

Here comes a new challenger!¹: Evaluaciones mediante juegos en la selección de personal

Dr. Pedro José Ramos Villagrasa

Profesor Titular del área de Psicología Social de la Universidad de Zaragoza

pjramos@unizar.es

Resumen

El desarrollo tecnológico de las últimas décadas ha puesto a disposición de los profesionales de los recursos humanos nuevas formas de realizar los procesos de selección, como es el uso de videojuegos para evaluar a las personas candidatas. Sus defensores plantean que, utilizando juegos, se realizan evaluaciones iguales o mejores que utilizando pruebas tradicionales en términos de fiabilidad y validez, se reduce la discriminación y el falseamiento de las respuestas, además de incrementar las reacciones favorables por parte de las personas candidatas. En esta conferencia nos apoyamos en la evidencia científica existente para comprobar la veracidad de estas aseveraciones y clasificar los diferentes tipos de evaluaciones relacionadas con juegos (ERJ) existentes en la selección de personal. Derivada de esta experiencia, se plantea la existencia de, al menos, cuatro grandes tipos de ERJ: (1) evaluaciones gamificadas; (2) evaluaciones diseñadas mediante gamificación; (3) evaluaciones basadas en juegos; y (4) juegos con finalidad lúdica utilizados con fines de evaluación. Esta clasificación permite comprender las diferencias encontradas en los diferentes estudios empíricos existentes. Finalmente, se presentan algunos resultados obtenidos en población española y se establecen qué garantías debemos buscar como profesionales de los recursos humanos antes de utilizar una ERJ en nuestros procesos de selección de personal.

Agradecimientos

El autor agradece el apoyo del Ministerio de Ciencia e Innovación al proyecto “*SErious GAMes* para una Selección de Personal Inclusiva” (SEGASPI) financiado a través de la convocatoria Proyectos de Generación del Conocimiento 2021 (ref. PID2021-122867NA-I00) y al grupo S31_23R del Gobierno de Aragón (Departamento de Ciencia, Universidad y Sociedad del Conocimiento). Además, también quiere dar las gracias a la Cátedra de Modelos y Aplicaciones Psicométricos (MAP) la invitación a impartir esta conferencia, y a sus compañeros Elena Fernández del Río, Ramón Hermoso y Jorge Cebrián los comentarios realizados sobre este texto que sin duda alguna han contribuido a mejorarlo sustancialmente.

1. Introducción²

Las Tecnologías de la Información y la Comunicación (TIC) hace mucho tiempo que dejaron de ser las “nuevas tecnologías” para pasar a formar parte de la gestión organizacional. La selección de personal no es una excepción, de manera que incluso las organizaciones más modestas recurren en mayor o menor medida a las TIC en sus procesos, por ejemplo, difundiendo una oferta de trabajo a través de portales de empleo, páginas web y redes sociales. La pandemia mundial causada por la COVID-19 potenció este efecto, forzando al uso de la tecnología para continuar trabajando de manera remota o con menor contacto social (Scheel et al., 2023) y, como efecto colateral, familiarizando rápidamente a muchas personas con tecnologías que, estando previamente disponibles, no siempre eran utilizadas con frecuencia, como las videoconferencias (Karl et al., 2022).

En este nuevo escenario que se abre ahora que finaliza el primer cuarto del siglo XXI, los profesionales de la selección de personal tienen la oportunidad de repensar su praxis en busca de mejoras en los procesos, reducción de plazos, optimización de recursos y, en definitiva, de aprovechar lo que las TIC pueden ofrecer para mejorar su trabajo. La investigación científica siempre ha sido una aliada para este fin, siendo posible encontrar ya investigaciones acerca de cuestiones punteras como las entrevistas asíncronas (Kleinlogel et al., 2023), la evaluación de los candidatos mediante *machine learning* (Hickman et al., 2023), y el uso de videojuegos como método de selección (Barends y De Vries, 2023). En España, es destacable el papel que los seminarios anuales organizados por la Cátedra MAP junto al Instituto de Ingeniería del Conocimiento han tenido para reflexionar y divulgar acerca de muchas de estas cuestiones, como la *HR analytics* (2018), el teletrabajo (2020), la inteligencia artificial aplicada a la gestión del talento (2022), o este mismo seminario.

La presente conferencia tiene como objetivo presentar el estado actual de las evaluaciones relacionadas con juegos (ERJ, *game-related assessments* en inglés) en la selección de personal. Comenzaremos definiendo este concepto y mostrando la nada desdeñable variedad de métodos de evaluación existentes, para después analizar qué sabemos gracias a la investigación científica sobre esta materia. Todo ello nos preparará para responder a la pregunta: ¿debería plantearme incluir videojuegos en el próximo proceso de selección de personal de mi organización?

2. Evaluaciones relacionadas con juegos (ERJ)

En los últimos años, *gamificación* ha pasado a ser un término conocido por el gran público³. La gamificación consiste en incorporar elementos propios de los juegos, tales como

reglas y metas, historias inmersivas y elementos de ambientación como música, en contextos distintos del juego (Nacke y Deterding, 2017). Dicho de otra forma, supone coger algunas o todas las características que definen a los juegos e introducirlas en un contexto diferente, generando en el proceso algunos beneficios, tales como hacer más divertida la actividad que hemos modificado, generar mayor implicación, etc.⁴

Las ERJ son métodos de evaluación de personas desarrollados mediante gamificación (Ramos-Villagrasa et al., 2022b). Su objetivo, por tanto, es el de estimar características personales de interés (aptitud cognitiva, competencias, personalidad, etc.) con métodos que, en mayor o menor medida, sean identificados por las personas evaluadas como juegos. Esta definición ha generado que algunos autores consideren ERJ los *business games* (Buil et al., 2020) e incluso los *assessment center* y las simulaciones (Landers y Sanchez, 2022).

Tanto por parte de las empresas que han comenzado a comercializar estos métodos como por la de los investigadores que han empezado a estudiarlos, se han apuntado al menos cuatro supuestos beneficios del uso de ERJ con respecto a las pruebas tradicionales como entrevistas, test y cuestionarios (Landers y Sanchez, 2022; Melchers y Basch, 2022; Wu et al., 2022): (1) la reducción de la propensión de los candidatos a falsear sus respuestas (*faking*), al estar más involucrados en el juego y tener mayores dificultades para percibir qué se está evaluando y cuál es la respuesta más deseable socialmente; (2) el incremento de las reacciones favorables hacia el método de evaluación, el proceso de selección y la propia organización por parte de los candidatos, especialmente entre los más jóvenes (*millennials*, generación Z); (3) la realización de medidas ecuanímes, que no discriminen a determinados colectivos (por ejemplo, por razón de sexo o edad); y (4) ser capaz de hacer todo esto sin una pérdida sustancial de validez predictiva⁵ respecto a las pruebas ya establecidas. Más adelante hablaremos acerca de esas supuestas ventajas de las ERJ.

El uso de las ERJ ha empezado a popularizarse ahora, pero no son en absoluto novedosas. A modo meramente ilustrativo pueden citarse algunos hitos: en 2004, la empresa Unilever utiliza en Alemania una ERJ basada en una ficticia marca de helados (“Indian Summer”) para evaluar la aptitud cognitiva en los candidatos a puestos de aprendiz (Kupka et al., 2004), el videojuego para móviles Wasabi Waiter de la compañía Knack, probablemente una de las ERJ más populares, fue lanzado en 2013. En 2018, la compañía Red Bull lanzó en cinco idiomas diferentes, incluido el español, Wingfinder, una ERJ que puede ser utilizada por cualquier persona con fines de orientación y que ellos mismos utilizan en sus procesos de selección. En el mercado hispanoparlante es posible encontrar alternativas propias como Nawaiam, lanzado en 2019. En 2020 las Fuerzas Aéreas de los Estados Unidos de América han comenzado a utilizar

Virus Slayer, una ERJ diseñado a su medida para seleccionar a los ocupantes de puestos de trabajo en ciberseguridad (Wiernik et al., 2022). El interés por parte de los profesionales de los recursos humanos en las ERJ tiene su reflejo en la investigación científica, que ha pasado de un único artículo publicado en 2012 como algo aislado, hasta 2016, cuando se produce un incremento progresivo hasta los catorce estudios de 2022, año en que se publica un monográfico sobre el tema en la prestigiosa revista *International Journal of Selection Assessment* (Ramos-Villagrasa et al., 2022b). Estamos, por tanto, ante un tipo de evaluaciones que no son nuevas, pero que sí han recibido un nuevo interés derivado de aprovechar al máximo la digitalización de los recursos humanos en la que nos encontramos inmersos en este momento (Nikolaou, 2021).

3. Tipos de ERJ

Al acercarse por primera vez a las ERJ el primer problema que aparece no es tan diferente al que encontramos en otros métodos de evaluación. Al igual que existen diferentes tipos de entrevistas de selección (Alonso y Moscoso, 2018) y de cuestionarios de personalidad (Martínez y Salgado, 2021), las ERJ no son un conjunto homogéneo, sino que aglutinan evaluaciones que pueden llegar a ser sustancialmente diferentes entre sí. A partir de una revisión sistemática de los artículos científicos al respecto y de la distinción terminológica realizada por Landers y Sanchez (2022), nuestro equipo de investigación ha propuesto una taxonomía que distingue entre cuatro grandes categorías de ERJ (Ramos-Villagrasa et al., 2022b). Esta taxonomía, presentada en la Figura 1, será la que explicaremos a continuación.



Figura 1. Tipos de ERJ (Ramos-Villagrasa et al., 2022b)

De acuerdo con nuestra clasificación, las ERJ se distribuyen en un continuo según su grado de ludismo, siendo uno de los extremos las evaluaciones tradicionales, que no tienen intención lúdica alguna, y el otro extremo los juegos convencionales, desarrollados con el único fin de divertir. Dependiendo de lo próximo que se encuentre una ERJ de uno u otro extremo hablaremos de: (1) evaluaciones gamificadas; (2) evaluaciones diseñadas mediante gamificación; (3) evaluaciones basadas en juegos; y (4) juegos con finalidad lúdica utilizados con

finés de evaluaci3n. De estas cuatro categorías, las tres primeras entrarían dentro de la denominaci3n de juegos serios⁶.

Las evaluaciones gamificadas son pruebas tradicionales (ya existentes o creadas expresamente) a las que se les añaden elementos propios de los juegos mediante gamificaci3n para generar la ERJ. Por ejemplo, se puede tomar un cuestionario de personalidad estándar, con formato de respuesta tipo Likert, y convertirlo en una evaluaci3n gamificada introduciendo una historia que relacione los ítems, música y/o minijuegos, pero se seguirán planteando ítems a las personas evaluadas y se responderán en formato Likert igualmente. Es decir, en las evaluaciones gamificadas podríamos eliminar las características propias de los juegos y la evaluaci3n continuaría siendo posible. Son ejemplos de evaluaciones gamificadas Wingfinder (Red Bull, 2023), Gamified Set-Shifting Task (Hommel et al., 2022) y VASSIP (Ramos-Villagrasa et al., 2023), siendo este último desarrollado por el equipo de investigaci3n del proyecto SEGASPI, al que pertenezco.

Las evaluaciones diseñadas mediante gamificaci3n constituyen un paso más en términos de ludismo. La incorporaci3n de características propias de los juegos es tal que, aunque recurren a formas de evaluaci3n tradicionales (por ejemplo, un test de juicio situacional), si eliminamos los elementos de juego la evaluaci3n no podría realizarse. Son ejemplos de evaluaciones diseñadas mediante gamificaci3n Owiwi (Georgiou et al., 2019), Social Shapes Test (Brown et al., 2022) y HEXACO-PI Rush (Katsadorakis y Katsadoraki, 2022).

En las evaluaciones basadas en juegos el estímulo que recibe la persona evaluada no recuerda a los propios de una evaluaci3n tradicional, sino a los juegos diseñados con finalidad lúdica. La diferencia radica en que aquéllos han sido desarrollados con la finalidad de evaluar, no de divertir. Dentro de las evaluaciones basadas en juegos coexisten aquéllas que se han diseñado a partir de un modelo teórico previo y buscan evaluar constructos concretos en las personas que juegan (evaluaciones basadas en juego guiadas por teoría) y las que han establecido relaciones entre el comportamiento de la persona evaluada en el juego y el desempeño laboral (evaluaciones basadas en juego guiadas por datos). De las evaluaciones de este tipo guiadas por la teoría son ejemplos Nawaiam, basado en el modelo de personalidad DISC (Ramos-Villagrasa y Fernández-del-Río, 2023), Cognify, para medir aptitud cognitiva (Van Lill et al., 2023), y Clicktown, diseñado para medir responsabilidad (Wu et al., 2022), el rasgo de los “Cinco Grandes” factores de personalidad con mayor relaci3n con el desempeño laboral. En cuanto a las evaluaciones guiadas por los datos existe un estudio sobre el juego Cognify en el que se emplean medidas adicionales al desempeño en el juego como la posici3n del rat3n y el

recorrido que realiza para mejorar la medida de la aptitud cognitiva que realiza el juego (Auer et al., 2022).

Los juegos con finalidad lúdica que se utilizan con fines de evaluación son juegos convencionales, como puede ser el conocido Tetris, en los que se ha mostrado la correspondencia entre los resultados obtenidos por el jugador y alguna variable de interés. Son, por tanto, ERJ, pero no juegos serios de acuerdo con nuestra clasificación. Este tipo de ERJ cuentan con una relevancia mucho más escasa en el ámbito profesional, aunque en el ámbito investigador sí existen estudios al respecto, como los realizados sobre videojuegos de realidad virtual (Sanchez et al., 2022; Simons et al., 2023) y los realizados en España por Quiroga y colaboradores que muestran la relación entre la aptitud cognitiva y juegos no necesariamente “cerebrales” como Space Invaders y Splatoon (Quiroga et al., 2019).

De este repaso a las diferentes ERJ podemos extraer una primera conclusión: no todas las ERJ son iguales, ni en cuanto a cómo se han diseñado, ni en los constructos que miden ni en cómo evalúan a los candidatos. La siguiente pregunta que podemos hacernos es: ¿Es el tipo de juego relevante para la selección de personal? La respuesta es sí, y a continuación explicaremos el por qué.

4. ERJ y selección de personal inclusiva

Para comenzar, debemos ser conscientes que la selección de personal actual ha cambiado sustancialmente respecto a la existente hace diez, quince o veinte años. Además de los desarrollos tecnológicos ya mencionados⁷, en muchos países como España se han desarrollado normativas que exigen a las organizaciones realizar procesos de selección que garanticen una evaluación justa de las personas candidatas y sin recabar más información que la estrictamente necesaria (Ramos-Villagrasa, 2020). Las propias organizaciones también son más conscientes del impacto que generan en las personas candidatas y sus consecuencias mucho más allá del propio proceso de selección (Bauer et al., 2020). Como consecuencia, las organizaciones actuales tratan de realizar lo que Salgado y colaboradores denominan *procesos de selección personal inclusiva*, que se definen como procesos en los que la selección cumple su objetivo de evaluar el ajuste al puesto de trabajo y la organización, pero asegurándose de respetar la diversidad y la no-discriminación (Salgado et al., 2017). En términos operativos, esto se traduce en que los métodos de selección empleados deben mostrar al menos tres cualidades: (1) propiedades psicométricas adecuadas (fiabilidad y validez)⁸; (2) ausencia de sesgos⁹; y (3) promover reacciones favorables en las personas candidatas¹⁰.

Si atendemos a las presuntas ventajas de las ERJ apuntadas anteriormente (i.e., Landers y Sanchez, 2022; Melchers y Basch, 2022; Wu et al., 2022), aparentemente son métodos adecuados para la selección de personal inclusiva, añadiendo además la prevención del falseamiento de las respuestas. Sin embargo, estas afirmaciones deben venir refrendadas por la investigación científica. A este respecto, nuestro equipo de investigación realizó una revisión de la literatura acerca de esta cuestión (i.e., Ramos-Villagrasa et al., 2022), cuyos resultados se resumen a continuación.

Con respecto a las propiedades psicométricas, los resultados con respecto a la fiabilidad de las ERJ muestran resultados adecuados, en la mayor parte de las ocasiones estimados mediante consistencia interna (α de Cronbach). Sin embargo, dada la amplia variedad de ERJ y que la fiabilidad de algunos de ellos difícilmente podría estimarse mediante consistencia interna (por ejemplo, en los casos en que la prueba incluye en la evaluación aspectos relacionados con el comportamiento del jugador), se recomienda la estimación por método test-retest (Sanchez et al., 2022).

Continuando con las propiedades psicométricas, nos centraremos ahora en la validez. En primer lugar, debemos considerar la validez de constructo, donde los resultados están lejos de ser concluyentes porque analizan ERJ de tipos muy distintos de acuerdo con la clasificación presentada en la Figura 1. El resultado más claro con la evidencia hasta la fecha es que, desde el punto de vista de la validez de constructo, es mejor emplear juegos diseñados específicamente con propósitos de evaluación y que se basen en constructos bien establecidos, como la aptitud cognitiva y la personalidad. El segundo tipo de validez en el que queremos detenernos es la validez predictiva, donde diversas investigaciones han encontrado relación entre las ERJ y diversos tipos de desempeño¹¹, principalmente el académico y el desempeño de tarea (e.g., Georgiou et al., 2019; Harman y Brown, 2022; Landers y Collmus, 2022). Es destacable que, en las investigaciones realizadas en España, también hemos encontrado relación de una evaluación gamificada con el desempeño contextual y los comportamientos contraproducidos en el trabajo (Ramos-Villagrasa et al., 2023) y de una evaluación basada en juegos con el desempeño adaptativo (Ramos-Villagrasa y Fernández-del-Río, 2023). Además, es interesante mencionar que Landers et al. (2021) han investigado si utilizar en un mismo proceso una evaluación basada en juegos y una prueba convencional, ambas para medir aptitud cognitiva, suponía alguna mejora. Sus resultados muestran que la ERJ no supone una mejora significativa respecto a usar solamente un test de aptitud cognitiva, pero sí al revés: el test sí mejora los resultados de la ERJ. De todas formas, al tratarse de un único estudio y con un único tipo de ERJ tampoco podemos ser taxativos a este respecto.

Antes de finalizar nuestra revisión sobre las propiedades psicométricas, hablaremos del falseamiento de las respuestas (*faking*). Hasta el momento únicamente un estudio ha analizado esta cuestión, pero sus resultados son positivos, encontrando un menor falseamiento en una prueba de personalidad realizada mediante evaluación gamificada que en una prueba convencional (Landers y Collmus, 2022).

Hablemos ahora de la ausencia de sesgos. Los estudios han analizado diferentes características, siendo las más habituales género, edad y experiencia con los videojuegos, partiendo del estereotipo de que los varones jóvenes tienden a jugar más a videojuegos que el resto de personas y, por tanto, quizá obtengan mejores resultados en las ERJ. Las investigaciones no apoyan esos resultados y, en los casos en los que lo hace, la diferencia parece ser irrelevante en un contexto real (Melchers y Basch, 2022). Sin embargo, sí encuentran que los resultados de las ERJ favorecen a las personas de mayor nivel educativo, que tienen experiencia con la tecnología y se perciben a sí mismos como capaces de jugar bien a videojuegos¹², aunque la magnitud de esa influencia probablemente varíe dependiendo del tipo de ERJ. Los estudios que se han centrado en el impacto adverso de las ERJ no encuentran resultados significativamente diferentes a los de las pruebas tradicionales (Brown et al., 2022; Hilliard et al., 2022; Landers et al., 2021). En síntesis, a tenor de los resultados existentes, parece que las ERJ no son ni más ni menos justas que las pruebas ya existentes.

Finalmente, hablemos de las reacciones favorables por parte de las personas candidatas. Éste es, considerablemente, el aspecto más investigado hasta el momento y el que ofrece resultados más concluyentes independientemente del tipo de ERJ del que se trate: las ERJ promueven reacciones significativamente más positivas que las pruebas de selección convencionales, especialmente en lo referido al atractivo organizacional¹³. Sin embargo, existe una excepción notable: la relación con el desempeño del puesto de trabajo, aunque se puede paliar dando explicaciones a las personas evaluadas acerca de esta cuestión (Georgiou, 2021).

5. Conclusiones

Es innegable que las ERJ han tenido un considerable desarrollo entre los últimos cinco y diez años, extendiéndose tanto al ámbito profesional como al investigador. En el ámbito profesional ya existen diversas empresas, tanto en el ámbito internacional como en el nacional, que ofrecen ERJ para procesos de selección y promoción de candidatos. En el ámbito investigador todavía queda mucho trabajo por hacer, pero se han analizado las características definitorias de estas nuevas formas de evaluación, se han establecido tipologías, y se han comenzado a investigar cuáles son sus características psicométricas, su impacto en las

reacciones de las personas candidatas y su posible imparcialidad, además de qué características de diseño son más apropiadas para promover una selección de personal inclusiva. Teniendo todo esto en cuenta podemos extraer algunas recomendaciones para los profesionales de los recursos humanos y los diseñadores de ERJ.

Con respecto a los profesionales de los recursos humanos:

- (1) Las ERJ pueden ser empleadas como un método de selección más. Los resultados de las investigaciones hasta el momento no nos permiten pensar que sean significativamente mejores o peores que otros métodos de selección existentes. Justamente por eso, los profesionales que deseen incorporar ERJ en sus procesos de selección deben plantearse qué valor añadido les aporta, en su caso concreto, respecto a métodos ya existentes. Por ejemplo, en sectores como el IT, en el que existe una elevada competencia por captar a los mejores candidatos, el efecto favorable que parecen tener las ERJ sobre el atractivo organizacional puede ser útil, pero sin comprometer la calidad de la medida (fiabilidad y validez).
- (2) El tipo de ERJ importa. Existe una considerable variedad de ERJ y las características psicométricas, de reacciones de los candidatos y de justicia parecen variar sustancialmente de unas a otras. Recorra solamente a juegos serios. Dentro de ellos, de manera general, parece que cuánto más nos acercamos a los métodos convencionales, la ERJ tenderá a tener unos mejores resultados psicométricos pero unas reacciones de las personas candidatas menos favorables. Es posible que la investigación sobre ERJ matice esto en el futuro.
- (3) La ERJ debe tener avales que apoyen su uso. De igual manera que un test o un cuestionario empleado en selección de personal debe ir refrendado por un estudio psicométrico que explique los constructos que mide y evidencias de que lo hace (validez de contenido), que los resultados de la evaluación se relacionan con el desempeño laboral (validez predictiva) y cualquier posible sesgo que influya en la evaluación imparcial de personas candidatas, las ERJ deben también ofrecer esta información. Algunas empresas los desarrollan internamente, otros colaboran con investigadores independientes para que avalen su uso, como Owiwi (Georgiou et al., 2019) y Nawaiam (Ramos-Villagrana y Fernández-del-Río, 2023). La persona que realiza la evaluación es la que debe responsabilizarse de exigir y comprobar esos avales en cualquier prueba que emplee.
- (4) Tener en cuenta las diferencias sociodemográficas. Relacionado con el punto anterior están algunas diferencias que han aparecido en varias investigaciones

relacionadas con que las personas de mayor nivel educativo, experiencia con tecnología y autoeficacia para jugar videojuegos tienden a obtener mejores resultados en las ERJ. Tener en cuenta esos tres aspectos en la evaluación de candidatos puede ayudar a hacer evaluaciones más justas.

Con respecto a los diseñadores de ERJ:

- (5) Contar con expertos en medición psicológica. Desarrollar una ERJ requiere tecnología, pero si las bases de medida no están claras (el constructo adecuadamente definido, la forma de analizar las respuestas, etc.), el resultado será, en el mejor de los resultados, pobre e inferior a las pruebas existentes.
- (6) Elegir el tipo de ERJ que se desea crear. La clasificación de la Figura 1 ofrece cuatro alternativas diferentes y cómo alcanzarlas. Si se es nuevo en el campo, mejor partir de evaluaciones gamificadas o evaluaciones diseñadas mediante gamificación.
- (7) Hacer un uso inteligente de la tecnología. Aunque no todas las ERJ deban basarse en tecnología, dado que la mayor parte lo hacen podemos aprovechar esa circunstancia de diversas formas: (a) tener baremos crecientes y dinámicos; (b) emplear los propios datos de desempeño durante la ERJ para mejorar la predicción del criterio (Auer et al., 2022); (c) servirse de los desarrollos en el ámbito de los algoritmos para la mejora de los modelos predictivos como, por ejemplo, mediante grafos de conocimiento (*knowledge graphs*)¹⁴; (d) dar *feedback* automático a las personas candidatas.
- (8) Buscar evidencias de resultados, primero internas (para calibrar el instrumento, alcanzar un producto mínimo viable) y después externas (por ejemplo, mediante la colaboración con equipos de investigación).

A modo de cierre, recordamos el título de esta conferencia, que aludía a la ERJ como un nuevo competidor. La moneda ya ha sido echada y la partida ha comenzado. Dependerá de los científicos, los diseñadores de pruebas de evaluación y los profesionales de recursos humanos si las ERJ alcanzan o no el *high score* de los métodos de selección de personal.

¹ La frase *Here comes a new challenger!* fue popularizada por el videojuego Street Fighter II (1991) para indicar que otro jugador desafiaba al que estaba jugando, de manera que el vencedor continuaba con la partida. La frase se escuchaba a menudo en las salas de recreativos cuando un jugador retaba a otro, y sucesivas entregas de la saga y otros videojuegos la han mantenido. En nuestro contexto, refleja muy bien la entrada de las evaluaciones relacionadas con juegos en la selección de personal.

² Si bien este capítulo ha sido escrito de forma inclusiva, puntualmente se utiliza el masculino como neutro, aludiendo a ambos sexos indistintamente, con objeto de facilitar la lectura.

³ Como muestra, en junio de 2023 el término *gamificación* arrojaba en Google aproximadamente 4.350.000 resultados. Si se utiliza el término en inglés *gamification* se obtienen 29.7000.000 resultados.

⁴ El observatorio de palabras de la Real Academia de la Lengua Española recoge el término *gamificación*, si bien desaconseja su uso a favor del término *ludificación*, que ya se encuentra en su diccionario y que se define como “acción o efecto de ludificar”, esto es, 1. tr. Transformar algo en un juego o fomentar sus aspectos lúdicos. 2. tr. Aplicar técnicas o dinámicas propias del juego a actividades o entornos no recreativos para potenciar la motivación y la participación, o facilitar el aprendizaje y la consecución de objetivos. Si bien es cierto que el término *ludificar* parece más apropiado, lo cierto es que ha quedado totalmente eclipsado por el préstamo del inglés.

⁵ La validez predictiva hace referencia al grado en que en una prueba de evaluación es capaz de predecir una conducta futura (Aguinis et al., 2001). En la selección de personal lo habitual es que la conducta que queramos predecir es el desempeño de la persona candidata en caso de que, efectivamente, se incorporase al puesto de trabajo.

⁶ Los juegos serios (*serious games* en inglés) son juegos diseñados para propósitos diferentes del lúdico. En la selección de personal todas las ERJ son juegos serios, salvo que se esté empleando juego lúdico con otros fines (el cuarto tipo de acuerdo con la clasificación propuesta).

⁷ De los cuales puede verse una revisión extensa en Woods et al. (2020).

⁸ La fiabilidad hace referencia al grado en que la medida de un instrumento está libre de errores, sean estos debidos a variaciones momentáneas (distracciones, fatiga, etc.), de interpretación de los candidatos o de las valoraciones que realizan los evaluadores (Moscoso et al., 2017). La validez, en cambio, se refiere a “la pertinencia de las inferencias que realizamos a partir de las calificaciones obtenidas en una prueba” (Schmitt y Klimoski, 1991, p.89) y puede ser de diferentes tipos. A efectos de esta conferencia nos interesa la validez de contenido (garantía de que la prueba de evaluación constituye una muestra adecuada y representativa de los contenidos que pretendemos evaluar) y la validez predictiva (grado en que las puntuaciones en la prueba de evaluación permiten predecir un criterio de interés, como el desempeño laboral).

⁹ Ausencia de sesgos se refiere a que el resultado de la prueba dependa de las características que queremos evaluar en la persona candidata y no de otras características tales como su sexo, edad, experiencia con videojuegos, competencia idiomática, etc.

¹⁰ Las reacciones de las personas candidatas hacen referencia a las impresiones que éstas se hacen acerca de cómo son evaluadas, del proceso de selección en sí y de la propia organización a partir de lo que sucede durante la selección. La investigación ha mostrado que estas reacciones impactan en la decisión de aceptar una oferta de trabajo, de volver a presentarse e incluso en el desempeño laboral y la reputación de la organización (Bauer et al., 2020).

¹¹ Los modelos actuales de desempeño laboral van más allá de los comportamientos directamente relacionados con el desempeño del puesto para incluir también (Ramos-Villagrasa et al., 2022a): cómo contribuye la persona trabajadora de manera voluntaria y extra-rol al ambiente organizacional y el logro de las metas (desempeño contextual) y sus comportamientos dirigidos en el sentido contrario, es decir, a menoscabar el ambiente y perjudicar las metas organizacionales (comportamientos contraproductivos). Aproximaciones todavía más recientes incorporan también la capacidad de modificar el ritmo y forma de trabajo de acuerdo a circunstancias cambiantes (desempeño adaptativo).

¹² Denominado de manera técnica: autoeficacia para jugar videojuegos.

¹³ Tanto es así que el mero hecho de presentar una prueba de evaluación tipo test o cuestionario como un juego, modificando levemente la presentación e instrucciones, ya consigue unas reacciones más favorables (Collmus y Landers, 2019; McChesney et al., 2022).

¹⁴ Los grafos de conocimiento se utilizan para crear relaciones inteligentes entre los conjuntos de datos que los componen. Por ejemplo, de la misma forma que estamos habituados a establecer una asociación estadística entre los resultados de una prueba de evaluación y el desempeño laboral, podríamos enriquecer el sistema con otros conjuntos de datos (descripciones de puestos de trabajo, evidencias de desempeño de la plantilla) para mejorar nuestros modelos predictivos. Ésta es una de las líneas de trabajo que estamos explorando en nuestro grupo de investigación con VASSIP.

Referencias

- Aguinis, H., Henle, C. A. y Ostroff, C. (2001). Measurement in Work and Organizational Psychology. En *Handbook of Industrial, Work and Organizational Psychology: Personnel Psychology* (pp. 27-50). SAGE Publications Ltd. <https://doi.org/10.4135/9781848608320.n3>
- Alonso, P. y Moscoso, S. (2018). Employment Interview Perceptions Scale. *Revista de Psicología Del Trabajo y de Las Organizaciones*, 34, 203-212. <https://doi.org/10.5093/jwop2018a22>
- Auer, E. M., Mersy, G., Marin, S., Blaik, J. y Landers, R. N. (2022). Using machine learning to model trace behavioral data from a game-based assessment. *International Journal of Selection and Assessment*, 30, 82-102. <https://doi.org/10.1111/ijsa.12363>
- Barends, A. J. y De Vries, R. E. (2023). Construct validity of a personality assessment game in a simulated selection situation and the moderating roles of the ability to identify criteria and dispositional insight. *International Journal of Selection and Assessment*, 31, 120-134. <https://doi.org/10.1111/ijsa.12404>
- Bauer, T. N., McCarthy, J., Anderson, N., Truxillo, D. M. y Salgado, J. F. (2020). *What We Know About the Candidate Experience: Research Summary and Best Practices for Applicant Reactions*. Society for Industrial and Organizational Psychology.
- Brown, M. I., Speer, A. B., Tenbrink, A. P. y Chabris, C. F. (2022). Using game-like animations of geometric shapes to simulate social interactions: An evaluation of group score differences. *International Journal of Selection and Assessment*, 30, 167-181. <https://doi.org/10.1111/ijsa.12375>
- Buil, I., Catalán, S. y Martínez, E. (2020). Understanding applicants' reactions to gamified recruitment. *Journal of Business Research*, 110, 41-50. <https://doi.org/10.1016/j.jbusres.2019.12.041>
- Collmus, A. B. y Landers, R. N. (2019). Game-Framing to Improve Applicant Perceptions of Cognitive Assessments. *Journal of Personnel Psychology*, 18, 157-162. <https://doi.org/10.1027/1866-5888/a000227>
- Georgiou, K. (2021). Can explanations improve applicant reactions towards gamified assessment methods? *International Journal of Selection and Assessment*, 29, 253-268. <https://doi.org/10.1111/ijsa.12329>
- Georgiou, K., Gouras, A. y Nikolaou, I. (2019). Gamification in employee selection: The development of a gamified assessment. *International Journal of Selection and Assessment*, 27, 91-103. <https://doi.org/10.1111/ijsa.12240>
- Harman, J. L. y Brown, K. D. (2022). Illustrating a narrative: A test of game elements in game-like personality assessment. *International Journal of Selection and Assessment*, 30, 157-166. <https://doi.org/10.1111/ijsa.12374>
- Hickman, L., Herde, C. N., Lievens, F. y Tay, L. (2023). Automatic scoring of speeded interpersonal assessment center exercises via machine learning: Initial psychometric evidence and practical guidelines. *International Journal of Selection and Assessment*, 31, 225-239. <https://doi.org/10.1111/ijsa.12418>
- Hilliard, A., Kazim, E., Bitsakis, T. y Leutner, F. (2022). Measuring Personality through Images: Validating a Forced-Choice Image-Based Assessment of the Big Five Personality Traits. *Journal of Intelligence*, 10, 12. <https://doi.org/10.3390/jintelligence10010012>

- Hommel, B. E., Ruppel, R. y Zacher, H. (2022). Assessment of cognitive flexibility in personnel selection: Validity and acceptance of a gamified version of the Wisconsin Card Sorting Test. *International Journal of Selection and Assessment*, 30, 126-144. <https://doi.org/10.1111/ijsa.12362>
- Karl, K. A., Peluchette, J. V. y Aghakhani, N. (2022). Virtual Work Meetings During the COVID-19 Pandemic: The Good, Bad, and Ugly. *Small Group Research*, 53, 343-365. <https://doi.org/10.1177/10464964211015286>
- Katsadorakis, A. y Katsadoraki, A. (2022). *HEXACO-PI Rush*. <https://hexacorush.com/>
- Kleinlogel, E. P., Schmid Mast, M., Jayagopi, D. B., Shubham, K. y Butera, A. (2023). "The interviewer is a machine!" Investigating the effects of conventional and technology-mediated interview methods on interviewee reactions and behavior. *International Journal of Selection and Assessment*, ijsa.12433. <https://doi.org/10.1111/ijsa.12433>
- Kupka, K., Diercks, J. y Kopping, N. (2004). Webbasierte Personalvorauswahl durch E-Assessment bei Unilever Deutschland [Preselección de personal por Internet mediante evaluación electrónica en Unilever Alemania]. *Wirtschaftspsychologie aktuel*, 3, 24-28.
- Landers, R. N., Armstrong, M. B., Collmus, A. B., Mujcic, S. y Blaik, J. (2021). Theory-driven game-based assessment of general cognitive ability: Design theory, measurement, prediction of performance, and test fairness. *Journal of Applied Psychology*. <https://doi.org/10.1037/apl0000954>
- Landers, R. N. y Collmus, A. B. (2022). Gamifying a personality measure by converting it into a story: Convergence, incremental prediction, faking, and reactions. *International Journal of Selection and Assessment*, 30, 145-156. <https://doi.org/10.1111/ijsa.12373>
- Landers, R. N. y Sanchez, D. R. (2022). Game-based, gamified, and gamefully designed assessments for employee selection: Definitions, distinctions, design, and validation. *International Journal of Selection and Assessment*, 30, 1-13. <https://doi.org/10.1111/ijsa.12376>
- Martínez, A. y 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>
- McChesney, J., Campbell, C., Wang, J. y Foster, L. (2022). What is in a name? Effects of game-framing on perceptions of hiring organizations. *International Journal of Selection and Assessment*, 30 182-192. <https://doi.org/10.1111/ijsa.12370>
- Melchers, K. G. y Basch, J. M. (2022). Fair play? Sex-, age-, and job-related correlates of performance in a computer-based simulation game. *International Journal of Selection and Assessment*, 30, 48-61. <https://doi.org/10.1111/ijsa.12337>
- Moscoso, S., Alonso, P. y Cuadrado, D. (2017). Métodos de investigación en ergonomía y psicología social aplicada a la prevención de riesgos laborales. En A. L. Garcia-Izquierdo (Ed.), *Ergonomía y Psicosociología Aplicada a la Prevención de Riesgos Laborales* (pp. 65-90). Servicio de Publicaciones de la Universidad de Oviedo.
- Nacke, L. E. y Deterding, S. (2017). The maturing of gamification research. *Computers in Human Behavior*, 71, 450-454. <https://doi.org/10.1016/j.chb.2016.11.062>
- Nikolaou, I. (2021). What is the Role of Technology in Recruitment and Selection? *The Spanish Journal of Psychology*, 24, e2. <https://doi.org/10.1017/SJP.2021.6>
- Quiroga, M. A., Diaz, A., Román, F. J., Privado, J. y Colom, R. (2019). Intelligence and video games: Beyond "brain-games". *Intelligence*, 75, 85-94. <https://doi.org/10.1016/j.intell.2019.05.001>

- Ramos-Villagrasa, P. J. (2020). Los procesos de reclutamiento y selección de personal con perspectiva de género. En E. Sierra Hernández y R. Vallejo Dacosta (Eds.), *Diseño e implementación de planes de igualdad en las empresas* (pp. 367-398). Aranzadi.
- Ramos-Villagrasa, P. J. y Fernández-del-Río, E. (2023). 'Keep trying!': Predictive validity, applicant reactions, and influence of personal characteristics of a gamefully designed assessment. *Artículo en revisión*.
- Ramos-Villagrasa, P. J., Fernández-del-Río, E. y Castro, Á. (2022a). Analysis of a brief biodata scale as a predictor of job performance and its incremental validity over the Big Five and Dark Tetrad personality traits. *PLOS ONE*, *17*, e0274878. <https://doi.org/10.1371/journal.pone.0274878>
- Ramos-Villagrasa, P. J., Fernández-del-Río, E. y Castro, Á. (2022b). Game-related assessments for personnel selection: A systematic review. *Frontiers in Psychology*, *13*, 952002. <https://doi.org/10.3389/fpsyg.2022.952002>
- Ramos-Villagrasa, P. J., Hermoso, R., Cebrián, J. y Fernández-del-Río, E. (2023). VASSIP: A gamified assessment to measure Big Five personality traits. *Artículo en revisión*.
- Red Bull. (2023). *The science behind Wingfinder*. <https://www.redbull.com/int-en/wingfinder/science>
- Salgado, J. F., Moscoso, S., García-Izquierdo, A. L. y Anderson, N. R. (2017). Inclusive and Discrimination-Free Personnel Selection. En A. Arenas, D. Di Marco, L. Munduate, y M. C. Euwema (Eds.), *Shaping Inclusive Workplaces Through Social Dialogue* (pp. 103-119). Springer International Publishing. https://doi.org/10.1007/978-3-319-66393-7_7
- Sanchez, D. R., Weiner, E. y Van Zelderren, A. (2022). Virtual reality assessments (VRAs): Exploring the reliability and validity of evaluations in VR. *International Journal of Selection and Assessment*, *30*, 103-125. <https://doi.org/10.1111/ijsa.12369>
- Scheel, T. E., Bendixen, L., Procházka, J. y Acquadro Maran, D. (2023). Working during the COVID-19 pandemic: Demands, resources, and mental wellbeing. *Frontiers in Psychology*, *13*, 1037866. <https://doi.org/10.3389/fpsyg.2022.1037866>
- Schmitt, N. y Klimoski, R. (1991). *Research Methods in Human Resources Management*. South-Western Pub.
- Simons, A., Wohlgenannt, I., Zelt, S., Weinmann, M., Schneider, J. y Vom Brocke, J. (2023). Intelligence at play: Game-based assessment using a virtual-reality application. *Virtual Reality*. <https://doi.org/10.1007/s10055-023-00752-9>
- Van Lill, X., McColl, L. y Neale, M. (2023). Cross-national applicability of a game-based cognitive assessment. *International Journal of Selection and Assessment*, *31*, 302-320. <https://doi.org/10.1111/ijsa.12425>
- Wiernik, B. M., Raghavan, M., Caretta, T. R. y Coovert, M. D. (2022). Developing and validating a serious game-based assessment for cyber occupations in the US Air Force. *International Journal of Selection and Assessment*, *30*, 27-47. <https://doi.org/10.1111/ijsa.12378>
- 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*, 64-77. <https://doi.org/10.1080/1359432X.2019.1681401>
- Wu, F. Y., Mulfinger, E., Alexander, L., Sinclair, A. L., McCloy, R. A. y Oswald, F. L. (2022). Individual differences at play: An investigation into measuring Big Five personality facets with game-based assessments. *International Journal of Selection and Assessment*, *30*, 62-81. <https://doi.org/10.1111/ijsa.12360>