Maydeu-Olivares, A. (2010). Issues that should not be overlooked in the dominance versus ideal point controversy. *Industrial and Organizational Psychology*, 3(4), 489-493. <https://doi.org/10.1111/j.1754-9434.2010.01277.x>
- Brown, A., & Maydeu-Olivares, A. (2012). Fitting a Thurstonian IRT model to forced-choice data using Mplus. *Behavior research methods*, 44(4), 1135-1147. <https://doi.org/10.3758/s13428-012-0217-x>
- Brown, A., & Maydeu-Olivares, A. (2018). Modelling forced-choice response formats. En P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing* (pp. 523-569). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118489772.ch18>
- Bürkner, P.-C., Schulte, N., & Holling, H. (2019). On the statistical and practical limitations of Thurstonian IRT models. *Educational and Psychological Measurement*, 79(5), 827-854. <https://doi.org/10.1177/0013164419832063>
- Cao, M., & Drasgow, F. (2019). Does forcing reduce faking? A meta-analytic review of forced-choice personality measures in high-stakes situations. *Journal of Applied Psychology*, 104(11), 1347-1368. <https://doi.org/10.1037/apl0000414>
- Chen, C., Wang, W., Chiu, M. M., & Ro, S. (2020). Item selection and exposure control methods for computerized adaptive testing with multidimensional ranking items. *Journal of Educational Measurement*, 57(2), 343-369. <https://doi.org/10.1111/jedm.12252>
- Cuadrado, D., Salgado, J. F., & Moscoso, S. (2021). Personality, intelligence, and counterproductive academic behaviors: A meta-analysis. *Journal of Personality and Social Psychology*, 120(2), 504-537. <https://doi.org/10.1037/pspp0000285>
- Drasgow, F., Stark, S., Chernyshenko, O. S., Nye, C. D., Hulin, C. L., & White, L. A. (2012). *Development of the tailored adaptive personality assessment system (TAPAS) to support army personnel selection and classification decisions*. Drasgow Consulting Group Urbana IL.
- Fisher, P., Robie, C., Christiansen, N., Speer, A., & Schneider, L. (2019). Criterion-related validity of forced-choice personality measures: A cautionary note regarding Thurstonian IRT versus classical test theory scoring. *Personnel Assessment and Decisions*, 5(1). <https://doi.org/10.25035/pad.2019.01.003>
- Frick, S., Brown, A., & Eunike Wetzel (2021) Investigating the normativity of trait estimates from multidimensional forced-choice data, *Multivariate Behavioral Research*, <https://doi.org/10.1080/00273171.2021.1938960>
- Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, 91(1), 9-24. <https://doi.org/10.1037/0021-9010.91.1.9>
- Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, 74(3), 167-184. <https://doi.org/10.1037/h0029780>
- Hontangas, P. M., de la Torre, J., Ponsoda, V., Leenen, I., Morillo, D., & Abad, F. J. (2015). Comparing traditional and IRT scoring of forced-choice tests. *Applied Psychological Measurement*, 39(8), 598-612. <https://doi.org/10.1177/0146621615585851>
- Hontangas, P. M., Leenen, I., & de la Torre, J. (2016). Traditional scores versus IRT estimates on forced-choice tests based on a dominance model. *Psicothema*, 28, 1, 76-82. <https://doi.org/10.7334/psicothema2015.204>
- Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, 98(6), 875-925. <https://doi.org/10.1037/a0033901>
- Kreitchmann, R. S., Abad, F. J., Ponsoda, V., Nieto, M. D., & Morillo, D. (2019). Controlling for response biases in self-report scales: Forced-choice vs. psychometric modeling of Likert items. *Frontiers in Psychology*, 10, 2309. <https://doi.org/10.3389/fpsyg.2019.02309>
- Kreitchmann, R. S., Abad, F. J., & Sorrel, M. A. (2021). A genetic algorithm for optimal assembly of pairwise forced-choice questionnaires. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-021-01677-4>
- Kreitchmann, R. S., Sorrel, M. A., & Abad, F. J. (enviado). On bank assembly and block selection in multidimensional forced-choice adaptive assessments.
- Lee, P., & Joo, S.-H. (2021). A new investigation of fake resistance of a multidimensional forced-choice measure: An application of differential item/test functioning. *Personnel Assessment and Decisions*, 7(1). <https://doi.org/10.25035/pad.2021.01.004>
- Lin, Y., & Brown, A. (2017). Influence of context on item parameters in forced-choice personality assessments. *Educational and Psychological Measurement*, 77(3), 389-414. <https://doi.org/10.1177%2F0013164416646162>
- Makransky, G., Mortensen, E. L., & Glas, C. A. W. (2013). Improving personality facet scores with multidimensional computer adaptive testing: An illustration with the Neo Pi-R. *Assessment*, 20(1), 3-13. <https://doi.org/10.1177/1073191112437756>

- Martínez, A., & Salgado, J. F. (2021). A meta-analysis of the faking resistance of forced-choice personality inventories. *Frontiers in Psychology*, 12, 732241. <https://doi.org/10.3389/fpsyg.2021.732241>
- Morillo, D. (2018). *Item response theory models for forced-choice questionnaires*. Doctoral dissertation, Universidad Autónoma de Madrid.
- Morillo, D., Abad, F. J., Kreitchmann, R. S., Leenen, I., Hontangas, P., & Ponsoda, V. (2019). The journey from Likert to forced-choice questionnaires: Evidence of the invariance of item parameters. *Journal of Work and Organizational Psychology*, 35(2), 75-83. <https://doi.org/10.5093/jwop2019a11>
- Morillo, D., Leenen, I., Abad, F. J., Hontangas, P., de la Torre, J., & Ponsoda, V. (2016). A dominance variant under the multi-unidimensional pairwise-preference framework: Model formulation and Markov Chain Monte Carlo estimation. *Applied Psychological Measurement*, 40(7), 500-516. <https://doi.org/10.1177/0146621616662226>
- Mulder, J., & van der Linden, W. J. (2009). Multidimensional adaptive testing with optimal design criteria for item selection. *Psychometrika*, 74(2), 273-296. <https://doi.org/10.1007/s11336-008-9097-5>
- Nieto, M. D., Abad, F. J., & Hernández-Camacho, A. (2017). Calibrating a new item pool to adaptively assess the Big Five. *Psicothema*, 29(3), 390-395. <https://doi.org/10.7334/psicothema2016.391>
- Nieto, M. D., Abad, F. J., & Olea, J. (2018). Assessing the Big Five with bifactor computerized adaptive testing. *Psychological Assessment*, 30(12), 1678-1690. <https://doi.org/10.1037/pas0000631>
- Olea, J., Abad, F. J., & Barrada, J. R. (2010). Tests informatizados y otros nuevos tipos de tests. *Papeles del psicólogo*, 31(1), 97-107.
- Otero, I., Cuadrado, D., & Martínez, A. (2020). Convergent and predictive validity of the Big Five factors assessed with single stimulus and quasi-ipsative questionnaires. *Journal of Work and Organizational Psychology*, 36(3), 215-222. <https://doi.org/10.5093/jwop2020a17>
- Pavlov, G., Shi, D., Maydeu-Olivares, A., & Fairchild, A. (2021). Item desirability matching in forced-choice test construction. *Personality and Individual Differences*, 183, 111114. <https://doi.org/10.1016/j.paid.2021.111114>
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135(2), 322-338. <https://doi.org/10.1037/a0014996>
- Reise, S. P., & Henson, J. M. (2000). Computerization and adaptive administration of the NEO PI-R. *Assessment*, 7(4), 347-364. <https://doi.org/10.1177/107319110000700404>
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353-387. <https://doi.org/10.1037/a0026838>
- Salgado, J. F. (2005). Personalidad y deseabilidad social en contextos organizacionales: implicaciones para la práctica de la psicología del trabajo y las organizaciones. *Papeles del psicólogo*, 92, 115-128.
- Salgado, J. F. (2016). A theoretical model of psychometric effects of faking on assessment procedures: Empirical findings and implications for personality at work: A Theoretical Model of faking psychometric effects. *International Journal of Selection and Assessment*, 24(3), 209-228. <https://doi.org/10.1111/ij-sa.12142>
- Salgado, J. F., Anderson, N., & Tauriz, G. (2015). The validity of ipsative and quasi-ipsative forced-choice personality inventories for different occupational groups: A comprehensive meta-analysis. *Journal of Occupational and Organizational Psychology*, 88(4), 797-834. <https://doi.org/10.1111/joop.12098>
- Salgado, J. F., & Tauriz, G. (2014). The Five-Factor Model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies. *European Journal of Work and Organizational Psychology*, 23(1), 3-30. <https://doi.org/10.1080/1359432X.2012.716198>
- Sass, R., Frick, S., Reips, U.-D., & Wetzel, E. (2020). Taking the test taker's perspective: Response process and test motivation in multidimensional forced-choice versus rating scale instruments. *Assessment*, 27(3), 572-584. <https://doi.org/10.1177/1073191118762049>
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise-preference model. *Applied Psychological Measurement*, 29(3), 184-203. <https://doi.org/10.1177/0146621604273988>
- Stark, S., Chernyshenko, O. S., Drasgow, F., Nye, C. D., White, L. A., Heffner, T., & Farmer, W. L. (2014). From ABLE to TAPAS: A new generation of personality tests to support military selection and classification decisions. *Military Psychology*, 26(3), 153-164. <https://doi.org/10.1037/mil0000044>
- Trent, J. D., Barron, L. G., Rose, M. R., & Carretta, T. R. (2020). Tailored Adaptive Personality Assessment System (TAPAS) as an indicator for counterproductive work behavior: Comparing validity in applicant, honest, and directed faking conditions. *Military Psychology*, 32(1), 51-59. <https://doi.org/10.1080/08995605.2019.1652481>
- Tsutsui, S. (2006). Node histogram vs. edge histogram: A comparison of probabilistic model-building genetic algorithms in permutation domains. 2006 *IEEE International Conference on Evolutionary Computation*, 1939-1946. <https://doi.org/10.1109/CEC.2006.1688a44>
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59(2), 197-210. <https://doi.org/10.1177/00131649921969802>

El uso de las nuevas tecnologías en las evaluaciones educativas: La lectura en un mundo digital

The use of new technologies in educational assessments: Reading in a digital world

Javier Suárez-Álvarez¹, Rubén Fernández-Alonso², Francisco J. García-Crespo³ y José Muñiz⁴

¹Universidad de Massachusetts Amherst, Estados Unidos. ²Universidad de Oviedo, España. ³Universidad Complutense de Madrid, España. ⁴Universidad Nebrija, España

Ser un lector competente en un mundo digital requiere una base sólida de Lectura, pero también la capacidad para pensar críticamente; una tarea pendiente para muchos estudiantes españoles. Las pruebas adaptativas informatizadas y los datos de proceso (información sobre las acciones que realizan los estudiantes al responder a la prueba) son especialmente importantes cuando se evalúan competencias como la Lectura. El objetivo de este trabajo es analizar cómo el uso de la tecnología está cambiando el concepto de Lectura y las formas para evaluarlo. Esto tiene implicaciones tanto para el alumnado español como para cualquier profesional encargado de interpretar y diseñar evaluaciones educativas. El investigador debe asegurar que el uso de los datos y la tecnología sea el adecuado para los objetivos de la evaluación y sirva de forma fiable, válida y justa a las personas involucradas, pero también del usuario saber cuándo, cómo y para qué utilizar los datos.

Palabras claves: Tecnología, Datos de proceso, Tests adaptativos informatizados, PISA, Competencia lectora.

Being a proficient reader in a digital world requires a strong reading foundation, but also the ability to think critically, which is a challenge for many students in Spain. Computerized adaptive tests and process data (information about students' actions when responding to the test) are especially important when assessing skills such as reading. This work aims to analyze how the use of technology is changing the concept of reading and the ways to evaluate it. This has implications for Spanish students and any professional in charge of interpreting and designing educational evaluations. The researcher must ensure that the use of data and technology is adequate for the objectives of the evaluation and that it works in a reliable, valid, and fair way for the people involved, but also the user must know when, how, and for what purposes to use the data.

Key words: Technology, Process data, Computerized adaptive tests, PISA, Reading.

Durante mucho tiempo se ha pensado que de los millenials para adelante hay que dar por supuestas las habilidades digitales. Esto se está empezando a cuestionar. No cabe duda de que el alumnado actual usa más la tecnología, pero ¿sabe usarla bien? Si algo muestran los datos es que tener más exposición a la tecnología no implica necesariamente una mayor competencia digital y las diferencias son enormes entre estudiantes de entornos favorecidos y desfavorecidos (OECD, 2021a). Numerosos estudios han demostrado que las generaciones más jóvenes pueden estar más familiarizadas con la tecnología que las precedentes; sin embargo, los "nativos digitales" no siempre están necesariamente equipados con las habilidades adecuadas en términos de acceso y uso de la información digital (Breakstone et al., 2018; Macedo-Rouet et al., 2020; McGrew et al., 2018; OECD, 2011). Los estudiantes necesitan utilizar las tecnologías de la información y la comunicación (TIC) para acceder a textos a través de buscadores, utilizar enlaces

y pestañas, procesar información de múltiples fuentes, evaluar la calidad de las fuentes de información, detectar posibles conflictos y resolverlos. Las definiciones de lo que significa ser un lector competente están en continua actualización para responder a cambios en la tecnología, la sociedad, la economía, y la cultura, y con ellas las formas de evaluar estas competencias. El objetivo de este trabajo es analizar la forma en que el uso de las Nuevas Tecnologías está cambiando el concepto de Lectura y, en consecuencia, las formas para evaluarlo. En este artículo abordamos algunas de las implicaciones más importantes que tienen estos cambios para el alumnado español, así como para cualquier profesional encargado de interpretar o diseñar evaluaciones educativas que aprovechen los avances de la tecnología. Para más información sobre el impacto de las nuevas tecnologías en otros ámbitos de evaluación como la Psicología y la evaluación ambulatoria ver Andrés et al. (2022), Fonseca-Pedrero et al. (2022) y Santamaría y Sánchez-Sánchez (2022) en este monográfico.

Recibido: 1 noviembre 2021 - Aceptado: 21 diciembre 2021

Correspondencia: Javier Suárez-Álvarez, Ph.D., Senior Research Fellow, Center for Educational Assessment, University of Massachusetts Amherst, College of Education, 813 North Pleasant Street, Amherst, MA 01003, USA. E-mail: suarezj@umass.edu

¹ PISA 2018 incluyó una tarea en que se invitaba a los estudiantes a que hicieran clic en el enlace de un correo electrónico de un conocido operador de telefonía móvil y que llenaran un formulario con sus datos para ganar un teléfono inteligente, también conocido como correos electrónicos de phishing.

LA LECTURA EN UN MUNDO DIGITAL: EL CONSTRUCTO EN CONTINUO CAMBIO

Según los últimos datos del Programa de Evaluación Internacional de Estudiantes (PISA 2018), el estudiante promedio de 15 años en España pasa unas 35 horas semanales conectado a Internet – lo que supone un incremento de un 66% en 6 años y son datos previos a la pandemia Covid-19 por lo que esta cifra posiblemente no ha llegado a su techo (OECD, 2021a). Sin embargo, aproximadamente uno de cada tres estudiantes en España respondió en las pruebas PISA que hacer clic en el enlace de un correo electrónico de phishing era algo apropiado o muy apropiado¹. Menos de la mitad de los jóvenes de 15 años en España

son capaces de distinguir hechos de opiniones en las tareas de PISA². España se encuentra por debajo de la media de la OCDE en ambos indicadores, aunque la diferencia es de menos de 10 puntos porcentuales (INEE, 2021). La buena noticia es que trabajar estas competencias en el aula – por ejemplo, aprender a detectar información sesgada – está relacionado con una mejor competencia lectora – por ejemplo, ser capaz de distinguir hechos y opiniones (Suarez-Alvarez, 2021). Estos datos confirman que la escuela puede jugar un papel fundamental en el aprendizaje de las competencias digitales y ayudar a reducir los riesgos asociados a ellas, aunque, por supuesto, no existen fórmulas mágicas.

Si bien es cierto que 29 de las 35 horas semanales que los estudiantes españoles pasan en Internet son fuera de la escuela, aquellos países cuyos estudiantes pasan más horas utilizando tecnología en el aula no siempre son los que mejor competencia digital muestran (OECD, 2021a). La relación entre el rendimiento en Lectura y el tiempo dedicado a utilizar dispositivos digitales para el trabajo escolar fue negativa en 36 países y economías, incluida España. En Australia, Dinamarca, Corea, Nueva Zelanda y Estados Unidos, esta relación fue positiva después de tener en cuenta la situación socioeconómica de los estudiantes y de las escuelas. Tampoco podemos olvidar que el uso de las tecnologías digitales responde a una realidad multidimensional: en ocasiones el uso de la tecnología en el aula se hace de manera complementaria a otras actividades que también son importantes, mientras que en otros casos pueden llegar a reemplazarlas. La clave puede estar, no tanto en el uso en sí, sino en el cómo se realiza una actividad concreta.

Por ejemplo, aún existe un beneficio claro de leer en papel, algo en lo que coincide la literatura que ha investigado este asunto (Clinton, 2019; Delgado et al., 2018). Estos resultados también son consistentes en los más de 70 países y economías que participaron en PISA 2018: los estudiantes que leen con más frecuencia libros en papel que en formato digital puntúan mejor en las pruebas de Lectura de PISA 2018 e invierten más tiempo leyendo por placer. Estos datos sugieren que no deberíamos desterrar al papel en favor de un monopolio digital y parece razonable buscar una compatibilidad entre formatos de Lectura tradicionales y analógicos. De hecho, los estudiantes más competentes en Lectura son aquellos capaces de optimizar las ventajas de ambos formatos. El alumnado con buenas competencias lectoras parece responder al siguiente perfil: lee libros en papel y utiliza dispositivos digitales, dependiendo del objetivo, es decir, puede leer una novela en papel, pero lee diariamente las noticias online.

Otro ejemplo claro es el tipo de estrategias de Lectura que los estudiantes aprenden en la escuela, hecho que en el caso de España es particularmente interesante. Los estudiantes en España obtuvieron una puntuación particularmente alta en dos de los tres índices de conocimiento de estrategias de Lectura incluidos en PISA 2018 (comprender y recordar un texto, y redactar un resumen), pero menos en el índice de estrategias de Lectura para evaluar la credibilidad de las fuentes de información (INEE, 2021). Los datos sugieren que los estudiantes en España tienen un conocimiento relativamente bueno

de los aspectos tradicionales y aún importantes de la Lectura. Sin embargo, todavía carecen, en promedio, del conocimiento y las habilidades relevantes para navegar en un mundo digital. Leer no consiste solamente en aprender a decodificar palabras escritas durante la infancia, sino en aprender un conjunto de competencias que van cambiando a lo largo de la vida. Para convertirse en lectores competentes en un mundo digital, los estudiantes necesitan una base sólida de Lectura, pero también la capacidad de pensar críticamente, y ajustar su comportamiento en función de la tarea, así como motivarse para perseverar frente a las dificultades.

En este sentido, España podría beneficiarse de fortalecer los conocimientos y habilidades de los estudiantes para navegar a través de la ambigüedad, contrastar y validar puntos de vista. Sin embargo, agregar asignaturas específicas sobre habilidades digitales en la escuela sin ajustar otras partes del currículo podría ser problemático. Es importante equilibrar el contenido y las competencias para abordar las nuevas demandas sociales sin sobrecargar el plan de estudios. El reto consiste en tratar de responder a las necesidades cambiantes y, al mismo tiempo, minimizar la expansión y la sobrecarga de contenido. Una posibilidad para encontrar el equilibrio entre la actualización curricular y la sobrecarga consiste en incorporar temas transversales o competencias en materias existentes (OECD, 2020), así como la utilización de métodos innovadores de enseñanza y aprendizaje (Paniagua y Istance, 2018; Pérez et al., 2018).

LA TECNOLOGÍA EN LA EVALUACIÓN EDUCATIVA: LAS PRUEBAS ADAPTATIVAS Y LOS DATOS DE PROCESO

Las definiciones de las competencias escolares, no solo de la Lectura como se ha descrito anteriormente, sino también de otras competencias como las matemáticas y las ciencias, están en continuo proceso de actualización para reflejar los cambios en la tecnología, la sociedad, la economía, y la cultura (Fraillon et al., 2019; Mullis y Martin, 2019; OECD, 2018). La navegación, se ha convertido en un componente transversal a cualquier tipo de aprendizaje en un mundo digital y, como describiremos a continuación, también a la forma en que se evalúan estas competencias.

Las evaluaciones a gran escala coordinadas por organismos internacionales como PISA, TIMSS (Estudio internacional de tendencias en Matemáticas y Ciencias) o PIRLS (Estudio Internacional para el Progreso de la Comprensión Lectora) se administran mayoritariamente por ordenador, siendo éste su método principal de administración también en España. Las administraciones públicas españolas llevan probando desde hace algún tiempo las pruebas informatizadas en sus evaluaciones de diagnóstico o de sistema, aunque su uso aún no está generalizado. La realidad de las aulas españolas no es muy diferente. Las evaluaciones informatizadas en el aula (exámenes, rúbricas, etc.) aunque habituales en muchos centros educativos aún no se ha convertido en el *status quo*. Sin embargo, la tendencia es clara: a medida que el acceso a internet y a dispositivos digitales

² La Pregunta 3 de la unidad Rapa Nui evalúa la capacidad de distinguir hechos de opiniones. Se trata de un ítem de respuesta parcial donde la respuesta nula se califica con 0, la respuesta parcial se califica con 0.5 y la respuesta completa se califica con 1. El porcentaje correcto estimado para la respuesta completa en este ítem es menor al 41% en España y menor al 47 % en promedio en Países de la OCDE. La pregunta 3 de Rapa Nui es un ítem de nivel 5. Esto significa que los estudiantes deben tener un nivel de competencia 5 para tener un 62 % de probabilidad de obtener la puntuación completa en este ítem.

aumentan también lo hace su uso. Los avances en la tecnología informática también han influido sustancialmente en las formas en que se realizan, administran, califican y se informa a los examinados (Zenisky y Sireci, 2002). En este artículo nos centramos en dos áreas emergentes que su imparable auge ha generado un cambio de paradigma en la evaluación educativa. Para más información sobre otras aplicaciones como *computational psychometrics* y *machine learning*, ver por ejemplo von Davier et al. (2019); en este monográfico Elosua (2022).

La primera y quizá más extendida es la posibilidad de realizar test adaptativos informatizados (Olea et al., 2010; en este monográfico Abad et al., 2022). Los tests adaptativos informatizados permiten aumentar la precisión en la medición utilizando menos preguntas o ítems por estudiante. Esto se logra presentando a los estudiantes ítems que están alineados con su nivel de competencia. A diferencia de las evaluaciones no adaptativas que habitualmente se centran en evaluar a estudiantes de rendimiento medio, las pruebas adaptativas también permiten una diferenciación más refinada de la capacidad del estudiante en los extremos superior e inferior del nivel de competencia del estudiante (Tabla 1). Este aspecto de las pruebas adaptativas es particularmente relevante cuando se comparan grupos de alto y bajo rendimiento o estudiantes de familias favorecidas y desfavorecidas. Al mismo tiempo, la experiencia de las pruebas para los estudiantes es mejor, ya que no necesitan responder preguntas que sean demasiado difíciles para ellos o demasiado fáciles que hagan que su interés en la prueba se reduzca. El uso de pruebas adaptativas ya se utiliza en la inmensa mayoría de evaluaciones educativas a gran escala como PISA, TIMSS, y PIRLS e incluso en la evaluación de la personalidad (Pedrosa et al., 2016; Postigo et al., 2020; en este monográfico Abad et al., 2022).

TABLA 1
RESUMEN DE LAS VENTAJAS E INCONVENIENTES DE LOS TESTS ADAPTATIVOS INFORMATIZADOS FRENTE A LOS CONVENCIONALES

Ventajas	Inconvenientes
Más eficiente, mejora la medición sin aumentar el tiempo de evaluación.	Requiere métodos de análisis más sofisticados (Teoría de Respuesta a los Ítems).
Más efectivo, optimiza la precisión de la medición en un rango más amplio de competencia de los examinados (por ejemplo, estudiantes de alto y bajo rendimiento).	Requiere bancos de preguntas más grandes (especialmente en pruebas adaptativas informatizadas a nivel de ítem, algo menos en los diseños de formato adaptativo de varias etapas).
Más equitativo, una mayor uniformidad en la precisión de la medida en los diferentes niveles de competencia de los examinados permite tomar decisiones mejor informadas (por ejemplo, al comparar estudiantes de entornos favorecidos y desfavorecidos).	Requiere muestras de participantes más grandes (por ejemplo, PISA utiliza un mínimo de 200 respuestas por ítem).
Mejor experiencia, no se presentan preguntas que sean ni muy difíciles ni muy fáciles.	Requiere administración por ordenador que puede suponer una barrera para algunos estudiantes (el modo de administración de la prueba no puede favorecer a determinados grupos frente a otros).

El segundo cambio revolucionario en la evaluación informatizada es la posibilidad de almacenar archivos de registro, también conocidos como datos de proceso. Estos datos contienen información sobre las acciones realizadas por los examinados cuando interactúan con las tareas que se le presentan en el ordenador y el tiempo dedicado a cada acción durante el proceso (Tabla 2). Este tipo de datos proporciona información adicional más allá de los datos de respuesta que, por lo general, muestran si la pregunta fue respondida de forma correcta o no (He et al., 2019, 2021; von Davier et al., 2019). Por ejemplo, la cantidad de tiempo que los estudiantes dedican a responder a las pruebas puede reflejar su nivel de compromiso (si están dando lo mejor de sí) y, en consecuencia, afectar al rendimiento en la prueba (Wise et al., 2021). Esta información puede ayudar a los responsables de la formulación de políticas, a los investigadores y los educadores a comprender mejor las estrategias cognitivas de los estudiantes y las causas subyacentes del bajo y alto rendimiento. Esto, a su vez, puede repercutir en una mejora del diseño de las evaluaciones y conducir a programas de formación y aprendizaje más eficaces (OECD, 2019a).

EJEMPLO ILUSTRATIVO DEL USO DE DATOS DE PROCESO EN PISA

PISA es una encuesta trienal de estudiantes de 15 años de todo el mundo que evalúa en qué medida han adquirido los conocimientos y las habilidades clave esenciales para una participación plena en las sociedades. PISA introduce por primera vez las pruebas adaptativas informatizadas en 2018 para evaluar la competencia lectora de los estudiantes de 15 años (últimos datos disponibles hasta la fecha). La naturaleza interactiva de las evaluaciones adaptativas informatizadas como PISA las convierte en candidatos ideales para análisis basados en datos de proceso (Goldhammer et al., 2016; Vörös et al., 2021).

La capacidad de los estudiantes para pensar, monitorear y ajustar su actividad a una tarea en particular son aspectos esenciales a la hora de leer en entornos digitales (OECD, 2019b). Los lectores digitales no solo necesitan seguir estructuras de información lineales, si-

TABLA 2
RESUMEN DE INDICADORES DE NAVEGACIÓN BASADOS EN LOS DATOS DE PROCESO

Indicadores de cantidad	Indicadores de calidad	Indicadores temporales
Número de páginas visitadas	Conductas de navegación y estrategia	Tiempo utilizado en la primera página (mediana)
	Utilización de hipervínculos	Tiempo utilizado en la primera página (proporción)
		Ratio de transición efectiva entre páginas (más de tres segundos)
		Tiempo utilizado en la página de instrucciones

Nota: Adaptado de (OECD, 2021a).

no también construir sus propios textos seleccionando y evaluando información de diferentes fuentes. Una buena navegación, por tanto, debería ser consistente con estos objetivos. La prueba de Lectura de PISA 2018 permite a los estudiantes navegar a través de las diferentes tareas de tal forma que el estudiante puede decidir cómo interactuar con el texto y qué estrategia es más eficaz. Por ejemplo, un estudiante puede decidir seguir estrictamente las instrucciones de cada tarea según aparecen mientras que otros pueden preferir explorar qué tareas vendrán a continuación para gestionar su tiempo y conocimientos de manera diferente.

El reciente informe PISA, *Lectores del Siglo XXI, desarrollando la competencia lectora en un mundo digital*, utiliza datos de proceso para agrupar a los estudiantes en función de su navegación en la unidad de Lectura basada en escenarios, Rapa Nui³ (CR551, ver Apéndice A):

- ✓ Sin navegación: estudiantes que no tienen actividades de navegación ni en ítems con textos de fuente única (ver Apéndice A ítems 1, 2, 3, 4 y 5) ni en ítems con textos de múltiples fuentes de información (ver Apéndice A ítems 6 y 7)
- ✓ Navegación limitada: estudiantes que navegan simplemente en ítems con textos de una sola fuente (ver Apéndice A ítems 1, 2, 3, 4 y 5), pero no a través de ítems con textos de múltiples fuentes (ver Apéndice A ítems 6 y 7)
- ✓ Navegación estrictamente enfocada: estudiantes que siguen estrictamente las instrucciones de las tareas para navegar activamente en ítems con textos de múltiples fuentes de información (ver Apéndice A ítems 6 y 7), pero limitaron la navegación en ítems con textos de fuente única (ver Apéndice A ítems 1, 2, 3, 4, y 5), y
- ✓ Navegación exploratoria activa: estudiantes que navegan activamente en ítems con textos de fuente única y múltiple (ver Apéndice A ítems 1, 2, 3, 4, 5, 6, y 7).

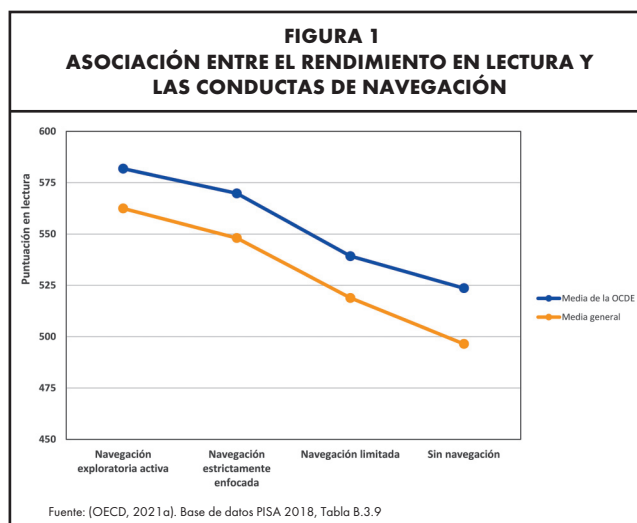
La Figura 1 muestra que el rendimiento en Lectura está fuertemente relacionado con el tipo de navegación de los estudiantes. En la mayoría de los países y economías se encontró un patrón consistente entre las puntuaciones en Lectura y los patrones de navegación. Los datos señalan una diferencia de 66 puntos entre los estudiantes que navegaron activamente entre las páginas y los que no ejecutaron actividades de navegación. El 11% de los estudiantes que pertenecían al grupo de navegación exploratoria activa obtuvieron las puntuaciones más altas en Lectura. Estos estudiantes navegaron activamente a través de tareas de fuente única (ver Apéndice A ítems 1, 2, 3, 4, y 5) y múltiple (ver Apéndice A ítems 6, y 7). Su navegación en tareas con textos de fuente única superó el número de páginas requerido para completar la tarea de lectura. Es decir, no solo leyeron la página requerida, sino también otras páginas accesibles. Este tipo de navegación podría ayudar a los estudiantes a obtener una idea general de toda la prueba de Lectura y prepararse mejor para las siguientes tareas. Esto también podría explicar, al menos en parte, la razón por la que el grupo de estudiantes de exploración activa puntuó mejor en la prueba de lectura que el grupo de estudiantes que siguió estrictamente las instrucciones de la tarea.

Las chicas presentan mejores estrategias de lectura que los chicos en los tres índices incluidos en PISA 2018 (comprender y recordar

un texto, redactar un resumen, y evaluar la credibilidad de las fuentes de información). Este hecho se repite en todos los grupos de navegación. Aun así, tanto los chicos como las chicas que navegaron y exploraron activamente la unidad Rapa Nui tienen un mejor conocimiento de las estrategias de lectura para evaluar la credibilidad de las fuentes que los estudiantes con una navegación limitada o nula. Además, las diferencias de género en las estrategias de lectura son más estrechas entre los estudiantes con una navegación más activa que aquellos sin navegación (Figura 2).

DISCUSIÓN

El uso de la tecnología en la evaluación educativa permite utilizar nuevas fuentes de datos tanto para el proceso de evaluación (el diseño de las pruebas) como para el producto (las puntuaciones en la prueba). Los test adaptativos informatizados y los datos de proceso son dos de las áreas emergentes que más beneficios tienen en la evaluación educativa. El uso de datos de proceso en pruebas informatizadas es particularmente valioso cuando se evalúan competencias que requieren tareas interactivas y el uso de tecnología. En el caso de la evaluación de la competencia lectora, ejemplo analizado en este artículo, los estudiantes necesitan utilizar las tecnologías de la información y la comunicación (TIC) para acceder a textos a través de buscadores, utilizar enlaces y pestañas, procesar información de múltiples fuentes, evaluar la calidad de las fuentes de información, detectar posibles conflictos y resolverlos. Los datos de proceso, más concretamente las secuencias de acciones realizadas por los estudiantes cuando responden a la prueba, permiten identificar las estrategias de navegación utilizadas por los estudiantes cuando interactúan con textos en un entorno digital. Estos datos se podrían utilizar para medir el grado en que los estudiantes se involucran en la tarea de la forma esperada y mejorar así la validez de las interpretaciones, pero también se podría utilizar junto a las respuestas de los estudiantes para estimar sus puntuaciones en las competencias examinadas (Wise et al., 2021).



³ La unidad de Lectura basada en escenarios, Rapa Nui (CR551) consta de tres textos: una página web del blog de un profesor, una reseña de un libro y un artículo de noticias de una revista de ciencia. En estas situaciones de lectura de múltiples textos, los lectores deben tomar decisiones sobre cuál de los fragmentos de texto disponibles es el más importante, relevante, preciso o veraz.

De esta forma, los datos de proceso, que a priori son aspectos inherentes al método de evaluación y diseño de la prueba (el proceso), se convierten en cierta medida en el objeto de la evaluación (el producto). Es decir, la tecnología ha calado en todas las capas de la evaluación educativa de tal forma que es casi imposible desenredar la medición del constructo. Las posibilidades que este tipo de datos tiene para la evaluación educativa son incontables (Jiao et al., 2021), pero como nos recuerda el refranero español: es importante no poner el carro delante de los bueyes. Existen al menos tres consideraciones importantes para optimizar el uso de este tipo de datos:

Diseñar la prueba con el objetivo de extraer y utilizar datos de registro. Los datos de proceso son un subproducto de características inherentes al software utilizado para realizar la evaluación. Es decir, en la inmensa mayoría de casos, las pruebas no están construidas y diseñadas con el objetivo de utilizar estos datos, sino que más bien se utilizan a modo de *serendipia* para mejorar el proceso y el producto de la evaluación una vez la prueba ha sido administrada y las respuestas a las preguntas codificadas. Si bien mejorar el proceso y el producto de la evaluación no tiene nada de malo, de hecho, los estándares para las pruebas educativas y psicológicas así lo recomiendan (AERA et al., 2014), es importante analizar las consecuencias que puedan tener en otras áreas también importantes de la evaluación. Por ejemplo, si el objetivo es mejorar el producto de las evaluaciones (las puntuaciones) pero el uso de datos de proceso no se ha tenido en cuenta en el diseño de la prueba, la validez de contenido (Pedrosa et al., 2014) puede verse alterada con respecto a su marco conceptual de evaluación. En otras palabras, el riesgo de dar significado a datos a *posteriori* es que puede desviar el objetivo y el contenido de la evaluación del originalmente propuesto en los marcos conceptuales de evaluación. Para una discusión sobre este tema ver, por ejemplo, cómo cambiarían los rankings de PISA si este

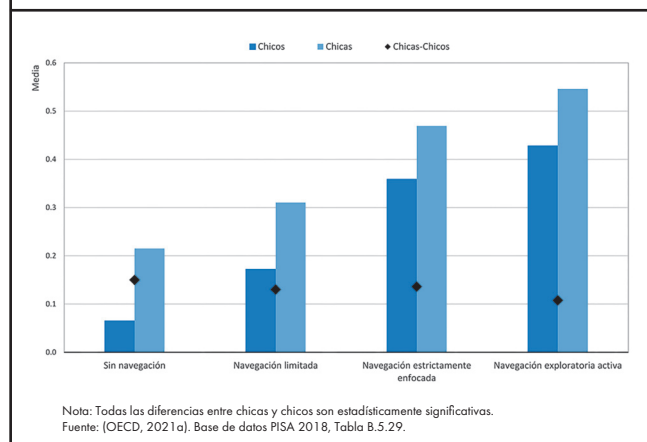
tipo de datos se tuviera en cuenta en la estimación de puntuaciones (Pohl et al., 2021).

La extracción de datos de proceso debe estar basada en consideraciones teóricas y analíticas. La elección de qué información se está registrando mientras el examinado responde a la prueba no responde a consideraciones teóricas ni analíticas. Es decir, el software registra ciegamente toda la actividad del examinado, sea ésta relevante o no. Por lo tanto, la extracción de datos de proceso debe estar guiada por consideraciones teóricas y analíticas para evitar interpretaciones espurias. Una aplicación particularmente interesante del uso de datos de proceso para mejorar la medición (el proceso) consiste en utilizar tiempos de respuesta para identificar diferencias en el proceso de respuesta entre grupos culturales y lingüísticos diferentes y así mejorar la validez de las interpretaciones que de lo contrario podrían pasar por alto en los análisis de Funcionamiento Diferencial de los Ítems⁴ (Ercikan et al., 2020).

La interpretación de los datos de registro debe estar guiada por modelos cognitivos. Es importante señalar que los procesos de respuesta cognitiva no son observables ni en las pruebas convencionales ni en las digitales. Al igual que sucede con otros métodos para obtener evidencia de validez de los procesos de respuesta como los laboratorios cognitivos, o los protocolos de pensamiento en voz alta (Padilla y Benítez, 2014), los datos de registro no reflejan procesos cognitivos en sí, sino rastros o huellas de los procesos cognitivos que los estudiantes utilizaron. La interpretación, por tanto, requiere la utilización de métodos mixtos que compaginen interpretaciones cualitativas y cuantitativas. En otras palabras, es fundamental que la interpretación de los datos de proceso esté guiada por modelos teóricos, en este caso, cognitivos. Por ejemplo, la evaluación de constructos complejos como el razonamiento o el pensamiento crítico puede involucrar procesos cognitivos como tiempos de respuesta explícitamente incluidos en la definición del constructo. Otro ejemplo, como veíamos en este artículo, es la inclusión de conductas de navegación como un aspecto transversal en el aprendizaje de la competencia lectora en un mundo digital.

La tecnología potencia el conocimiento a niveles inimaginables, sea este conocimiento el adecuado o no. De hecho, es habitual que el conocimiento inadecuado o erróneo tenga una mayor capacidad de expansión. Véase como ejemplo la transmisión de información en las redes sociales; una noticia falsa se extiende a una velocidad más rápida que una noticia verdadera (Vosoughi et al., 2018). *Mutatis mutandis*, los estudios que son más difíciles de replicar tienden a ser más citados (Serra-García y Gneezy, 2021). En España, además, el uso de los test en el ámbito profesional es enorme y la formación y conocimiento sobre el uso de ellos relativamente bajo (Muñiz et al., 2020; en este monográfico Hernández et al., 2022). Es la responsabilidad del investigador asegurar que el uso de los datos y la tecnología sea el adecuado para los objetivos de la evaluación y sirva de forma fiable, válida y justa a las personas involucradas, pero también del usuario saber cuándo, cómo y para qué utilizar los datos. Esperamos que este artículo ayude a reflexionar sobre ambas cosas.

FIGURA 2
ASOCIACIÓN ENTRE LAS ESTRATEGIAS DE LECTURA PARA EVALUAR LA CREDIBILIDAD DE LAS FUENTES DE INFORMACIÓN Y LAS CONDUCTAS DE NAVEGACIÓN



⁴ La existencia de funcionamiento diferencial del ítem (DIF) indica que personas que tienen igual grado de habilidad en la competencia evaluada presentan diferente probabilidad de responder correctamente a un ítem, lo que se relaciona con el sesgo, es decir, favorecer a un grupo sobre otro en la evaluación

CONFLICTO DE INTERESES

No existe conflicto de intereses.

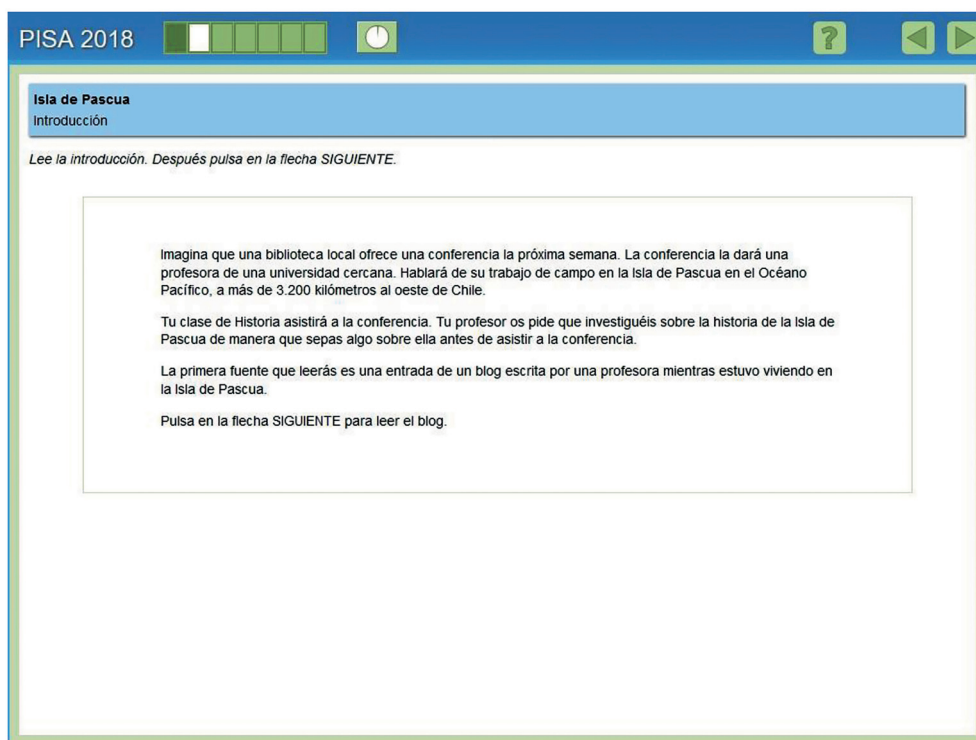
REFERENCIAS

- 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>
- AERA, APA, y NCME. (2014). *Standards for educational and psychological testing (2nd ed.)*. Washington, DC: American Educational Research Association (AERA).
- 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>
- Breakstone, J., McGrew, S., Smith, M., Ortega, T., y Wineburg, S. (2018). Why we need a new approach to teaching digital literacy. *Phi Delta Kappan*, 99(6), 27-32. <https://doi.org/10.1177/0031721718762419>
- Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading*, 42(2), 288-325. <https://doi.org/10.1111/1467-9817.12269>
- Delgado, P., Vargas, C., Ackerman, R., y Salmerón, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review*, 25, 23-38. <https://doi.org/10.1016/j.edurev.2018.09.003>
- 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>
- Ercikan, K., Guo, H., y He, Q. (2020). Use of response process data to inform group comparisons and fairness research. *Educational Assessment*, 25(3), 179-197. <https://doi.org/10.1080/10627197.2020.1804353>
- 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>
- Fraillon, J., Ainley, J., Schulz, W., Duckworth, D., y Friedman, T. (2019). IEA International Computer and Information Literacy Study 2018 Assessment Framework. *IEA International Computer and Information Literacy Study 2018 Assessment Framework*. <https://doi.org/10.1007/978-3-030-19389-8>
- Goldhammer, F., Martens, T., Christoph, G., y Lüdtke, O. (2016). *Test-taking engagement in PIAAC*. OECD Education Working Papers, No. 133, OECD Publishing, Paris. <https://doi.org/10.1787/5j1z1f6fhs2-en>
- He, Q., Borgonovi, F., y Paccagnella, M. (2019). *Using process data to understand adults' problem-solving behaviour in the Programme for the International Assessment of Adult Competencies (PIAAC): Identifying generalised patterns across multiple tasks with sequence mining*. OECD Publishing, Paris. https://www.oecd-ilibrary.org/education/using-process-data-to-understand-adults-problem-solving-behaviour-in-the-programme-for-the-international-assessment-of-adult-competencies-piaac_650918f2-en
- He, Q., Borgonovi, F., y Paccagnella, M. (2021). Leveraging process data to assess adults' problem-solving skills: Using sequence mining to identify behavioral patterns across digital tasks. *Computers & Education*, 166, 104170. <https://doi.org/10.1016/j.compedu.2021.104170>
- 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>
- INEE. (2019). PISA 2018. Items liberados. Secretaría General Técnica. Subdirección General de Atención al Ciudadano, Documentación y Publicaciones. https://sede.educacion.gob.es/publiventa/descarga.action?f_codigo_agc=20232
- INEE. (2021). Nota país: Lectores del siglo XXI. PISA 2018. Secretaría General Técnica. Subdirección General de Atención al Ciudadano, Documentación y Publicaciones.
- Jiao, H., He, Q., y Veldkamp, B. P. (2021). Editorial: Process data in educational and psychological measurement. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.793399>
- Macedo-Rouet, M., Salmerón, L., Ros, C., Pérez, A., Stadler, M., y Rouet, J.-F. (2020). Are frequent users of social network sites good information evaluators? An investigation of adolescents' sourcing abilities. *Journal for the Study of Education and Development*, 43(1), 101-138. <https://doi.org/10.1080/02103702.2019.1690849>
- McGrew, S., Breakstone, J., Ortega, T., Smith, M., y Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education*, 46(2), 165-193. <https://doi.org/10.1080/00933104.2017.1416320>
- Mullis, I. V. S., y Martin, M. O. (2019). *Assessment Frameworks PIRLS 2021*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- 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.psicol.2020.2921>
- OECD. (2011). *PISA 2009 Results: Students On Line*. PISA, OECD Publishing, Paris. <https://doi.org/10.1787/9789264112995-en>
- OECD. (2018). *PISA 2021 Mathematics Framework*. PISA, OECD Publishing, Paris.
- OECD. (2019a). *Beyond proficiency: Using log files to understand respondent behaviour in the survey of adult skills*. OECD Publishing, Paris. <https://doi.org/10.1787/0b1414ed-en>
- OECD. (2019b). *PISA 2018 Assessment and analytical framework*. PISA, OECD Publishing, Paris. <https://doi.org/10.1787/b25efab8-en>
- OECD. (2020). *Curriculum Overload*. PISA, OECD Publishing, Paris. <https://doi.org/10.1787/3081ceca-en>
- OECD. (2021a). *21st-Century readers: Developing literacy skills in a digital world*. PISA, OECD Publishing, Paris. <https://doi.org/10.1787/a83d84cb-en>
- 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.
- Padilla, J. L., y Benítez, I. (2014). Evidencia de validez basada en los procesos de respuesta. *Psicothema*, 26(1), 136-144. <https://doi.org/10.7334/PSICOHEMA2013.259>
- Paniagua, A., y Istance, D. (2018). *Teachers as designers of learning environments*. OECD Publishing, Paris. <https://doi.org/10.1787/9789264085374-en>
- Pedrosa, I., Suárez-Álvarez, J., y García-Cueto, E. (2014). Evidencias sobre la validez de contenido: Avances teóricos y métodos para su estimación. *Acción Psicológica*, 10(2), 3. <https://doi.org/10.5944/ap.10.2.11820>

- Pedrosa, I., Suárez-Álvarez, J., García-Cueto, E., y Muñiz, J. (2016). A computerized adaptive test for enterprising personality assessment in youth. *Psicothema*, 28(4). <https://doi.org/10.7334/psicothema2016.68>
- Pérez, A., Potocki, A., Stadler, M., Macedo-Rouet, M., Paul, J., Salmerón, L., y Rouet, J.-F. (2018). Fostering teenagers' assessment of information reliability: Effects of a classroom intervention focused on critical source dimensions. *Learning and Instruction*, 58, 53–64. <https://doi.org/10.1016/j.learninstruc.2018.04.006>
- Pohl, S., Ulitzsch, E., y von Davier, M. (2021). Reframing rankings in educational assessments. *Science*, 372(6540), 338–340. <https://doi.org/10.1126/science.abd3300>
- Postigo, Á., Cuesta, M., Pedrosa, I., Muñiz, J., y García-Cueto, E. (2020). Development of a computerized adaptive test to assess entrepreneurial personality. *Psicología: Reflexão e Crítica*, 33(1), 6. <https://doi.org/10.1186/s41155-020-00144-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>
- Serra-Garcia, M., y Gneezy, U. (2021). Nonreplicable publications are cited more than replicable ones. *Science Advances*, 7(21). https://doi.org/10.1126/SCIADV.ABD1705/SUPPL_FILE/SCIADV.ABD1705_SM.PDF
- Suarez-Alvarez, J. (2021). "Are 15-year-olds prepared to deal with fake news and misinformation?". *PISA in Focus*, No. 113, OECD Publishing, Paris, <https://doi.org/10.1787/6ad5395e-en>
- von Davier, A. A., Deonovic, B., Yudelso, M., Polyak, S. T., y Woo, A. (2019). Computational psychometrics approach to holistic learning and assessment systems. *Frontiers in Education*, 4. <https://doi.org/10.3389/feduc.2019.00069>
- von Davier, M., Khorramdel, L., He, Q., Shin, H. J., y Chen, H. (2019). Developments in psychometric population models for technology-based large-scale assessments: An overview of challenges and opportunities. *Journal of Educational and Behavioral Statistics*, 44(6), 671–705. <https://doi.org/10.3102/1076998619881789>
- Vörös, Z., Kehl, D., y Rouet, J.-F. (2021). Task characteristics as source of difficulty and moderators of the effect of time-on-task in digital problem-solving. *Journal of Educational Computing Research*, 58(8), 1494–1514. <https://doi.org/10.1177/0735633120945930>
- Vosoughi, S., Roy, D., y Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. https://doi.org/10.1126/SCIENCE.AAP9559/SUPPL_FILE/AAP9559_VOUGHI_SM.PDF
- Wise, S. L., Im, S., y Lee, J. (2021). The Impact of Disengaged Test Taking on a State's Accountability Test Results. *Educational Assessment*, 26(3), 163–174. <https://doi.org/10.1080/10627197.2021.1956897>
- Zenisky, A., y Sireci, S. (2002). Technological Innovations in Large-Scale Assessment. *Applied Measurement in Education*, 15(4), 337–362.

Apéndice A

Unidad CR551: Isla de Pascua (INEE, 2019)



PISA 2018

Isla de Pascua
Pregunta 1 / 7

Consulta el blog de la profesora a la derecha. Pulsa en una opción para responder a la pregunta.

Según el blog, ¿cuándo empezó la profesora su trabajo de campo?

☐ Durante los años 90.

☐ Hace nueve meses.

☐ Hace un año.

☐ A principios de mayo.

Blog

www.elblogdelaprofesora.com/trabajodecampo/IsladePascua

 El blog de la profesora

Publicado el 23 de mayo a las 11:22 h

Esta mañana, mientras miro por mi ventana, veo el paisaje que he aprendido a amar aquí en Rapa Nui, conocida en algunos lugares como la Isla de Pascua. La hierba y los arbustos son verdes, el cielo es azul y los viejos volcanes extintos se alzan en el horizonte.

Me siento un poco triste por ser esta mi última semana en la isla. He terminado mi trabajo de campo y vuelvo a casa. Más tarde, iré a dar un paseo por las colinas y a despedirme de los moáis que he estado estudiando durante los últimos nueve meses. Esta es una foto de algunas de estas enormes estatuas.



Si has estado siguiendo mi blog durante este año, sabrás que los habitantes de la Isla de Pascua esculpieron estos moáis hace cientos de años. Estos impresionantes moáis se han esculpido en una sola cantera de la parte oriental de la isla. Algunos de ellos pesan toneladas. Aun así los habitantes de la Isla de Pascua pudieron trasladarlos a lugares que quedaban lejos de la cantera sin grúas ni maquinaria pesada.

Durante años, los arqueólogos no supieron cómo se trasladaron estas estatuas enormes. Fue un misterio hasta los años 90, cuando un equipo de arqueólogos y habitantes de la Isla de Pascua demostraron que los moáis habrían podido transportarse y levantarse usando cuerdas hechas de plantas, rodillos de madera y rampas hechas de los grandes árboles que en otra época florecían en la isla. El misterio de los moáis al fin pudo resolverse.

Sin embargo, aún quedaba otro misterio. ¿Qué pasó con esas plantas y los grandes árboles que una vez se usaron para trasladar los moáis? Como digo, cuando miro por mi ventana, veo hierba y arbustos y uno o dos árboles pequeños, pero nada que hubiera podido utilizarse para trasladar estas enormes estatuas. Es un misterio fascinante, y lo estudiaré en futuras publicaciones y clases. Hasta entonces, quizás quieras investigar el misterio por ti mismo. Te recomiendo que empieces con un libro llamado *Colapso* de Jared Diamond. [Esta reseña de Colapso es un buen punto para comenzar.](#)

 Viajero_14 24 de mayo a las 16:31 h

¡Hola, profesora! Me encanta seguir su trabajo sobre la Isla de Pascua. ¡No veo la hora de poder leer *Colapso*!

 Carlos_Isla 25 de mayo a las 9:07 h

A mí también me encanta leer sus experiencias en la Isla de Pascua. Sin embargo, creo que existe otra teoría que debería tener en cuenta. Mire este artículo: www.noticiascientificas.com/Ratas_polinesias_Isla_de_Pascua

PISA 2018

Isla de Pascua

Pregunta 2 / 7

Consulta el blog de la profesora a la derecha. Escribe la respuesta a la pregunta.

En el último párrafo del blog, la profesora escribe: "Aún quedaba otro misterio..."

¿A qué misterio se refiere?

Blog

www.elblogdelaprofesora.com/trabajodecampo/IsladePascua

 El blog de la profesora

Publicado el 23 de mayo a las 11:22 h

Esta mañana, mientras miro por mi ventana, veo el paisaje que he aprendido a amar aquí en Rapa Nui, conocida en algunos lugares como la Isla de Pascua. La hierba y los arbustos son verdes, el cielo es azul y los viejos volcanes extintos se alzan en el horizonte.

Me siento un poco triste por ser esta mi última semana en la Isla. He terminado mi trabajo de campo y vuelvo a casa. Más tarde, iré a dar un paseo por las colinas y a despedirme de los moáis que he estado estudiando durante los últimos nueve meses. Esta es una foto de algunas de estas enormes estatuas.



Si has estado siguiendo mi blog durante este año, sabrás que los habitantes de la Isla de Pascua esculpieron estos moáis hace cientos de años. Estos impresionantes moáis se han esculpido en una sola cantera de la parte oriental de la isla. Algunos de ellos pesan toneladas. Aun así los habitantes de la Isla de Pascua pudieron trasladarlos a lugares que quedaban lejos de la cantera sin grúas ni maquinaria pesada.

Durante años, los arqueólogos no supieron cómo se trasladaron estas estatuas enormes. Fue un misterio hasta los años 90, cuando un equipo de arqueólogos y habitantes de la Isla de Pascua demostraron que los moáis habrían podido transportarse y levantarse usando cuerdas hechas de plantas, rodillos de madera y rampas hechas de los grandes árboles que en otra época florecían en la isla. El misterio de los moáis al fin pudo resolverse.

Sin embargo, aún quedaba otro misterio. ¿Qué pasó con esas plantas y los grandes árboles que una vez se usaron para trasladar los moáis? Como digo, cuando miro por mi ventana, veo hierba y arbustos y uno o dos árboles pequeños, pero nada que hubiera podido utilizarse para trasladar estas enormes estatuas. Es un misterio fascinante, y lo estudiaré en futuras publicaciones y clases. Hasta entonces, quizás quieras investigar el misterio por ti mismo. Te recomiendo que empieces con un libro llamado *Colapso* de Jared Diamond. [Esta reseña de Colapso es un buen punto para comenzar.](#)



Viajero_14

24 de mayo a las 16:31 h

¡Hola, profesora! Me encanta seguir su trabajo sobre la Isla de Pascua. ¡No veo la hora de poder leer *Colapso*!



Carlos_Isla

25 de mayo a las 9:07 h

A mí también me encanta leer sus experiencias en la Isla de Pascua. Sin embargo, creo que existe otra teoría que debería tener en cuenta. Mire este artículo: www.noticiascientificas.com/Ratas_polinesias_Isla_de_Pascua

PISA 2018

Isla de Pascua
Pregunta 3 / 7

Consulta la reseña de *Colapso* a la derecha. Pulsa en las opciones de la tabla para responder a la pregunta.

A continuación hay una lista de afirmaciones de la Reseña de *Colapso*. ¿Estas afirmaciones son hechos u opiniones? Pulsa en **Hecho** u **Opinión** en cada afirmación.

¿Es la afirmación un hecho o una opinión?	Hecho	Opinión
En el libro, el autor describe el colapso de varias civilizaciones debido a las decisiones tomadas y su impacto en el medio ambiente.	☐	☐
Uno de los ejemplos más llamativos mencionados en el libro es la Isla de Pascua.	☐	☐
Esculpiron los moáis, las famosas estatuas, y usaron los recursos naturales que tenían a su alcance para trasladar esos enormes moáis a diferentes lugares por toda la isla.	☐	☐
Cuando los primeros europeos llegaron a la Isla de Pascua en 1722, los moáis todavía estaban allí, pero los árboles habían desaparecido.	☐	☐
El libro está bien escrito y merece ser leído por cualquiera con cierta preocupación por el medio ambiente.	☐	☐

Reseña de *Colapso*

El nuevo libro de Jared Diamond, *Colapso*, es una clara llamada de atención sobre las consecuencias de dañar nuestro medio ambiente. En el libro, el autor describe el colapso de varias civilizaciones debido a las decisiones tomadas y su impacto en el medio ambiente. Uno de los ejemplos más llamativos mencionados en el libro es la Isla de Pascua.

Según el autor, los polinesios se establecieron en la Isla de Pascua alrededor del 700 D.C. Desarrollaron una sociedad próspera de probablemente unas 15.000 personas. Esculpiron los moáis, las famosas estatuas, y usaron los recursos naturales que tenían a su alcance para trasladar esos enormes moáis a diferentes lugares por toda la isla. Cuando los primeros europeos llegaron a la Isla de Pascua en 1722, los moáis aún estaban allí, pero los árboles habían desaparecido. La población se redujo a unos pocos miles de personas que se esforzaban por sobrevivir. Diamond escribe que los habitantes de la Isla de Pascua despejaron la tierra para usos agrícolas y otros fines y sobreexplotaron la caza de numerosas especies de aves acuáticas y terrestres que vivían en la isla. Especula con la idea de que los recursos naturales menguantes derivaron en guerras civiles y en la desaparición de la sociedad de la Isla de Pascua.

La moraleja de este maravilloso pero aterrador libro es que en el pasado, los seres humanos destruyeron su medio ambiente cortando todos los árboles y cazando las especies animales hasta el punto de la extinción. Con cierto optimismo, el autor puntualiza que nosotros podemos elegir **no** cometer los mismos errores en la actualidad. El libro está bien escrito y merece ser leído por cualquiera con cierta preocupación por el medio ambiente.

PISA 2018

Isla de Pascua
Pregunta 4 / 7

Consulta el artículo "¿Fueron las ratas polinesias las que destruyeron los árboles de la Isla de Pascua?" a la derecha. Pulsa en una opción para responder a la pregunta.

¿En qué coincidían los científicos mencionados en el artículo y Jared Diamond?

- ☐ Los humanos se asentaron en la Isla de Pascua hace cientos de años.
- ☐ Han desaparecido grandes árboles de la Isla de Pascua.
- ☐ Las ratas polinesias se comieron todas las semillas de los grandes árboles de la Isla de Pascua.
- ☐ Los europeos llegaron a la Isla de Pascua en el siglo XVIII.

NOTICIAS CIENTÍFICAS

¿Fueron las ratas polinesias las que destruyeron los árboles de la Isla de Pascua?
Por Marcos Kamat, periodista científico

En 2005, Jared Diamond publicó *Colapso*. En el libro, describió el asentamiento humano de la Isla de Pascua (también llamada Rapa Nui).

El libro provocó una tremenda polémica poco después de su publicación. Muchos científicos cuestionaron la teoría de Diamond de lo que pasó en la Isla de Pascua. Coincidían en que cuando los primeros europeos llegaron a la Isla en el siglo XVIII, ya habían desaparecido los enormes árboles, pero no estaban de acuerdo con la teoría de Jared Diamond sobre la causa de la desaparición.

Ahora bien, dos científicos, Carl Lipo y Terry Hunt, han publicado una nueva teoría. Ellos creen que la rata polinesia se comió todas las semillas de los árboles, evitando que crecieran otros nuevos. La rata, según creen, llegó allí accidentalmente o con algún propósito en las canoas que los primeros colonizadores humanos usaron para llegar a la Isla de Pascua.

Los estudios muestran que una población de ratas puede duplicarse cada 47 días. Eso son muchas ratas que alimentar. Para justificar su teoría, Lipo y Hunt señalan los restos de nueces de palma que muestran las marcas roídas hechas por las ratas. Por supuesto, reconocen que los humanos jugaron un papel fundamental en la destrucción de los bosques de la Isla de Pascua, pero creen que la rata polinesia fue la principal culpable entre una serie de factores.

PISA 2018

Isla de Pascua
Pregunta 5 / 7

Consulta el artículo "¿Fueron las ratas polinesias las que destruyeron los árboles de la Isla de Pascua?" a la derecha. Pulsa en una opción para responder a la pregunta.

¿Qué prueba presentan Carl Lipo y Terry Hunt para justificar su teoría de la razón por la que los grandes árboles de la Isla de Pascua desaparecieron?

- ☐ Las ratas llegaron a la isla en las canoas de los colonizadores.
- ☐ Los colonizadores pudieron haber llevado las ratas con algún propósito.
- ☐ Las poblaciones de ratas pueden duplicarse cada 47 días.
- ☐ Los restos de nueces de palma muestran las marcas roídas hechas por las ratas.

Blog Reseña del libro Noticias científicas

www.noticiascientificas.com/Ratas_polinesias_Isla_de_Pascua

NOTICIAS CIENTÍFICAS

¿Fueron las ratas polinesias las que destruyeron los árboles de la Isla de Pascua?

Por Marcos Kamat, periodista científico

En 2005, Jared Diamond publicó *Colapso*. En el libro, describió el asentamiento humano de la Isla de Pascua (también llamada Rapa Nui).

El libro provocó una tremenda polémica poco después de su publicación. Muchos científicos cuestionaron la teoría de Diamond de lo que pasó en la Isla de Pascua. Coincidían en que cuando los primeros europeos llegaron a la isla en el siglo XVIII, ya habían desaparecido los enormes árboles, pero no estaban de acuerdo con la teoría de Jared Diamond sobre la causa de la desaparición.

Ahora bien, dos científicos, Carl Lipo y Terry Hunt, han publicado una nueva teoría. Ellos creen que la rata polinesia se comió todas las semillas de los árboles, evitando que crecieran otros nuevos. La rata, según creen, llegó allí accidentalmente o con algún propósito en las canoas que los primeros colonizadores humanos usaron para llegar a la Isla de Pascua.

Los estudios muestran que una población de ratas puede duplicarse cada 47 días. Eso son muchas ratas que alimentar. Para justificar su teoría, Lipo y Hunt señalan los restos de nueces de palma que muestran las marcas roídas hechas por las ratas. Por supuesto, reconocen que los humanos jugaron un papel fundamental en la destrucción de los bosques de la Isla de Pascua, pero creen que la rata polinesia fue la principal culpable entre una serie de factores.

PISA 2018

Isla de Pascua
Pregunta 6 / 7

Lee las tres fuentes de la derecha pulsando en cada pestaña.

Arrastra y suelta las causas, y el efecto que tienen en común, en los lugares correctos dentro de la tabla sobre de las teorías.

Las teorías

Causa	Efecto	Defensores de la teoría
		Jared Diamond
		Carl Lipo y Terry Hunt
Los moáis fueron esculpidos en la misma cantera.	Las ratas polinesias se comieron las semillas de los árboles y como resultado no pudieron crecer nuevos árboles.	Los colonizadores usaron canoas para llevar ratas polinesias a la Isla de Pascua.
Los grandes árboles desaparecieron de la Isla de Pascua.	Los habitantes de la Isla de Pascua necesitaban recursos naturales para trasladar los moáis.	Los humanos cortaron árboles para limpiar el terreno para uso agrícola y por otras razones.

Blog Reseña del libro Noticias científicas

www.noticiascientificas.com/Ratas_polinesias_Isla_de_Pascua

NOTICIAS CIENTÍFICAS

¿Fueron las ratas polinesias las que destruyeron los árboles de la Isla de Pascua?

Por Marcos Kamat, periodista científico

En 2005, Jared Diamond publicó *Colapso*. En el libro, describió el asentamiento humano de la Isla de Pascua (también llamada Rapa Nui).

El libro provocó una tremenda polémica poco después de su publicación. Muchos científicos cuestionaron la teoría de Diamond de lo que pasó en la Isla de Pascua. Coincidían en que cuando los primeros europeos llegaron a la isla en el siglo XVIII, ya habían desaparecido los enormes árboles, pero no estaban de acuerdo con la teoría de Jared Diamond sobre la causa de la desaparición.

Ahora bien, dos científicos, Carl Lipo y Terry Hunt, han publicado una nueva teoría. Ellos creen que la rata polinesia se comió todas las semillas de los árboles, evitando que crecieran otros nuevos. La rata, según creen, llegó allí accidentalmente o con algún propósito en las canoas que los primeros colonizadores humanos usaron para llegar a la Isla de Pascua.

Los estudios muestran que una población de ratas puede duplicarse cada 47 días. Eso son muchas ratas que alimentar. Para justificar su teoría, Lipo y Hunt señalan los restos de nueces de palma que muestran las marcas roídas hechas por las ratas. Por supuesto, reconocen que los humanos jugaron un papel fundamental en la destrucción de los bosques de la Isla de Pascua, pero creen que la rata polinesia fue la principal culpable entre una serie de factores.

PISA 2018

Isla de Pascua
Pregunta 7 / 7

Lee las tres fuentes de la derecha pulsando en cada pestaña. Escribe la respuesta a la pregunta.

Tras leer las tres fuentes, ¿qué crees que causó la desaparición de los grandes árboles de la Isla de Pascua? Proporciona información concreta de las fuentes para justificar tu respuesta.

NOTICIAS CIENTÍFICAS

¿Fueron las ratas polinesias las que destruyeron los árboles de la Isla de Pascua?

Por Marcos Kamat, periodista científico

En 2005, Jared Diamond publicó *Colapso*. En el libro, describió el asentamiento humano de la Isla de Pascua (también llamada Rapa Nui).

El libro provocó una tremenda polémica poco después de su publicación. Muchos científicos cuestionaron la teoría de Diamond de lo que pasó en la Isla de Pascua. Coincidían en que cuando los primeros europeos llegaron a la Isla en el siglo XVIII, ya habían desaparecido los enormes árboles, pero no estaban de acuerdo con la teoría de Jared Diamond sobre la causa de la desaparición.

Ahora bien, dos científicos, Carl Lipo y Terry Hunt, han publicado una nueva teoría. Ellos creen que la rata polinesia se comió todas las semillas de los árboles, evitando que crecieran otros nuevos. La rata, según creen, llegó allí accidentalmente o con algún propósito en las canoas que los primeros colonizadores humanos usaron para llegar a la Isla de Pascua.

Los estudios muestran que una población de ratas puede duplicarse cada 47 días. Eso son muchas ratas que alimentar. Para justificar su teoría, Lipo y Hunt señalan los restos de nueces de palma que muestran las marcas roídas hechas por las ratas. Por supuesto, reconocen que los humanos jugaron un papel fundamental en la destrucción de los bosques de la Isla de Pascua, pero creen que la rata polinesia fue la principal culpable entre una serie de factores.

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 NNTT; 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 NNTT 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 NNTT 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 NNTT 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 NNTT Y PARA UN MUNDO SIN BARRERAS?

Al mismo tiempo que las NNTT 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 NNTT 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 vídeos 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 necesidad del 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 accesorio, 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 NNTT, 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 NNTT EN LA INDUSTRIA DE LOS TEST

Uno de los actores implicados de forma preminente en la aplicación de las NNTT 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 NNTT y de la industria de los test.

En cierto sentido, ambos resultan contrapuestos; el horizonte temporal de las NNTT 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 NNTT 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 NNTT. 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 NNTT 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 NNTT: 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, NNTT 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 NNTT 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 NNTT en general, lo que complica aún más la ecuación para la creación de test basados en NNTT: 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 NNTT, 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). MATRICES-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 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>
- 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.psicol.2020.2921>
- 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). *MATRICES. 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.

Comisión de Test: Veinticinco años velando por la calidad de los test Test Commission: Twenty five years working on test quality

Ana Hernández¹, Paula Elosua², José Ramón Fernández-Hermida³ y José Muñiz⁴

¹Universidad de Valencia, España. ²Universidad del País Vasco, España. ³Universidad de Oviedo, España. ⁴Universidad Nebrija, España

El objetivo del presente trabajo es mostrar las actividades y proyectos llevados a cabo por la Comisión de Test del Consejo General de Psicología de España para mejorar la calidad y el uso de los test. Se repasan los proyectos más relevantes de la comisión durante sus veinticinco años de existencia, describiendo los trabajos llevados a cabo para a) conocer las opiniones y actitudes de los psicólogos hacia los test, b) evaluar la calidad de los test editados en España, y c) desarrollar estándares, directrices y códigos éticos y deontológicos que mejoren la práctica de los test. También se resaltan otras actividades de la comisión de carácter formativo, así como sus relaciones con otras organizaciones internacionales con fines similares. Finalmente se llevan a cabo algunas reflexiones sobre los retos futuros de la Comisión de Test, haciendo hincapié en el papel clave que tendrán las nuevas tecnologías en la evaluación psicológica.

Palabras clave: Test, Psicometría, Evaluación, Calidad de los test, Uso adecuado de los test.

The objective of this paper is to present the activities and projects carried out by the Spanish Test Commission, established by the Spanish Psychological Association to improve the quality and use of tests. The most relevant projects of the commission during its twenty-five years of existence are reviewed, describing the work carried out: a) to understand the opinions and attitudes of psychologists towards the tests, b) to evaluate the quality of the tests published in Spain, and c) to develop standards, guidelines, and ethical and deontological codes to improve testing practices. The commission's other activities of a formative nature are also highlighted, as well as its relationships with other international organizations with similar purposes. Finally, several reflections on the future challenges of the Test Commission are shared, emphasizing the key role new technologies will have in psychological assessment.

Key words: Test, Psychometrics, Evaluation, Test quality, Appropriate test use.



RÍGENES Y OBJETIVOS

La Comisión de Test fue creada en 1995 por la organización colegial de la Psicología española (Colegio Oficial de Psicólogos, después Consejo General de Colegios Oficiales de Psicólogos) y actúa como órgano asesor de la junta de gobierno. Su objetivo central es llevar a cabo todo un conjunto de actividades y proyectos encaminados a mejorar la calidad y el uso de los test en España. Los test constituyen una de las herramientas más utilizadas por los psicólogos, tanto en su labor profesional como investigadora, por tanto, velar por su calidad psicométrica y por su uso adecuado es fundamental para ofrecer un servicio de calidad a los usuarios de la psicología y a la sociedad en general (Hernández et al. 2015; Muñiz et al., 2020). El uso adecuado de un instrumento de medida requiere, en primer lugar, que este tenga calidad psicométrica probada, pero también que se use de forma adecuada, para lo cual los profesionales e investigadores que lo utilizan deben disponer de una formación apropiada. La Comisión de Test durante sus más de veinticinco años de existencia ha desarrollado distintas actividades encaminadas a mejorar esos tres aspectos: calidad de los test, uso adecuado y formación de quienes utilizan los test. En la fecha en la que se estableció la Comisión de Test, la Psicología ya estaba bien

asentada en España, tanto profesional como académicamente. Para entonces la carrera de Psicología incluye la *Psicometría* y la *Evaluación Psicológica* como asignaturas obligatorias en todas las universidades españolas, hay una regulación profesional de la psicología, y se dispone de una gran variedad de test psicológicos en el mercado. Puede decirse, en suma, que se daban unas condiciones adecuadas para el establecimiento de la comisión de test, como complemento a la formación universitaria reglada.

Con el fin de recoger todos los puntos de vista implicados en la calidad y uso de los test, la Comisión siempre estuvo formada, a lo largo de estos veinticinco años, por expertos en Psicometría y Evaluación Psicológica y por los representantes de las empresas españolas editoras de test. Esta convergencia de los expertos profesionales y académicos y los editores es fundamental para mejorar la calidad y el uso de los test, pues se debe trabajar de forma convergente. También incluye un representante del COP, dado que muchas de las decisiones de la Comisión tienen repercusiones profesionales e institucionales. Desde su fundación hasta 2020 la Comisión fue presidida por José Muñiz, Catedrático de Psicometría, a partir de 2020 la preside Ana Hernández, profesora de Psicometría de la Universidad de Valencia. En la tabla 1 se incluyen todas las personas que formaron parte de la Comisión desde 1995 hasta nuestros días.

Para llevar a cabo sus objetivos la Comisión de Test propone y emprende numerosas iniciativas, colaborando con organizaciones internacionales que comparten objetivos, tales como la Comisión de Test de la Federación Europea de Asociaciones de Psicología (EFPA) (actualmente denominada Board of Assessment), o la Comisión International de Test (International Test Commission -ITC). Cabe señalar

Recibido: 1 noviembre 2021 - Aceptado: 10 diciembre 2021

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

E-mail: Ana.Hernandez@uv.es

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

que la Comisión Europea de test fue creada a propuesta del COP, y su primer presidente fue José Muñiz, que también presidía la comisión española.

PROYECTOS DE LA COMISIÓN DE TEST

Las actividades y proyectos encaminados a velar por la calidad y el buen uso de los test pueden encuadrarse dentro de dos grandes estrategias, una denominada *restrictiva* y otra *informativa*. La *restrictiva* recoge todas las iniciativas dirigidas a restringir el uso de los test a los profesionales que han recibido formación específica para hacerlo, y que pueden ir desde restricciones de compra y uso de pruebas, hasta disposiciones legales que impidan el uso de test a determinados profesionales. La estrategia *informativa* incluye la difusión de información para fomentar el buen uso de los test, tales como códigos éticos y profesionales, directrices y recomendaciones, o la divulgación de información sobre la calidad y características de los test disponibles en el mercado (Muñiz y Bartram, 2007; Muñiz y Fernández-Hermida, 2010; Muñiz et al., 2015). Estas dos estrategias se complementan con una tercera igualmente importante, e íntimamente ligada a las anteriores, la *formativa*, orientada a la actualización de los conocimientos y competencias de los profesionales y a su formación continua (Hernández et al., 2021). Se mencionan a continuación algunos de los proyectos llevados a cabo por la Comisión de Test, los cuales se encuadran en alguna de las estrategias citadas, sobre todo en la formativa e informativa.

TABLA 1
PERSONAS QUE HAN FORMADO PARTE EN ALGÚN
MOMENTO DE LA COMISIÓN DE TEST DEL COP

NOMBRE	AFILIACIÓN
Ana Hernández Baeza	Universidad de Valencia
Ana Martínez Dorado	GiuntiEOS
Eduardo Montes Velasco	COP
Francisco José Abad	Universidad Autónoma de Madrid
Frederique Vallar	Pearson
Gerardo Prieto	Universidad de Salamanca
Jaime Pereña	Tea Ediciones
Javier Rubio	SHL Group
José Luis Galve	CEPE
José Muñiz	Universidad Nebrija
José Ramón Fernández Hermida	COP
Miguel Martínez	EOS
Milagros Antón	TEA Ediciones
Nicolás Seisdedos	Tea Ediciones
Paula Elosua Oliden	Universidad del País Vasco
Rocío Fernández Ballesteros	Universidad Autónoma de Madrid
Rosario Martínez Arias	GiuntiEOS
Vicente Ponsoda	Universidad Autónoma de Madrid
Viviana Gutman Mariach	Pearson

OPINIONES Y ACTITUDES DE LOS PSICÓLOGOS HACIA LOS TEST

Si se pretende mejorar el uso de los test es fundamental conocer cuáles son las opiniones y las actitudes que tienen los profesionales sobre ellos, pues eso nos permitirá hacer hincapié en aquellos aspectos que se perciben como más deficitarios. Con ese fin, la Comisión de test lleva a cabo de forma regular una encuesta entre los profesionales. Hasta la fecha se han llevado a cabo tres ediciones, la primera en el año 1999, la segunda en el año 2009, y la tercera en 2019. Los resultados (Muñiz y Fernández-Hermida, 2000, 2010; Muñiz et al., 2020; Hernández et al., 2021) permiten establecer un diagnóstico preciso sobre la percepción de los psicólogos profesionales acerca del uso de los test, que resulta esencial para promover acciones de mejora continua.

El cuestionario usado en estos estudios permite recoger información sobre varias dimensiones, y además se pide a los participantes que indiquen los tres test que más utilizan en su práctica profesional. Con la finalidad de analizar la evolución de las opiniones y actitudes de los psicólogos hacia los test a través del tiempo, el cuestionario empleado siempre ha sido el mismo, aunque con pequeñas variaciones ajustadas a cada momento temporal. Por ejemplo, en el último estudio (Muñiz et al., 2020, Hernández et al., 2021) el cuestionario estuvo compuesto por 31 ítems agrupados en 5 dimensiones: Actitudes hacia los test, Formación y conocimientos sobre los test, Necesidad de control y regulación de los test y su uso, Influencia de las nuevas tecnologías en el uso de los test, y Problemas en el uso de los test. Además, se pidió a los participantes que señalaran los tres test que más se utilizaban en la práctica profesional, y en la edición del 2019 se incluyeron por primera vez ítems sobre el proceso de evaluación anual de test que lleva a cabo la comisión desde el año 2011, tal como se detalla en el apartado siguiente.

Los resultados más relevantes de la última edición indican (a) que los psicólogos reconocen que la formación recibida en el grado de Psicología puede no ser suficiente para la correcta utilización de la mayoría de los test, y reconocen la necesidad de seguir formándose para usar los test adecuadamente; (b) que las actitudes de los psicólogos respecto al uso de los test son positivas; (c) que, en general, los psicólogos están a favor de incrementar el control y la regulación de los test y su uso; (d) que existe cierto escepticismo sobre la incorporación de las nuevas tecnologías en la práctica de los test y, (e) que en su ámbito profesional se reconocen problemas, como el uso indebido de fotocopias, o no estar al día en materia de test, si bien estos problemas se dan con una frecuencia moderada. El trabajo de Muñiz et al. (2020) analiza las diferencias existentes por especialidad y Hernández et al. (2021) profundizan en aquellas diferencias vinculadas a la edad, el sexo y el sector laboral. En cuanto a la evolución de las opiniones y actitudes a lo largo del tiempo, la comparativa ofrece resultados bastante estables, aunque se aprecian cambios positivos relacionados con el incremento del uso de los test en el ejercicio de la profesión, la mejora de la información disponible sobre la calidad de los test, y la disminución de problemas o malas prácticas, como el uso de fotocopias (Muñiz et al., 2020).

Finalmente cabe destacar que los resultados sobre los test más utilizados en España muestran que, con el paso de los años, los test proyectivos han quedado completamente relegados, predominando claramente el uso de test psicométricos. De hecho, en el último estudio, los 25 test más utilizados por los psicólogos españoles son todos psicométricos, bien sean pruebas adaptadas al contexto español, o test de elaboración nacional (24%). Estos datos son un indicador claro de la gran actividad que se lleva a cabo en nuestro país en materia de construcción, adaptación y publicación de test (Muñiz et al., 2020).

EVALUACIÓN DE LA CALIDAD DE LOS TEST EDITADOS EN ESPAÑA

A la hora de seleccionar un test para su uso es importante que el profesional evalúe o pueda acceder a evaluaciones externas sobre la calidad de los test disponibles. Con esta finalidad la Comisión de Test inició el año 2010 un proceso de revisión de los test utilizados en España. Se trata de evaluaciones tanto cuantitativas como cualitativas realizadas por expertos, cuyo objetivo es ayudar a los profesionales a evaluar el test que mejor se adecúa a sus necesidades.

El proceso de revisión ha tenido como referente el modelo de evaluación de test (Cuestionario de Evaluación de Test, CET) propuesto inicialmente por Prieto y Muñiz (2000). El CET se ha utilizado en las tres primeras ediciones del proyecto, con alguna pequeña modificación en la tercera edición. Para la cuarta edición se revisó por Hernández et al. (2016), generando una nueva versión: CET-R. En esta revisión se incorporaron los avances psicométricos y tecnológicos más relevantes recogidos en el modelo de evaluación de test de la EFPA en su revisión del 2013 (Evers et al., 2013). Ambos modelos (CET y CET-R) están disponibles en la página web de la Comisión de Test: <https://www.cop.es/index.php?page=evaluar-calidad>. Los dos contienen tres grandes apartados: (a). Descripción técnica de la prueba, que incluye aspectos como la finalidad del test, su clasificación, el modo de corrección, la posibilidad de obtener informes automatizados, o el precio, entre otros. (b) Evaluación técnica de las características de la prueba, que incluye cuestiones generales como la calidad del modelo sustantivo, o la calidad de los materiales e informes generados, y cuestiones relacionadas con su calidad psicométrica: análisis psicométrico de ítems, evidencias de validez, fiabilidad y adecuación de las normas o baremos. Para todas estas cuestiones se incluyen preguntas abiertas que permiten razonar las puntuaciones asignadas y proporcionar alguna otra información que pudiera resultar relevante. Y, finalmente, (c) valoración global cuantitativa y cualitativa del test, cuyo resultado cuantitativo queda plasmado en una ficha técnica.

El proceso de revisión se inicia con el nombramiento de un coordinador o coordinadora por parte de la Comisión de Test, y con la selección de las pruebas a evaluar. Cada test es evaluado por dos revisores independientes, habitualmente un psicólogo y un profesional experto en el constructo que mide el test. Las revisiones independientes son integradas por el coordinador en un informe que es enviado a las editoriales para que hagan los comentarios que estimen oportunos. Finalmente, con toda la información a su disposición el coordinador de la evaluación elabora un informe final que se

cuelga en la página web de la Comisión de Test, y está a libre disposición de todos los profesionales. En la tabla 2 aparecen los coordinadores de las diferentes evaluaciones.

Hasta la fecha (noviembre 2021) se han llevado a cabo nueve ediciones de este proceso de revisión, con un total de 89 pruebas revisadas, en sus distintas versiones. En este momento ya está en marcha la décima edición. Todas las revisiones detalladas pueden consultarse en la página web del COP: <http://www.cop.es/index.php?page=evaluacion-test-editados-en-espana>. A partir de ellas se han llevado a cabo diferentes trabajos que resumen tanto los resultados como el proceso de evaluación, realizando también interesantes aportaciones complementarias e innovadoras (por edición sucesiva: Muñiz et al., 2011; Ponsoda y Hontangas, 2013; Hernández et al., 2015; Elosua y Geisinger, 2016; Fonseca y Muñiz, 2017; Hidalgo y Hernández, 2019; Gómez-Sánchez, 2019; Viladrich et al., 2021).

Creemos que la evaluación de la calidad de los test es un proyecto muy relevante, con repercusiones muy positivas sobre la mejora del uso de los test y, por ende, sobre la práctica profesional de la Psicología. Prueba de ello es que, en la última encuesta de opinión de los psicólogos sobre los test (Muñiz et al., 2020), aquellos que conocen las evaluaciones, en general las consideran importantes y necesarias. Indican además que consultan los informes y que éstos les ayudan a tomar decisiones sobre los test a utilizar. No obstante, se constata que únicamente un 22.5% de los encuestados dijeron conocerlas (Muñiz, et al., 2020), por tanto, queda pendiente una clara labor de difusión.

ESTÁNDARES, DIRECTRICES Y CÓDIGOS ÉTICOS Y PROFESIONALES

Dentro de la labor de diseminación de información, la comisión se encarga de propagar, y para ello en muchos casos ha de traducir y resumir, las directrices internacionales más relevantes en materia de evaluación mediante test. En su página web pone a disposición de los profesionales las siguientes directrices:

- ✓ Una adaptación de los códigos éticos y de buenas prácticas de la Asociación Americana de Psicología (APA, 1992).
- ✓ Las normas mínimas para el uso adecuado de los test (adaptadas de Moreland et al., 1995).

TABLA 2
COORDINADORES DE LAS EVALUACIONES DE TEST
LLEVADAS A CABO POR LA COMISIÓN DE TEST DEL COP

EDICIÓN	NOMBRE	AFILIACIÓN
1	José Muñiz	Universidad Nebrija
2	Vicente Ponsoda	Universidad Autónoma de Madrid
3	Ana Hernández Baeza	Universidad de Valencia
4	Paula Elosua Oliden	Universidad del País Vasco
5	Eduardo Fonseca Pedrero	Universidad de la Rioja
6	M. Dolores Hidalgo Montesinos	Universidad de Murcia
7	Laura E. Gómez Sánchez	Universidad de Oviedo
8	Carme Viladrich	Universidad Autónoma de Barcelona
9	Luis Manuel Lozano	Universidad de Granada

- ✓ Las directrices de la ITC referidas a: (a) uso ético y adecuado de los test; (b) traducción y adaptación de test de unas culturas a otras (ver Muñiz et al., 2013 y Hernández et al., 2020); (c) control de calidad de las puntuaciones de los test, su análisis e informes; (d) seguridad de los test, exámenes y otras evaluaciones; y (e) uso de los test y otros instrumentos de evaluación en investigación. En Muñiz et al. (2015) puede consultarse un resumen de las tres últimas directrices.

Por lo que se refiere al proceso de evaluación psicológica, que incluye a los test como una herramienta fundamental, cabe destacar la participación de la Comisión de test en la elaboración de la Guía europea del Proceso de Evaluación (Fernández-Ballesteros et al., 2001), traducida al español en la web de la comisión, y de la Norma ISO 10667 (ISO, 2011) en cuya redacción colaboró el entonces presidente de la comisión, José Muñiz. La Norma, que ha sido revisada en 2020, entró en funcionamiento en España el año 2013 y es especialmente relevante para el campo de la Psicología del Trabajo y de las Organizaciones, ya que regula todo lo relativo al proceso de evaluación de las personas en el ámbito laboral y organizacional. Aunque, según información proporcionada por AE-NOR, la agencia responsable de la norma en España (4 de noviembre de 2021), en este momento solo hay 10 empresas certificadas, la encuesta llevada a cabo por el Grupo de Trabajo de Psicología y Buenas Prácticas en Reclutamiento y Selección de Personas del COP de Madrid (COP-Madrid, 2015), sugiere que la norma puede estar siendo adoptada a nivel menos formal por un número mayor de organizaciones. Según esta encuesta, y tal como recomienda la Norma: (a) la mayoría de empresas llevan un registro de la documentación técnica de los métodos de evaluación que utilizan en sus procesos de reclutamiento y selección (56%), (b) las personas que realizan las evaluaciones poseen la formación técnica/específica requerida (74%), (c) se informa a las personas evaluadas sobre las distintas fases del proceso y las implicaciones de cada una de estas fases (90%), y (d) se garantiza la seguridad y la confidencialidad de la información recabada durante el proceso de evaluación (91%). Estos números indican que se va por el buen camino, si bien, en algunos aspectos, todavía queda espacio para la mejora.

OTRAS ACCIONES Y PROYECTOS

La organización de mesas redondas y simposios centrados en la evaluación mediante test en congresos nacionales e internacionales, o la organización de acciones formativas puntuales sobre construcción, utilización y evaluación de pruebas son acciones complementarias organizadas por la comisión de test. En concreto, destacan los cursos organizados por el Consejo General del COP a través del Programa de Formación Continuada a Distancia (FOCAD) (Elosua, 2019; Muñiz y Fonseca, 2017), o los ofrecidos por las principales editoriales de test que operan en España (TEA, Pearson, GiuntiEOS o CEPE). Son estas últimas las que velan por aplicar la única acción restrictiva que se ejerce en nuestro país para el uso de los test: exigir que únicamente las personas cualificadas tengan acceso a los test según su clasificación de acuerdo con las categorías APA (A, B y C) (APA, 2014). Aquellos test clasificados como B

(pruebas colectivas de carácter cognoscitivo y de personalidad) o C (pruebas individuales y test proyectivos) únicamente pueden ser adquiridas por psicólogos.

INTERNACIONALIZACIÓN DE LA COMISIÓN DE TEST

Como no podría ser de otra manera en un mundo globalizado, los proyectos en los que trabaja la Comisión traspasan los límites de nuestras fronteras, y se enmarcan en una visión internacional de colaboración y cooperación para la mejora en el uso de los test. La Comisión de Test trabaja regularmente con organismos internacionales como la Federación Europea de Asociaciones de Psicólogos (EFPA, *European Federation of Psychologists' Associations*), y la Comisión Internacional de Test (ITC, *International Test Commission*) que ofrecen un marco de referencia y un ámbito de aplicación internacional en la regulación del uso de los test (Bartram, 2011; Muñiz y Bartram, 2007; Muñiz et al., 2001).

En este sentido es importante destacar el trabajo de la ITC en el desarrollo y difusión de directrices relacionadas con diferentes aspectos relacionados con la construcción, adaptación y uso de test. Las directrices tratan de dar respuesta a las demandas y retos a los que se enfrenta el uso correcto de test. Además de las directrices ya mencionadas, la ITC ha realizado recomendaciones sobre: (a) evaluaciones a gran escala en poblaciones lingüísticamente y/o culturalmente diversas (ITC, 2018), (b) uso de actualizaciones de test y test obsoletos (ITC, 2015), (c) evaluación informatizada y a través de internet (ITC, 2005) y a publicar próximamente, (d) evaluaciones apoyadas en la tecnología, desarrolladas de forma conjunta con la *Association of Test Publishers* (ATP).

La Comisión de Test en su calidad de miembro del *Committee on Tests and Testing* (desde 2011 denominado *Board of Assessment*) de la EFPA, ha participado en los estudios internacionales sobre las actitudes y opiniones de los psicólogos hacia los test. En la primera edición del proyecto participaron 6 países europeos (Muñiz et al., 2001), y en la segunda el número de participantes se incrementó hasta 17 (Evers et al., 2012), a los que se añadieron posteriormente 12, que incluían países de África, América y Oceanía, como Brasil, Líbano, Indonesia, Nigeria o Nueva Zelanda (Evers et al., 2017). En este momento se continúa recopilando información para abordar una tercera edición del proyecto. Los resultados de estos estudios permiten comparar las opiniones de los psicólogos de distintos países. Los últimos datos concluyen que en todos los países en general la opinión hacia los test es positiva, apreciándose un ligero escepticismo con relación al uso de Internet, con excepción de China que puntúa significativamente alto en este aspecto (Evers et al., 2017). Una vez finalice el estudio será interesante analizar la evolución de las opiniones y actitudes durante los últimos 30 años.

Otro proyecto importante de la EFPA que ha calado en la Comisión es el desarrollo y actualización de un modelo para la evaluar la calidad de los test. El primer modelo fue publicado en la página web de la EFPA en 2002 (Bartram, 2002), y posteriormente ha sido revisado en dos ocasiones: 2008 (Lindley et al., 2008) y 2013 (Evers et al., 2013). La última actualización sirvió de inspiración al modelo español CET-R. En este momento la EFPA ha formado un nuevo grupo de trabajo para estudiar una nueva actualización, que incorpore

aspectos novedosos, como la gamificación, que están siendo incorporados en la evaluación psicológica y educativa.

Finalmente, destacamos la creación de los estándares de competencias requeridas para poder usar test en los distintos campos de la psicología. Estos estándares son de obligado cumplimiento para los países miembros de EFPA que estén interesados en obtener certificaciones europeas de usuarios de test. Este proyecto se puso en marcha junto con la EAWOP (*European Association of Work and Organizational Psychology*) para el ámbito de la Psicología del Trabajo y las Organizaciones y, posteriormente, se generalizó a las áreas de Educación y de Salud. Estas certificaciones europeas, que acreditan las competencias de los usuarios de test, están siendo ya emitidas en países como Reino Unido o Suecia.

Toda documentación referida a los distintos proyectos de la EFPA en materia de test, puede consultarse en <http://assessment.efpa.eu/documents/>. Las directrices y recomendaciones de la ITC pueden consultarse en <https://www.intestcom.org/page/28>.

MIRANDO HACIA EL FUTURO

La Comisión de Test ha cumplido veinticinco años, y en sus bodas de plata, además de la mirada a lo ya hecho, queremos hacer algunas reflexiones sobre el futuro. Lo hacemos con suma prudencia, sabedores de que el futuro, como bien nos advirtió Séneca (2013), yace en la incertidumbre, y que su naturaleza es líquida, evanescente y está plagada de cisnes negros, es decir, sucesos altamente improbables e impredecibles que cambian nuestras vidas, la pandemia por COVID-19, fue el último (Bauman, 2002; Taleb, 2008). En los años venideros la comisión va a continuar fortaleciendo y potenciando las líneas de trabajo que ya tienen una sólida trayectoria, y para potenciar su eficacia y utilidad hay que tratar de mejorar su difusión entre los profesionales de la psicología. Como hemos señalado, en el último estudio de opinión sobre los test por parte de los psicólogos (Muñiz, et al., 2020), solo un 22.5% de los encuestados dijeron conocer una actividad central de la comisión como es la evaluación de la calidad de los test que se hacen públicas en la página web del COP. Tampoco la publicación de estándares y recomendaciones sobre el uso de los test parece, en sí misma, suficiente para cambiar las prácticas evaluativas (Rios y Sireci, 2014). Una vía para mejorar el impacto de las directrices relacionadas con el uso de test sería su inclusión de forma explícita en los planes de estudio de Psicología, principalmente a través de las asignaturas de Psicometría y Evaluación Psicológica. Algunas universidades ya lo hacen, incorporando el modelo CET-R en las prácticas de Psicometría (Viladrich et al., 2021). En cualquier caso, se debe seguir trabajando en aumentar la visibilidad y el impacto de las acciones de la Comisión de Test entre los profesionales, lo que redundará en un acortamiento de la distancia entre el mundo académico y la investigación y la práctica profesional (Elosua, 2012).

Pero seguramente los principales retos futuros de la comisión tienen que ver con la creciente incorporación de las nuevas tecnologías a la evaluación. Los avances tecnológicos están influyendo de forma determinante en las distintas fases de la evaluación y en los propios instrumentos de medida. Por citar solo algunos ejemplos, la elaboración de los ítems, con contenidos multimedia, o basados en juegos, la obtención de puntuaciones y la elaboración de informes, que pue-

de realizarse de forma automatizada, o el modo de aplicación de las pruebas, a través de internet, mediante aplicaciones adaptativas, etc. (Nieto et al., 2018; Marshall et al., 2010; Sanz et al., 2020; Selow et al., 2019; Sorrel et al., 2021; Wan y Henly, 2012). Todo ello genera nuevas necesidades y situaciones a las que hay que responder. La reciente pandemia originada por la COVID-19 y su impacto en la evaluación psicológica, por ejemplo, muestran la necesidad continua de adaptación a nuevos entornos y el papel que puede jugar la tecnología en ese proceso. El confinamiento y el distanciamiento social forzaron la necesidad de realizar evaluaciones a distancia y de administrar test de forma remota (Elosua, 2021), algo que habría sido impensable hace décadas. Pero no solo eso, la tecnología, especialmente a través de teléfonos móviles y otros portátiles, también está contribuyendo al aumento de la evaluación ambulatoria, que recoge información sobre las conductas, emociones, pensamientos, etc., de las personas, de una forma personalizada, dinámica, contextual y ecológica (Muñiz y Fonseca-Pedrero, 2019). Y estos avances permiten el uso de nuevos indicadores, tales como respuestas fisiológicas, indicadores físicos, movimientos oculares, o información proporcionada por las redes sociales. Este tipo de indicadores generan tal cantidad de información que las técnicas tradicionales de análisis de datos no son capaces de manejarlos, por lo que emergen nuevas metodologías de análisis agrupadas bajo la denominación de *big data* (Kosinski et al., 2013). Este tipo de datos también requieren nuevos modelos psicométricos, como los modelos de redes (Borsboom y Cramer, 2013; Fonseca-Pedrero, 2018), o los modelos de sistemas dinámicos (Nelson et al., 2017).

Todos estos avances suponen grandes oportunidades para la evaluación psicológica y tienen importantes ventajas. Tal y como señalan Simmering et al. (2019), en primer lugar, facilitan la recogida de datos dinámicos, en tiempo real, y permiten incluir información contextual. Todo esto puede contribuir a aumentar la validez ecológica de las evaluaciones, y evitar la necesidad de responder de forma retrospectiva o de generalizar a partir de una evaluación puntual mediante un test tradicional. En segundo lugar, facilita el registro de información sobre procesos (tiempos de respuesta a los ítems, mediciones de la pupila, interacciones sociales) que puede resultar muy útil para obtener evidencias de la validez de las medidas de ciertos constructos, como el control de la atención o, a nivel grupal, la confianza o el liderazgo. En tercer lugar, permite no solo crear test adaptativos, sino que también puede dar a los individuos feedback personalizado en función de sus respuestas o de las puntuaciones globales en el test, lo que en sí mismo es una intervención que puede mejorar el aprendizaje y ciertos comportamientos. Asimismo, con la evaluación ambulatoria, la identificación casi inmediata de estados o situaciones de riesgo para la persona permitirá realizar acciones inmediatas que incrementen las posibilidades de éxito de la intervención. Finalmente, la posibilidad de presentar ítems audiovisuales o de realidad virtual permite crear situaciones más realistas que aumenten la validez ecológica de las evaluaciones y disminuyan ciertos sesgos de respuesta como la deseabilidad social (Woods et al., 2020). Además, estos formatos suelen ser más motivadores para las personas, que se involucran más en las pruebas, lo que podría contribuir a reducir el error de medida.

Algunos de los avances tecnológicos mencionados ya se están implementando en nuestro país. Por ejemplo, hay más test adaptativos comercializados (Barrada, 2012), y se están creando aplicaciones móviles para la evaluación ambulatoria, como, por ejemplo, la app para realizar el seguimiento telemático de niños y jóvenes en riesgo de padecer problemas psicológicos durante el reciente confinamiento por COVID-19: <https://www.fbbva.es/noticias/una-app-para-el-seguimiento-de-jovenes-con-problemas-de-salud-mental-durante-el-confinamiento/>. Sin embargo, según la encuesta realizada por Muñiz et al. (2020), la mayoría de los psicólogos colegiados siguen siendo algo reticentes a la incorporación de este tipo avances tecnológicos. Dicha reticencia puede estar justificada por las cuestiones que todavía no están bien resueltas y lo retos que suponen todos los avances mencionados (Iliescu y Greiff, 2019; Simmering et al., 2019; Tonidandel et al., 2002).

Uno de los principales retos es garantizar que estos avances tecnológicos tan prometedores no pongan en peligro lo esencial: la validez de constructo (Iliescu y Greiff, 2019; Simmering et al., 2019; Woods et al., 2020). Por muy motivadoras y realistas que sean las pruebas, por más información objetiva que incorporen, es necesario garantizar la calidad psicométrica de las puntuaciones generadas. Algunas reflexiones y ejemplos del rigor psicométrico que debe acompañar a los test innovadores tecnológicamente pueden encontrarse en los trabajos de este monográfico (Abad et al., 2022; Andrés, et al., 2022; Elosua, 2022; Fonseca-Pedrero et al., 2022; Santamaría y Sánchez-Sánchez, 2022; Suárez-Álvarez et al., 2022). Por su parte, la Psicometría deberá avanzar en el diseño de nuevas formas de evaluar la calidad psicométrica. Además, la incorporación de la tecnología abre toda una serie de cuestiones éticas y de equidad, dependiendo de los recursos tecnológicos y del grado de familiaridad con la tecnología de las personas evaluadas, sin descartar posibles reacciones negativas ante la evaluación. Otro foco de nuevos problemas son las cuestiones de seguridad, tales como el control de ítems expuestos online, la identidad de participantes que son evaluados online, la privacidad (por ejemplo, cuando se emplean datos de redes sociales), el mantenimiento y almacenamiento de gran cantidad de datos, etc.

A pesar de estos problemas y riesgos, los test que emplean algunos de los avances metodológicos presentados (los test informatizados, la obtención automática de puntuaciones, los test adaptativos, los informes automatizados, la evaluación ambulatoria) están teniendo cada vez más presencia en nuestro país. Y no nos cabe duda de que se irán incorporando muchos más. Sin embargo, antes de adoptar una tecnología concreta, los psicómetras, los autores y los editores de test deben valorar los posibles costes y beneficios de dicha tecnología para evaluar un constructo determinado en la población de interés. Y, sobre todo, las pruebas que implementen estos avances deben ir acompañadas de garantías de calidad psicométrica, con estudios rigurosos que las avalen.

La Comisión debe estar muy atenta a cómo los distintos avances se implementan en nuestro país y son adoptados por las casas editoriales y por los profesionales. En la medida en que las distintas tecnologías se incorporen a la práctica profesional de la Psicología, será necesario actualizar el modelo de evaluación de test (CET-R). Por

otra parte, la traducción y difusión de las directrices que regulan la evaluación informatizada y a través de internet (ITC, 2005) y de las próximas directrices de la ITC que regulan las evaluaciones basadas en la tecnología, y su seguimiento, permitirán aumentar el rigor en la construcción e implementación de estos test de nueva generación. Sin embargo, hay que tener en cuenta que muchas de estas nuevas formas de evaluación utilizan complejos algoritmos y modelos que escapan con frecuencia a las competencias y conocimientos de los psicólogos, siendo necesario el trabajo colaborativo con otros profesionales, como programadores e ingenieros de software y hardware. La tecnología por sí sola no es garantía de nada (Liem et al., 2018), es crucial que la incorporación de la tecnología al ámbito de la evaluación se haga siempre con garantías psicológicas y psicométricas, no dejándose deslumbrar por los fuegos artificiales de la tecnología por la tecnología. Ni los datos ni la tecnología saben psicología, ese saber lo ponen los psicólogos; la definición clara del constructo psicológico evaluado, la fiabilidad y la validez no son negociables, por citar solo tres cuestiones esenciales. La Comisión de Test del COP, cuan Argos de cien ojos, pondrá todo su empeño para velar para que se cumplan dichas garantías psicológicas y psicométricas.

AGRADECIMIENTOS

Los autores desean expresar su máximo agradecimiento a todas las personas que han colaborado con la Comisión de Test en el desarrollo de sus proyectos y actividades, con especial mención para quienes formaron parte de ella en algún momento. Muchas gracias a todos, sin su ayuda no hubiera sido posible el desarrollo de las actividades descritas en el artículo.

CONFLICTO DE INTERESES

Los autores declaran que 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>
- American Psychological Association (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, 47(12), 1597-1611. <https://doi.org/10.1037/0003-066X.47.12.1597>
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Psychological Association.
- 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>
- Barrada, J. R. (2012). Tests adaptativos informatizados: una perspectiva general. *Anales de Psicología*, 28(1), 289-302.
- Bartram, D. (2002). *Review model for the description and evaluation*

- of psychological tests. Brussels: European Federation of Psychologists' Associations (EFPA)
- Bartram, D. (2011). Contributions of the EFPA Standing Committee on Test and Testing (SCTT) to standards and good practice. *European Psychologist*, 16(2), 149-159. <https://doi.org/10.1027/1016-9040/a000093>
- Bauman, Z. (2002). *Modernidad líquida*. México: FCE.
- Borsboom D., y Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91-121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>
- Colegio Oficiales de Psicólogos de Madrid (COP-Madrid) (2015). *II Encuesta del Observatorio de Reclutamiento y Selección de Personas*. https://issuu.com/colegiooficialpsicologosmadrid/docs/presentacion_de_resultados_2014_de
- Elosua, P. (2012). Tests publicados en España: Usos, Costumbres y Asignaturas Pendientes. *Papeles del Psicólogo*, 33(1), 12-21.
- Elosua, P. (2019). Los test psicológicos en la práctica profesional. FOCAD. Consejo General de Psicología de España.
- Elosua, P. (2021). Aplicación remota de test: riesgos y recomendaciones. *Papeles del Psicólogo*, 42, 33-37. <https://doi.org/10.23923/pap.psicol2021.2952>
- 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>
- Elosua, P., y Geisinger, K. F. (2016). Cuarta evaluación de tests editados en España: Forma y fondo. *Papeles del Psicólogo*, 37(2), 82-88.
- Evers, A., Muñiz, J., Bartram, D., Boben, D., Egeland, J., Fernández-Hermida, J.R., y col. (2012). Testing practices in the 21st Century: Developments and European psychologists' opinions. *European Psychologist*, 17(4), 300-319. <https://doi.org/10.1027/1016-9040/a000102>
- Evers, A., Muñiz, J., Hagemester, C., Høstmælingen, A., Lindley, P., Sjöberg, A., y Bartram, B. (2013). Assessing the quality of test: Revision of the EFPA review model. *Psicothema*, 25(3), 283-291. <https://doi.org/10.7334/psicothema2013.97>
- Evers, A., McCormick, C. M., Hawley, L. R., Muñiz, J., Balboni, G., Bartram, D., ... & Zhang, J. (2017). Testing practices and attitudes toward tests and testing: An international survey. *International Journal of Testing*, 17(2), 158-190. <https://doi.org/10.1080/15305058.2016.1216434>
- Fernández-Ballesteros, R., DeBruyn, E.E.J., Godoy, A., Hornke, L., Ter Laak, J. Vizcarro, C., Westhoff, K., Westmeyer H. y Zacagnini, J.L. (2001) Guidelines for the Assessment Process (GAP): A proposal for Discussion. *European Journal of Psychological Assessment*, 17(3), 187-200. <https://doi.org/10.1027/1015-5759.17.3.187>
- Fonseca-Pedrero, E. (2018). Análisis de redes en psicología. *Papeles del Psicólogo*, 39(1), 1-12. <https://doi.org/10.23923/pap.psicol2018.2852>
- Fonseca, E., y Muñiz, J. (2017). Quinta evaluación de tests editados en España: mirando hacia atrás, construyendo el futuro. *Papeles del Psicólogo*, 38(3), 161-168. 7. <https://doi.org/10.23923/pap.psicol2017.2844>
- Fonseca-Pedrero, E., Ródenas, Gabriel, Pérez-Albéniz, A., Al-Halabí, S., Pérez, M., y Muñiz, J. (2022). La hora de la evaluación ambulatoria. *Papeles del Psicólogo*, 43(1), 21-28 <https://doi.org/10.23923/pap.psicol.2983>
- Gómez Sánchez, L. E. (2019). Séptima evaluación de test editados en España. *Papeles del Psicólogo*, 40(3), 205-210. <https://doi.org/10.23923/pap.psicol2019.2909>
- Hernández, A., Hidalgo, M. D., Hambleton, R. K. y Gómez-Benito, J. (2020). International Test Commission guidelines for test adaptation: A criterion checklist. *Psicothema*, 32(3), 390-398. <https://doi.org/10.7334/psicothema2019.306>
- Hernández, A., Hidalgo, M. D., Muñiz, J., y Fernández-Hermida, J. R. (2021). Diferencias de opinión sobre el uso de los test: su relación con la edad el género y el sector laboral. *Papeles del Psicólogo*, 42(2), 94-104. <https://doi.org/10.23923/pap.psicol.2960>
- Hernández, A., Tomás, I., Ferreres, A., y Lloret, S. (2015). Tercera evaluación de tests editados en España. *Papeles del Psicólogo*, 36(1), 1-8.
- Hernández, A., Ponsoda, V., Muñiz, J., Prieto, G., y Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 37(3), 192-197.
- Hidalgo, M. D., y Hernández, A. (2019). Sexta evaluación de test editados en España: resultados e impacto del modelo en docentes y editoriales. *Papeles del Psicólogo*, 40(1), 21-30. <https://doi.org/10.23923/pap.psicol2019.2886>
- Iliescu, D., y Greiff, S. (2019). The impact of technology on psychological testing in practice and policy. *European Journal of Psychological Assessment*, 35(2), 151-155. <https://doi.org/10.1027/1015-5759/a000532>
- International Test Commission (2005). *International Guidelines on Computer-Based and Internet Delivered Testing*. [<http://www.intestcom.org>].
- International Test Commission (2015). *International Guidelines for Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal*. [<http://www.intestcom.org>].
- International Test Commission. (2018). *ITC Guidelines for the Large-Scale Assessment of Linguistically and Culturally Diverse Populations*. [<http://www.InTestCom.org>]
- ISO (2011). *Procedures and methods to assess people in work and organizational settings (part 1 and 2)*. Geneva: Author.
- Kosinski, M., Stillwell, D., y Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behaviour. *Proceedings of the National Academy of Sciences (PNAS)*, 110, 5802-5805. <https://doi.org/10.1073/pnas.1218772110>
- Liem, C. C. S., Langer, M., Demetriou, A., Wicaksana, A.S., Hiemstra, A. M. F., Born, M. P., y König, C. J. (2018). Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. En H. J. Escalante, S. Escalera, I. Guyon, X. Baró, Y. Güçlü-türk, U. Güçlü, y M. A. J. v. Gerven (Eds.), *Explainable and interpretable models in computer vision and machine learning* (pp. 197-251). Springer. https://doi.org/10.1007/978-3-319-98131-4_9
- Lindley, P.A., Bartram, D., & Kennedy, N. (2008). *EFPA Review Model for the description and evaluation of psychological tests: Test review form and notes for reviewers: Version 3.42*. Brussels: EFPA

- Standing Committee on Tests and Testing (September, 2008).
- Moreland, K. L., Eyde, L. D., Robertson, G. J., Primoff, E. S., y Most, R. B. (1995). Assessment of test user qualifications: A research-based measurement procedure. *American Psychologist*, 50(1), 14–23. <https://doi.org/10.1037/0003-066X.50.1.14>
- Muñiz, J., y Bartram, D. (2007). Improving international test and testing. *European Psychologist*, 12(3), 206–219. <https://doi.org/10.1027/1016-9040.12.3.206>
- Muñiz, J., y Fernández-Hermida, J.R. (2000). La utilización de los test en España. *Papeles del Psicólogo*, 76, 41–49.
- Muñiz, J., y Fernández-Hermida, J.R. (2010). La opinión de los psicólogos españoles sobre el uso de los test. *Papeles del Psicólogo*, 31(1), 108–121.
- Muñiz, J., y Fonseca, E. (2017). *Construcción de instrumentos de medida en psicología*. Madrid: FOCAD. Consejo General de Psicología de España
- Muñiz, J., Elosua, P., y Hambleton, R.K. (2013). Directrices para la traducción y adaptación de los test: segunda edición. *Psicothema*, 25(2), 151–157. <https://doi.org/10.7334/psicothema2013.24>
- Muñiz, J., Hernández, A., y Ponsoda, V. (2015). Nuevas directrices sobre el uso de los tests: investigación, control de calidad y seguridad. *Papeles del Psicólogo*, 36(3), 161–173.
- 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://dx.doi.org/10.23923/pap.psicol2020.2921>
- Muñiz, J., Fernández-Hermida, J.R., Fonseca-Pedrero, E., Campillo-Álvarez, A., y Peña-Suárez, E. (2011). Evaluación de test editados en España. *Papeles del Psicólogo*, 32(2), 113–128.
- Muñiz, J., Bartram, D., Evers, A., Boben, D., Matesic, K., Glabeke, K., Fernández-Hermida, J.R., y Zaal, J. (2001). Testing practices in European countries. *European Journal of Psychological Assessment*, 17(3), 201–211. <https://doi.org/10.1027/1015-5759.17.3.201>
- Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T., y Hartmann, J. A. (2017). Moving from static to dynamic models of the onset of mental disorder. *JAMA Psychiatry*, 74(5), 528–534. <https://doi.org/10.1001/jamapsychiatry.2017.0001>
- Nieto, M. D., Abad, F. J., y Olea, J. (2018). Assessing the Big Five with bifactor computerized adaptive testing. *Psychological Assessment*, 30(12), 1678–1690. <https://doi.org/10.1037/pas0000631>
- Parshall, C. G., Harmes, J. C., Davey, T., y Pashley, P. J. (2010). Innovative Items for Computerized Testing. En W. J. van der Linden, y C. A. W. Glas (Eds.), *Elements of Adaptive Testing, Statistics for Social and Behavioral Sciences* (pp. 215–230). Springer. <https://doi.org/10.1007/978-0-387-85461-8>
- Ponsoda, V., y Hontangas, P. (2013). Segunda evaluación de test editados en España. *Papeles del Psicólogo*, 34(2), 82–90.
- Prieto, G., y Muñiz, J. (2000). Un modelo para evaluar la calidad de los test utilizados en España. *Papeles del Psicólogo*, 77, 65–71.
- Rios, J. A., y Sireci, S. G. (2014). Guidelines versus practices in cross-lingual assessment: A disconcerting disconnect. *International Journal of Testing*, 14(4), 289–312. <https://doi.org/10.1080/15305058.2014.924006>
- 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>
- Sanz, S., Luzardo, M., García, C., y Abad, F. J. (2020). Detecting cheating methods on unproctored Internet tests. *Psicothema*, 32(4), 549–558. <https://doi.org/10.7334/psicothema-2020.86>
- Seelow, D. (2019). The art of assessment: Using game based assessments to disrupt, innovate, reform and transform testing. *Journal of Applied Testing Technology*, 20(S1), 1–16.
- Séneca, L. A. (2013). *Sobre la brevedad de la vida, el ocio y la felicidad*. Acanfilado.
- Simmering, V. R., Ou, L., y Bolsinova, M. (2019) What Technology Can and Cannot Do to Support Assessment of Non-cognitive Skills. *Frontiers of Psychology*, 10: 2168. <https://doi.org/10.3389/fpsyg.2019.02168>
- Sorrel, M. A., Abad, F. J., y Nájera, P. (2021). Improving Accuracy and Usage by Correctly Selecting: The Effects of Model Selection in Cognitive Diagnosis Computerized Adaptive Testing. *Applied Psychological Measurement*, 45(2), 112–129. <https://doi.org/10.1177/0146621620977682>
- Suárez-Álvarez, J., Fernández-Alonso, R., García-Crespo, F. J., y Muñiz, 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>
- Taleb, N. (2008). *El cisne negro*. Paidós.
- Tonidandel, S., Quiñones, M. A., y Adams, A. A. (2002). Computer-adaptive testing: The impact of test characteristics on perceived performance and test takers' reactions. *Journal of Applied Psychology*, 87(2), 320–332. <https://doi.org/10.1037/0021-9010.87.2.320>
- Viladrich, C., Doval, E., Penelo, E., Aliaga, J., Espelt, A., García-Rueda, R., y Angulo-Brunet, A. (2021). Octava evaluación de test editados en España: Una experiencia participativa. *Papeles del Psicólogo*, 42(1), 1–9. <https://doi.org/10.23923/pap.psicol2020.2937>
- Wan, L., y Henly, G. A. (2012). Measurement properties of two innovative item formats in a computer-based test. *Applied Measurement in Education*, 25(1), 58–78. <https://doi.org/10.1080/08957347.2012.635507>
- Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., y Anderson, N. R. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, 29(1), 64–77. <https://doi.org/10.1080/1359432X.2019.1681401>



La consecución del funcionamiento psicológico pleno: La dimensión de crecimiento personal

Achieving full psychological functioning: The dimension of personal growth

Carlos Freire¹, María del Mar Ferradás¹, José Carlos Núñez² y Antonio Valle¹

¹Universidade da Coruña, España. ²Universidad de Oviedo, España

Los grandes avances médicos logrados en las últimas décadas han posibilitado un inusitado aumento de la esperanza de vida en buena parte de la población mundial, especialmente en los países desarrollados. Sin embargo, este notable incremento de la salud física de las personas contrasta con las elevadas cifras de enfermedades mentales que asolan en numerosas sociedades contemporáneas. En aras de prevenir y reducir la alta incidencia de problemas de salud mental, en los últimos años estamos asistiendo a un creciente interés por el estudio científico de los recursos psicológicos personales como garante de un funcionamiento psicológico pleno. En consonancia con este planteamiento, numerosos enfoques psicológicos enfatizan la importancia de promover el desarrollo de las capacidades y potencialidades individuales, en la medida en que esta dimensión de crecimiento personal se erige en uno de los factores más genuinos y representativos del bienestar psicológico. El presente trabajo pretende profundizar en la comprensión de este pilar esencial del bienestar humano. En primer lugar, se realiza un sucinto recorrido histórico del constructo crecimiento personal, desde su pionera concepción aristotélica hasta su operativización por parte de psicólogos relevantes como Maslow, Rogers, Jung o Frankl, entre otros. Estas aportaciones conforman la raíz por la que se nutre la psicología contemporánea a la hora de explicar y definir este tópico. En concreto, y apoyándonos en diferentes teorías, enfoques y corrientes psicológicas ampliamente aceptadas en la actualidad, analizamos en este artículo aquellos recursos más estrechamente asociados al crecimiento personal, entre los que se encuentran la autodeterminación, el estado de flujo, el mindfulness, la compasión y el capital psicológico.

Palabras clave: Crecimiento personal, Bienestar psicológico, Funcionamiento psicológico positivo, Recursos personales, Prevención.

The great medical advances achieved in the last decades have allowed an unusual increase in life expectancy in a large part of the world population, especially in developed countries. However, this remarkable improvement in the physical health of people contrasts with the high numbers of mental illnesses that plague in many contemporary societies. In order to prevent and reduce the high incidence of mental health problems, in recent years we are witnessing a growing interest in the scientific study of personal psychological resources as a guarantor of full psychological functioning. In line with this approach, numerous psychological proposals emphasize the importance of promoting the development of individual capacities and potentialities, to the extent that this dimension of personal growth stands as one of the most genuine and representative factors of psychological well-being. This paper aims to deepen the understanding of this core pillar of human well-being. Firstly, there is a brief historical journey of the personal growth construct, from its pioneering Aristotelian conception to its operationalization by relevant psychologists such as Maslow, Rogers, Jung or Frankl, among others. These contributions constitute the root for which contemporary psychology is nourished when explaining and defining this topic. In particular, and relying on different theories, approaches, and psychological currents widely accepted today, secondly, we analyze in this paper those resources most closely associated with personal growth, among which are self-determination, flow, mindfulness, compassion, and psychological capital.

Key words: Personal growth, Psychological well-being, Positive psychological functioning, Personal resources, Prevention.

INTRODUCCIÓN

La buena salud y el bienestar se erigen en uno de los ejes vertebradores de la Agenda 2030 para el Desarrollo Sostenible promulgada por Naciones Unidas en 2015. La consecución de este objetivo, sin embargo, tiene entre sus grandes desafíos redu-

cir las elevadas cifras de enfermedades mentales y trastornos por consumo de sustancias que asolan en numerosas sociedades contemporáneas. Según estimaciones recientes, aproximadamente una de cada seis personas en el mundo (entre el 15 y el 20% de la población) padece al menos un trastorno mental o un trastorno por abuso de sustancias, encontrándose los trastornos de ansiedad (3.76% de la población mundial), depresión (3.44%), consumo de alcohol (1.4%) y otras drogas (e.g., tabaco, cannabis, cocaína) (0.94%) entre los más prevalentes (Global Burden of Disease, 2018).

Históricamente, las intervenciones en el ámbito de la salud mental han adoptado un enfoque reactivo, centrado en la enfermedad y en el déficit y, por ende, en el alivio y la reducción del malestar y del sufrimiento. En las últimas décadas, empero, estamos asistiendo a un progresivo cambio en el concepto de salud, en el que los aspectos vinculados al funcionamiento positivo óptimo gozan de una atención

Recibido: 12 agosto 2021 - Aceptado: 2 diciembre 2021

Correspondencia: Carlos Freire. Departamento de Psicología. Universidad de la Coruña. 15071 A Coruña. España.

E-mail: carlos.freire.rodriguez@udc.es

Financiación: Este trabajo ha sido desarrollado con la financiación de los proyectos EDU2013-44062-P (MINECO), EDU2017-82984-P (MEIC), y Gobierno del Principado de Asturias, España. Fondo Europeo de Desarrollo Regional (Programa de Grupos de Investigación FC-GRUPIN-IDI/2018/000199)



creciente. Este neófito planteamiento tiene su génesis con la proliferación de los Estados de Bienestar tras la Segunda Guerra Mundial, cuando la salud empieza a ser conceptualizada más allá de la mera ausencia de enfermedad, esto es, como un recurso para favorecer la calidad de vida de las personas (Organización Mundial de la Salud, 1986). Sin embargo, no es hasta comienzos de este siglo cuando se produce un verdadero interés por el estudio científico del bienestar como garante de la salud de las personas.

Los estudios sobre bienestar se han organizado en torno a dos grandes tradiciones de investigación (Ryan y Deci, 2001): una relacionada fundamentalmente con el disfrute de una vida placentera (bienestar hedónico) y otra ligada al desarrollo del potencial humano (bienestar eudaimónico). Si bien existe evidencia de que ambos tipos de bienestar, hedónico y eudaimónico, previenen la aparición de estados psicopatológicos o, cuando menos, contribuyen a mitigar sus síntomas (Chakhssi et al., 2018; Kimiecik, 2016), en el presente artículo nos centraremos en el papel desempeñado por el bienestar eudaimónico y, específicamente, en una de sus dimensiones más representativas como exponente del funcionamiento positivo óptimo (Díaz et al., 2006), el crecimiento personal.

2. BIENESTAR EUDAIMÓNICO Y CRECIMIENTO PERSONAL

En su *Ética a Nicómaco* (Ross, 1925), Aristóteles señala que, de entre todos los bienes alcanzables por el ser humano, el fin supremo es lograr la virtud, la eudaimonia. La eudaimonia supone la consecución de la excelencia o la perfección en uno mismo (McDowell, 1980), de acuerdo con las capacidades y el potencial individual.

El crecimiento personal, por tanto, constituye una dimensión central dentro de la perspectiva eudaimonista del bienestar. Ryff (1989) define el crecimiento personal como el afán que experimenta el individuo por continuar creciendo y desarrollando al máximo todas sus potencialidades y capacidades, por ser inconformista con los logros obtenidos en el pasado, tratando de asumir nuevos desafíos personales que le permitan crecer como persona y alcanzar su plena autorrealización. Se trata, como señala Robitschek (1998), de un compromiso activo e intencional con el crecimiento psicológico de uno mismo, lo que conllevaría una implicación tanto cognitiva (e.g., "sé cómo modificar cosas que quiero cambiar en mi vida") como

conductual (e.g., "si quiero modificar algo en mi vida, inicio el proceso de cambio") en pos del desarrollo intraindividual.

La investigación constata que el crecimiento personal se erige en un robusto predictor de la salud mental (Robitschek y Keyes, 2009; Ryff, 2017), por lo que en las siguientes páginas profundizaremos en el análisis de este constructo, realizando un breve recorrido histórico desde su operativización por parte de diferentes corrientes psicológicas, hasta abordar su relación con diversas variables psicológicas nucleares para la consecución de un funcionamiento positivo óptimo.

3. EL CONCEPTO DE CRECIMIENTO PERSONAL EN LA PSICOLOGÍA MODERNA

El afán por seguir creciendo y desarrollando al máximo todas nuestras potencialidades individuales hasta lograr la excelencia, que Aristóteles y, por extensión, el enfoque eudaimonista, señalan como fin supremo del ser humano ha adquirido una notable repercusión en la Psicología moderna de la mano de la corriente humanista, las teorías del desarrollo adulto y el enfoque sociológico (Figura 1).

3.1. El crecimiento personal en la Psicología humanista

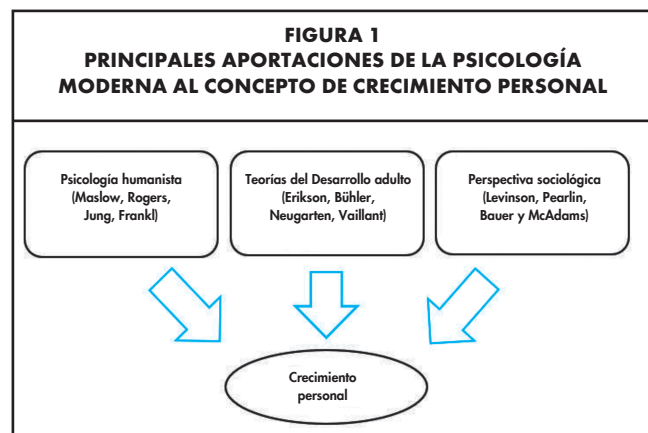
Diversos autores adscritos a la Psicología humanista han hecho aportaciones de consideración al concepto de crecimiento personal. Para Maslow (1968), el ser humano está biológicamente provisto de tendencias que le impulsan de forma natural hacia un desarrollo constante y a la superación personal (autoactualización). El proceso de crecimiento personal se llevaría a cabo una vez que el individuo tiene satisfechas sus necesidades más básicas (i.e., necesidades fisiológicas) y, por consiguiente, enfoca todas sus capacidades hacia su pleno desarrollo individual. Rogers (1961) comparte con Maslow la idea de que el crecimiento personal constituye una necesidad central presente en todos los individuos. La tendencia a la autoactualización, afirma Rogers, es inherente al ser humano.

Para Jung (2003), las personas deben someterse a un proceso de individuación, en función del cual lograrán la plena consciencia de quiénes son realmente. La individuación, por tanto, constituye un proceso básico de maduración personal, de autoconocimiento, con el fin de lograr llegar a ser uno mismo. Por su parte, Frankl (1962) considera que, más que la búsqueda de sí mismo, el individuo persigue fundamentalmente encontrar un significado a su vida. Así, el ser humano alcanza la autotrascendencia, la plenitud, cuando es completamente consciente de su relación con el mundo que le rodea y, fruto de ello, logra hallar un propósito a su existencia más allá de sí mismo.

3.2. Contribuciones de otras corrientes psicológicas a la comprensión del crecimiento personal

Al margen de los planteamientos humanistas, otros desarrollos teóricos como los de Erikson, Bühler, Neugarten o Vaillant, encuadrados en las teorías sobre el desarrollo adulto, sostienen –en la línea de la perspectiva suscrita por Maslow y Rogers– que durante la adultez existe una tendencia natural hacia el crecimiento personal, determinado por el desarrollo de la madurez personal y la sabiduría experiencial (Ryff, 1982). Desde esta perspectiva, madurez y bie-

FIGURA 1
PRINCIPALES APORTACIONES DE LA PSICOLOGÍA MODERNA AL CONCEPTO DE CRECIMIENTO PERSONAL



nestar irían de la mano, erigiéndose ambos en las dos facetas de lo que King (2001) denomina la “buena vida”.

En contraposición a esta perspectiva, los enfoques sociológicos (Levinson, 1986; Pearlin, 1982) defienden que la vida adulta no conduce necesariamente a una mayor madurez. Será el modo de afrontar los retos vitales el que determinará si el individuo experimenta un avance, un estancamiento o un retroceso en su crecimiento personal. De esta manera, no todas las personas lograrían desarrollar al máximo sus capacidades y potencialidades individuales. Al contrario, la posibilidad de que las personas sean capaces de superar sus propios estándares autorreferidos y perfeccionarse como individuos se verá condicionada tanto por factores internos como socio-ambientales. A esta corriente se adscribirían planteamientos como el de Bauer y McAdams (2010). Para estos investigadores, el crecimiento eudaimónico se alcanzará solo si el individuo logra un equilibrio entre su crecimiento intelectual y socioemocional.

Los individuos que se orientan hacia su desarrollo personal intelectual conceden gran importancia a incrementar su autoconocimiento (o el conocimiento de otras personas sobre sí mismas), si bien no necesariamente buscan sentir en profundidad estas experiencias. Esto último sí es relevante para quienes se orientan hacia el crecimiento personal socioemocional, en la medida en que su prioridad es el fortalecimiento y la profundización de los sentimientos y emociones que subyacen a sus propias vidas psicosociales (e.g., sentirse plenamente vivos, cultivar relaciones personales significativas, contribuir al desarrollo de la sociedad y de las nuevas generaciones, etc.). La comprensión de las mismas, sin embargo, sería secundaria.

A su vez, ambas tipologías de crecimiento personal (intelectual y socioemocional) pueden expresarse en un plano intra e interpersonal. De acuerdo con McAdams et al. (1996), el crecimiento intrapersonal se experimenta cuando el individuo percibe una elevada capacidad para influir en su entorno, logra resultados importantes para él mismo y alcanza un nivel óptimo de autogestión. Por su parte, el crecimiento interpersonal vendría determinado por el grado en que el individuo ve satisfechas sus necesidades de amistad y de amor; de relación con diferentes grupos de su entorno, con la sociedad en general, con la humanidad y/o con la divinidad; y está en disposición de ayudar y cuidar a otras personas. Estos cuatro aspectos del crecimiento personal (intelectual, socioemocional, intrapersonal e interpersonal) serían susceptibles de combinarse de forma ortogonal, conformando una matriz 2x2 (Tabla 1).

4. EL CRECIMIENTO PERSONAL EN EL MARCO DE LAS INTERVENCIONES EN PSICOLOGÍA POSITIVA

Tomando como referencia las aportaciones efectuadas por la corriente humanista, así como por las teorías del desarrollo adulto y los enfoques sociológicos, pero, fundamentalmente, al abrigo de la noción de bienestar eudaimónico, el concepto de crecimiento personal constituye en la actualidad uno de los pilares de la denominada Psicología positiva. La Psicología positiva, centrada en comprender los factores que favorecen el florecimiento humano y analizar su contribución a la salud de las personas (Seligman, 2011), ha adquirido un notable interés en los últimos años como campo de intervención

eficaz en la prevención y mejora de diversos trastornos mentales, tanto en población clínica como no clínica. Entre ellos se encuentran la depresión, la ansiedad o el consumo abusivo de sustancias adictivas (Bolier et al., 2013; Rashid, 2015; Schotanus-Dijkstra, 2018).

Las intervenciones en Psicología positiva se basan en la potenciación de un amplio y heterogéneo volumen de recursos psicológicos personales, muchos de los cuales contribuyen de manera significativa a la experimentación, por parte del individuo, de un alto grado de crecimiento personal. Algunos de estos recursos, ampliamente considerados en los programas de intervención, serían la autodeterminación, el estado de flujo, el mindfulness, la compasión y el capital psicológico. A continuación, profundizaremos en el análisis de estas variables.

4.1. Autodeterminación

La teoría de la autodeterminación (Ryan y Deci, 2017) postula que el ser humano tiene la necesidad de implicarse en actividades que posibiliten el pleno desarrollo de sus capacidades y potencialidades individuales hasta lograr la autorrealización. Esta teoría defiende la existencia de tres necesidades psicológicas innatas y universales que conducen al logro de nuestro crecimiento personal: la relación social, la competencia y la autonomía. Toda conducta dirigida a la satisfacción de estas tres necesidades sería intrínsecamente motivante y, por ende, plenamente autodeterminada, en la medida en que es libremente escogida, sin interferencias ni influencias externas.

Asimismo, este tipo de conductas constituyen un factor de protección frente a los problemas de salud mental (Dreison et al., 2018; Taylor et al., 2016), tal y como muestra el Modelo de cambio conductual relacionado con la salud (Ryan et al., 2008) (Figura 2). De acuerdo con este modelo, las intervenciones que promueven (a) un clima de apoyo hacia la adopción de conductas autónomas (i.e.,

TABLA 1
TIPOLOGÍAS DE CRECIMIENTO PERSONAL
(BAUER Y MCADAMS, 2010)

	CRECIMIENTO INTRAPERSONAL	CRECIMIENTO INTERPERSONAL
CRECIMIENTO INTELECTUAL	Incremento del conocimiento conceptual sobre el self Exploración conceptual y aprendizaje en beneficio del autoconocimiento	Incremento del conocimiento conceptual de otras personas y de las relaciones sociales Impulsar el desarrollo intelectual de los demás
CRECIMIENTO SOCIOEMOCIONAL	Experimentación profunda del self y de las propias acciones y situaciones vitales Fortalecimiento de las potencialidades individuales significativas	Experimentación profunda de los demás y de las relaciones sociales Potenciar las capacidades para actuar justa y solidariamente Impulsar el crecimiento de otras personas

aquellas conductas que se realizan porque son en sí mismas importantes o valiosas para la persona, y no por presiones sociales, búsqueda de recompensas extrínsecas o evitación de consecuencias aversivas), así como (b) la búsqueda de aspiraciones vitales intrínsecas (i.e., lograr un crecimiento personal) favorecen la satisfacción de las necesidades de autonomía, competencia y relación. A su vez, estas tres necesidades satisfechas se relacionarían con una mejor salud mental (e.g., menos tendencia a la depresión, a la somatización, a la ansiedad, mayor calidad de vida percibida).

El modelo representado en la Figura 2 refleja la estrecha relación existente entre la motivación intrínseca y el crecimiento personal, tal y como han corroborado empíricamente diversos estudios. Por ejemplo, Carver y Baird (1998) demostraron que la orientación a metas basada en la motivación intrínseca se halla positivamente asociada a la autorrealización, mientras que las metas sostenidas en consonancia con estilos motivacionales extrínsecos correlacionan negativamente con ella. Desde esta consideración, la motivación o interés intrínseco se erigiría en uno de los factores que contribuirían en mayor medida al proceso de búsqueda y desarrollo del potencial humano (Schwartz y Waterman, 2006), siendo esta búsqueda su principal objetivo (Vitterso y Sohlt, 2011). En este sentido, Waterman et al. (2008) señalan que las actividades intrínsecamente motivantes son aquellas que se realizan, no solo por el placer que proporcionan, sino también por la sensación de autorrealización que experimenta el individuo al llevarlas a cabo. En concreto, las actividades que se realizan por motivación intrínseca reúnen una serie de características que las vinculan a la experimentación de altos niveles de crecimiento personal:

- ✓ Relación de equilibrio entre desafío y habilidad/capacidad.
- ✓ Percepción de que las propias actividades, por sí mismas, promueven el desarrollo del máximo potencial individual.
- ✓ Disposición a invertir mucho esfuerzo en el auto perfeccionamiento.
- ✓ Creencia de que esas actividades son realmente importantes para uno mismo.

4.2. Estado de flujo

Las actividades intrínsecamente motivantes han sido conceptualiza-

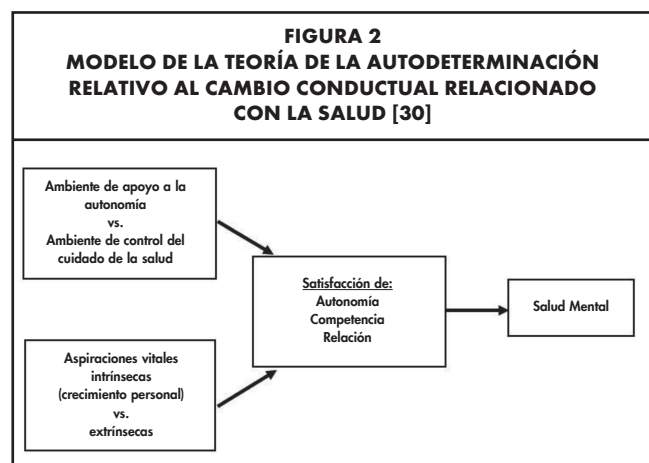
das también en términos de autotéticas o autorreforzantes, dado que el individuo alcanza un grado de implicación máxima durante la realización de las mismas. Esta completa absorción, en la que el individuo se involucra hasta el extremo de olvidarse del tiempo, la fatiga o cualquier otro aspecto ajeno a la propia actividad, es lo que Csikszentmihalyi (1990) ha denominado flujo o experiencia óptima.

Si bien motivación intrínseca y estado de flujo se hallan íntimamente relacionadas, la vinculación entre ambas es desigual, por cuanto una persona intrínsecamente motivada mostrará siempre un gran interés, pero solo en ocasiones alcanzará el estado de flujo (Deci y Ryan, 1985). Este último, a su vez, constituye una experiencia mucho más intensa (Waterman et al., 2003). De acuerdo con ello, Sansone y Harackiewicz (1996) han caracterizado el estado de flujo como arquetipo del interés intrínseco.

El flujo no constituye un estado estático y momentáneo, sino dinámico, en la medida en que la satisfacción se experimenta durante la realización de la actividad, no solo al final de la misma (Salanova et al., 2005). Diversos rasgos permiten identificar la experiencia de flujo (Csikszentmihalyi, 1990):

- ✓ Existencia de metas claras y con gran significación personal.
- ✓ Implicación en actividades que conlleven un equilibrio entre un alto desafío y un alto nivel de capacidad.
- ✓ La atención se focaliza plenamente en lo que se está haciendo.
- ✓ El individuo experimenta una sensación de control sobre sus propias acciones y sobre su entorno inmediato.
- ✓ Completa ausencia de preocupación en relación a un posible fracaso.
- ✓ Percepción distorsionada de la duración del tiempo (habitualmente, referida a una sensación de que el tiempo pasa más rápido de lo normal).

La implicación en actividades que nos permiten experimentar un estado de flujo no solo constituye un factor de protección frente a la adversidad y la patología (Nakamura y Csikszentmihalyi, 2009), sino también una vía para el crecimiento personal (Bonaiuto et al. 2016; Csikszentmihalyi, 1990; Seligman, 2002). Ahora bien, Waterman y colaboradores (Waterman et al., 2008; Waterman et al., 2003) señalan que existe una relación asimétrica entre experiencia de flujo y crecimiento personal, de tal manera que aquella se experimenta con más frecuencia. La principal diferencia entre ambos constructos radica en que el estado de flujo se vincula a la realización de una determinada actividad, pero únicamente se experimenta con respecto a dicha tarea. Por su parte, el crecimiento personal constituye una experiencia de desarrollo individual y autoexpresión mucho más amplia y a largo plazo, resultado de la implicación en actividades intrínsecamente motivantes. En dichas actividades, el individuo puede alcanzar el estado de flujo, pero este no es condición suficiente para que la persona perciba que sus capacidades y potencialidades individuales alcanzan su máximo desarrollo. No obstante, cuantas más actividades cotidianas lleven a experimentar un estado de flujo, mayor número de momentos vitales con alta significación personal percibirá el individuo (Nakamura y Csikszentmihalyi, 2009), lo que contribuirá notablemente a su crecimiento personal.





4.3. Mindfulness

Inspirado en los principios de la meditación budista, el mindfulness es en la actualidad un recurso terapéutico ampliamente utilizado en la psicología occidental. Durante las últimas décadas, las técnicas de mindfulness han evidenciado su eficacia en multitud de programas de mejora del bienestar psicológico (Brown y Ryan, 2003; Burke, 2010), así como en tratamientos psicopatológicos para contrarrestar la ansiedad (Kabat-Zinn, 2003), la depresión (Romero et al., 2020) o el estrés (Martín-Asuero y García-Banda, 2010). Su uso se ha extendido por todo el mundo y se ha ampliado también a otros ámbitos como el laboral (Goilean et al., 2020) o el educativo (García Campayo y Demarzo, 2015).

Pese a que no existe un acuerdo unánime en torno a la definición de este constructo, en general el mindfulness es caracterizado como un estado de la mente humana caracterizado por la atención y conciencia plena, a través del cual el individuo se centra de forma activa y reflexiva en la realidad presente, aceptándola sin pretensión de juzgarla, controlarla o modificarla (Kabat-Zinn, 2003; Shapiro et al., 2016).

El mindfulness ha sido definido también en términos de una capacidad autorregulatoria (Brown y Ryan, 2003), de una habilidad de aceptación (Linehan, 1994), o en términos de control atencional (Teasdale et al., 1995). El mindfulness, por consiguiente, genera autoc conocimiento, es decir, facilita la discriminación de nuestro propio comportamiento. También implica observar nuestros eventos privados sin juzgarlos ni juzgarnos a nosotros mismos, aceptándolos como son, abandonando la lucha y el control, manteniendo una actitud serena y atenta.

Baer et al. (2006) señalan los siguientes elementos clave del mindfulness:

- ✓ Actuar con atención. A diferencia de nuestra tendencia natural a tener la mente divagando, esto es, saltando de un pensamiento a otro de forma inconsciente, mindfulness implica prestar atención momento a momento, pero de una manera relajada, sin una vigilancia estresante.
- ✓ Observar en el momento presente. Se pretende "observar" nuestros propios pensamientos y emociones tal y como suceden, sin buscar su control. El objetivo no es modificar, reprimir o reemplazar un pensamiento que no nos gusta. Se busca desidentificarnos con ellos, tomándolos como simples experiencias.
- ✓ Describir: Los pensamientos y sensaciones de los que estamos siendo conscientes se describen tal y como son experimentados, tanto de una forma verbal como no verbal.
- ✓ Aceptar sin juzgar. Se acepta la experiencia de forma radical, sin valorarla ni etiquetarla. Lo positivo o negativo, lo perfecto o lo imperfecto, en sus diversos grados, son aceptados como experiencias naturales.
- ✓ No reactividad. Generalmente, tendemos a reaccionar de forma inmediata y precipitada a los estímulos, lo que en muchas ocasiones produce una sensación de arrepentimiento ante la respuesta dada. Mindfulness posibilita que el individuo genere un espacio entre los estímulos (internos y externos) y las respuestas proporcionadas (internas y externas), proporcionando un tiempo adicional para pensar cuál es la respuesta más adecuada.

Diversos trabajos han evidenciado que mantenerse centrados en el "aquí y el ahora" constituye una importante vía para el crecimiento personal (Brown y Ryan, 2003; Chang et al., 2015; Iani et al., 2017). De hecho, la búsqueda de la autoliberación con el objetivo de lograr un crecimiento personal constituye una de las principales razones de las personas para practicar la meditación con mindfulness (Pepping et al., 2016; Shapiro, 1994). En efecto, durante la práctica formal del mindfulness, el individuo pasa del "modo hacer" al "modo ser" (Segal et al., 2002). En el "modo hacer", nuestros esfuerzos se enfocan hacia la consecución de una determinada meta, para lo cual tratamos de reducir la brecha existente entre nuestra realidad presente y el cómo nos gustaría que fuesen las cosas. Por tanto, si nuestros esfuerzos se orientan hacia el logro de resultados, con toda probabilidad nuestras preocupaciones se centrarán fundamentalmente en resolver cuestiones tales como "¿soy capaz?" o "¿qué ocurrirá si no soy capaz?", generando estados de ansiedad ante el temor a un posible fracaso. En cambio, en el "modo ser" se acepta la realidad presente tal y como es, desarrollando un estado de atención plena sobre cada una de las experiencias vividas en cada momento. Desaparece, de esta manera, cualquier posible conflicto entre los estados deseados y los reales. Al contrario, nuestro interés se centra en el proceso, en el "¿cómo puedo hacerlo?", esto es, la atención se dirige plenamente a la definición de los pasos necesarios para conseguirlo.

Asimismo, las personas que se sienten autorrealizadas muestran algunas características consideradas también como elementos clave en el mindfulness (Ellis, 2004): disfrutan de las acciones como fines o placeres en sí mismos, son plenamente conscientes de los sentimientos propios sin tratar de reprimirlos, manifiestan una alta autoaceptación, toleran las dificultades y son flexibles y abiertos al cambio.

4.4. Compasión

La compasión ha adquirido una gran atención en la última década como complemento al mindfulness. De hecho, en la tradición budista, mindfulness y compasión son consideradas las dos alas del pájaro de la sabiduría, siendo ambas, por tanto, imprescindibles para poder volar, de ahí que se suelen practicar de forma conjunta aprovechando sus respectivas aportaciones específicas (García Campayo y Demarzo, 2015) (Tabla 2).

Desde la perspectiva psicológica, la compasión hace referencia a la sensibilidad hacia el sufrimiento propio y ajeno, unido al deseo de ayudar a evitar dicho sufrimiento (Gilbert, 2015). Así pues, las terapias centradas en la compasión parten de la tesis de que el dolor es consustancial al ser humano (e.g., el fallecimiento de un ser querido, una enfermedad, una ruptura amorosa) y, como tal, es inevitable. Pero, por lo común, este dolor primario suele acompañarse de un sufrimiento secundario (Young, 2005), que deviene cuando nos resistimos a aceptar el dolor, cuando nos culpabilizamos a nosotros mismos o a otras personas por la situación y cuando anticipamos el futuro (e.g., pensar que nadie nos va a volver a querer, que lo hemos perdido todo, que no teníamos que haber hecho esto o dicho esto otro). El sufrimiento secundario sería el principal responsable de nuestro malestar, pues mantiene vivo el dolor en nuestra mente, y es evitable (García Campayo y De-



marzo, 2015), por lo que es el centro de atención de las terapias centradas en la compasión.

Uno de los modelos conceptual y empíricamente más aceptados sobre compasión es el que ha desarrollado Neff (2003). De acuerdo con este modelo, tres cualidades definirían a las personas compasivas:

- ✓ Autoamabilidad, entendida como propensión a tratarse bien a uno mismo, con afecto y comprensión, en vez de autocriticarse o autoculpabilizarse recurrentemente por los errores.
- ✓ Humanidad compartida, asumiendo que el dolor y el malestar son parte de la vida y, como tales, inherentes al ser humano, en la medida en que todos somos imperfectos y vulnerables. Lo contrario nos llevaría a aislarnos, pensando que nuestra situación y nuestro sufrimiento son únicos.
- ✓ Atención y conciencia plena, atributo que alude a la capacidad para detectar el sufrimiento en nosotros y en los demás, sin pretender juzgarlo, negarlo o huir de él. Según Neff, el mindfulness ayudaría a la persona a no sobreidentificarse con el sufrimiento, quedándose atrapada por este.

La compasión no solo se revela como un recurso psicológico eficaz para prevenir y reducir el afecto negativo, la depresión, el estrés y la ansiedad (Muris et al., 2016; Pauley and McPherson, 2010; Trompeter et al., 2017). También constituye un predictor significativo del funcionamiento psicológico positivo (Neff et al., 2018; Sun et al., 2016). En este sentido, uno de los principales beneficios de la compasión es que promueve en el individuo el afán por lograr su crecimiento personal (Aranha et al., 2019; Neff et al., 2007). Lejos de llevar a la autocom-

placencia y a la autoindulgencia, la compasión propicia una autoevaluación más realista (Leary et al., 2007), lo que conduce a una mayor aceptación y disposición a asumir los errores cometidos, y a enfocarse hacia la automejora y la búsqueda de la mejor versión de uno mismo (Breines y Chen, 2012; Zhang y Chen, 2016).

4.5. Funcionamiento positivo óptimo en el ámbito organizacional: capital psicológico

El interés científico por el funcionamiento positivo óptimo ha trascendido el ámbito clínico, alcanzando también gran relevancia en otros ámbitos como el organizacional. En este sentido, en las últimas dos décadas se observa una atención creciente hacia el estudio y aplicación de los recursos, capacidades y potencialidades humanas que contribuyen a lograr un funcionamiento más adaptativo e incrementar el bienestar del trabajador en el seno de la organización (Luthans, 2002). Se pretende, en definitiva, identificar aquellas variables psicológicas que puedan constituir recursos laborales eficaces para facilitar la consecución de los objetivos organizacionales; reducir las demandas y presiones laborales que puedan entrañar un coste importante para la salud física y psicológica del trabajador; y estimular su crecimiento personal (Demerouti et al., 2001).

Con esta finalidad, Luthans et al. (2007) propusieron el constructo capital psicológico como exponente paradigmático de la conducta organizacional positiva. El término capital psicológico alude a un estado de desarrollo psicológico positivo caracterizado por:

- ✓ Confiar en la capacidad propia para esforzarse y tener éxito en tareas que entrañen un desafío (autoeficacia).
- ✓ Perseverar en los objetivos y, cuando sea necesario, reorientarlos para alcanzar el éxito (esperanza).
- ✓ Hacer atribuciones positivas respecto al rendimiento presente y futuro (optimismo).
- ✓ Resistir y recuperarse de las adversidades para lograr el éxito (resiliencia).

Se trata de un recurso psicológico eminentemente plástico y, como tal, puede ser potenciado y desarrollado (Luthans y Youssef-Morgan, 2017). De hecho, el capital psicológico resulta adaptativo, no solo desde el punto de vista productivo, favoreciendo el compromiso del trabajador con la organización, su rendimiento y su satisfacción laboral (Datu et al., 2018; Djorouva et al., 2019; Newman et al., 2014), sino también para su salud mental. Así, por un lado, numerosos trabajos han encontrado una relación negativa significativa entre el capital psicológico y la depresión, la ansiedad, el estrés, el burnout o el abuso de sustancias (Krasikova et al., 2015; Rahimnia et al., 2013; Rehman et al., 2017). Por otro lado, el capital psicológico se erige en un predictor del bienestar psicológico (Baron et al., 2016; Hernández-Varas y García-Silgo, 2021), contribuyendo significativamente al crecimiento personal del individuo (Culbertson et al., 2010; Manzano-García y Ayala, 2017).

Los cuatro componentes del capital psicológico (autoeficacia, esperanza, optimismo y resiliencia) no solo ejercen una acción sinérgica en su relación con el crecimiento personal, sino que, de forma individual, también parecen contribuir a su consecución, tal y como se expone en los siguientes apartados.

TABLA 2
APORTACIONES ESPECÍFICAS DE LAS TERAPIAS DE
MINDFULNESS Y COMPASIÓN
(GARCÍA CAMPAYO Y DEMARZO, 2015)

	MINDFULNESS	COMPASIÓN
Pregunta a la que responde	¿Cuál es la experiencia aquí y ahora?	¿Qué necesita la persona ahora para estar bien y disminuir el sufrimiento?
Objetivo	Tomar conciencia de la experiencia real y aceptar su naturaleza	Reconfortar a la persona frente al sufrimiento, entendiendo que el dolor primario es inherente al ser humano
Riesgo de cada terapia si no se equilibra con la otra	Aceptar el malestar de la persona, olvidándose de sus necesidades (afecto), centrándose exclusivamente en la experiencia. Eventual ausencia de motivación y actitud ética y compasiva hacia uno mismo y hacia el mundo	No aceptar la experiencia del sufrimiento, que es inevitable. No centrarse en el aquí y ahora, en la naturaleza real de las cosas, y querer buscar soluciones para el futuro



4.5.1. Autoeficacia

La autoeficacia desempeña un papel central dentro de la teoría socio-cognitiva propuesta por Bandura (1997). Este constructo hace referencia a la creencia del individuo en relación a su competencia para ejecutar una tarea, alcanzar una meta o superar un determinado obstáculo con éxito, siendo uno de los determinantes más importantes de la motivación y de la conducta humana. Cuanta más confianza muestre el individuo en su propia competencia ante una determinada situación, mayores serán las probabilidades de que intente afrontarla y se esfuerce y persista por superar las dificultades surgidas. Dada su influencia sobre los procesos cognitivos, afectivos y motivacionales (Schunk y DiBenedetto, 2020), un abundante cuerpo de trabajos ha destacado el papel de la autoeficacia como promotora y protectora de la salud y del bienestar de las personas (Chenoweth et al., 2017; Schönfeld et al., 2016). En este sentido, la autoeficacia favorecería la iniciativa del individuo por potenciar su crecimiento personal (Luszczynska et al., 2005; Ogunyemi y Mabeke, 2007).

4.5.2. Esperanza

De acuerdo con Snyder (2000), la esperanza constituye un estado motivacional positivo definido por la combinación de un elevado nivel de energía orientada a la consecución de un objetivo concreto (agency) y la capacidad para planificar las rutas que llevan a lograrlo (pathways). En otras palabras, la esperanza requiere que el individuo confíe en que los objetivos pretendidos sean asequibles, así como en su capacidad para planificar las estrategias adecuadas para alcanzarlos. De esta forma, las personas que tienden a percibir su día a día con esperanza suelen albergar expectativas generalizadas de éxito (Luthans y Jensen, 2002) y son más proclives a la búsqueda de su crecimiento personal (Meyers et al., 2015; Tong et al., 2010).

4.5.3. Optimismo

El optimismo ha sido conceptualizado desde dos planteamientos complementarios. Uno, más circunscrito a una situación específica y otro, centrado en una percepción más generalizada de los eventos vitales. En consonancia con la primera perspectiva, Seligman (1998) define el optimismo como un estilo o una tendencia a explicar los acontecimientos positivos que nos ocurren aludiendo a factores internos, estables y generalizados (e.g., la inteligencia, la capacidad) y los acontecimientos negativos en términos de factores externos, inestables y específicos (e.g., mala suerte, un mal día). Este estilo se opondría al pesimismo, caracterizado por juzgar los acontecimientos positivos atribuyéndolos a causas externas, inestables y específicas (e.g., la suerte, la ayuda de otra persona) y los acontecimientos negativos con base en factores internos, estables y generalizados (e.g., torpeza, falta de capacidad). Desde una perspectiva más global, el optimismo ha sido definido también como la expectativa generalizada respecto a que el futuro deparará resultados exitosos (Carver y Scheier, 2001), en contraposición a la visión pesimista, cuya expectativa respecto a los eventos futuros tiende a la negatividad.

Si bien el componente de optimismo que forma parte del capital psicológico toma en consideración los dos enfoques anteriormente descritos (Luthans y Youssef-Morgan, 2017), parece que las personas que poseen esta característica son ciertamente flexibles y realistas en sus valoraciones (Snyder, 2000), de tal modo que son capaces de juzgar con objetividad si una determinada situación, sea de éxito o de fracaso, es atribuible a factores internos o externos.

Las características inherentes al optimismo –expectativas futuras positivas, flexibilidad y adaptabilidad– hacen de este recurso psicológico un predictor significativo de la salud mental, actuando como factor de protección frente al estrés (Puig et al., 2021) y la depresión (Alarcon et al., 2013) y favoreciendo la experimentación de un alto crecimiento personal en diferentes poblaciones (Augusto-Landa et al., 2011; Segovia et al., 2015).

4.5.4. Resiliencia

Equiparándolo a ciertos materiales capaces de soportar colisiones sin consecuencias, el término resiliencia ha sido adoptado en nuestros días en la Psicología para hacer referencia a aquellos individuos que, ante situaciones denodadamente negativas como una guerra, una catástrofe natural, una enfermedad grave e incluso experiencias como la pérdida de empleo o una ruptura amorosa, no solo resisten su impacto, sino que logran una adaptación positiva. La capacidad resiliente, como nos expone Lemay (2000), entraña, pues, mucho más que resistir el golpe. Se trata de metamorfosearlo, en tanto que el individuo saca lo mejor de sí mismo, incluso aun no siendo inicialmente consciente de sus propios recursos personales, y aprende a convivir con el sufrimiento de manera constructiva, desarrollando nuevas capacidades de adaptación. Supone la habilidad de prosperar, madurar y mejorar las competencias en condiciones adversas (Gordon, 1996).

Es precisamente esa capacidad de transformación, crecimiento y mejora la que vincula la resiliencia al crecimiento personal, de tal manera que el individuo se vuelve más fuerte y exitoso ante la adversidad (Masten, 2001), evaluando la misma como una oportunidad para desarrollar las potencialidades individuales (Malaguti, 2005). Desde esta consideración, la resiliencia se erige en un poderoso recurso psicológico para favorecer el funcionamiento positivo óptimo de las personas (Di Fabio y Pelazzeschi, 2015; Seaton y Beaumont, 2015).

5. CONCLUSIONES

Los Objetivos de Desarrollo Sostenible tienen entre sus prioridades la salud y el bienestar de los seres humanos. En consonancia con esta ambiciosa meta, el planteamiento del presente trabajo se adscribe a la prometedora línea de investigaciones que reivindican el rol protector del crecimiento personal, conceptualizado como afán por llegar a ser lo mejor de uno mismo, en la salud mental. A la luz de los planteamientos de la Psicología positiva, sucintamente expuestos a lo largo de este artículo, el desarrollo de las potencialidades y capacidades humanas parece jugar un papel relevante tanto en la prevención y recuperación de condiciones y patologías altamente prevalentes en la actualidad, como la depresión, la ansiedad o el abuso de sustancias con potencial adictivo, como en la promoción



del bienestar individual. En este sentido, recursos como la autodeterminación, el estado de flujo, el mindfulness, la compasión y el capital psicológico han demostrado ampliamente su eficacia a la hora de contribuir a la consecución de un funcionamiento positivo óptimo en diferentes poblaciones y grupos etarios.

La promoción del crecimiento personal de los seres humanos, como máximo exponente del bienestar eudaimónico, se convierte, así pues, en un objetivo social y políticamente deseable, con especial prioridad en aquellos colectivos y contextos más vulnerables. Desde esta consideración, el diseño y aplicación de intervenciones basadas en Psicología positiva, bien desde un prisma preventivo-educativo, bien desde un enfoque clínico-rehabilitador, se erigen en una herramienta altamente adaptativa para garantizar una vida saludable y promover el bienestar universal.

CONFLICTO DE INTERESES

No existe conflicto de intereses.

REFERENCIAS

- Alarcon, G. M., Bowling, N. A., y Khazon, S. (2013). Great expectations: A meta-analytic examination of optimism and hope. *Personality and Individual Differences*, 54(7), 821-827. <https://doi.org/10.1016/j.paid.2012.12.004>
- Aranha, A. M., Raj, L., y Premanand, V. (2019). The role of personal growth initiative in the relationship between self-compassion and self-esteem among college students. *International Journal of Indian Psychology* 7(1), 196-203. <https://doi.org/10.25215/0701.021>
- Augusto-Landa, J. M., Pulido-Martos, M., y Lopez-Zafra, E. (2011). Does perceived emotional intelligence and optimism/pessimism predict psychological well-being? *Journal of Happiness Studies*, 12, 463-474. <https://doi.org/10.1007/s10902-010-9209-7>
- Baer, R. A., Hopkins, J., Krietemeyer, J., Smith, G. T., y Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, 13(1), 27-45. <https://doi.org/10.1177/1073191105283504>
- Bandura, A. (1997). *Self-Efficacy. The exercise of control*. Freeman.
- Baron, R. A., Franklin, R. J., y Hmieleski, K. M. (2016). Why entrepreneurs often experience low, not high, levels of stress: The joint effects of selection and psychological capital. *Journal of Management*, 42(3), 742-768. <https://doi.org/10.1177/0149206313495411>
- Bauer, J. J., y McAdams, D. P. (2010). Eudaimonic growth: Narrative growth goals predict increases in ego development and subjective well-being 3 years later. *Developmental Psychology*, 46(4), 761-772. <https://doi.org/10.1037/a0019654>
- Bolier, L., Haverman, M., Westerhof, G. J., Riper, H., Smit, F., y Bohlmeijer, E. (2013). Positive psychology interventions: A meta-analysis of randomized controlled studies. *BMC Public Health*, 13, 119. <https://doi.org/10.1186/1471-2458-13-119>
- Bonaiuto M., Mao, Y., Roberts, S., Psalti, A., Aricchio, S., Ganucci, U., y Csikszentmihalyi, M. (2016). Optimal experience and personal growth: Flow and the consolidation of place identity. *Frontiers in Psychology*, 7: 1654. <https://doi.org/10.3389/fpsyg.2016.01654>
- Breines, J. G., y Chen, S. (2012). Self-compassion increases self-improvement motivation. *Personality and Social Psychology Bulletin*, 38(9), 1133-1143. <https://doi.org/10.1177/0146167212445599>
- Brown, K. W., y Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, 84(4), 822-848. <https://doi.org/10.1037/0022-3514.84.4.822>
- Burke, C. A. (2010). Mindfulness-based approaches with children and adolescents: A preliminary review of current research in an emergent field. *Journal of Child and Family Studies*, 19, 133-144. <https://doi.org/10.1007/s10826-009-9282-x>
- Carver, C. S., y Baird, E. (1998). The American dream revisited: Is it what you want or why you want it that matters? *Psychological Sciences*, 9(4), 289-292. <https://doi.org/10.1111/1467-9280.00057>
- Carver, C. S., y Scheier, M. F. (2001). Optimism, pessimism, and self-regulation. En E. C. Chang (Ed.), *Optimism & pessimism: Implications for theory, research, and practice* (pp. 31-51). American Psychological Association.
- Chakhssi, F., Kraiss, J. T., Sommers-Spijkerman, M., y Bohlmeijer, E. T. (2018). The effect of positive psychology interventions on well-being and distress in clinical samples with psychiatric or somatic disorders: A systematic review and meta-analysis. *BMC Psychiatry*, 18: 211. <https://doi.org/10.1186/s12888-018-1739-2>
- Chang, J.-H., Huang, C.-L., y Lin, Y.-C. (2015). Mindfulness, basic psychological needs fulfillment, and well-being. *Journal of Happiness Studies*, 16(5), 1149-1162. <https://doi.org/10.1007/s10902-014-9551-2>
- Chenoweth, L., Stein-Parbury, J., White, D., McNeill, G., Jeon, Y.-H., y Zaratan, B. (2017). Coaching in self-efficacy improves care responses, health and well-being in dementia carers: A pre/post-test/follow-up study. *BMC Health Services Research*, 16: 166. <https://doi.org/10.1186/s12913-016-1410-x>
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal performance*. Cambridge University Press.
- Culbertson, S. S., Fullagar, C. J., y Mills, M. J. (2010). Feeling good and doing great: the relationship between psychological capital and well-being. *Journal of Occupational Health Psychology*, 15(4), 421-433. <https://doi.org/10.1037/a0020720>
- Datu, J. A. D., King, R. B., y Valdez, J. P. M. (2018). Psychological capital bolsters motivation, engagement, and achievement: Cross-sectional and longitudinal studies. *Journal of Positive Psychology*, 13(3), 260-270. <https://doi.org/10.1080/17439760.2016.1257056>
- Deci E. L., y Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press.
- Demerouti, E., Bakker, A. B., Nachreiner, F., y Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology*, 86(3), 499-512. <https://doi.org/10.1037/0021-9010.86.3.499>
- Di Fabio, A., y Palazzeschi, L. (2015). Hedonic and eudaimonic well-being: The role of resilience beyond fluid intelligence and personality traits. *Frontiers in Psychology*, 6: 1367. <https://doi.org/10.3389/fpsyg.2015.01367>
- Díaz, D., Rodríguez-Carvajal, R., Blanco, A., Moreno-Jiménez, B.,



- Gallardo, I., Valle, C., y van Dierendock, D. (2006). Adaptación española de las Escalas de Bienestar Psicológico de Ryff. *Psicothema*, 18(3), 572-577.
- Djorouva, N. P., Rodríguez, I., y Lorente-Prieto, L. (2019). Individual profiles of psychological capital in a Spanish sample. *Journal of Social Science Research*, 14, 3029-3047. <https://doi.org/10.24297/jssr.v14i0.8042>
- Dreison, K. C., White, D. A., Bauer, S. M., Salyers, M. A., y McGuire, A. B. (2018). Integrating self-determination and job demands-resources theory in predicting mental health provider burnout. *Administration and Policy in Mental Health and Mental Health Services Research*, 45, 121-130. <https://doi.org/10.1007/s10488-016-0772-z>
- Ellis, P. (2004). Australian and New Zealand College of Psychiatrists clinical practice guidelines team for depression. *Australian and New Zealand Journal of Psychiatry*, 38(6), 389-407. <https://doi.org/10.1080/j.1440-1614.2004.01377.x>
- Frankl, V. E. (1962). *Man's search for meaning*. Beacon Press.
- García Campayo, J., y Demarzo, M. (2015). *Mindfulness y auto-compasión. La nueva revolución*. Siglantana.
- Gilbert, P. (2015). The evolution and social dynamics of compassion. *Social and Personality Psychology Compass*, 9(6), 239-254. <https://doi.org/10.1111/spc3.12176>
- Global Burden of Disease Collaborative Network (2018). *Global Burden of Disease Study 2017 (GBD 2017) Results*. Institute for Health Metrics and Evaluation.
- Goilean, C., Gracia, F. J., Tomás, I., y Subirats, M. (2020). Mindfulness en el ámbito del trabajo y de las organizaciones. *Papeles del Psicólogo*, 41(2), 139-146. <https://doi.org/10.23923/pap.psicol2020.2929>
- Gordon, K. A. (1996). Resilient Hispanic youths' self-concept and motivational patterns. *Hispanic Journal of Behavioral Sciences*, 18(1), 63-73. <https://doi.org/10.1177/07399863960181007>
- Hernández-Varas, E., y García-Silgo, M. (2021). Benefits of PsyCap training on the wellbeing in military personnel. *Psicothema*, 33(4), 536-543. <https://doi.org/10.7334/psicothema2021.151>
- Iani, L., Lauriola, M., Cafaro, V., y Didonna, F. (2017). Dimensions of mindfulness and their relations with psychological well-being and neuroticism. *Mindfulness*, 8(3), 664-676. <https://doi.org/10.1007/s12671-016-0645-2>
- Jung, C. G. (2003). *The spirit in man, art, and literature*. Routledge.
- Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: Past, present, and future. *Clinical Psychology: Science and Practice*, 10(2), 144-156. <https://doi.org/10.1093/clipsy.bpg016>
- Kimiecik, J. (2016). The eudaimonics of health: Exploring the promise of positive well-being and healthier living. En J. Vitterso (Ed.), *Handbook of eudaimonic well-being* (pp. 349-370). Springer.
- King, L. A. (2001). The hard road to the good life: The happy, mature person. *Journal of Humanistic Psychology*, 41(1), 51-72. <https://doi.org/10.1177/0022167801411005>
- Krasikova, D. V., Lester, P. B., y Harms, P. D. (2015). Effects of psychological capital on mental health and substance abuse. *Journal of Leadership and Organizational Studies*, 22(3), 280-291. <https://doi.org/10.1177/1548051815585853>
- Leary, M. R., Tate, E. B., Adams, C. E., Batts Allen, A., y Hancock, J. (2007). Self-compassion and reactions to unpleasant self-relevant events: The implications of treating oneself kindly. *Journal of Personality and Social Psychology*, 92(5), 887-904. <https://doi.org/10.1037/0022-3514.92.5.887>
- Lemay, M. (2000). What is resilience? *Virage* 6, 1-4.
- Levinson, D. J. (1986). A conception of adult development. *American Psychologist*, 41(1), 3-13. <https://doi.org/10.1037/0003-066X.41.1.3>
- Linehan, M. M. (1994). Acceptance and change: The central dialectic in psychotherapy. En S. C. Hayes, N. C. Jacobson, V. M. Follette, y M. Dougher, *Acceptance and change: Content and context in psychotherapy* (pp. 73-86). Context Press.
- Luszczynska, A., Mohamed, N. E., y Schwarzer, R. (2005). Self-efficacy and social support predict benefit finding 12 months after cancer surgery: The mediating role of coping strategies. *Psychology, Health & Medicine*, 10(4), 365-375. <https://doi.org/10.1080/13548500500093738>
- Luthans, F. (2002). The need for and meaning of positive organizational behavior. *Journal of Organizational Behavior*, 23(6), 695-706. <https://doi.org/10.1002/job.165>
- Luthans, F., y Jensen, S. M. (2002). Hope: A new positive strength for human resource development. *Human Resource Development Review*, 1(3), 304-322. <https://doi.org/10.1177/1534484302013003>
- Luthans, F., Youssef, C. M., y Avolio, B. J. (2007). *Psychological capital: Developing the human competitive edge*. Oxford University Press.
- Luthans, F., y Youssef-Morgan, C. M. (2017). Psychological capital: An evidence-based positive approach. *Annual Review of Organizational Psychology and Organizational Behavior*, 4, 339-366. <https://doi.org/10.1146/annurev-orgpsych-032516-113324>
- Malaguti, E. (2005). *Educate yourself on resilience: How to deal with crises and difficulties and improve*. Erickson.
- Manzano-García, G., y Ayala, J. C. (2017). Relationship between psychological capital and psychological well-being of direct support staff of specialist autism services. The mediator role of burnout. *Frontiers in Psychology*, 8, 2277. <https://doi.org/10.3389/fpsyg.2017.02277>
- Martín-Asuero, A., y García-Banda, G. (2010). The Mindfulness-Based Stress Reduction Program (MBSR) reduces stress-related psychological distress in healthcare professionals. *Spanish Journal of Psychology*, 13(2), 897-905. <https://doi.org/10.1017/s1138741600002547>
- Maslow, A. H. (1968). *Toward a psychology of being*. D. van Nostrand.
- Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist*, 56(3), 227-238. <https://doi.org/10.1037/0003-066X.56.3.227>
- McAdams, D. P., Hoffman, B. J., Mansfield, E. D., y Day, R. (1996). Themes of agency and communion in significant autobiographical scenes. *Journal of Personality*, 64(2), 339-377. <https://doi.org/10.1111/j.1467-6494.1996.tb00514.x>
- McDowell, J. (1980). The role of eudaimonia in Aristotle's ethics. En A. O. Rorty (Ed.), *Essays on Aristotle's ethics* (pp. 359-376). University of California Press.



- Meyers, M. C., van Woerkom, M., de Reuver, R. S. M., Bakk, Z., y Oberski, D. L. (2015). Enhancing psychological capital and personal growth initiative: Working on strengths or deficiencies. *Journal of Counseling Psychology*, 62(1), 50-62. <https://doi.org/10.1037/cou0000050>
- Muris, P., Meesters, C., Pierik, A., y de Kock, B. (2016). Good for the self: Self-compassion and other self-related constructs in relation to symptoms of anxiety and depression in non-clinical youths. *Journal of Child and Family Studies*, 25(2), 607-617. <https://doi.org/10.1007/s10826-015-0235-2>
- Nakamura, J., y Csikszentmihalyi, M. (2009). Flow theory and research. En C. R. Snyder y S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 195-206). Oxford University Press.
- Neff, K. D. (2003). Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. *Self and Identity*, 2(2), 85-101. <https://doi.org/10.1080/15298860309032>
- Neff, K. D., Long, P., Knox, M. C., Davidson, O., Kuchar, A., Costigan, C., Williamson, Z., Rohleder, N., Tóth-Király, I., y Breines, J. G. (2018). The forest and the trees: Examining the association of self-compassion and its positive and negative components with psychological functioning. *Self and Identity*, 17(6), 627-645. <https://doi.org/10.1080/15298868.2018.1436587>
- Neff, K. D., Rude, S. S., y Kirkpatrick, K. L. (2007). An examination of self-compassion in relation to positive psychological functioning and personality traits. *Journal of Research in Personality*, 41(4), 908-916. <https://doi.org/10.1016/j.jrp.2006.08.002>
- Newman, A., Ucbasaran, D., Zhu, F., y Hirst, G. (2014). Psychological capital: A review and synthesis. *Journal of Organizational Behavior*, 35(S1), 120-138. <https://doi.org/10.1002/job.1916>
- Ogunyemi, A. O., y Mabekoje, S. O. (2007). Self-efficacy, risk-taking behavior and mental health as predictors of personal growth initiative among university undergraduates. *Electronic Journal of Research in Educational Psychology*, 5(2), 349-362.
- Organización Mundial de la Salud (1986). *Ottawa charter for health promotion: First International Conference on Health Promotion*. Organización Mundial de la Salud.
- Pauley, G., y McPherson, S. (2010). The experience and meaning of compassion and self-compassion for individuals with depression or anxiety. *Psychology and Psychotherapy*, 83(2), 129-143. <https://doi.org/10.1348/147608309X471000>
- Pearlin, L. I. (1982). The social contexts of stress. En L. Goldberger y S. Brenznitz (Eds.), *Handbook of stress* (pp. 367-379). Free Press.
- Pepping, C. A., Walters, B., Davis, P. J., y O'Donovan, O. (2016). Why do people practice mindfulness? An investigation into reasons for practicing mindfulness meditation. *Mindfulness*, 7, 542-547. <https://doi.org/10.1007/s12671-016-0490-3>
- Puig, S., Aliño, M., Duque, A., Cano, I., Martínez, P., Almela, M., y García-Rubio, M. J. (2021). Optimismo disposicional y estrés: Claves para promover el bienestar psicológico. *Papeles del Psicólogo*, 42(2), 135-142. <https://doi.org/10.23923/pap-psicol2021.2953>
- Rahimnia, F., Mazidi, A., y Mohammadzadeh, Z. (2013). Emotional mediators of psychological capital on well-being: The role of stress, anxiety, and depression. *Management Science Letters*, 3(3), 913-926. <https://doi.org/10.5267/j.msl.2013.01.029>
- Rashid, T. (2015). Positive psychotherapy: A strength-based approach. *Journal of Positive Psychology*, 10(1), 25-40. <https://doi.org/10.1080/17439760.2014.920411>
- Rehman, S., Qingren, C., Latif, Y., y Iqbal, P. (2017). Impact of psychological capital on occupational burnout and performance of faculty members. *International Journal of Educational Management*, 31(4), 455-469.
- Robitschek, C. (1998). Personal growth initiative: The construct and its measure. *Measurement and Evaluation in Counseling and Development*, 30(4), 183-198.
- Robitschek, C., y Keyes, C. L. (2009). Keyes's model of mental health with personal growth initiative as a parsimonious predictor. *Journal of Counseling Psychology*, 56(2), 321-329. <https://doi.org/10.1037/a0013954>
- Rogers, C. R. (1961). *On becoming a person*. Houghton Mifflin.
- Romero, R., Penas, R., García-Conde, A., Llobart, P., Pérez, S., y Marco, J. H. (2020). Mindfulness-Based Cognitive Therapy for Spanish oncology patients: The Bartley Protocol. *Psicothema*, 32(4), 508-515. <https://doi.org/10.7334/psicothema-2019.320>
- Ross, D. (1925). *Aristotle: The Nicomachean Ethics: Translated with an Introduction*. Oxford University Press.
- Ryan, R. M., y Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, 52, 141-166. <https://doi.org/10.1146/annurev.psych.52.1.141>
- Ryan, R. M., y Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.
- Ryan, R. M., Patrick, H., Deci, E. L., y Williams, G. C. (2008). Facilitating health behaviour change and its maintenance: Interventions based on Self-Determination Theory. *European Health Psychologist*, 10(1), 2-5.
- Ryff, C. D. (1982). Successful aging: A developmental approach. *Gerontologist*, 22(2), 209-214. <https://doi.org/10.1093/geront/22.2.209>
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57(6), 1069-1081. <https://doi.org/10.1037/0022-3514.57.6.1069>
- Ryff, C. D. (2017). Eudaimonic well-being, inequality, and health: Recent findings and future directions. *International Review of Economics*, 64(2), 159-178. <https://doi.org/10.1007/s12232-017-0277-4>
- Salanova, M., Martínez, I. M., Cifre, E., y Schaufeli, W. B. (2005). ¿Se pueden vivir experiencias óptimas en el trabajo? Analizando el Flow en contextos laborales. *Revista de Psicología General y Aplicada*, 58, 89-100.
- Sansone, C., y Harackiewicz, J. M. (1996). "I don't feel like it": The function of interest in self-regulation. En L. L. Martin y A. Tesser



- (Eds.), *Striving and feeling. Interactions among goals, affect, and self-regulation* (pp. 203-228). Lawrence Earlbaum Associates.
- Schönfeld, P., Brailovskaia, J., Bieda, A., Zhang, X. C., y Margraf, J. (2016). The effects of daily stress on positive and negative mental health: Mediation through self-efficacy. *International Journal of Clinical and Health Psychology*, 16(1), 1-10. <https://doi.org/10.1016/j.ijchp.2015.08.005>
- Schunk, D. H., y DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary Educational Psychology*, 60, 101832. <https://doi.org/10.1016/j.cedpsych.2019.101832>
- Segal, Z. V., Williams, M. G., y Teasdale, J. D. (2002). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse*. Guilford Press.
- Segovia, F., Moore, J. L., Linnville, S. E., y Hoyt, R. E. (2015). Optimism predicts positive health in repatriated prisoners of war. *Psychological Trauma: Theory, Research, Practice, and Policy*, 7(3), 222-228. <https://doi.org/10.1037/a0037902>
- Seligman, M. E. P. (1998). *Learned Optimism*. Pocket Books.
- Seligman, M. E. P. (2002). *Authentic happiness*. Free Press.
- Seligman, M. E. P. (2011). *Flourish: A visionary new understanding of happiness and well-being*. Nicholas Brealey Publishing.
- Schotanus-Dijkstra, M., Drossaert, C. H. C., Pieterse, M. E., Walburg, J. A., Bohlmeijer, E. T., y Smit, F. (2018). Towards sustainable mental health promotion: Trial-based health-economic evaluation of a positive psychology intervention versus usual care. *BMC Psychiatry*, 18, 265. <https://doi.org/10.1186/s12888-018-1825-5>
- Schwartz, S. J., y Waterman, A. S. (2006). Changing interests: A longitudinal study of intrinsic motivation for personally salient activities. *Journal of Research in Personality*, 40(6), 1119-1136. <https://doi.org/10.1016/j.jrp.2005.12.003>
- Seaton, C. L., y Beaumont, S. L. (2015). Pursuing the good life: A short-term follow-up study of the role of positive/negative emotions and ego-resilience in personal goal striving and eudaimonic well-being. *Motivation and Emotion*, 39(5), 813-826. <https://doi.org/10.1007/s11031-015-9493-y>
- Shapiro, D. H. (1994). Examining the content and context of meditation: a challenge for psychology in the areas of stress management, psychotherapy, and religion/values. *Journal of Humanistic Psychology*, 34(4), 101-135. <https://doi.org/10.1177/00221678940344008>
- Shapiro, S., Rechtschaffen, D., y de Sousa, S. (2016). Mindfulness training for teachers. En K. A. Schonert-Reichl y R. W. Roeser (Eds.), *Mindfulness in behavioral health. Handbook of mindfulness in education: Integrating theory and research into practice* (pp. 83-97). Springer-Verlag Publishing.
- Snyder, C. R. (2000). *Handbook of hope*. Academic.
- Sun, X., Chan, D. W., y Chan, L.-K. (2016). Self-compassion and psychological well-being among adolescents in Hong Kong: Exploring gender differences. *Personality and Individual Differences*, 101, 288-292. <https://doi.org/10.1016/j.paid.2016.06.011>
- Taylor, E. K., Moxham, L., Perlman, D. J., Patterson, C. F., Brighton, R. M., y Liersch, S. (2016). Self-determination in the context of mental health recovery. *Australian Nursing and Midwifery Journal*, 23(10), 41.
- Teasdale, J. D., Zindel, V., y Williams, M. G. (1995). How does cognitive therapy prevent depressive relapse and why should attentional control (mindfulness) training help? *Behavioral Research and Therapy*, 33(1), 25-39. [https://doi.org/10.1016/0005-7967\(94\)E0011-7](https://doi.org/10.1016/0005-7967(94)E0011-7)
- Tong, E. M. W., Fredrickson, B. L., Chang, W., y Lim, Z. X. (2010). Re-examining hope: The roles of agency thinking and pathways thinking. *Cognition and Emotion*, 24(7), 1207-1215. <https://doi.org/10.1080/02699930903138865>
- Trompetter, H. R., Kleine, E., y de Bohlmeijer, E. (2017). Why does positive mental health buffer against psychopathology? An exploratory study on self-compassion as a resilience mechanism and adaptive emotion regulation strategy. *Cognitive Therapy and Research*, 41(3), 459-468. <https://doi.org/10.1007/s10608-016-9774-0>
- Vitterso, J., y Sohlt, Y. (2011). Life satisfaction goes with pleasure and personal growth goes with interest: Further arguments for separating hedonic and eudaimonic well-being. *Journal of Positive Psychology*, 6(4), 326-335. <https://doi.org/10.1080/17439760.2011.584548>
- Waterman, A. S., Schwartz, S. J., y Conti, R. (2008). The implications of two conceptions of happiness (hedonic enjoyment and eudaimonia) for the understanding of intrinsic motivation. *Journal of Happiness Studies*, 9(1), 41-79. <https://doi.org/10.1007/s10902-006-9020-7>
- Waterman, A. S., Schwartz, S. J., Goldbacher, E., Green, H., Miller, C., y Philip, S. (2003). Predicting the subjective experience of intrinsic motivation: The roles of self-determination, the balance of challenges and skills, and self-realization values. *Personality and Social Psychology Bulletin*, 29(11), 1447-1458. <https://doi.org/10.1177/0146167203256907>
- Young, S. (2005). *Break through pain: A step-by-step mindfulness meditation program for transforming chronic and acute pain*. Sounds True.
- Zhang, J. W., y Chen, S. (2016). Self-compassion promotes personal improvement from regret experiences via acceptance. *Personality and Social Psychology Bulletin*, 42(2), 244-258. <https://doi.org/10.1177/0146167215623271>



Using functional analytic psychotherapy strategies for supporting Latinas victims of interpersonal partner violence

Amanda M. Muñoz-Martínez¹ and Renzo Aguilar-Cacho²

¹Universidad de los Andes, Colombia. ²Universidad Peruana de Ciencias Aplicadas, Perú. Instituto Peruano de Psicología y Ciencia Contextual -VALUES, Perú

La violencia de pareja (IPV, por sus siglas en inglés) es un problema de salud pública mundial que genera altas tasas de discapacidades físicas y conductuales. Los comportamientos para buscar ayuda son claves para brindar soporte en el afrontamiento y el manejo de la violencia de género. El machismo, marianismo y familismo son prácticas culturales relacionadas con conductas de guardar silencio y permanecer en una relación abusiva en mujeres latinas. Asimismo, las dificultades para compartir emociones, establecer límites interpersonales y comunicar necesidades también son barreras para buscar ayuda. Estos factores están asociados a la desconexión social que reduce las posibilidades de que las víctimas sean apoyadas en situaciones de riesgo. Aunque se han desarrollado diversos programas psicológicos y sociales para ayudarlas, se necesita más evidencia sobre adaptaciones que respondan a los factores culturales e interpersonales propios de las mujeres Latinas víctimas de IPV. La Psicoterapia Analítica Funcional (FAP) propone que las relaciones terapéuticas basadas en la intimidad pueden ayudarlas a identificar características de relaciones que brindan ayuda cuando ellas la necesitan. En este artículo teórico se describen diferentes estrategias basadas en FAP con el fin de que los terapeutas que trabajan con Latinas víctimas de IPV puedan ayudarlas.

Palabras clave: Violencia de pareja, Psicoterapia Analítica Funcional, Latinas, Prácticas culturales, Efectividad interpersonal.

Interpersonal partner violence (IPV) is a worldwide public health problem. As a result of IPV, victims report high rates of physical and behavioral disabilities. Help-seeking behaviors are key to obtaining support on facing and managing IPV. Cultural practices such as machismo, marianismo, and familismo are related to keeping quiet and remaining in an abusive relationship in Latinas. Difficulties in sharing emotions, setting interpersonal boundaries, and communicating needs are also behavioral barriers to seeking help in this population. All these factors are associated with social disconnection which reduces victims' chances of obtaining support in risky situations. Although several programs have been developed for helping victims of IPV, more evidence on adaptations to particular cultural and interpersonal struggles faced by Latinas' victims is needed. Functional Analytic Psychotherapy (FAP) proposes that therapeutic relationships based on intimacy can help victims of IPV to identify characteristics of those relationships that provide help when they need it. In this theoretical article, several FAP-based strategies aid therapists working with Latinas' victims of IPV in strengthening clients' interpersonal functioning and defeating cultural barriers when seeking help.

Key words: Interpersonal partner violence, Functional Analytic Psychotherapy, Latinas, Cultural practices, Interpersonal effectiveness.

Interpersonal partner violence (IPV) involves situations in which an intimate partner behaves physically, sexually, and psychologically violent against a current or former partner (Breiding et al., 2015). In physical abuse, perpetrators use objects or their bodies to hurt another person intentionally, producing minor injuries to death. Coercion is another type of abuse, in which perpetrators force their partners to involve in behaviors they do not want to. Some forms of sexual violence involve coerciveness in which perpetrators force their partners to engage in sexual acts they do not agree with. Unauthorized physical touch and penetration are also well-known forms of sexual violence. Psychological abuse is a frequent and underreported type of abuse that includes verbal (i.e., blame, threat, insult) and non-verbal (i.e., ignore, neglect) aggressions

as well as stalking. As a result, perpetrators get control over the emotions, finances, and behaviors of their partners.

One out of three women between 15 and 49-years old have experienced some form of interpersonal partner violence (IPV) at least once in their lives in the American continent (Bott et al., 2019). Between 2003 and 2017 an average of 175.864 cases reported sexual, physical, or both types of abuse in Latin America (Bott et al., 2019). The severity of violent acts perpetrated against women in this region varies among countries. In Perú is reported the highest rate of violence against women (25.5%) while Haiti statistics are the lowest (7.4%; Bott et al., 2019). In general, the prevalence of moderate violent acts (i.e., being slapped, pushed, shoved) in Latino countries ranged between 14.9% and 37.9% (Jaitman et al., 2017).

Global costs associated with IPV are higher every year. While Global Gross Domestic Product (GDP) to compensate IPV effects is 0,05%, Latin American countries have higher costs, reaching an average of 0,09% (Fearon & Hoeffler, 2014). Particularly, Colombia GPD costs are around 4,22% (Ribero & Sánchez, 2005), Chile 2%, and Nicaragua 1,6% (Morrison & Biehl, 1999).

Received: 12 septiembre 2021 - Accepted: 10 diciembre 2021

Correspondence: Renzo Aguilar-Cacho. Universidad Peruana de Ciencias Aplicadas, Perú. Instituto Peruano de Psicología y Ciencia Contextual -VALUES. Jirón Los Conquistadores cuadra 3. Urbanización Las Lomas de La Molina. 15024 Lima. Peru.

E-mail: renzoaguilar84@gmail.com



BARRIERS TO SEEKING AND GETTING HELP IN INTERPERSONAL PARTNER VIOLENCE

Women face several barriers to seeking help and withdrawing from IPV relationships. Anderson and Saunders (2003) conducted a meta-analysis of factors associated with leaving an abusive partner, finding that holding traditional religious values and experiencing love towards the perpetrator reduced the odds of withdrawing an abusive partner. Contrary to expectations, women who witnessed domestic violence in their childhood do more likely to escape IPV relationships (Anderson & Saunders, 2003; Kim & Gray, 2008). In addition, marital status and length of the relationship did not predict staying in an abusive relationship.

From social determinants of health perspective, research has found that the lack of employment and income are other factors that hold back women in IPV relationships. Kim and Gray (2008) identified that financial dependence was associated with low rates of leaving perpetrators. Indeed, women's income and employment status were powerful predictors to leave (Anderson & Saunders, 2003). Behavioral variables such as fear to be harmed, shame, poor perception of control, minimizing and denying the abuse are linked to difficulties in acknowledging and withdrawing IPV relationships. Rather, strong social networks characterized by providing informational, instrumental, and emotional support were significantly associated with sustainable wellbeing in women that have escaped abusive relationships (Anderson et al., 2003).

While social determinants factors of health such as employment, income, and social networks are crucial to overcoming barriers to withdraw IPV relationships, survivors seldom report saving money or asking for support from others (e.g., family, other survivors, health providers, or police) as coping strategies (Rizo et al., 2017). In this line, medical and mental health providers face the challenge of finding strategies to aid IPV survivors in engaging in healthy emotional support and finding sustainable sources of income.

IPV IN LATINAS: BARRIERS FOR SEEKING HELP

Factors associated with IPV in Latinas have some similarities with those described in other ethnic groups. A systematic review among Hispanic women living in the U.S. identified that low SES and income are tightly linked to IPV in this population. Contrary to IPV studies conducted with mainly Caucasian women living in the US, inter-generational violence is a relevant risk factor for Latinas (Ali & Naylor, 2013; González-Guarda & Becerra, 2012).

Studies have identified that some socio-economic factors such as access to a retirement fund, education, and salary differ based on gender in Latinx communities. Comparing time spent over the age of 15 years of paid and unpaid work in four Latin-American countries (Colombia, Mexico, Ecuador, Uruguay), disparities are marked between men and women of all ages in those countries. While paid hours per week fluctuate between 10 and 20 hours for women, labor hours paid to men ranged between 20 and 48 hours. The prospect is bleak when contrasting unpaid hours. Women used to work between 20 and 50 hours weekly without getting paid for it while men only spend 10 hours on average in unpaid jobs. Additionally, the unemployment rate is higher in women than in men across Latin-American countries. Women also face challenges associated with motherhood. For instance, the average duration in school for Latinas

who were adolescents' mothers is lower than those who were not, reducing the chances of getting a well-paid job and financial independence (Bárcena et al., 2016). These socioeconomic factors make Latinas financially vulnerable that is a main contributing factor to staying in violent relationships. Public health policies that reduce gender-based disparities and protect women's human rights need to be on the agenda of governments to reduce the impact of these factors that greatly contribute to IPV (Milosavljevic, 2007).

Other factors such as cultural values are also related to IPV in Latinas. Some qualitative studies have identified a negative contribution of gender roles to relationships' balance and IPV in Latinas. Hispanic women consider that *machismo* (a cultural belief that men are superior to women and masculinity should be exaggerated) and *marianismo* (a cultural practice that encourages women's behaviors of submission and demure) put themselves at higher risk to IPV (González-Guarda & Becerra, 2012; Márquez-Espinoza, 2019). *Machismo* and *marianismo* are considered two cultural phenomena based on gender-based behaviors from men and women. While *machismo* is defined as an aggressive attitude in interpersonal relationships and an exaggeration of virility, *marianismo* is conceptualized as a practice in which women's worth relies on spirituality, superior morality, and having the ability to give birth (Moisés, 2012). Such cultural practices facilitate abusive behaviors in romantic relationships as they promote men's power over women by holding control of women's actions, money, sexuality, and so forth (Moral de la Rubia & Ramos Basurto, 2016; Terrazas-Carrillo & Sabina, 2019). These gender roles also normalize the cycle of violence, equating love with emotional dependence which holds back victims from leaving and perpetuating violence and (Mariana Mateo, 2015).

Familismo (family idealization) is a sociological term used to understand the family cultural practice in which members communicate often and should maintain reciprocal ties in order to get mutual help and feel socially connected (Garrido & Chulía, 2021). However, *familismo* can have a dark side, particularly, when leading women to believe that a real family should remain together and forgive faults because love heals everything; thus, leaving a partner is not an option and could lead to marginalization (Echeburúa-Odriozola et al., 2002). This is linked to other cultural practices around romantic love that promote women's sacrifice for their partners as proof of real love (Echeburúa & Muñoz, 2002).

CONSEQUENCES OF IPV IN LATINAS VICTIMS

Victims of IPV faced several consequences in different life domains. At the work level, absenteeism and problems to be productive are frequent. These affect their ability to keep a stable wage and job (Ministerio de la Mujer y Poblaciones Vulnerables; MIMP, 2017). Further, IPV victims have a lower income compared to women who are not in abusive relationships. Reproductive health is another area impacted by IPV; a high rate of sexually transmitted infections has been found in this population. Pregnancy could be also risky as complications during childbirth have been observed. In addition, mental health consequences such as anxiety, hypervigilance, fear of being abused, depression, and traumatic symptoms are often reported (MIMP, 2017). Social isolation, conflicting interpersonal relationships, problems with self-care, self-esteem, feelings of guilt, justification of violence, and dissociation are other psychological effects of IPV (Vaca-Ferrer et al., 2020). IPV does not only impact



direct victims, but it also affects members of the nuclear and extended family. For instance, children from IPV victims have higher chances of being late into school, getting sick (e.g., respiratory, fever, diarrhea, anemia), or facing developmental delays.

Research on interpersonal and social consequences of IPV in Latinas living in the US showed that severely battered women often have a weak social network and low social contact which reduces their chance to leave (Briones-Vozmediano et al., 2016; O'Neal & Beckman, 2017). Studies conducted with health providers at shelters for battered women in Spain identified that weak and ambivalent social and family networks are a result of IPV that turned into a barrier to leaving abusive relationships (Moriana Mateo, 2015). On a behavioral level, a systematic review on the socioemotional effects of IPV in Hispanics showed that victims frequently presented difficulties on interpersonal effectiveness, particularly, a limited repertoire of problem-solving and emotional communication (Blázquez-Alonso & Moreno-Manso, 2008).

The combination of socio-economic and cultural barriers, as well as interpersonal and behavioral consequences of IPV, contribute Latinas to being held in abusive relationships which result in high levels of isolation and marginalization. Individual interventions for this particular population may be boosted by including strategies to strengthen the following interpersonal repertoires: (a) expressing emotions and needs, (b) establishing and maintaining interpersonal limits, and (c) asking for help. In addition, to overcome interpersonal and cultural barriers, aiding victims to identify nurturing environments in which those effective interpersonal behaviors would be acknowledged and maintained is crucial to strengthen social networks and contact alternative cultural practices.

PSYCHOLOGICAL INTERVENTIONS FOR IPV VICTIMS

Interventions for reducing the effects of IPV have been conducted from different approaches. Eckhardt et al. (2013) conducted a review of intervention programs for IPV victims that were published in English and without a particular analysis over participants' ethnicity. This study found that several treatment outcomes were evaluated in studies conducted with IPV victims, including post-traumatic stress disorder (PTSD), self-esteem, and social support. Examining treatment outcomes, cognitive-behavioral therapy (CBT) showed good evidence in reducing PTSD and depression symptoms, as well as promising results on helping victims to transition from shelters to an independent life. Supportive group interventions showed promising but limiting evidence on social support and emotional distress.

Therapeutic approaches based on contextual behavioral science have also shown positive effects on IPV victims' functioning. Iverson et al. (2009) adapted Dialectical Behavioral Therapy (DBT) to a 12-sessions group modality for women victims of IPV, including the following strategies: (a) pre-treatment, (b) agreement of schedules and groups, (c) group skills (mindfulness, interpersonal effectiveness, emotional regulation, distress tolerance), and (d) transversal validation. This program decreased participants' depressive symptoms and hopelessness while increasing social functioning and general psychiatric distress. This program shows promising results for continuing implementing contextual behavioral intervention with this population.

CBT and DBT programs have mainly focused on reducing psychological symptoms and improving victims' assertiveness.

However, there is not enough information on how those treatments shape specific interpersonal behaviors hindered by IPV (DBT program mainly focused on assertion needs and interpersonal boundaries). They neither offered a culturally adapted intervention for reducing the impact of social practices on victims. Cultural variables have proven to be boosters and buffers of evaluation and intervention processes in psychological therapies (Bernal et al., 2009). In addition, they do not provide a rationale for therapists to manage interpersonal and cultural challenges when clients are hesitant to complete treatment and generalize behavioral changes. In this context, this paper offers some strategies based on Functional Analytic Psychotherapy that can be implemented by therapists struggling with particular cultural and interpersonal barriers presented by Latina's victims of IPV.

FUNCTIONAL ANALYTIC PSYCHOTHERAPY: A CONTEXTUAL BEHAVIOR THERAPY

Functional Analytic Psychotherapy (FAP) is a contextual behavioral therapy focused on strengthening effective interpersonal repertoires through modifying behavioral mechanisms within the context of the therapeutic relationship (Kohlenberg & Tsai, 1991; Callaghan & Follette, 2020). In the latest years, some FAP authors have proposed models for interpersonal functioning (i.e., social connection, intimacy) by extending behavioral principles into middle-level terms (Maitland et al., 2017; Marin-Vila et al., 2020); however, those models are still under development and require more supporting evidence. Because of that, this article will be developed based on the original FAP model developed by Kohlenberg & Tsai (1991).

In FAP, therapists provide natural social reinforcement within the therapeutic setting to shape clients' effective interpersonal behaviors (Clinically Relevant Behaviors 2; CRB2s) such as emotional communication, asserting needs, conflict resolution, and bidirectional communication (Callaghan, 2006). In addition, FAP seeks to reduce problematic interpersonal repertoires (Clinically Relevant Behaviors 1; CRB1) that include hiding emotions, minimizing needs and opinions, establishing weak interpersonal limits, isolating, and concealing information from meaningful relationships (Holman et al., 2017; Kanter et al., 2020).

FAP procedures include implementing five therapeutic rules by therapists (Kohlenberg & Tsai, 1991). Rule 1 is therapists' active observations in order to identify clients' Clinically Relevant Behaviors (CRBs) within the therapeutic interaction. Rule 2 involves therapists' behaviors in the services of evoking CRBs. Rule 3 includes therapists' implementation of contingent reinforcement. When clients present CRB2s, therapists provide positive natural reinforcers (e.g., support, validation, acknowledgment, warmth) to enhance clients' interpersonal effective behavior (TCRB2). In the case that clients emit problematic interpersonal behaviors (CRB1s) such as shutting down, minimizing themselves, concealing, etc., therapists apply differential reinforcement of alternative behaviors to reduce ineffective behaviors and create the opportunity for CRB2s (TCRB1). Rule 4 is the action by which therapists check whether their implementation of contingent reinforcement was or not effective. Finally, through rule 5 therapists promote generalization and discrimination of interpersonal behaviors from the therapeutic setting to the everyday life environment (Kohlenberg & Tsai, 1991; Kanter et al., 2010; Holman et al., 2017).

FAP has been implemented as a booster intervention for several



contextual behavioral therapies (Kanter et al., 2010). A particular area in which FAP has gained support as a standalone intervention is interpersonal and social functioning. Some studies with Latinx populations have shown positive outcomes when implementing FAP for improving interpersonal repertoires. Esparza-Lizarazo et al. (2015) found FAP procedures reduced emotional dysregulation patterns in- and out-of-sessions in two women and one man in Colombia. In the same country, Rincón et al. (2021) implemented FAP with three Latino self-identified gay men who presented difficulties in intimacy-related behaviors, finding a large effect size in intimacy after finishing the intervention. In Spain, Ferro-García et al. (2021) used FAP for treating a wide variety of behavioral problems such as anxiety, depression, obsessive-compulsive disorder, etc., presented by 10 participants. They formulated cases using FAP interpersonal functional categories rather than using syndromal categories. In this study, FAP was associated with a reduction in behavioral problems as well as improvements in interpersonal functional classes.

FAP emphasizes that therapeutic relationships are a model of healthy relationships (e.g., bidirectional, supportive, connected) by which this therapy results in a good fit for reducing interpersonal barriers in victims of IPV. There is not a standalone FAP intervention for IPV victims; however, Vaca-Ferrer et al. (2020) developed a group program with Hispanic victims of IPV that included procedures from various behavioral contextual therapies such as FAP, acceptance and commitment therapy (ACT), and behavioral activation. In this study, initial sessions were focused on strengthening the therapeutic relationship through implementing FAP. In addition, they focused on promoting several effective interpersonal repertoires (CRB2s) such as involvement in protective behaviors, identification of indicators of violent behaviors within relationships, and participation in pleasant activities and relationships in the program. As a result, they found a reduction in distress and risk behaviors in IPV victims. In addition, participants enhanced closeness and openness with others. These are promising results on the utility of including FAP within treatments for Latinas' victims of IPV.

CULTURAL PRACTICES, CONTEXTUAL BEHAVIORAL SCIENCE, AND FAP

From a contextual behavioral perspective, cultural practices are constituted by group-shared behavioral patterns susceptible to be modified (Glenn, 2004). Particularly, Glenn (2004) specifies that macrocontingencies, a type of cultural practice, are the result of cumulative effects of individual behaviors and their consequences within a group. When individuals systematically interact with other groups who offer alternative patterns of behaviors, their old cultural practices can be altered, and learn new cultural rules. For instance, if IPV victims who hold machismo and marianismo cultural practices interact with others who promote equalitarian gender roles, it is likely that their macrocontingencies turn over and new cultural patterns emerge.

FAP therapists' emphasis on shaping and modeling a reciprocal therapeutic relationship with equalitarian gender roles makes this therapeutic approach a good fit for Latinas' victims of IPV. In addition, FAP offers an alternative avenue to develop intimate behavioral patterns fundamental in establishing and maintaining reciprocal relationships. Intimacy refers to the interchange of vulnerable

behaviors (responses punished in the past) within alternative contexts that offer validation, care, and understanding (Cordova & Scott, 2001). In the following section, a description of how FAP promotes key interpersonal repertoires that are at a low rate in Latinas due to IPV is presented. This therapy framework provides some ideas for therapists interested in helping clients seek help timely, stretch social support networks, and identify alternative cultural practices in relationships.

FAP APPROACH TO PROMOTE EFFECTIVE INTERPERSONAL BEHAVIORS IN LATINAS VICTIMS OF IPV

A basic repertoire to establish relationships and maintain meaningful relationships is emotional communication. Cordova and Scott (2001) described how expressing vulnerable emotions and experiences is key to building up intimacy. Concealing emotional experiences associated with being a victim of IPV is common in Latinas, as well as shame and fear of rejection from family and friends which deteriorates social support (Blázquez-Alonso & Moreno-Manso, 2008). Thus, reducing the impact of emotional concealment is crucial to help victims to maintain strong and safe relationships.

Difficulties expressing emotions would be the result of ongoing exposure to punishment and invalidation within the abusive relationship. Thus, creating a genuine, caring, and honest therapeutic relationship is the first step to promoting effective emotional communication. To do so, FAP therapists intentionally evoke emotional expression from their clients. In the case of victims of interpersonal violence every emotional expression would count, that is why asking for their experience, their deepest sense, or their sensations when storytelling would help evoke emotional expressions from the client. Once clients disclose emotions and feelings, therapists are encouraged to reinforce them in a natural supportive way (Callaghan, 2006). For instance, a client who left her partner after an episode of physical abuse may isolate from family so that they would not criticize her for leaving her partner (a cultural practice associated with *familismo*). A therapist may ask for her feelings after making such a decision and acknowledge the meaning of sharing her shame and bravery for protecting herself. Interchanging emotional support and social reinforcement would likely strengthen the client's emotional expressiveness and an intimate therapeutic relationship.

Establishing and maintaining interpersonal limits is another repertoire often thwarted in Latinas' victims of IPV. Interpersonal boundaries are tightly related to identifying and communicating needs and opinions. Often, history of abuse shapes ineffective communication patterns from victims. They learned to send confusing or unclear messages about what they want and need from relationships that result in difficulties from others to be attentive and responsive, perpetuating a loop of misunderstandings and indirect communication (Callaghan, 2006). In this case, FAP encourages therapists to identify those circumstances in which clients show indirect or unclear messages that may lead to difficulties in setting interpersonal boundaries within the therapeutic interaction. Natural opportunities can come up through therapy, for instance, when rescheduling a session, setting the agenda, or being late. Therapists are also encouraged to explicitly ask clients to make requests, describe their needs, and state their opinions regarding therapy and



the therapeutic relationship. When those evoking opportunities happen and clients behave assertively, therapists offer natural reinforcement to needs communication by providing what clients need. Then, it is expected that therapists establish parallels between in-session and daily life behaviors to promote generalization to other relationships.

A more complex repertoire embedded within assertion needs in victims of IPV is asking for help from others. Intertwined characteristics of Latino social factors may contribute to maintaining low rates of seeking help by IPV victims in this culture. A pervasive history of punishment when requesting support and understanding by social contexts likely shapes patterns of avoidance intimacy promoting behaviors such as asking for help (Callaghan, 2006). In addition, the cultural practice of *familismo*, in which what happens at home should remain inside the home, perpetuates a pact of silence among Latinas' victims of IPV. In therapy, it is common that clients to avoid asking for help from their therapist. For this reason, noticing subtle requests of help is crucial to enhance this repertoire. Directly evoking petitions from clients and inviting them to talk about it would normalize asking for help. Thus, naturally reinforcing small steps towards disclosing personal matters outside the family would increase support-seeking from friends and community, defeating barriers associated with *familismo*.

In FAP, natural social reinforcement is the mechanism that produces behavioral change (Callaghan & Follette, 2020; Muñoz-Martínez, 2019). Identifying the most valuable social reinforcers for clients is highly important to increase the chance of modifying interpersonal repertoires. While showing support, validation, and closeness are effective ways of reinforcing emotional communication in the majority of cases, they may be not such effective when targeting assertion of needs behaviors. For enhancing communication of needs and setting boundaries, providing or eliminating what clients state may be a more powerful reinforcer. For instance, if a client asks a therapist to be in the hearing at the court (needs assertion), the natural reinforcer would be accepting the invitation and going to court (reinforcer matches client's request), if possible, rather than showing their deep appreciation for her request (consequence does not match client's petition).

Helping Victims to Discriminate Interpersonal Safe Environments

Ongoing interactions in invalidating and abusive environments hinder individuals' ability to discriminate safety and caring social context. Discrimination implies people's skills to identify what particular "consequence is administered for a particular behavior when a specific stimulus is present but not when another stimulus is present" (Sarafino, 2012, p. 214). In the context of IPV, victims likely present difficulties to clearly distinguish people's characteristics that indicate that they are safe or dangerous.

Discrimination repertoires for identifying safety environments can be promoted in the context of supporting the development of interpersonal effective behaviors. For instance, when clients assert their needs or address conflicts effectively within the therapeutic relationship, the therapist can prompt clients to identify what factors from the relationship, and the therapist facilitates their engagement with effective behaviors. This would help clients to discriminate factors that favor their skillful behavior. Likewise, it is important to ask for the difference between the therapeutic environment and outside social

contexts, particularly, those features that would indicate danger vs safety.

BEING CONNECTED: CREATING MEANINGFUL RELATIONSHIP

Strong and reliable social networks are crucial for Latinas leaving IPV relationships. Asking for help is unlikely when no social support is available. The therapeutic relationship can be a fruitful context to learn various forms of engaging in effective social support and connection (Holman et al., 2017). Connectedness as people's sense of closeness to others by being actively involved with another person, groups, and social world is a relevant factor associated with healthy interpersonal functioning and wellbeing (Dunkel Schetter, 2017; Townsend & McWhirter, 2005). Victims of traumatic experiences are more willing to engage in interpersonal risks, such as sharing shameful and difficult experiences, when feeling socially connected within the interaction (Gabriel et al., 2017).

Kanter et al. (2020) proposed a model based on FAP to promote social connection based on the development of an intimate and reciprocal therapeutic relationship. To do so with IPV clients, therapists are encouraged to initiate therapeutic interactions in which they reinforce clients' vulnerable behaviors such as emotional expressions, self-disclosures, ask personal questions, and so forth (e.g., sharing memories of abuse). As a response to clients' disclosure, therapists are prompted to be vulnerable and sensitive (e.g., expressing emotions evoked by clients' narratives), reinforcing intimacy and closeness. Building up a reciprocal interchange of vulnerable experiences between therapists and clients creates a sense of connection and belongingness in the therapeutic relationship, which can aid victims to ask for help from others as the therapeutic relationship has shaped intimate behaviors and discriminate contextual features that favor such repertoires.

CONCLUSION

Latinas' victims of IPV face several barriers and consequences that hinder their ability of help-seeking. Growing within a culture in which gender roles (*machismo*, *marianismo*) and family duties (*familismo*) regulate what people are allowed to share with others increases the chances of being held in abusive relationships. These cultural practices plus invalidating and emotionally neglectful social contexts are a breeding ground for ineffective interpersonal behaviors in Latinas.

FAP's emphasis on building an intimate and caring therapeutic relationship aid Latinas' victims of IPV to defeat interpersonal barriers such as concealment of shame and guilt, following gender roles promoted by Latino culture, and isolation. From this approach, social support to leave IPV relationships would be seeded in the context of an intimate therapeutic interaction. Doing so implies that therapists actively engage in an ongoing vulnerable interchange with clients, creating a sense of belonging and connection within the therapeutic relationship.

As this is a theoretical article, empirical research is needed to test whether the implementation of the strategies proposed here is useful and effective in aiding Latina victims' help-seeking and therapist support to them. While many of the FAP recommendations are focused on supporting victims to strengthen interpersonal repertoires



and creating a safe therapeutic environment which opposed to problematic cultural practices, it would be useful to test whether they could have an additive effect to other therapies that have demonstrated positive effects in symptom reduction in other populations (e.g., CBT). Blending FAP with other contextual behavioral therapists' approaches will be also useful, particularly with DBT that has shown good outcomes in social functioning. It would be interesting to evaluate whether blending DBT and FAP in individual sessions can boost other interpersonal repertoires such as emotional communication and refine relational behaviors by discriminating nurturing environments.

As noted in this article, Vaca-Ferrer et al. (2020) found improvements in closeness, openness, and distress of Hispanic IPV victims who completed a program that combined FAP, ACT, and behavioral activation. Future research would dismantle this program to evaluate whether FAP-alone could have similar effects or event extended if including modules that target discrimination of safe environment and alternative cultural practices in Latinas as proposed in this manuscript.

In sum, FAP invites therapists to keep the therapeutic relationship as the core to model more functional interpersonal repertoires and interactions when working with Latinas' victims. Its functional nature conceptualizes the lack of help-seeking behaviors as the result of victims' interaction with problematic social environments and cultural practices. This perspective reduces bias over the reasons why Latinas' victims stay quiet and remain in IPV relationships as it does not pathologize them. In this line, FAP therapists would focus on transforming victims' interpersonal dynamics. Doing so would enable victims of IPV to build up social connections and discriminate characteristics of people who could offer help when they need it.

CONFLICT OF INTEREST

There is not conflict of interest

REFERENCES

- Ali, P.A. & Naylor, P.B. (2013). Intimate partner violence: A narrative review of the feminist, social and ecological explanations for its causation. *Aggression and Violent Behavior*, 18(6), 611-619. <https://doi.org/10.1016/j.avb.2013.07.009>
- Anderson, D.K., & Saunders, D.G. (2003). Leaving an abusive partner: An empirical review of predictors, the process of leaving, and psychological well-being. *Trauma, Violence, & Abuse*, 4, 163-191. <https://doi.org/10.1177/1524838002250769>
- Anderson, D.K., Saunders, D.G., Yoshihama, M., Bybee, D.I., & Sullivan, C.M. (2003). Long-term trends in depression among women separated from abusive partners. *Violence Against Women*, 9, 807-838. <https://doi.org/10.1177/1077801203009007004>
- Bárcena, A., Prado, A., Abramo, L. & Pérez, R. (2016). The social inequality matrix in Latin America. Recuperado de: https://repositorio.cepal.org/bitstream/handle/11362/40710/1/S1600945_en.pdf
- Bernal, G., Jiménez-Chafey, M. I., & Domenech Rodríguez, M. M. (2009). Cultural adaptation of treatments: A resource for considering culture in evidence-based practice. *Professional Psychology: Research and Practice*, 40(4), 361-368. <https://doi.org/10.1037/a0016401>
- Blázquez-Alonso, M. & Moreno-Manso, J.M. (2008). Análisis de la inteligencia emocional en la violencia de género. *Electronic Journal of Research in Educational Psychology*, 6(2), 475-500. <https://doi.org/10.25115/ejrep.v6i15.1281>
- Bott, S., Guedes, A., Ruiz-Celis, A.P., & Mendoza, J.A. (2019). Intimate partner violence in the Americas: A systematic review and reanalysis of national prevalence estimates. *Revista Panamericana de Salud Pública*, 43, 1-12. <https://doi.org/10.26633/RPSP.2019.26>
- Breiding, M.J., Basile, K.C., Smith, S.G., Black, M.C. & Mahendra, R. (2015). Intimate partner violence surveillance: uniform definitions and recommended data elements, Version 2.0. <https://www.cdc.gov/violenceprevention/pdf/ipv/intimatepartnerviolence.pdf>
- Briones-Vozmediano, E., Davó-Blanes, M.C., García-de la Hera, M., Goicolea, I. & Vives-Cases, C. (2016). Discursos profesionales sobre la violencia del compañero íntimo: Implicación en la atención de las mujeres inmigrantes en España. *Gaceta Sanitaria*, 30(5), 326-332. <https://doi.org/10.1016/j.gaceta.2016.04.022>
- Callaghan, G.M. (2006). The functional idiographic assessment template FIAT system: For use with interpersonally-based interventions including functional analytic psychotherapy (FAP) and FAP-enhanced treatments. *The Behavior Analyst Today* 7(3), 357-398. <https://doi.org/10.1037/h0100160>
- Callaghan, G.M. & Follette, W.C. (2020). Interpersonal behavior therapy (IBT), functional assessment, and the value of principle-driven behavioral case conceptualizations. *The Psychological Record*, 70, 625-635. <https://doi.org/10.1007/s40732-020-00395-1>
- Cordova, J.V., & Scott, R.L. (2001). Intimacy: A behavioral interpretation. *The Behavior Analyst*, 24(1), 75-86. <https://doi.org/10.1007/BF03392020>
- Dunkel Schetter, C. (2017). Moving research on health and close relationships forward—a challenge and an obligation: Introduction to the special issue. *American Psychologist*, 72(6), 511-516. <https://doi.org/10.1037/amp0000158>
- Esparza Lizarazo, N., Muñoz-Martínez, A. M., Santos, M. M., & Kanter, J. W. (2015). A within-subjects evaluation of the effects of functional analytic psychotherapy on in-session and out-of-session client behavior. *The Psychological Record*, 65(3), 463-474. <https://doi.org/10.1007/s40732-015-0122-7>
- Echeburúa-Odrizola, E., Corral-Gargallo, P. & Amor-Andrés, P.J. (2002). Mujeres maltratadas en convivencia prolongada con el agresor: Variables relevantes. *Revista de Acción Psicológica*, 1(2), 135-150.
- Echeburúa, E. & Muñoz, J.M. (2002). Boundaries between psychological intimate partner violence and dysfunctional relationships: psychological and forensic implications. *Anales de Psicología*, 33(1), 18-25. <https://doi.org/10.6018/analesps.33.1.238141>
- Eckhardt, C.I., Whitaker, D.J., Sprunguer, J., Dykstra, R. & Woodard, K. (2013). The effectiveness of interventions programs for perpetrators and victims of intimate partner violence. *Partner Abuse*, 4(2), 196-231. <http://doi.org/10.1891/1946-6560.4.2.196>
- Fearon, J.D. & Hoeffler, A. (2014). Benefits and costs of the conflict and violence targets for the post-2015 development agenda.



- https://www.copenhagenconsensus.com/sites/default/files/conflict_assessment_-_hoeffler_and_fearon.pdf
- Gabriel, S., Read, J. P., Young, A. F., Bachrach, R. L., & Troisi, J. D. (2017). Social surrogate use in those exposed to trauma: I get by with a little help from my (fictional) friends. *Journal of Social and Clinical Psychology*, 36(1), 41-63. <https://doi.org/10.1521/jscp.2017.36.1.41>
- Garrido, L. & Chuliá, E. (2021). La pandemia y las familias: refuerzo del familismo y declive de la institución familiar. *Panorama Social*, 33, 95-108.
- Glenn, S.S. (2004). Individual behavior, culture, and social change. *The Behavior Analyst*, 27(2), 133-151. <https://doi.org/10.1007/BF03393175>
- González-Guarda, R.M. & Becerra, M.M. (2012). Violencia de pareja en mujeres hispanas: Implicancias para la investigación y la práctica. *Horizonte de enfermería*, 23(2), 27-38. https://doi.org/10.7764/Horiz_Enferm.23.2
- Holman, G., Kanter, J., Tsai, M. & Kohlenberg, R. (2017). *Functional analytic psychotherapy made simple: A practical guide to therapeutic relationships*. New Harbinger Publications.
- Iverson, K.M., Shenk, C., & Fruzzetti, A.E. (2009). Dialectical behavior therapy for women victims of domestic abuse: A pilot study. *Professional Psychology Research and Practice*, 40(3), 242-248. <https://doi.org/10.1037/a0013476>
- Jaitman, L., Capriolo, D., Granguillhome Ochoa, R., Keefer, P., Leggett, T., Lewis, J.A., Mejía-Guerra, J., Mello, M., Sutton, H., & Torre, I. (2017). *The costs of crime and violence: New evidence and insights in Latin America and the Caribbean*. <https://publications.iadb.org/publications/spanish/document/Los-costos-del-crimen-y-de-la-violencia-Nueva-evidencia-y-hallazgos-en-Am%C3%A9rica-Latina-y-el-Caribe.pdf>
- Kanter, J.W., Kuczynski, A.M., Manbeck, K.E., Corey, M.D., & Wallace, E.C. (2020). An integrative contextual behavioral model of intimate relations. *Journal of Contextual Behavioral Science*, 18, 75-91. <https://doi.org/10.1016/j.jcbs.2020.09.001>
- Kanter, J.W., Tsai, M. & Kohlenberg, R.J. (2010). Introduction to the practice of functional analytic psychotherapy. In M. Tsai, R.J., Kohlenberg, J.W., Kanter, B. Kohlenberg, W.C. Follette & G.M. Callaghan (Eds.). *The practice of functional analytic psychotherapy* (pp. 1-7). Springer.
- Kim, J., & Gray, K.A. (2008). Leave or stay? Battered women's decision after intimate partner violence. *Journal of Interpersonal Violence*, 23, 1465-1482. <https://doi.org/10.1177/0886260508314307>
- Kohlenberg, R.J. & Tsai, M. (1991). *Functional analytic psychotherapy: Creating intense and curative therapeutic relationships*. Springer.
- Maitland, D.W., Kanter, J.W., Manbeck, K.E. & Kuczynski, A.M. (2017). Relationship science-informed clinically relevant behavior in functional analytic psychotherapy: The awareness, courage, and love model. *Journal of Contextual Behavioral Science*, 6, 347-359. <https://doi.org/10.1016/j.jcbs.2017.07.002>
- Marin-Vila, M., Ortiz-Fune, C. & Kanter, J.K. (2020). Modelo conductual-contextual de las relaciones íntimas: Aplicabilidad en psicoterapia. *International Journal of Psychology and Psychological Therapy*, 20(3). 373-394.
- Márquez-Espinoza, D.I. (2019). Violencia intrafamiliar contra la mujer, un problema sociocultural. *El Cotidiano*, 217, 33-46.
- Milosavljevic, M. (2007). *Estadísticas para la equidad de género. Magnitudes y tendencias para América Latina*. Naciones Unidas.
- Ministerio de la Mujer y Poblaciones Vulnerables. (2017). *Impacto y consecuencias de la violencia contra las mujeres*. Recuperado de: <https://observatorioviolencia.pe/wp-content/uploads/2018/11/impacto-y-consecuencias-violencia-1.pdf>
- Moisés, A. (2012). Sexualidad en Mesoamérica machismo y marianismo. *Revista Científica*, 1(1), 45-53.
- Moral de la Rubia, J. & Ramos Basurto, S. (2016). Machismo, victimización y perpetración en mujeres y hombres mexicanos. *Estudios sobre las Culturas Contemporáneas*, 22(43), 37-66.
- Moriana Mateo, G. (2015). Barreras para escapar de la violencia de género: la mirada de las profesionales de los centros de protección de mujeres. *Cuadernos de Trabajo Social*, 28, 93-102. https://doi.org/10.5209/rev_CUTS.2015.v28.n1.44401
- Morrison, A.R. & Biehl, M.L. (1999). Too close to home: Domestic violence in the Americas. <https://publications.iadb.org/publications/spanish/document/El-costo-del-silencio-Violencia-dom%C3%A9stica-en-las-Am%C3%A9ricas.pdf>
- Muñoz-Martínez, A. M. (2019). Assessing the mechanism of change of functional analytic psychotherapy: an exploratory analysis of a process-based behavioral intervention (Doctoral dissertation). Recuperado de: <https://scholarworks.unr.edu/handle/11714/5707>
- O'Neal, E. N., & Beckman, L. O. (2016). Intersections of race, ethnicity, and gender. *Violence Against Women*, 23(5), 643-665. <https://doi.org/10.1177/1077801216646223>
- Ribero, R. & Sánchez, F. (2005). Determinants, effects and costs of domestic violence. [Documento CEDE, Universidad de los Andes]. <https://hdl.handle.net/1992/40995>
- Rincón, C. L., Muñoz-Martínez, A. M., Hoefflein, B., & Skinta, M. D. (2021). Enhancing interpersonal intimacy in Colombian gay men using functional analytic psychotherapy: An experimental nonconcurrent multiple baseline design. *Cognitive and Behavioral Practice*. Advance Online Publication. <https://doi.org/10.1016/j.cbpra.2021.10.003>
- Rizo, C. F., Givens, A., & Lombardi, B. (2017). A systematic review of coping among heterosexual female IPV survivors in the United States with a focus on the conceptualization and measurement of coping. *Aggression and Violent Behavior*, 34, 35-50. <https://doi.org/10.1016/j.avb.2017.03.006>
- Sarafino, E.P. (2012). *Applied behavior analysis: Principles and procedures for modifying behavior*. Wiley.
- Terrazas-Carrillo, E. & Sabina, C. (2019). Dating violence attitudes among Latino college students: An examination of gender, machismo, and marianismo. *Violence & Victims*, 34(1), 194-210. <https://doi.org/10.1891/0886-6708.34.1.194>
- Townsend, K. C., & McWhirter, B. T. (2005). Connectedness: a review of the literature with implications for counseling, assessment, and research. *Journal of Counseling & Development*, 83(2), 191-201. <https://doi.org/10.1002/j.1556-6678.2005.tb00596.x>
- Vaca-Ferrer, R., Ferro-García, R. & Valero-Aguayo, L. (2020). Efficacy of a group intervention program with women victims of gender violence in the framework of contextual therapies. *Anales de psicología*, 36(2), 189-199. <https://doi.org/10.6018/analesps.396901>

Normas de Publicación

Papeles del Psicólogo / Psychologist Papers

- 1 Papeles del Psicólogo / Psychologist Papers es una revista bilingüe que publica en español e inglés trabajos referidos al campo profesional de la psicología, principalmente en su vertiente aplicada y profesional.
- 2 Los trabajos remitidos para publicación habrán de ser inéditos y no enviados simultáneamente a otra revista.
- 3 La preparación de los originales ha de atenerse a las normas de publicación de la APA (2010). *Publication Manual of the American Psychological Association (6th ed.)*. Washington, DC: Author. <http://www.apastyle.org>.
- 4 Los trabajos tendrán una extensión máxima de 6000 palabras (que incluirá las referencias bibliográficas, figuras y tablas).
La primera página debe contener: título en español e inglés. Nombre, apellidos, profesión y lugar de trabajo de cada autor. Nombre y dirección del autor al que dirigir la correspondencia. Teléfono, Fax, Correo electrónico de contacto. La segunda página ha de incluir un resumen de no más de 150 palabras en español e inglés, así como las palabras clave en ambos idiomas.
- 5 Papeles del Psicólogo / Psychologist Papers acusará recibo inmediato de todo artículo recibido. Los originales no serán devueltos ni se mantendrá correspondencia sobre los mismos. En un plazo máximo de 90 días se contestará acerca de la aceptación o no para su publicación.
- 6 Todos los documentos que publique Papeles del Psicólogo / Psychologist Papers serán previamente evaluados de forma anónima por expertos, para garantizar la calidad científica y el rigor de los mismos, así como su interés práctico para los lectores. Podrían no ser evaluados aquellos trabajos empíricos y de corte experimental, más apropiados para las revistas especializadas; manuscritos cuyo estilo de redacción fuese también muy especializado y que no se ajustase al espectro de los lectores de la revista; originales que aborden cuestiones que hubieran sido ya recientemente tratadas y que no ofrecieran contribuciones relevantes a lo publicado; o trabajos cuya redacción no estuviera a la altura de la calidad exigible.
- 7 Los trabajos serán enviados a través de esta página web www.papelesdelpsicologo.es / www.psychologistpapers.com (siguiendo las instrucciones que se encuentran en el apartado “envío”). Su recepción se acusará de inmediato y, en el plazo más breve posible, se contestará acerca de su aceptación.

Una vez que un manuscrito ha sido remitido con éxito a través del sistema online, los autores pueden hacer un seguimiento de su estado mediante el sistema de envío online (por correo electrónico se proporcionarán los detalles al respecto). Una vez completado el envío, el sistema genera automáticamente un documento electrónico (PDF), que será utilizado para la revisión. Toda correspondencia, incluida la decisión del Director y la propuesta de modificaciones se procesarán por el sistema y serán recibidas por el autor, mediante correo electrónico. Los autores podrán enviar dudas acerca del proceso de envío o los procesos de publicación a los Editores mediante el siguiente formulario:
<http://www.papelesdelpsicologo.es/contacto>
- 8 Los trabajos sometidos a revisión para Papeles del Psicólogo / Psychologist Papers podrán abordar cualquiera de las siguientes cuestiones:
 - ✓ Implicaciones prácticas de investigaciones empíricas (investigación + ejercicio profesional); implementación de cuestiones generalmente ignoradas por los investigadores.
 - ✓ Investigación y desarrollo; desarrollo e innovaciones (I+D+I) (soluciones prácticas novedosas o de elección en áreas específicas); evaluaciones (análisis y crítica de tendencias emergentes, desde la perspectiva de su aplicación práctica).
 - ✓ Revisiones, estados de la cuestión, actualizaciones y meta-análisis de temáticas de la Psicología aplicada.
 - ✓ Contraste de opiniones, debates, políticas profesionales y cartas al editor (Forum).Por otra parte, el Comité Editorial podrá encargar trabajos específicos a autores reconocidos o proponer números especiales monográficos.
- 9 La aceptación de un trabajo para su publicación implica la cesión, por el/los autor/es, de los derechos de *copyright* al Consejo General de la Psicología de España.
Otro aspecto de la nueva política de la revista tiene que ver con **Conflicto de Intereses**. Todos los autores deben manifestar si existe algún conflicto de intereses potencial de tipo económico o de otras relaciones con personas y organizaciones.
Los puntos de vista, opiniones, hallazgos, conclusiones y recomendaciones planteados en cualquier artículo remitido a Papeles del Psicólogo / Psychologist Papers pertenecen a sus autores y no necesariamente reflejan posiciones de la Revista o de sus Editores.

