

Test de Atención d2: Consistencia interna, estabilidad temporal y evidencias de validez

The d2 Test of Attention: Internal Consistency, Temporal Stability and Evidence of Validity

Josiane Pawlowski

Universidad de Iberoamérica, Costa Rica

Resumen

El objetivo de esta investigación fue analizar la adecuación psicométrica del Test d2 para su uso en Costa Rica. Una muestra de 445 individuos con edades entre 18 y 47 años, 58% mujeres y 42% hombres, contestaron la prueba d2; 50 de ellos completaron la prueba una segunda vez y 40 respondieron también las pruebas Test del Trazo, Test de Colores y Palabras de Stroop, Span de Dígitos, Búsqueda de Símbolos, Claves y Cancelación. Fueron evaluadas la consistencia interna, la estabilidad temporal y las evidencias de validez basadas en relaciones con otras variables y en la estructura interna. Los puntajes más fiables fueron los siguientes: total de respuestas (TR), total de aciertos (TA), índice de efectividad en la prueba (TOT) e índice de concentración (CON), con resultados del alfa de Cronbach y de fiabilidad test-retest en el rango de .90 a .97. Las correlaciones con pruebas que evalúan atención sostenida y selectiva, velocidad de procesamiento, precisión y control inhibitorio sugieren evidencias de validez para los puntajes CON, TR, TA y O (errores de omisión). Además, el aumento de edad está relacionado a la disminución de los puntajes TA y CON, en especial. Se discuten los hallazgos respecto a la estructura factorial de la prueba. Los resultados corroboran estudios previos e indican puntajes válidos del d2 para evaluar individuos de nivel universitario de Costa Rica.

Palabras clave: Atención sostenida, Atención selectiva, Test d2, validez, fiabilidad

Josiane Pawlowski, Universidad de Iberoamérica, Costa Rica.

La correspondencia en relación con este artículo se dirige a Josiane Pawlowski. Facultad de Psicología, Universidad de Iberoamérica, Tibás, San José, Costa Rica, apartado postal 11303. Tel.: +506 2297-2242 ext. 2209, dirección electrónica: josiane.pawlowski@unibe.academy



J. PAWLOWSKI

Abstract:

In this research, some psychometric properties of the d2 test were analyzed for its use in Costa Rica. A total of 445 participants aged between 18 and 47 years, 58% women and 42% men, completed the d2 test; a subset of 50 participants completed the test a second time, and 40 participants answered the additional tests: Trail Making Test, Stroop Test, Digit Span, Symbol Search, Coding and Cancellation. Internal consistency, temporal stability and validity evidences based on relations to other variables and based on the internal structure were evaluated. The most reliable scores were total number of responses (TR), total number of correct characters canceled (TA), effectiveness index (TOT) and concentration index (CON), with Cronbach's Alpha results and test-retest reliability ranged from .90 to .97. Correlations with tests that evaluate sustained and selective attention, processing speed, accuracy, and inhibitory control suggest validity evidence for the CON, TR, TA, and O (omission errors) scores. In addition, aging is especially related to the decrease in TA and CON scores. The findings regarding the factorial structure of the test are discussed. The results corroborate previous studies and indicate validity evidence for d2 test scores for evaluating Costa Rican university students and graduates.

Keywords: Sustained Attention, Selective Attention, d2 Test, Validity, Reliability

La atención es una función neurocognitiva básica y necesaria para llevar a cabo toda actividad de tipo comportamental o cognitiva, que permite seleccionar, priorizar, procesar y supervisar la información (Lezak et al., 2012). Se entiende como un sistema funcional de filtrado de información, de naturaleza compleja, multimodal, jerárquica y dinámica (García & Portellano, 2014), por lo que resulta un concepto difícil de definir (Rosenberg et al. 2017). Por ende, una multiplicidad de definiciones puede ser encontrada. Ello se relaciona con la cantidad de modelos de atención producidos en las últimas décadas, de los que se destacan dos, el de Mirsky y Duncan y el de Posner y Petersen, en virtud de que ofrecen datos sólidos que los apoyan (Fernández, 2014). Frecuentemente, los modelos distinguen entre subtipos de atención, con énfasis a los clásicos, que separan las funciones de focalización, sostenimiento y cambio (Fortenbaugh et al., 2017).

Tal como los modelos de atención involucran diferentes subtipos y son necesarias distintas pruebas para que se logren evaluar todos los componentes o funciones de la atención, también es difícil encontrar una prueba que mida solamente la atención (Ardila & Rosselli, 2007). Además de otras funciones cognitivas, la subprueba Dígitos de la Escala Wechsler de Inteligencia (Wechsler, 2014) evalúa la amplitud de la atención, el Test de Símbolos y Dígitos o *Symbol Digit Modalities Test* (Smith, 2002) mide la atención dividida, el Test del Trazo o *Trail Making Test* (Army Individual Test Battery, 1944) evalúa la atención visual selectiva, el Test de Colores y Palabras de Stroop (Stroop, 1935) mide la atención selectiva y la inhibición de la atención, el Continuous Performance Test (Rosvold et al., 1956) evalúa la atención sostenida o concentración y el Test de Atención d2 (Seisdedos, 2002) mide tanto la atención sostenida como la atención selectiva. Para desempeñarse adecuadamente en cualquier prueba de atención, concentración o rastreo, es necesaria la atención sostenida (Lezak et al., 2012).

Para la medida de la atención sostenida y selectiva, ambos componentes de velocidad y precisión son tomados en cuenta en el sistema de puntuación del Test de Atención d2 en su versión de papel y

TEST DE ATENCIÓN d2

lápiz (Baghaei et al., 2019). Como una prueba de cancelación, las personas evaluadas deben tachar las variables relevantes entre estímulos similares que no deben ser seleccionados. Las variables relevantes son las letras d con dos rayitas, una encima y otra por debajo, o dos por encima o por debajo de la letra, las cuales se intercalan aleatoriamente entre otros caracteres irrelevantes constituidos por letras d o p con una, tres o cuatro rayitas. Estos caracteres se presentan en 14 líneas consecutivas. A cada una de las líneas se le asigna el límite de tiempo de 20 segundos sin pausa entre ellas. Debido a que la prueba d2 está cronometrada y requiere centrarse en los estímulos objetivo entre los irrelevantes, ella también se considera una medida de velocidad, precisión del rastreo y atención selectiva (Baghaei et al., 2019).

Al considerar estas características de la prueba, los dos principales constructos evaluados por el Test d2 son la atención sostenida y la atención selectiva. La atención sostenida se refiere a una capacidad para permanecer en tareas largas, aburridas, repetitivas y/o caracterizadas por largos intervalos entre eventos relevantes (Grahn & Manly, 2012). Ella se diferencia de otras clases de atención por la duración de la actividad requerida; es decir, la atención sostenida requiere que un observador mantenga una participación en una tarea específica durante un período prolongado de tiempo (Fortenbaugh et al., 2017), mientras la atención selectiva permite a los individuos extraer información relevante mientras ignoran los estímulos distractores (Gomez-Ramirez et al., 2016). Por lo tanto, muchos instrumentos que evalúan la atención, también miden otros constructos relacionados, tal como el test de Atención d2, que posibilita la evaluación de la velocidad de procesamiento, del control atencional y de la precisión (Seisdedos, 2002).

Respecto a las características psicométricas de la prueba, además de los estudios conducidos en Alemania con la versión original, Zilmer y Kennedy (1999) investigaron, en una muestra de 506 estudiantes universitarios estadounidenses de 18 a 32 años, la validez de constructo del test d2 basada en la relación con las pruebas *Stroop Test*, *Trail Making Test* y *Symbol Digit Modalities Test*. Los principales resultados indicaron que los índices de efectividad en la prueba y el índice de concentración correlacionaron de manera significativa con todas las otras medidas de atención evaluadas. Bates y Lemay Jr. (2004) estudiaron la consistencia interna e indicadores de validez del d2 en una muestra de 364 adultos estadounidenses. Los autores señalaron que la prueba es una medida válida de la precisión y de la velocidad de la exploración visual. Además, fueron realizados estudios de fiabilidad y validez en España (Seisdedos, 2002) y Brasil (Centro Editor de Testes e Pesquisa em Psicologia, 2002), los cuales presentaron resultados de correlaciones entre los puntajes de la prueba e indicaron adecuados índices de fiabilidad, en especial para los puntajes de velocidad o rapidez (TR) y precisión (TA).

Sobre la estructura factorial del d2, en la versión española del manual de la prueba, que presenta datos para adolescentes y adultos (Seisdedos, 2002), y en el estudio de Izquierdo et al. (2007), conducido con niños, se efectuó el análisis de componentes principales (ACP) con la rotación oblicua. Los autores lo presentaron como método de análisis de la estructura factorial del test. Sin embargo, los investigadores en análisis factorial (AF) no están de acuerdo con el hecho de que el ACP sea un método de AF, pues consideran que no explique la correlación entre las variables, debido a que los componentes se calculan al utilizar toda la varianza de las variables manifiestas y el análisis es computado sin relación con cualquier estructura subyacente causada por las variables latentes (Brown, 2006; Costello & Osborne, 2005; Martínez-Arias et al, 2006). El método de ACP se considera adecuado cuando el objetivo del

investigador consiste en la reducción de una cantidad grande de variables a un número más fácilmente manejable sin buscar un modelo que explique la correlación entre variables observadas.

Respecto a los resultados, los dos estudios coinciden en el número de componentes; sin embargo, no están de acuerdo con la distribución de las variables por componente. Mientras la versión española del manual presenta los puntajes TR (velocidad) y O (error de tipo omisión) en el componente 1, los puntajes C (error de tipo comisión) en el componente 2 y el puntaje TA (precisión) en el componente 3, el estudio de Izquierdo et al. (2007) incluye las variables TR y TA en el componente 1 y las variables O y C en los componentes 2 y 3, respectivamente. Tal como los anteriores, el estudio de Bates y Lemay Jr. (2004) aplicó el análisis de componentes principales con la rotación Promax, pero otras medidas de desempeño cognitivo fueron inseridas en los análisis, además de los puntajes del test d2. Es importante recalcar que, según Lloret-Segura et al. (2014), los componentes no son variables latentes, tampoco los ítems constituyen una medida indirecta de ellos. Así, no es correcto hablar de factores o de variables latentes en un análisis de componentes principales, pero puede ser útil para trabajar con sumas de variables que expresen contenidos semejantes. Por consiguiente, estos estudios no presentan un modelo teórico factorial definido para la prueba d2.

Otros estudios que pretenden investigar una estructura factorial examinaron muestras en las que se aplicó una batería de instrumentos, tal como la investigación de Zillmer y Kennedy (1999), en que los autores realizaron un análisis factorial exploratorio, aplicando el método de extracción de máxima verosimilitud y rotación oblicua. El estudio reveló una estructura de dos factores denominados atención selectiva y flexibilidad mental, en los cuales los puntajes de velocidad (TR) y concentración (CON) del d2 estaban en el factor 1, mientras los errores cargaban en el segundo factor. Ambos factores también incluyeron otras pruebas. Se destaca mencionar que “aunque existe una comprensión común de las características esenciales de las pruebas de atención sostenida, hasta la fecha, no existe un modelo formal o un examen sistemático de los subcomponentes que pueden impulsar el rendimiento en estas pruebas” (Blotenberg & Schmidt-Atzert, 2019, p. 02). Los autores proponen e investigan un modelo de proceso de pruebas de atención sostenida al aplicar diferentes pruebas que miden esta función cognitiva y, mediante análisis de ecuaciones estructurales, demuestran que dos subcomponentes principales, denominados velocidad de la operación mental y perceptual explican una gran cantidad de la varianza en las pruebas de atención, con una tendencia a poca influencia de la velocidad motora en el desempeño. Aunque su estudio es considerado un importante paso para la comprensión de los subcomponentes de la atención sostenida, los datos investigados no confirman todo el modelo propuesto por los autores.

Si bien algunos estudios sugieren las posibilidades de dos o tres factores o componentes para el test d2, ante la incertidumbre respecto al modelo factorial de la prueba y la carencia de un modelo teórico explicativo de las pruebas de atención sostenida, algunas investigaciones más actuales buscan verificar la validez de determinados puntajes del d2. Steinborn et al. (2018) examinaron los efectos de pruebas repetidas sobre la velocidad y la precisión del rendimiento (errores de omisión y comisión) y verificaron alta confiabilidad para la medida de velocidad, que se conserva también con la mitad de la longitud de la prueba. Las puntuaciones de variabilidad y error presentaron más problemas con respecto a la confiabilidad en su investigación, en especial el puntaje de comisión, pero menos para los errores de omisión. En conclusión, los autores recomiendan a los investigadores y los profesionales que elijan el puntaje de

TEST DE ATENCIÓN d2

velocidad para una evaluación altamente confiable y que los puntajes de error se utilicen como una medida secundaria, mientras que los de variabilidad no deberían ser utilizados por razones de asimetría de la distribución de la variable e inestabilidad test-retest.

Otro estudio actual de relevancia en la determinación de la validez de puntajes de la prueba d2 fue conducido por Baghaei et al. (2019). La investigación buscó analizar el ajuste del test d2 a los modelos de teoría de respuesta al ítem (TRI). Según los autores, estos modelos de TRI buscan “modelar la relación entre el rendimiento de un examinado en un elemento y la ubicación general del examinado en el rasgo latente” (Baghaei et al., 2019, p.72). Mediante el examen del ajuste de seis técnicas de puntuación diferentes, Baghaei et al. (2019) evaluaron el ajuste de la prueba d2 al modelo *Rasch Poisson Counts* (RPCM) y encontraron que solamente dos técnicas de puntuación, los escores del índice de concentración y el número total de caracteres cancelados, se adaptan al RPCM. Además, el modelo no se ajustaba cuando se aplicaba a los errores de omisión, errores de comisión, errores totales y total procesado correctamente, hallazgos que son consistentes con los resultados reportados por Steinborn et al. (2018), quienes recomiendan utilizar el puntaje de velocidad y el puntaje de velocidad corregido por error para una evaluación confiable, mientras los puntajes de error podrían ser utilizados como una medida secundaria.

En lo que concierne a la utilidad de la prueba d2 en áreas específicas, en el ámbito educativo, algunas investigaciones buscan evaluar el comportamiento de puntajes específicos a lo largo del desarrollo y ofrecer baremos para diferentes poblaciones. En la evaluación de una muestra de escolares de segundo a sexto grado, con edad entre siete y trece años, Wassenberg et al. (2008) encontraron que la velocidad de procesamiento mejora hasta el sexto grado, la impulsividad, medida por el porcentaje de errores de comisión, reduce hasta el cuarto grado y la falta de atención, medida por el porcentaje de errores de omisión, se presentó estable en todos los grados. Estos resultados sugieren que la atención selectiva continúa por desarrollarse, al menos, hasta el final de la escuela primaria. Jiménez et al. (2012) también analizaron el desarrollo evolutivo de la atención en escolares de educación primaria. Señalaron cambios significativos en las variables de productividad y efectividad en esta muestra y obtuvieron datos normativos de la prueba d2 para escolares de edades entre los 6 años y los 12 años. Rivera et al. (2017) generaron datos normativos para población infantil de España y de nueve países de América Latina (Chile, Cuba, Ecuador, Guatemala, Honduras, México, Paraguay, Perú y Puerto Rico).

En la actividad clínica, el test de atención d2 puede ser utilizado en evaluaciones neuropsicológicas de pacientes con diferentes edades siempre que existan evidencias de validez y datos normativos para tal población. Por ejemplo, Zancada-Menéndez et al. (2013) aplicaron el test d2 para evaluar la atención sostenida o la concentración en una muestra de individuos de 66 a 96 años, que presentaba deterioro cognitivo leve y demencia. La prueba también puede contribuir al estudio de consecuencias de enfermedades autoinmunes, tal como la investigación de Leyhe et al. (2013) con pacientes que presentaban Tiroiditis de Hashimoto, en que se encuentra, mediante evaluación con la prueba d2, que la disminución de la atención se asocia con la reducción de la densidad de la sustancia gris en el lóbulo frontal.

Las características psicométricas de una prueba psicológica deben ser estudiadas en el país donde se pretende utilizarla, así como debe ser realizada la producción de estándares en el país, porque, al existir diferentes oportunidades educativas y aspectos distintos de la propia cultura, pueden ser generadas

interpretaciones equivocadas si se utilizan las normas estandarizadas en otros países. Por esta razón, el objetivo de este estudio fue investigar la fiabilidad y las evidencias de validez del Test de Atención d2 en adultos de nivel universitario de Costa Rica. Más específicamente, se examinó la consistencia interna, la estabilidad temporal y las evidencias de validez basadas en relaciones con otras variables y en la estructura interna (American Educational Research Association et al. [AERA] et al., 2018). Como la prueba d2 puede ser útil en las áreas educativa y neuropsicológica, la investigación sobre su adecuación psicométrica a una muestra costarricense puede contribuir futuramente a la evaluación cognitiva en estas áreas en Costa Rica, las cuales dependen de instrumentos de medida válidos y confiables.

Método

Tipo de estudio

La presente investigación corresponde a un estudio de tipo instrumental (Ato et al., 2013), debido a que analiza las propiedades psicométricas de la prueba d2.

Participantes

Para la recolección de los datos, se utilizó un muestreo no probabilístico por conveniencia. Se define como un procedimiento cuantitativo en que los participantes son seleccionados porque están dispuestos y disponibles para ser estudiados (Creswell, 2008), los cuales se invitaron a participar en centros educativos de nivel universitario, mediante explicación de la investigación en diferentes aulas de carreras distintas de dos universidades públicas y una privada, en las que fue facilitado y permitido el ingreso de la investigadora responsable. Fueron evaluados 445 participantes con edad entre 18 y 47 años, 58% mujeres y 42% hombres, 92.3% eran costarricenses y los demás, nacionalizados o residentes en Costa Rica, provenientes de 17 diferentes países. Los participantes eran estudiantes universitarios o graduados: 54.7% correspondían a Psicología, 28.9%, Física, Meteorología o Química, 11.2% de Ingeniería y 5.2% de otras carreras. El tiempo de escolaridad de la muestra evaluada, en media de años de estudio formal, fue de 14.55 ($DE = 2.40$). Fue criterio de inclusión de la muestra haber nacido en Costa Rica, ser nacionalizado o residente, por lo mínimo, hace cinco años en el país. Fueron excluidos cinco participantes, quienes presentaron un resultado de valor extremo en las variables básicas del test d2, TR, TA, O y C, que podrían afectar el cálculo de las correlaciones. Para la detección de puntuaciones extremas, se calcularon las puntuaciones típicas de cada variable y se consideraron como potenciales casos atípicos aquellos que presentaban puntajes Z fuera del rango ± 3 (Tabachnick & Fidell, 2007).

Instrumentos

Test de Atención d2, versión española de Seisdedos (2002). Es un test de tiempo limitado para medir la atención sostenida, la atención selectiva y la concentración mental. Se puede administrar de forma individual o colectiva con una duración entre 8 y 10 minutos. La prueba presenta 14 líneas con 47 estímulos (letras “d” o “p” acompañadas de rayitas). La tarea consiste en revisar cada línea y marcar toda letra “d” acompañada de dos rayitas. Las puntuaciones resultantes del test d2 son presentadas en la Tabla 1. Según la versión española del test (Seisdedos, 2002), en la mayoría de los estudios, las puntuaciones son muy fiables ($r > .90$).

TEST DE ATENCIÓN d2

Tabla 1
Puntuaciones resultantes del Test de Atención d2

Puntuación	Descripción	¿Qué mide?
TR: Total de Respuestas	Número de elementos intentados en las 14 líneas, tanto los relevantes como los irrelevantes	Velocidad de procesamiento, cantidad de trabajo realizado y motivación
TA: Total de Aciertos	Número de elementos relevantes correctos	Cantidad de trabajo y precisión del procesamiento
O: Omisiones	Número de elementos relevantes intentados, pero no marcados	Control atencional, cumplimiento de una regla, precisión de la búsqueda visual y calidad de la actuación
C: Comisiones	Número de elementos irrelevantes marcados	Control inhibitorio, cumplimiento de una regla, precisión del procesamiento, minuciosidad y flexibilidad cognitiva
TOT: Efectividad total en la prueba	Número de elementos procesados menos el número total de errores cometidos: TR – (O+C)	Control atencional e inhibitorio y relación entre la velocidad y la precisión
CON: Índice de concentración	Número de elementos relevantes marcados menos el número de comisiones: TA-C	Índice de concentración o del equilibrio entre velocidad y precisión de la actuación
VAR: Índice de variación	Diferencia entre la mayor y la menor productividad: (TR+) – (TR-)ª	Estabilidad y consistencia en el tiempo, variación o fluctuación en el modo de trabajar

Nota. Basado en la versión Española del Manual de la prueba traducida al castellano por Seisdedos (2002).

ª TR+ = Línea con mayor número de elementos intentados y TR- = Línea con menor número de elementos intentados.

Parte A (TMT-A) y Parte B (TMT-B) del Test del Trazo o Trail Making Test (Army Individual Test Battery, 1944). TMT-A evalúa atención, exploración visual, velocidad de la coordinación visomotora y procesamiento de la información (Cavaco et al., 2013). También mide memoria de trabajo, funciones ejecutivas y control ejecutivo (Sánchez-Cubillo et al., 2009). La administración de las Partes A y B fue realizada de acuerdo con las directrices presentadas por Strauss et al. (2006). El participante debe conectar con una línea números en orden ascendente en la Parte A y debe conectar alternativamente números y letras en orden alfabético y ascendente en la Parte B. El tiempo total en segundos fue registrado para las partes A y B, en representación de las puntuaciones directas TMT-A y TMT-B. En el estudio de Wagner et al. (2011), la confiabilidad por el método test-retest del TMT-A y B fue entre .76 y .89 y .86 y .94, respectivamente.

Test de Colores y Palabras de Stroop (Stroop, 1935). Valora aspectos de atención selectiva, velocidad de procesamiento y capacidad de inhibir una respuesta automática a favor de otra menos habitual (Stroop, 1935). Se utilizó la adaptación española del test de Stroop (Golden, 1994), que consta de tres páginas de cinco columnas con 20 ítems cada columna. La primera página contiene las palabras “azul,

rojo y verde” impresas en tinta negra, la segunda está formada por filas de X impresas en los colores rojo, verde o azul y la tercera incluye nombres de estos tres colores impresos en un color distinto al que corresponde a la palabra. Durante 45 segundos para cada página, el evaluado debe leer lo más rápidamente posible las palabras en la primera, denominar colores en la segunda y designar colores al inhibir la lectura de palabras en la tercera página. Se obtiene la suma del número de palabras o colores en cada página y se calcula una puntuación de resistencia a la interferencia. Sobre la fiabilidad temporal del test, Golden (1994) obtuvo valores de 0.73 o más tanto en la versión colectiva como en la individual; asimismo, Rodríguez et al. (2016) también encontraron confiabilidad superior a .70 en todas las láminas del Stroop.

Subpruebas de la Escala de Inteligencia para Adultos – WAIS-IV Español, versión Mexicana (Wechsler, 2014):

Retención de Dígitos. Evalúa la atención y la resistencia a la distracción, la memoria auditiva inmediata y la memoria de trabajo. Está formada por tres tareas: dígitos directos (se lee una secuencia de números que el evaluado debe repetir en el mismo orden que se presentan), dígitos inversos (repetir una serie de dígitos en orden inverso al presentado) y dígitos secuenciación (repetir los dígitos en orden ascendente).

Búsqueda de símbolos. Evalúa velocidad de la búsqueda y discriminación visual, precisión perceptiva, rapidez del procesamiento mental, memoria visual a corto plazo y la coordinación visomotora. El test presenta dos grupos de símbolos: un grupo objetivo y un grupo de búsqueda. La persona examina dos símbolos en el primer grupo y debe decidir si alguno de ellos está presente o no dentro del otro grupo. El tiempo límite es de 120 segundos.

Claves. Evalúa la atención sostenida, la rapidez y destreza visomotora y la capacidad de aprendizaje asociativo. El evaluado mira una tabla con números y símbolos emparejados. La tarea consiste en copiar en otra tabla, dentro de un tiempo límite de 120 segundos, los símbolos correspondientes a cada número.

Cancelación. Mide la vigilancia, la atención sostenida, la velocidad de procesamiento visual y la coordinación visomotora. La tarea consiste en examinar una planilla con figuras geométricas y tachar, dentro de un tiempo límite de 45 segundos, aquellas que coincidan con el criterio de búsqueda, o sea, de la misma forma y el color que las proporcionadas como ejemplos.

Respecto a las evidencias de consistencia interna y confiabilidad, el manual de la versión mexicana reporta el coeficiente alfa de Cronbach para las diferentes subpruebas evaluadas. Se encontraron los resultados de .87, .81, .86 y .78 para Retención de dígitos, Búsqueda de símbolos, Claves y Cancelación, respectivamente.

Procedimiento

Se realizaron evaluaciones colectivas con la duración de 30 minutos e individuales con un tiempo de una hora por participante, lo que dependía del objetivo planteado en la presente investigación, estudio de confiabilidad o de validez convergente, respectivamente. La recolección fue conducida por la investigadora principal, psicóloga y doctora en Psicología, certificada por el CONIS (Consejo Nacional de Investigación en Salud) de Costa Rica. Todas las evaluaciones fueron realizadas en ambiente adecuado al propósito de la investigación, aulas de las universidades o espacios apropiados en centros de atención psicológica o de investigación. Todos los participantes firmaron un consentimiento informado.

TEST DE ATENCIÓN d2

La investigación se realizó bajo las disposiciones señaladas en el Código de Ética y Deontológico del Colegio de Profesionales en Psicología de Costa Rica (Colegio de Profesionales en Psicología de Costa Rica, 2019). Para el estudio de estabilidad temporal, una muestra de cincuenta participantes, seleccionados por la técnica de conveniencia, fue evaluada en un tiempo no inferior a un mes y no superior a cuatro meses. Dicho criterio se estableció para evitar los efectos de recuerdo por el aprendizaje o de cambios por el tiempo prolongado hasta la segunda evaluación (Cohen & Swerdlik, 2006). Para los análisis de correlación con otras pruebas, se utilizó el programa G*power (Faul et al., 2009) para estimar si el tamaño muestral de 40 participantes sería suficiente para la finalidad y los objetivos del estudio, lo que se verificó como adecuado.

Análisis de datos

El análisis de datos se realizó con el SPSS 23.0 (*Statistical Package for Social Sciences*). La normalidad de las variables se evaluó al utilizar como parámetro los resultados de asimetría y curtosis y los resultados de la prueba de Kolmogorov-Smirnov. Para evaluar la consistencia interna, la estabilidad temporal y la evidencia de validez basada en la relación con otras variables, se analizaron los resultados de los puntajes TR, TA, TOT, CON, VAR, O y C de la prueba d2 mediante análisis descriptivos, α de Cronbach, correlación de Spearman o Pearson dependiendo de la distribución de las variables y prueba *t* de Student para muestras independientes. La estructura interna del test d2 fue evaluada mediante el Análisis Factorial Confirmatorio con el Método de Máxima Verosimilitud del programa estadístico AMOS 26.0. Como indicadores de ajuste de los modelos evaluados, fueron verificados el chi-cuadrado (χ^2), la raíz del cuadrado medio del error de aproximación (RMSEA), el índice de ajuste comparativo (CFI) y el índice de bondad de ajuste (GFI), en que $RMSEA < .08$, CFI y GFI $> .90$ son criterios para un ajuste aceptable del modelo (Hu & Bentler, 1999) y $RMSEA < .05$, CFI y GFI $> .95$ para un buen ajuste (Schreiber et al., 2006).

Resultados

Fiabilidad

Para la evaluación de la consistencia interna de los datos en la muestra total, los cuatro puntajes básicos obtenidos en la prueba d2 (TR, TA, O y C) fueron sometidos al procedimiento de α de Cronbach y de dos mitades con la finalidad de comparación a los resultados presentados en la estandarización española del test d2. El procedimiento α de Cronbach resultó en .96 para la variable TR, .97 para TA, .93 para O y .68 para C. Para el procedimiento de dos mitades, se obtuvieron las puntuaciones totales en la suma de las 7 filas pares y las puntuaciones en la suma de las 7 filas impares. Las variables TA y TR presentaban distribución normal, pero O y C no, debido a que se realizó la correlación de Pearson entre las mitades para los puntajes TA y TR, mientras para O y C se aplicó la correlación de Spearman y al resultado se le aplicó la corrección de Spearman-Brown para estimar la fiabilidad de una prueba doble (14 filas). El resultado de la correlación de Pearson para las dos mitades fue: .92 para TR y .94 para TA y el resultado de la correlación de Spearman fue de .83 para O y .44 para C. Al aplicar la corrección de Spearman-Brown para estimar la fiabilidad de una prueba doble se obtuvieron los resultados de .96 para TR, .97 para TA, .91 para O y .61 para C.

Estabilidad temporal

La estabilidad temporal se analizó por el método del *test-retest* en una muestra de 50 estudiantes universitarios, 25 mujeres y 25 hombres, con media de 23.16 años de edad ($DE = 5.0$) y 14.02 años de estudio ($DE = 1.32$). La segunda evaluación con la prueba d2 se efectuó después del intervalo de tiempo medio aproximado de dos meses ($M = 1.88$; $DE = .75$) de la primera evaluación. El análisis de normalidad indicó que las puntuaciones TR, TA, TOT y CON resultantes de la primera y la segunda evaluaciones presentaban distribución normal, mientras las puntuaciones de los errores O y C no la presentaban. El resultado de la correlación de Pearson fue de .91 para TA, .90 para TR, TOT y CON y de la correlación de Spearman fue de .68 para O y .56 para C, todos a $p < .01$.

Evidencias de validez

Basada en relaciones con otras variables

Se evaluó una muestra de 40 adultos de 20 a 47 años, 55% mujeres, universitarios o licenciados con tiempo de estudio formal de 12 a 23 años ($M = 17.12$; $DE = 2.98$). Los puntajes TR, TA, O, C, TOT, CON y VAR fueron correlacionados a diferentes medidas de las pruebas Test del Trazo (TMT-A y TMT-B), Stroop, Dígitos, Búsqueda de símbolos, Claves y Cancelación. Para esta muestra, se efectuaron correlaciones de Spearman, porque las variables estudiadas, en su mayoría, no presentaron distribución normal. Los resultados pueden ser observados en la Tabla 2. Las correlaciones más altas y positivas fueron observadas al comparar la prueba Claves con las variables TR, TA, TOT y CON. La tarea de dígitos orden directa presentó correlaciones moderadas con los puntajes TA y CON. Las variables CON y TA se correlacionaron negativamente con la prueba TMT-A, en especial. Aunque más baja, es importante mencionar la correlación negativa de la variable O con la prueba Claves y tres de las medidas del test Stroop. La única correlación del puntaje de errores C fue con la prueba Búsqueda de símbolos y el puntaje VAR presentó correlación negativa con la prueba Cancelación.

TEST DE ATENCIÓN d2

Tabla 2

Correlaciones entre puntajes en la prueba d2 y de las pruebas cognitivas Teste del Trazo (TMT-A y B), Stroop, Dígitos, Búsqueda de Símbolos, Claves y Cancelación

Puntajes d2 ^a	TR	TA	O	C	TOT	CON	VAR
Pruebas cognitivas							
Teste del Trazo (TMT-A)	-.42**	-.55**	.26	-.003	-.50**	-.57**	.27
Teste del Trazo (TMT-B)	-.39*	-.46**	.26	.126	-.46**	-.48**	.28
Dígitos Orden Directa	.40*	.51**	-.16	-.003	.48**	.52**	-.20
Dígitos Orden Inversa	.30	.35*	-.15	-.013	.35*	.35*	-.05
Dígitos Secuenciación	.29	.43**	-.26	.039	.38*	.44*	-.05
Stroop Palabra	.29	.28	.05	.029	.25	.26	-.01
Stroop Cor	.27	.46**	-.34*	-.28	.37*	.48**	-.30
Stroop Palabra-Cor	.35*	.49**	-.35*	-.06	.44**	.50**	-.16
Stroop Interferencia	.21	.32*	-.33*	.08	.28	.32*	-.03
Búsqueda de símbolos	.39*	.41**	.01	.33*	.41**	.41*	-.31*
Claves	.58**	.63**	-.32*	.16	.59**	.62**	-.25
Cancelación	.51**	.47**	-.05	.14	.51**	.48**	-.35*

Nota. TR: Total de Respuestas, TA: Total de Aciertos, O: Omisiones, C: Comisiones, TOT: Efectividad total en la prueba, CON: Índice de concentración y VAR: Índice de variación.

^a Puntajes en la prueba d2

* $p < .05$. ** $p < .01$.

Para verificar la evidencia de validez relacionada a los cambios con la edad, la muestra fue dividida en dos grupos extremos en relación con esta variable: 18-19 años ($n = 75$) y 29-47 años ($n = 75$). La prueba t de Student indicó diferencia significativa entre los grupos en las medidas de TR, TA, TOT, CON y O, y tamaño de efecto medio para los puntajes TA y CON, tal como se puede observar en la Tabla 3. La diferencia entre los grupos para los errores C no se consideró por el incumplimiento del supuesto de homocedasticidad (Test de Levene con $p < .05$).

J. PAWLOWSKI

Tabla 3

Diferencias entre los grupos etarios en los puntajes del test d2

Puntajes del test d2	18-18 años		29-47 años		<i>t</i> (148)	<i>p</i>	<i>d</i> de Cohen
	<i>M</i>	<i>DE</i>	<i>M</i>	<i>DE</i>			
TR	504.427	67.167	480.947	74.529	2.027	.044	0.331
TA	194.240	34.023	173.627	31.094	3.873	< .001	0.632
O	21.600	19.176	30.680	23.660	-2.582	.011	-0.422
TOT	479.400	65.677	448.440	66.400	2.871	.005	0.469
CON	190.813	34.506	171.800	30.886	3.556	< .001	0.581
VAR	14.613	4.926	13.093	4.880	1.898	.060	0.310

Basada en la estructura interna

Tres modelos hipotéticos fueron evaluados mediante AFC, fundamentados en: (a) los resultados de análisis factoriales de otras investigaciones que sugieren la estructura de dos factores para la prueba, (b) los resultados de otras investigaciones que indicaron la confiabilidad de las medidas de velocidad y del índice de concentración del test d2 y, (c) la existencia de un factor en estudios previos compuesto por los dos errores evaluados en la prueba. Los modelos hipotéticos analizados fueron los siguientes: (a) Modelo 1: Factor 1, conformado por la variable del índice de concentración ($CON = TA - C$), que es el número de elementos relevantes marcados en cada línea menos el número de elementos irrelevantes marcados, y el Factor 2, constituido por los errores de tipo O (número de elementos relevantes que no fueron marcados); (b) Modelo 2: Factor 1, que inclui las variables TR (total de respuestas) y Factor 2, conformado por los errores de tipo O; Modelo 3: Factor 1, con las variables TA, que son los caracteres cancelados correctamente, y Factor 2 con los errores de tipo O; y Modelo 4: Factor 1 con las variables CON y Factor 2 con los errores tipo O y tipo C. Los resultados encontrados para cada modelo evaluado son presentados en la Tabla 4, en que se observa, según los criterios de ajuste, la adecuación aceptable del Modelo 1 a la estructura factorial propuesta.

TEST DE ATENCIÓN d2

Tabla 4

Índices de Ajuste de los Análisis Factoriales Confirmatorios para los Modelos Hipotéticos

Modelos hipotéticos	Índices de ajuste						
	χ^2	gl	p^a	χ^2/gl	RMSEA	GFI	CFI
Modelo 1	1017,12	349	< .001	2.91	.066	0,86	0,93
Modelo 2	2812,20	349	< .001	8.06	.126	0,70	0,76
Modelo 3	1137,48	349	< .001	3.26	.071	0,84	0,92
Modelo 4	3314,87	818	< .001	4.05	.08	0,74	0,77

Nota. RMSEA = raíz del cuadrado medio del error de aproximación; CFI = índice de ajuste comparativo; índice de bondad de ajuste (GFI). Valores de RMSEA < .08, CFI y GFI > .90 son criterios para un ajuste aceptable del modelo (Hu & Bentler, 1999) y de χ^2/gl < 3, RMSEA < .05, CFI y GFI > .95 indican buena adecuación del ajuste (Schreiber et al., 2006).

^a Para χ^2 significativo, utilízase la fórmula corregida χ^2/gl .

Discusión

El estudio buscó presentar características psicométricas de la prueba d2 en una muestra de estudiantes universitarios y graduados residentes en Costa Rica. Se evaluaron la consistencia interna, la estabilidad temporal y las evidencias de validez con base en relaciones con otras variables y en la estructura interna de la prueba.

Respecto a la consistencia interna del test, el resultado de fiabilidad de .96 para la variable TR también fue exhibido en el estudio con una muestra americana presentado en la versión española del manual del test (Seisdedos, 2002). La investigación de Izquierdo et al. (2007) con una muestra de niños demostró resultados de consistencia interna en torno a .90 para todas las variables básicas del test d2, lo que está de acuerdo con los resultados para las variables TR, TA y O, pero no para C, que reveló un resultado más bajo de consistencia interna en la presente muestra.

En lo relativo al procedimiento de dos mitades, los resultados confirman la alta fiabilidad para las variables TR, TA y O, lo que también fue demostrado en la adaptación española de la prueba para estas variables. Sin embargo, los resultados para la variable C presentados en el manual varían entre .89 y .99 en muestras de 19 a 39 años (Seisdedos, 2002), mientras en la presente investigación se encontró el resultado de .61. Caglar y Koruç (2006) también encontraron coeficientes que variaron entre .91 a .96 en la evaluación por el método de dos mitades. De esta manera, las dos mitades del test fueron bastante consistentes entre sí en las puntuaciones de TA, TR y O, mientras muy poco consistentes en la puntuación de C.

En la evaluación de la estabilidad temporal por el procedimiento del test-retest, se encontraron resultados más altos para las correlaciones de las puntuaciones TR, TA, TOT y CON, lo que está de acuerdo con los resultados de los estudios originales y en otros países con el test, según la versión española del manual. En cuanto a las medidas de O y C, en general, presentan asimetría en la distribución y son poco fiables temporalmente, pues cambian mucho entre evaluaciones de la misma persona (Seisdedos, 2002; Steinborn et al., 2018).

En lo que se refiere a la confiabilidad de la prueba, los datos encontrados en la presente investigación están según los resultados de Steinborn et al. (2018). La medida de velocidad del procesamiento resulta la más confiable, mientras los puntajes de los errores no son tan estables, aunque el error de tipo omisión haya presentado más confiabilidad temporal en ambas investigaciones. En el caso del puntaje de la velocidad de rendimiento, la investigación de Steinborn et al. (2018) señala que la duración de la prueba puede reducirse hasta un 50%, si es necesario, por razones de economía de tiempo, dado que se encontró alta confiabilidad para el puntaje.

Respecto a la validez basada en la relación con otras variables, también denominada “evidencia convergente” según AERA et al. (2018, p. 17), se encontró en el presente estudio que la relación entre el puntaje de Claves y las medidas TR, TA, TOT y CON de la prueba d2 fue la más relevante. Debido a que la prueba Claves es considerada una medida de atención sostenida, rapidez y destreza visomotora (Wechsler, 2014), consecuentemente, se relaciona con los puntajes de velocidad de procesamiento (TR), precisión del procesamiento (TA), control atencional, velocidad y precisión (TOT) y al índice de concentración (CON). Referente a la prueba de Retención de dígitos, la correlación entre los puntajes TA, CON del d2 y la tarea de dígitos orden directo está relacionada con la precisión, la atención sostenida y la resistencia a la distracción necesarios a la realización de esta tarea. También, fueron observadas correlaciones con dígitos orden inverso y secuenciación, pero estas medidas son más relacionadas con la memoria de trabajo (Egeland, 2015).

En lo que concierne a los resultados de comparación del d2 con el Test del Trazo o *Trail Making Test*, las correlaciones más altas fueron observadas entre la medida del TMT-A y los puntajes TA y CON del d2, referentes a la precisión y a la relación entre la velocidad y la precisión, respectivamente. El rendimiento en el TMT-A está más acorde con las medidas de exploración visual, velocidad grafomotora y velocidad de procesamiento visomotor, mientras que la memoria de trabajo y el control de inhibición se asocian principalmente con el TMT-B y los puntajes TMT derivados (Llinàs-Reglà et al., 2017). Consecuentemente, la rapidez de la habilidad visoperceptiva puede ser el factor más influyente de la relación entre las variables TA y CON del d2 y el TMT-A, pues esta tarea requiere principalmente tal tipo de habilidad (Sánchez-Cubillo et al., 2009). La correlación del test d2 con el TMT-B un poco más reducida puede ser debido a que TMT-B refleja, en primer plano, la memoria de trabajo y en segundo la capacidad de cambiar la atención entre conjuntos de estímulos (Sánchez-Cubillo et al., 2009).

Los resultados correspondientes al test Stroop indican correlaciones más altas de los puntajes TA, TOT y CON del test d2 y la medida del Stroop en que la persona evaluada necesita decir el color de la palabra impreso y no leer el color que está redactado. Al tener que inhibir la lectura del color, la persona requiere más concentración, control atencional y precisión. Los resultados están de acuerdo con el estudio de Zillmer y Kennedy (1999). Además, la correlación inversa del puntaje O y el Stroop indica la presencia del factor control atencional o inhibitorio en estas medidas. Según Scarpina y Tagini (2017), el Stroop es una prueba utilizada para evaluar la capacidad de inhibir la interferencia cognitiva, que ocurre cuando el procesamiento de un estímulo específico impide el procesamiento simultáneo de un segundo atributo de estímulo, conocido como Efecto Stroop. De esta manera, se confirma que la prueba d2 también contribuye a la medida del control atencional o inhibitorio.

TEST DE ATENCIÓN d2

Con referencia a la prueba Búsqueda de símbolos, además de las correlaciones significativas con las variables TA y TOT del d2, que están asociadas a la precisión y al control inhibitorio, se encontró correlación positiva con el puntaje de errores C, en que el individuo marca estímulos irrelevantes. Tal resultado, al ser el único con significancia estadística para los errores C, requiere más estudios para su comprensión, visto que este puntaje también es considerado una medida de control inhibitorio, precisión del procesamiento, minuciosidad y flexibilidad cognitiva (Seisedos, 2002). En lo relativo a la prueba de Cancelación, se encontraron correlaciones positivas con los puntajes TR, TA, TOT y CON, relacionados con la velocidad del procesamiento, la atención selectiva y la precisión. Esta relación puede estar fundamentada en el hecho de que al realizar una tarea de búsqueda visual los estímulos de la tarea se almacenan en la memoria de trabajo visual y se utilizan para guiar la atención durante las tareas de búsqueda visual como Cancelación, así como lo han demostrado en el estudio de Woodman et al. (2007). Las pruebas de Búsqueda de Símbolos y de Cancelación también presentaron correlación con el puntaje VAR. Esta correlación fue inversa, lo que sugiere que entre más vigilancia y velocidad de procesamiento visual, menor será la fluctuación en el modo de trabajar, o sea, más estabilidad. En efecto, los individuos más rápidos presentan resultados más estables en el test d2.

Respecto a las bajas correlaciones entre los puntajes de errores (O y C) y el puntaje del índice de variación (o variabilidad en la velocidad), es importante recalcar que aquellos puntajes (O, C y VAR) presentan asimetría en la distribución y son muy inestables en los análisis de fiabilidad temporal (Seisedos, 2002; Steinborn et al., 2018). Por esta razón, es difícil encontrar confiabilidad en los resultados de estos puntajes, además de que la validez de su interpretación puede ser cuestionable.

En lo concerniente a las comparaciones entre grupos etarios, los participantes con más edad presentaron resultados más bajos en velocidad del procesamiento (TR), precisión del procesamiento (TA), concentración (CON) y efectividad en la prueba (TOT), así como menor control atencional, representado por la cantidad de errores de tipo O más elevada. Estos resultados reflejan una evidencia de validez de la prueba d2 visto que la atención sostenida reduce con el aumento de la edad (Fortenbaugh et al., 2015). Resultados semejantes fueron evidenciados en el estudio de Reyes-Parra et al. (2017), en que conforme aumentaba la edad de la muestra de trabajadores evaluada decrecía significativamente la puntuación obtenida en atención selectiva y sostenida.

Respecto al análisis de la estructura factorial de la prueba d2, los indicadores de ajuste verificados en el AFC sugieren que el Modelo 1 es aceptable (Hu & Bentler, 1999), puesto que no alcanzó los indicadores necesarios (RMSEA < .05, CFI y GFI > .95) para un buen ajuste (Schreiber et al., 2006). Este modelo incluyó las variables CON en el Factor 1 y O en el Factor 2, que son indicadores del índice de concentración y del control atencional, respectivamente. El Modelo 2 fue probado debido a que estudios previos han mencionado que la variable TR sería altamente confiable (Steinborn et al., 2018). El resultado en AFC sugiere que, aunque aceptable en estudios de fiabilidad temporal, esta variable no se ajusta a un posible modelo propuesto que incluye la evaluación de la atención mediante la velocidad y el control atencional. Comparado a este modelo, es más adecuado verificar la cantidad de respuestas marcadas correctamente o el nivel de precisión medido por TA, que resultó con mejores indicadores en el Modelo 3, aunque también solamente aceptable. El Modelo 4, que incluye los dos tipos de errores en el segundo

J. PAWLOWSKI

factor, tampoco revela adecuados índices de ajuste, lo que no corrobora la investigación de Zillmer y Kennedy (1999), que menciona la inclusión de los errores en un mismo factor.

La presente investigación fue un primer paso en el estudio de la adecuación psicométrica de la prueba de atención d2 en el país, en que se buscó replicar algunos métodos de análisis realizados en otros países para verificar el comportamiento de la prueba en una muestra costarricense. Su importancia también se refleja en la discusión de la carencia de datos psicométricos previos bien establecidos para la prueba, así como de modelos teóricos de la atención sostenida.

El presente estudio sugiere que las variables más fiables del test d2 son TR, TA, TOT y CON, mientras que los puntajes de errores, especialmente el C, son muy inestables. Los errores dependen mucho de las condiciones del evaluado o del momento en que se realizó la prueba y deben, por lo tanto, ser interpretados con cautela.

Las relaciones entre los puntajes de la prueba d2 y las pruebas Claves, Dígitos orden directa, TMT-A, test Stroop (página palabra-color) y Cancelación que se proponen a evaluar constructos similares, especialmente atención sostenida o selectiva, velocidad de procesamiento y precisión, proporcionan evidencia de validez de la prueba d2 basada en relaciones con otras variables o “evidencia convergente” (AERA et al., 2018, p. 17). La relación encontrada de los puntajes TR, TA, TOT, CON y O con la variable edad también sugiere este tipo de evidencia. El estudio de la estructura interna del test en modelos hipotéticos, debido a que no se han encontrado estudios previos conclusivos respecto a la estructura factorial de la prueba, indicó ajuste aceptable para el modelo bifactorial en que se incluyen los puntajes obtenidos en cada línea del test para las variables CON en un factor y de la variable O en otro factor. Estos hallazgos sumados sugieren evidencias de la validez de algunos puntajes del test d2 y que la prueba mida en primer plano la atención sostenida o la concentración y la precisión por los puntajes CON y TA y, en segundo plano, también, la atención selectiva y el control atencional, que pueden ser identificados por el puntaje O.

Como limitaciones del estudio, se puede mencionar el tipo de análisis empleado y el tipo de muestra evaluado. Relacionado con el primero, existen recomendaciones bastante actuales con respecto al uso de análisis factoriales confirmatorias o de ecuaciones estructurales para el estudio de la estructura interna de una prueba psicológica. No obstante, en la evaluación cognitiva, algunas veces el empleo de estos análisis, aunque mejores para evidenciar la validez, no se ajustan al tipo de variables de las pruebas cognitivas estudiadas o se encuentra carencia de modelos teóricos bien establecidos para algunos constructos, como fue lo mencionado en la presente investigación. Por consiguiente, la evidencia de relaciones con otras medidas cognitivas que presentan constructos teóricamente relacionados con el d2 puede auxiliar en la conducción de investigaciones futuras que utilicen, por ejemplo, ecuaciones estructurales para evaluar el modelo teórico de la atención sostenida. Respecto a la muestra evaluada, conformada, en especial, por estudiantes universitarios de la carrera de Psicología, recalcase la necesidad de nuevos estudios con muestras que presenten niveles educativos distintos, así como rangos de edad más amplios. Por lo cual, se recomienda ampliar las investigaciones con la prueba d2 en que se investiguen los resultados encontrados con muestras de adolescentes y adultos mayores, además de estudios con muestras que presenten condiciones neuropsicológicas específicas.

TEST DE ATENCIÓN d2

Debido a que, hasta donde se sabe, no existen estudios semejantes con esta prueba en el país, la presente investigación puede contribuir para el avance de áreas que requieren el uso de instrumentos para una evaluación más precisa, tal como en la evaluación neuropsicológica. Además, el estudio con una muestra de Costa Rica fomenta la difusión de la conducta ética y responsable en el uso de pruebas psicológicas y demuestra la preocupación por las posibles diferencias culturales que pueden aparecer reflejadas en los resultados de las pruebas.

Referencias

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2018). *Estándares para pruebas educativas y psicológicas*. American Educational Research Association.
- Ardila, A., & Rosselli, M. (2007). *Neuropsicología clínica*. El Manual Moderno.
- Army Individual Test Battery (AITB) (1944). *Manual of directions and scoring*. War Department, Adjutant General's Office.
- Ato, M., López-García, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038-1059. <https://doi.org/10.6018/analesps.29.3.178511>
- Bates, M. E., & Lemay Jr., E. P. (2004). The d2 Test of Attention: Construct validity and extensions in scoring techniques. *Journal of the International Neuropsychological Society*, 10(3), 392-400. <https://doi.org/10.1017/S135561770410307X>
- Baghaei, P., Ravand, H., & Nadri, M. (2019). Is the d2 Test of Attention Rasch Scalable? Analysis with the Rasch Poisson Counts Model. *Perceptual and Motor Skills*, 126(1), 70–86. <https://doi.org/10.1177/0031512518812183>
- Blotenberg, I., & Schmidt-Atzert, L. (2019). Towards a Process Model of Sustained Attention Tests. *Journal of Intelligence*, 7(1), 3. <https://doi.org/10.3390/jintelligence7010003>
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. The Guilford Press.
- Caglar, E., & Koruç, Z. (2006). d2 dikkat testinin sporcularda güvenilirliği ve geçerliği [Reliability and Validity of d2 Test of Attention for Athletes]. *Spor Bilimleri Dergisi*, 17(2), 58-80. http://www.sbd.hacettepe.edu.tr/fulltext/2006_2_3.pdf
- Cavaco, S., Gonçalves, A., Pinto, C., Almeida, E., Gomes, F., Moreira, I., Fernandes, J., & Teixeira-Pinto, A. (2013). Trail Making Test: regression-based norms for the Portuguese population. *Archives of clinical neuropsychology: the official journal of the National Academy of Neuropsychologists*, 28(2), 189–198. <https://doi.org/10.1093/arclin/acs115>
- Centro Editor de Testes e Pesquisa em Psicologia (CETEPP) (2002). *Teste d2: Atenção concentrada. Manual: Instruções, avaliação, interpretação*. CETEPP.

J. PAWLOWSKI

- Cohen, R. J., & Swerdlik, M. (2006). *Pruebas y evaluación psicológica: Introducción a las pruebas y a la medición*. (4a.ed.). Mc Graw-Hill.
- Colegio de Profesionales en Psicología de Costa Rica (2019). Código de Ética y Deontológico del CPPCR (Reforma 2019). https://psicologiacr.com/sdm_downloads/codigo-de-etica-y-deontologico-del-cppcr-reforma-2019/
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7). <http://www.sciencedirect.com/science/article/pii/S1539306205000701>
- Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Pearson Education, Inc.
- Egeland, J. (2015). Measuring working memory with Digit Span and the Letter-Number Sequencing Subtests from the WAIS-IV: Too low manipulation load and risk for underestimating modality effects. *Applied Neuropsychology. Adult*, 22(6), 445-451. <https://doi.org/10.1080/23279095.2014.992069>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149-60. <https://doi.org/10.3758/BRM.41.4.1149>
- Fernández, A. (2014). Neuropsicología de la atención. Conceptos, alteraciones y evaluación. *Revista argentina de neuropsicología*, 25(1), 1-28. https://www.researchgate.net/publication/273970215_Neuropsicologia_de_la_atencion_Conceptos_alteraciones_y_evaluacion
- Fortenbaugh, F. C., DeGutis, J., & Esterman, M. (2017). Recent theoretical, neural, and clinical advances in sustained attention research. *Annals of the New York Academy of Sciences*, 1396(1), 70-91. <https://doi.org/10.1111/nyas.13318>
- Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., & Esterman, M. (2015). Sustained Attention Across the Life Span in a Sample of 10,000: *Dissociating Ability and Strategy*. *Psychological Science*, 26(9), 1497-1510. <https://doi.org/10.1177/0956797615594896>
- García, J., & Portellano, J. A. (2014). *Neuropsicología de la atención, las funciones ejecutivas y la memoria*. Síntesis.
- Golden, C. J. (1994). *Stroop: Test de colores y palabras*. TEA Ediciones.
- Gomez-Ramirez, M., Hysaj, K., & Niebur, E. (2016). Neural mechanisms of selective attention in the somatosensory system. *Journal of Neurophysiology*, 116(3), 1218-1231. <https://journals.physiology.org/doi/pdf/10.1152/jn.00637.2015>
- Grahn, J. A., & Manly, T. (2012). Common neural recruitment across diverse sustained attention tasks. *PLoS ONE*, 7(11), e49556. <https://doi.org/10.1371/journal.pone.0049556>
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>

TEST DE ATENCIÓN d2

- Izquierdo, M. C., de Iscar Pérez, M. J., Losa, M. A. B., López, M. M., Pérez, L. Á., Solís, G., Colomer, B. F., & Pérez, J. L. A. (2007). Psychometric properties of the d2 selective attention test in a sample of premature and born-at-term babies. *Psicothema*, 19(4), 706–710.
- Jiménez, J. E., Hernández, S., García, E., Díaz, A., Rodríguez, C., & Martín, R. (2012). Test de atención D2: Datos normativos y desarrollo evolutivo de la atención en educación primaria. *European Journal of Education and Psychology*, 5(1), 93-106. <http://www.redalyc.org/articulo.oa?id=129324775008>
- Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). Oxford University Press.
- Leyhe, T., Ethofer, T., Bretscher, J., Künle, A., Säuberlich, A. L., Klein, R., Gallwitz, B., Häring, H. U., Fallgatter, A., Klingberg, S., Saur, R., & Müssig, K. (2013). Low performance in attention testing is associated with reduced grey matter density of the left inferior frontal gyrus in euthyroid patients with Hashimoto's thyroiditis. *Brain, behavior, and immunity*, 27(1), 33–37. <https://doi.org/10.1016/j.bbi.2012.09.007>
- Llinàs-Reglà, J., Vilalta-Franch, J., López-Pousa, S., Calvó-Perxas, L., Torrents Rodas, D., & Garre-Olmo, J. (2017). The Trail Making Test. *Assessment*, 24(2), 183–196. <https://doi.org/10.1177/1073191115602552>
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada [Exploratory Item Factor Analysis: A practical guide revised and updated]. *Anales de Psicología*, 30(3), 1151–1169. <http://dx.doi.org/10.6018/analesps.30.3.199361>
- Martínez-Arias, M. R., Hernández-Lloreda, M. J., & Hernández-Lloreda, M. V. (2006). *Psicometría*. Alianza Editorial.
- Rivera, D., Salinas, C., Ramos-Usuga, D., Delgado-Mejía, I. D., Vasallo Key, Y., Hernández Agurcia, G. P., Valencia Vásquez, J., García-Guerrero, C. E., García de la Cadena, C., Rabago Barajas, B. V., Romero-García, I., Campos Varillas, A. I., Sánchez-SanSegundo, M., Galvao-Carmona, A., Lara, L., Granja Gilbert, E. J., Martín-Lobo, P., Velázquez-Cardoso, J., Caracuel, A., & Arango-Lasprilla, J. C. (2017). Concentration Endurance Test (d2): Normative data for Spanish-speaking pediatric population. *NeuroRehabilitation*, 41(3), 661–671. <https://doi.org/10.3233/NRE-172248>
- Reyes-Parra, P., Jiménez-Martínez, M., & Lara-Báez, C. (2017). Atención selectiva y sostenida de trabajadores en dos jornadas del día. *Psychologia*, 11(2), 113-120. <https://dx.doi.org/10.21500/19002386.3059>
- Rosenberg, M. D., Finn, E. S., Scheinost, D., Constable, R. T., & Chun, M. M. (2017). Characterizing Attention with Predictive Network Models. *Trends in cognitive sciences*, 21(4), 290–302. <https://doi.org/10.1016/j.tics.2017.01.011>
- Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., Jr., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, 20(5), 343–350. <https://doi.org/10.1037/h0043220>
- Rodríguez, L. C., Pulido, N., & Pineda, C. A. (2016). Psychometric properties of the Stroop color-word test in non-pathological Colombian population. *Universitas Psychologica*, 15(2), 255-272. <http://dx.doi.org/10.11144/Javeriana.upsy15-2.ppst>
-

J. PAWLOWSKI

- Sánchez-Cubillo, I., Periañez, J. A., Adrover-Roig, D., Rodríguez-Sánchez, J. M., Ríos-Lago, M., Tirapu, J., & Barceló, F. (2009). Construct validity of the Trail Making Test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society: JINS*, 15(3), 438–450. <https://doi.org/10.1017/S1355617709090626>
- Scarpina, F., & Tagini, S. (2017). The Stroop Color and Word Test. *Frontiers in psychology*, 8, 557. <https://doi.org/10.3389/fpsyg.2017.00557>
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: a review. *The Journal of Educational Research*, 99(6), 323-337. <https://doi.org/10.3200/JOER.99.6.323-338>
- Seisdedos, N. (2002). *D2, test de atención. Adaptación española*. Tea Ediciones.
- Smith, A. (2002). Test de Símbolos y Dígitos (Symbol and Digit Test). TEA Ediciones S.A.
- Steinborn, M. B., Langner, R., Flehmig, H. C., & Huestegge, L. (2018). Methodology of performance scoring in the d2 sustained-attention test: Cumulative-reliability functions and practical guidelines. *Psychological Assessment*, 30(3), 339–357. <https://doi.org/10.1037/pas0000482>
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms and Commentary*. Oxford University Press.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662. <https://doi.org/10.1037/h0054651>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon/Pearson Education.
- Wagner, S., Helmreich, I., Dahmen, N., Lieb, K., & Tadic, A. (2011). Reliability of three alternate forms of the trail making tests a and B. *Archives of clinical neuropsychology: the official journal of the National Academy of Neuropsychologists*, 26(4), 314–321. <https://doi.org/10.1093/arclin/acr024>
- Wassenberg, R., Hendriksen, J. G., Hurks, P. P., Feron, F. J., Keulers, E. H., Vles, J. S., & Jolles, J. (2008). Development of inattention, impulsivity, and processing speed as measured by the d2 Test: results of a large cross-sectional study in children aged 7-13. *Child neuropsychology: a journal on normal and abnormal development in childhood and adolescence*, 14(3), 195–210. <https://doi.org/10.1080/09297040601187940>
- Wechsler, D. (2014). *Escala de Inteligencia para Adultos: WAIS-IV español*. El Manual Moderno.
- Woodman, G. F., Luck, S. J., & Schall, J. D. (2007). The role of working memory representations in the control of attention. *Cerebral Cortex (New York, N.Y.: 1991)*, 17(Suppl 1), i118–i124. <http://doi.org/10.1093/cercor/bhm065>
- Zacada-Menéndez, C., Sampedro-Piquero, P., Begega, A., López, L., & Arias, J. L. (2013). Attention and inhibition in Mild Cognitive Impairment and Alzheimer's Disease. *Escritos de Psicología*, 6(3), 43-50. <https://www.redalyc.org/articulo.oa?id=271030534006>
- Zillmer, E. A., & Kennedy, C. H. (1999). Construct validity for the d2 Test of Attention. *Archives of Clinical Neuropsychology*, 14(8), 728. [https://doi.org/10.1016/S0887-6177\(99\)80213-3](https://doi.org/10.1016/S0887-6177(99)80213-3)

TEST DE ATENCIÓN d2

Recibido: 20 de Septiembre de 2018
Revisión recibida: 16 de Julio de 2020
Aceptado: 20 de Septiembre de 2020

Sobre la autora:

Josiane Pawlowski es doctora en Psicología por la Universidade Federal do Rio Grande do Sul (UFRGS), Brasil. Fue docente de cursos de especialización en evaluación psicológica y neuropsicológica, miembro del Grupo de Estudo, Aplicação e Pesquisa em Avaliação Psicológica en la UFRGS de 2005 a 2010 y profesora nivel adjunto de la Universidade Federal do Rio de Janeiro (UFRJ), Brasil, de 2011 a 2015. Actualmente, es investigadora acreditada por el Consejo Nacional de Investigación en Salud (CONIS), miembro de la Comisión de Investigación y profesora en la Facultad de Psicología de la Universidad de Iberoamérica, Costa Rica. Lattes iD: <http://lattes.cnpq.br/2359349002629206>

Publicado en línea: 21 de diciembre de 2020