

TEST REVISADOS Y NUEVAS VERSIONES. ¿MERECE LA PENA CAMBIAR? UN EJEMPLO BASADO EN EL D2-R

Fernando Sánchez-Sánchez¹ y Belén Ruiz¹
¹Dpto. de I+D+i de Hogrefe TEA Ediciones

INTRODUCCIÓN Y OBJETIVOS

El uso de test estandarizados requiere una importante inversión en formación, entrenamiento y recursos (p. ej., materiales, equipos...). La aparición de nuevas versiones de los test —revisadas o renovadas— implica a menudo una renovación de estos esfuerzos, así como una adaptación de los protocolos de evaluación o de los criterios para la derivación consensuados en los centros.

Por este motivo, algunas cuestiones emergentes entre los profesionales son: ¿Hasta qué punto está justificado el cambio de un test cuando aparece una nueva versión de este? ¿Qué aspectos habría que tener en cuenta para tomar esta decisión? ¿Cuándo se puede justificar el mantenimiento de una versión obsoleta?

El objetivo del estudio fue analizar comparativamente algunos aspectos esenciales de la calidad de un test y de su versión revisada para proporcionar elementos de valoración para los profesionales.

Costes asociados al cambio de un test a una nueva versión

Inversión económica

Inversión en formación

Adaptación de protocolos

Pérdida de evidencia acumulada

MÉTODO

Se procedió a la comparación de la versión original del **d2**, *Test de Atención* publicada en español (Brickenkamp, 2002) y la versión revisada **d2-R**, *Test de Atención-Revisado* (Brickenkamp et al., 2022).

En concreto, se analizaron 4 apartados relevantes para el uso profesional de la prueba en diferentes aspectos:

- Aspectos formales que condicionan la evaluación y los resultados
- Aspectos relacionados con la calidad psicométrica de los instrumentos
- Aspectos relacionados con la calidad, variedad y actualización de los baremos
- Aspectos relacionados con la usabilidad y utilidad de los resultados

En cada apartado se describen las características de ambas versiones y la justificación o relevancia de las modificaciones introducidas.

RESULTADOS

Aspectos formales que condicionan la evaluación y los resultados

d2	Modificación	d2-R	Justificación
8 – 79 años	Rango de aplicación	6 – 80 años	El rango de aplicación se amplía para cubrir toda la etapa escolar
47 por línea	Aumento del n.º de estímulos	57 por línea	Mejora la precisión de la evaluación en los niveles altos y evita efecto techo en las puntuaciones.
20 segundos x línea 15 segundos x línea 7 minutos en total	Formatos de aplicación	20 segundos x línea 7 minutos en total	Se elimina la versión de 15 segundos porque ya no hay efecto techo.
TA, TR, E, E%, TOT, CON, VAR y SO	Depuración de puntuaciones	VT, E% y CON	Solo se mantienen las puntuaciones con propiedades óptimas y no redundantes.
Curva de trabajo visual a partir de TR	Nuevas puntuaciones	Curva de trabajo a partir de CON en cada bloque	Se sustituye la interpretación visual cualitativa por un sistema de contraste estadístico de la variación en el rendimiento de cada bloque.

Aspectos relacionados con la calidad psicométrica de los instrumentos

d2	Modificación	d2-R	Justificación
TA alfa = 0,97 O alfa = 0,94 C alfa = 0,98 TR/VT alfa = 0,95	Consistencia interna (alfa)	VT alfa = 0,98 E% alfa = 0,91 CON alfa = 0,98	Fiabilidad muy similar (excelente) en ambas versiones.
Ausencia de estudios test-retest	Estabilidad temporal (Test-retest)	VT r = 0,87 E% r = 0,66 CON r = 0,86	Estabilidad temporal excelente en las puntuaciones del d2-R. En consonancia con estudios internacionales.
Orden d2 → d2-R CON r = ,88 TR/VT r = ,85 E% r = ,80 N = 111	Equivalencia entre versiones	Orden d2-R → d2 CON r = ,86 TR/VT r = ,81 E% r = ,78 N = 142	La equivalencia entre las puntuaciones de las dos versiones es muy elevada después de controlar el efecto de la práctica. La equivalencia interpretativa de las puntuaciones permite aprovechar las evidencias recabadas con la versión anterior.

Aspectos relacionados con la calidad, variedad y actualización de los baremos

d2	Modificación	d2-R	Justificación
2002	Fecha	2022	Se constatan diferencias relevantes entre los baremos. Necesidad de actualización.
N = 1.541	Tamaño	N = 2.984	Muestra más amplia para cubrir adecuadamente todos los grupos de edad y los cruces de las variables de estratificación.
Muestreo incidental en función de la edad y sexo.	Composición y estratificación	Muestreo ajustado al censo en función de la edad, el sexo, la región y el nivel educativo.	Necesidad de un muestreo controlado y ajustado a censo de todo el rango de aplicación de la prueba, controlando nivel educativo en la muestra de adultos.
Baremos en grupos amplios 8-10 años (n = 217) 11-12 años (n = 115) 13-14 años (n = 214) 15-16 años (n = 157) 17-18 años (n = 105) 19-29 años (n = 485) 30-39 años (n = 136) 40-79 años (n = 112)	Tramos de comparación	• Tramos ajustados a las tendencias evolutivas. • Tramos de 1 año de 6 a 18 años. • de 5 años de 19 a 80 años. • Tamaño muestral de cada tramo: • n de 111 a 199.	Disponer de tramos más pequeños y ajustados a las tendencias evolutivas de la variable medida incrementan la precisión de las comparaciones y, por tanto, de los resultados.
Baremación tradicional	Técnica de baremación	Baremación continua (continuous norming no paramétrico)	La baremación continua aprovecha toda la potencia de la muestra para obtener tendencias más robustas y precisas y corregir el error muestral y los artefactos de los métodos tradicionales.

Aspectos relacionados con la usabilidad y utilidad de los resultados

d2	Modificación	d2-R	Justificación
Instrucciones básicas	Instrucciones y ejemplos	Instrucciones detalladas con ejemplos y entrenamiento	Las instrucciones originales resultaban complejas para los niños y las personas mayores.
Puntuación: manual, recuento de todas las marcas. Corrección: manual, sumatorios y uso de tablas.	Corrección	Puntuación: manual, recuento solo de dianas. Corrección: automatizada online.	El proceso de puntuación manual se ha simplificado para ahorrar tiempo y reducir errores. La corrección online realiza todo los cálculos, transformaciones y contrastes.
Hoja de resultados autocorregible. Resultados en Pc	Resultados	Perfil gráfico y comentario interpretativo de los resultados. Resultados en Pc y CI	Los resultados se presentan de una forma más clara y manejable. Las representaciones gráficas facilitan una interpretación más rápida y ajustada.
Normas de interpretación poco detalladas y complejas.	Interpretación	Capítulo de interpretación con información sobre cada variable e índice. Interpretación apoyada en gráficos y análisis de significación.	Normas de interpretación detalladas y apoyadas en la evidencias disponible sobre su uso en diferentes poblaciones. Se incluye automáticamente breve comentario de los resultados.
Información no disponible	Índice de cambio fiable (ICF) y estimación del error (ETM)	Se incluyen tablas y procedimiento para calcular ICF y ETM.	La estimación del ETM y del ICF —para valorar los cambios significativos entre las puntuaciones de una misma persona tras la intervención— son requeridos habitualmente para la toma de decisiones en la práctica profesional.

CONCLUSIONES

Diversas guías de buenas prácticas o códigos éticos profesionales explicitan que los profesionales de la psicología no deben basar sus decisiones de evaluación o intervención o sus recomendaciones en resultados de pruebas que estén obsoletas (i. e., Directriz 9.08 de los Principios Éticos y código de conducta de los Psicólogos de la APA). No obstante, en la actualidad no existe una referencia clara en las guías éticas ni en las guías de buenas prácticas acerca de cuándo resulta conveniente o necesario abandonar una versión de un test y pasar a una versión más actual. Por tanto, resulta necesario realizar un análisis detallado de los cambios introducidos en la nueva versión y en qué medida estos implican una mejora real en los aspectos sustantivos.

El tiempo transcurrido entre dos versiones de un mismo test (usualmente lustros o décadas, como en el caso del d2 y el d2-R), implica generalmente un avance paralelo en el conocimiento sobre la utilidad de la prueba en concreto, en el conocimiento teórico de los constructos que pretende evaluar y en el desarrollo de las técnicas y procedimientos psicométricos disponibles (p. ej., TRI, SEM, Continuos norming...). En la medida que las nuevas versiones incorporen estos avances en medición se producirá una mejora significativa en la calidad científica la evaluación frente a las versiones previas y se justificará su adopción por parte de los profesionales.

Referencias bibliográficas
American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. American psychologist, 57(12), 1060–1073.
Brickenkamp, R. (2002). d2-R. Test de Atención (N. Seisdedos, adaptador). TEA Ediciones.
Brickenkamp, R., Schmidt-Atzert, L. y Liepmann, D. (2022). d2-R. Test de Atención – Revisado (B. Ruiz-Fernández, adaptadora). Hogrefe TEA Ediciones.