



ORIGINAL

Normalización del Test Your Memory y evaluación de su concordancia con los resultados del examen psicométrico



J. Ferrero-Arias* y M.A. Turrión-Rojo

Centro Médico Orense, Madrid, España

Recibido el 3 de diciembre de 2014; aceptado el 24 de marzo de 2015

Accesible en línea el 6 de junio de 2015

PALABRAS CLAVE

Test Your Memory;
Deterioro cognitivo ligero;
Psicometría;
Demencia;
Enfermedad de Alzheimer;
Prueba cognitiva de cribado

Resumen

Objetivo: Explorar la relación entre las puntuaciones del Test Your Memory (TYM) y los hallazgos observados en una valoración neurocognitiva más extensa.

Métodos: Se administraron el TYM y 14 pruebas psicométricas a 84 sujetos de 50 o más años que acudieron a una consulta neurológica ambulatoria con sintomatología de estirpe cognitiva. Se determinó su estado cognitivo con independencia del resultado del TYM (CDR 0: 25, CDR 0,5: 45, CDR 1: 14). Se analizaron la validez concurrente del TYM con los hallazgos de la psicometría, así como el grado de concordancia entre ambos tipos de medida, confrontando datos normalizados y tipificados de los instrumentos.

Resultados: Aunque el coeficiente de correlación intraclase fue de 0,67 (intervalo de confianza del 95%, 0, 53-0,77), el análisis mediante el gráfico de Bland-Altman y la curva de supervivencia-acuerdo de Luiz et al. demuestran que las distancias individuales entre ambos métodos tienen una dispersión excesiva desde el punto de vista clínico. Los valores predichos por el TYM de la puntuación z media de las pruebas psicométricas discrepan sustancialmente en el 30% de los sujetos y solo aceptando como valores idénticos diferencias absolutas de hasta 0,87 puntos z entre ambas medidas se alcanza un 95% de concordancia. Además, el TYM subestima los rendimientos cognitivos cuando sus valores son bajos y los sobreestima si son elevados.

Conclusiones: El TYM es una prueba cognitiva de cribado que no debe utilizarse para predecir el resultado de la psicometría o determinar la ocurrencia de cambio cognitivo en ensayos clínicos
© 2014 Sociedad Española de Neurología. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

* Autor para correspondencia.

Correo electrónico: jferreroarias@gmail.com (J. Ferrero-Arias).

KEYWORDS

Test Your Memory;
Mild cognitive
impairment;
Psychometrics;
Dementia;
Alzheimer disease;
Cognitive screening
test

Standardization of the Test Your Memory and evaluation of their concordance with the outcome of the psychometric examination

Abstract

Objective: To explore the relationship between scores on the Test Your Memory (TYM) battery and findings from a more exhaustive neurocognitive assessment.

Methods: The TYM and fourteen psychometric tests were administered to 84 subjects aged 50 or older who attended an outpatient neurology clinic due to cognitive symptoms. Each patient's cognitive state was determined independently from his/her score on the TYM (CDR 0, n = 25; CDR 0.5, n = 45; CDR 1, n = 14). We analysed concurrent validity of TYM scores and results from the psychometric tests, as well as the degree of concordance between the two types of measurement, by contrasting normalised data from each instrument.

Results: Although the intraclass correlation coefficient was 0.67 (confidence interval 95%, 0.53-0.77), analysis of the Bland-Altman plot and the curve on the survival-agreement plot (Luiz et al. method) demonstrates that the individual distances between the two methods exhibit excessive dispersion from a clinical viewpoint. TYM-based predictions of the mean z-score on psychometric tests differed substantially from real results in 30% of the subjects. Concordance of 95% can only be achieved by accepting absolute inter-instrument differences of up to 0.87 as identical values. Furthermore, the TYM underestimates cognitive performance for low values and overestimates it for high values.

Conclusions: The TYM is a cognitive screening test which should not be used to predict results on psychometric tests or to detect cognitive changes in clinical trials.

© 2014 Sociedad Española de Neurología. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

En algunos estudios se ha observado que las puntuaciones obtenidas en determinadas pruebas cognitivas de cribado, por ejemplo el Mini-Examen Mental¹, predicen aceptablemente bien los resultados de la psicometría y que sus resultados son representativos del estado cognitivo global de los sujetos^{2,3}. Sin embargo, es posible que un buen test cognitivo de cribado tenga una adecuada capacidad de detección sin que refleje fielmente el estado cognitivo de un sujeto. Para comprobar si es correcto extender la utilidad de las pruebas de cribado para «medir el estado cognitivo», tal como se deriva del examen neurocognitivo, hemos utilizado los hallazgos observados en una versión en español del Test Your Memory (TYM), recientemente validada⁴. En este trabajo se presentan, en primer lugar, los resultados de la normalización del TYM, realizada en la muestra original de sujetos. Posteriormente, se aplica la norma generada a una nueva serie de sujetos y se confrontan las puntuaciones del TYM observadas en la nueva muestra con los resultados de una evaluación neurocognitiva más extensa realizada de forma simultánea.

Sujetos, material y métodos**Procedimiento de normalización y tipificación**

Utilizamos la muestra original de sujetos para proceder a la normalización y tipificación del TYM⁴. Se trataba de

739 sujetos cognitivamente intactos de 50 o más años de edad, con al menos 4 años de instrucción formal y en un estadio del Clinical Dementia Rating (CDR)^{5,6} de 0.

La metodología utilizada es muy similar a la aplicada en otros procedimientos de normalización^{7,8}. En el estudio de validación se había observado que el sexo no influye en la puntuación TYM⁴, por lo que esa variable no se tuvo en cuenta en la normalización. Se agrupó a los sujetos por quinquenios de edad y se ordenaron las puntuaciones TYM directas por percentiles en cada grupo etario. Posteriormente, se convirtieron los valores a puntuaciones escalares de media 10 y desviación típica 3, de modo que la distribución de los mismos se atuviera a una distribución normal. Para corregir por el nivel educacional (la variable más influyente en la puntuación TYM directa) se realizó una regresión lineal, siendo la variable dependiente la puntuación TYM escalar, ajustada por grupos etarios, y la predictiva los años de escolarización, segmentados en 4 niveles (0: 4-8 años, 1: 9 a 12 años, 2: 13 a 15 años, 3: más de 15 años de escolarización). Finalmente, se ajustó la puntuación escalar por años de instrucción mediante la siguiente ecuación:

$$TYM_{ec} = TYM_e - \beta * \text{estudios}$$

donde TYM_{ec} es la puntuación TYM escalar corregida por años de instrucción, TYM_e es la puntuación TYM escalar correspondiente al quinquenio etario del sujeto, β es el coeficiente de regresión obtenido en la regresión lineal y «estudios» es el número correspondiente al nivel de escolarización (4 segmentos).

Las puntuaciones escalares, ajustadas por edad y escolarización, se convirtieron a puntuaciones z ($TYM_z = [TYM_{ec} - 10] / 3$) para facilitar las comparaciones con las puntuaciones z de las variables psicométricas.

Grupo de comparación

Se seleccionó a pacientes de 50 o más años y al menos 4 de escolarización, que acudieron a la consulta neurológica de uno de los autores (JFA), que fueron evaluados cognitivamente y a los que se les administraron, además del TYM, pruebas neurocognitivas estandarizadas en el mismo día. Se descartó a los sujetos con lesiones focales (p. ej., tumor o infarto cerebral) y únicamente se incluyó a sujetos con CDR inferior a 2 (CDR 0 = 25, CDR 0,5 = 45, CDR 1 = 14). La inclusión en un estadio CDR se efectuó a partir de un cuestionario estructurado que se puede descargar en <http://alzheimer.wustl.edu/cdr/PDFs/Translations/Spanish%20Spain.pdf>.

Los sujetos que integraron este grupo no habían formado parte de la muestra de normalización. En cada sujeto se calculó la puntuación TYM_z correspondiente a su puntuación TYM directa, quinquenio de edad y grado de escolarización.

Evaluación neurocognitiva

A todos los sujetos del grupo de comparación se les administraron, además del TYM, al menos las siguientes pruebas: retención de dígitos directos e inversos⁹, fluidez verbal semántica (animales) y fonológica (palabras por «p»)¹⁰, test de recuerdo libre y selectivamente facilitado (recuerdo libre y total, y recuerdo libre y total diferido)¹¹, test de vocabulario de Boston¹², test de trazo A y B⁹, copia y reproducción de memoria de la figura compleja de Rey (tiempo de copia, precisión en la copia y de la reproducción de memoria a los 3 min)¹¹. Los resultados de las pruebas aplicadas se expresaron como puntuaciones z . Como indicador del rendimiento cognitivo general, se calculó la media de las puntuaciones z de las pruebas neurocognitivas aplicadas, excepto el TYM (PzMedia). Algunos estudios han mostrado que una combinación de pruebas similar a las administradas en este trabajo constituye una estimación adecuada de los rendimientos cognitivos generales^{13,14}.

Análisis estadístico

Para probar la validez concurrente, se calculó el coeficiente de correlación intraclase entre las puntuaciones TYM_z y PzMedia. Se calculó posteriormente la puntuación PzMedia predicha por el TYM_z (PzMediaP), aplicando los resultados obtenidos en una ecuación de regresión lineal en la que la variable criterio fue la PzMedia y la predictiva el TYM_z . Además, se evaluaron la concordancia y los límites de acuerdo en un gráfico de Bland-Altman¹⁵. Este método también permitió valorar mejor los sesgos. Calculada la PzMediaP, se procedió a determinar si los límites de acuerdo con la PzMedia eran adecuados desde el punto de vista clínico. Además, este último aspecto se evaluó también mediante la técnica gráfica descrita por Luiz et al. de «supervivencia-acuerdo»¹⁶. Se determinó el índice de cambio fiable (ICF) de

ambos instrumentos de medida con la fórmula de Jacobson y Truax¹⁷, a partir del error estándar de la medida (EEM) y mediante el cálculo del intervalo de confianza al 95% ($\pm EEM \times \sqrt{2} \times 1,96$):

$EEM =$ desviación típica $\times \sqrt{(1-r)}$, donde r es el coeficiente de fiabilidad test-retest de cada prueba.

En el caso de la media de las pruebas psicométricas (PzMedia) el coeficiente de fiabilidad test-retest utilizado en el cálculo del ICF es de 0,84 (obtenido por los autores en una muestra independiente de 49 sujetos control). La consistencia interna del conjunto de 14 pruebas (integrantes de una «escala virtual»), calculada en esa serie de 49 sujetos, es adecuada (alfa de Cronbach 0,85 [límite inferior del IC del 95% 0,80]). Por su parte, el ICF del TYM_z se calculó sobre la submuestra normativa en la que se efectuó un test-retest (42 sujetos)⁴. Para definir lo que podría constituir el mínimo cambio detectable, se tomó el promedio de ambos cálculos: el ICF de la PzMedia y del TYM_z (ICFm).

En los análisis se asumió un riesgo alfa bilateral de 0,05. Cuando pareció necesario, las estimaciones se dan con sus IC del 95%. Los análisis estadísticos se realizaron con los paquetes estadísticos Stata/IC 12.1TM y MedCal 14.10.2TM.

Resultados

Normalización y tipificación

En la [tabla 1](#) aparecen por subgrupos de CDR las características sociodemográficas y las pruebas administradas, organizadas por dominios cognitivos. La [tabla 2](#) refleja la conversión de las puntuaciones TYM directas a puntuaciones escalares por quinquenios de edad. La corrección por edad se realizó posteriormente aplicando la ecuación $TYM_{ec} = TYM_e - 0,811 \times$ estudios, obtenida de la regresión. El ICF del TYM_z fue de 0,57 y el de la PzMedia de 0,61, y por tanto el ICFm de $\pm 0,59$ puntos z (límites del IC del 95%). Como material suplementario a este artículo, se incluye una pequeña aplicación en Excel® que permite calcular las puntuaciones escalar y « z » a partir de la edad, los años de escolarización y los puntos TYM directos ([anexo](#)).

Muestra de comparación

La [tabla 3](#) muestra los coeficientes de correlación de Pearson entre la puntuación TYM_z y las puntuaciones z de cada prueba psicométrica. Como se ve, las pruebas más correlacionadas con el TYM son las que evalúan la memoria verbal y el lenguaje, mientras que las pruebas ejecutivas tienen una correlación más débil, aunque estadísticamente significativa. El coeficiente de correlación intraclase entre el TYM_z y la PzMedia fue de 0,67 (IC del 95%, 0,53-0,77) (R^2 0,45). La regresión lineal PzMedia – TYM_z deparó la siguiente ecuación: $PzMedia = -0,1754 + 0,3773 \times TYM_z$. En el gráfico de Bland-Altman elaborado con los datos de la PzMedia y el TYM_z ([fig. 1](#)) se observa que la diferencia entre ambas medidas supera el IC del 95% para una diferencia media de 0 y que los límites de acuerdo se sitúan manifiestamente por encima de una desviación típica en ambos sentidos. En la misma [figura 1](#) se aprecia que el TYM_z sobrestima el rendimiento

Tabla 1 Características sociodemográficas y pruebas administradas por CDR

	CDR 0	CDR 0,5	CDR 1
Número	25	45	14
Edad (media ± DT)	71,0 ± 10,1	71,1 ± 9,2	73,3 ± 6,4
Sexo (% mujeres)	67,0	69,2	30,0
Años de instrucción (media ± DT)	14,1 ± 4,5	12,2 ± 4,8	15,1 ± 4,4
Dominios cognitivos			
<i>Atención</i>			
Dígitos directos (media ± DT)	0,30 ± 0,80	0,10 ± 0,74	-0,43 ± 1,55
Dígitos inversos	0,49 ± 0,74	0,15 ± 0,71	0,10 ± 0,96
<i>Funcionamiento ejecutivo</i>			
Test de trazo A	-0,05 ± 0,95	-0,71 ± 1,00	-1,13 ± 1,25
Test del trazo B	-0,16 ± 0,99	-0,89 ± 0,88	-1,78 ± 1,25
Copia figura de Rey, tiempo	0,31 ± 0,94	-0,16 ± 0,79	-0,96 ± 0,56
Fluidez verbal fonológica («p»)	0,34 ± 0,81	-0,21 ± 0,92	-0,57 ± 1,08
<i>Lenguaje</i>			
Fluidez verbal semántica (animales)	0,02 ± 0,92	-0,81 ± 0,89	-2,15 ± 0,91
Vocabulario de Boston	0,57 ± 0,83	-0,09 ± 1,13	-0,99 ± 0,83
<i>Memoria</i>			
TRLSF recuerdo libre total (1 + 2 + 3)	-0,21 ± 0,73	-1,05 ± 0,77	-2,22 ± 0,91
TRLSF recuerdo total (1 + 2 + 3)	-0,16 ± 0,87	-0,93 ± 0,93	-2,11 ± 1,03
TRLSF libre diferido	-0,21 ± 0,71	-1,73 ± 1,13	-2,43 ± 1,17
TRLSF total diferido	-0,35 ± 0,78	-1,28 ± 1,36	-2,30 ± 0,93
<i>Visuoespacial</i>			
Copia figura de Rey, precisión	0,04 ± 1,05	-0,30 ± 0,93	-0,39 ± 1,37
Memoria figura de Rey	-0,40 ± 0,68	-1,21 ± 0,70	-2,13 ± 1,01
PzMedia	0,03 ± 0,42	-0,65 ± 0,36	-1,35 ± 0,56
TYM (Pz)	-0,02 ± 0,78	-1,82 ± 0,86	-2,00 ± 0,90

CDR: Clinical Dementia Rating; PzMedia: promedio de las puntuaciones z de las pruebas, excepto el TYM; TRLSF: test de recuerdo libre y selectivamente facilitado de Grober y Buschke.

Los valores referidos a las pruebas se expresan como medias de puntuaciones z ± DT.

Tabla 2 Conversión de puntuaciones TYM directas a escalares por quinquenios de edad

Escalares	Percentiles	Grupos etarios ^a							
		50-54	55-59	60-64	65-69	70-74	75-79	80-84	85+
2	< 1	< 39	< 36	< 36	< 35	< 32	< 32	< 31	< 30
3	1	—	—	—	35	32-35	32	31	—
4	2	—	36	36	36	36	33-34	32-35	—
5	3-5	39-40	37-38	37-38	37-38	37-38	35-37	36	30
6	6-10	41-42	39-40	39-40	39	39	38	37-38	31-33
7	11-18	43	41	41-42	40-41	40-41	39-40	39	34-35
8	19-28	44	42-43	43	42	42	41	40-41	36-38
9	29-40	45	44	44	43-44	43	42	42	39-40
10	41-59	46-47	45-46	45-46	45-46	44	43-44	43-44	41-42
11	60-71	48	47	47	47	45-46	45	45-46	43
12	72-81	—	—	48	48	47	46	47	44
13	82-89	49	48	49	49	48	47-48	48	45
14	90-94	—	49	—	—	49	49	49	46-47
15	95-97	—	—	—	—	—	—	—	—
16	98	—	—	—	—	—	—	—	48
17	99	—	—	—	—	—	—	—	49
18	> 99	50	50	50	50	50	50	50	50
Muestra	739	40	50	116	129	122	121	104	57

^a Datos derivados de los sujetos control del estudio citado en la referencia 4.

Tabla 3 Correlaciones entre las pruebas psicométricas y el Test Your Memory

	CC
TRLS recuerdo libre total (1 + 2 + 3)	0,685****
TRLSF recuerdo total (1 + 2 + 3)	0,684****
TRLSF recuerdo diferido total	0,639****
TRLSF recuerdo diferido libre	0,593****
Vocabulario de Boston	0,548****
Reproducción de memoria de la Figura de Rey (3 min)	0,509****
Fluidez verbal semántica (animales)	0,464****
Dígitos directos	0,416****
Copia de la Figura de Rey-Exactitud	0,412****
Test del Trazo B (tiempo)	0,405****
Dígitos inversos	0,372***
Fluidez verbal fonológica (palabras por «p»)	0,314**
Test del Trazo A (tiempo)	0,285**
Copia de la Figura de Rey-Tiempo	0,265*

Número: 84 sujetos.

CC: coeficiente de correlación de Pearson entre las pruebas psicométricas y el Test Your Memory (puntuaciones z); TRLSF: test de recuerdo libre y selectivamente facilitado de Grober-Buschke.

- * p < 0,05.
- ** p < 0,01.
- *** p < 0,001.
- **** p < 0,0001.

cognitivo general, medido con pruebas cognitivas, en valores altos y lo subestima en valores bajos, demostrando que la calibración es imprecisa en los casos en que hay valores extremos. En el gráfico de Bland-Altman confeccionado con

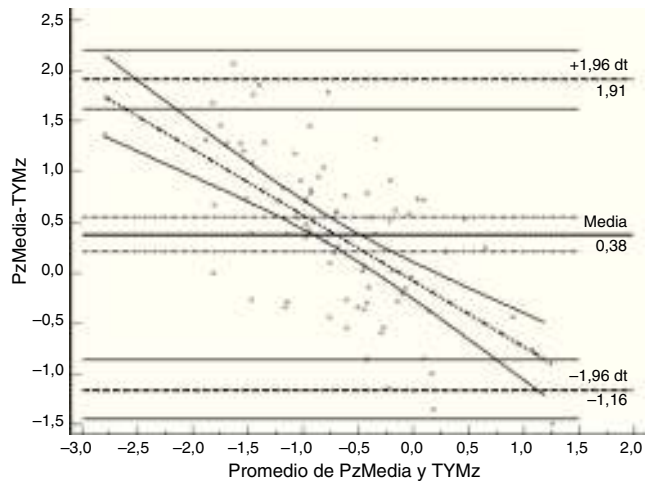


Figura 1 Gráfico de Bland-Altman elaborado con la PzMedia y el TYMz. En el eje de abscisas se representan los valores promedio de la PzMedia y el TYMz, y en las ordenadas, las diferencias entre los valores de los mismos parámetros. Se han trazado la línea del 0 y $\pm 1,96$ DT de las diferencias (con sus respectivos IC del 95%), así como la recta de regresión, cuya pendiente muestra claramente que en valores bajos el TYMz subestima los rendimientos obtenidos en pruebas neurocognitivas (diferencias positivas) y en valores altos los sobrestima (diferencias negativas).

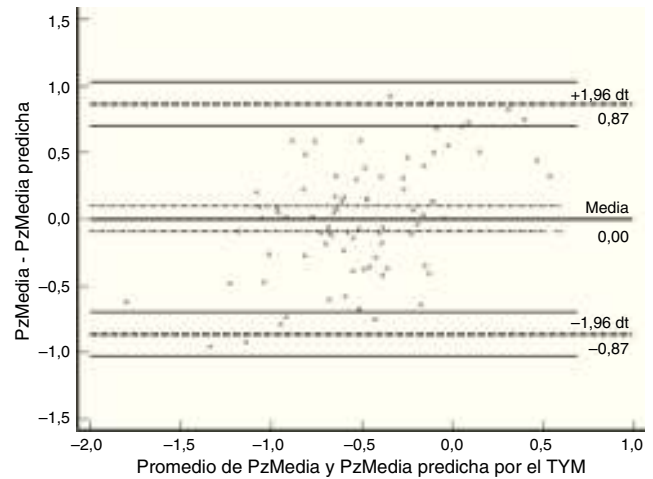


Figura 2 Gráfico de Bland-Altman confeccionado con la PzMedia y la PzMedia predicha calculada aplicando la ecuación de regresión $PzMedia = -0,1754 + 0,3773 \times TYMz$. Obsérvese que la diferencia media está bien centrada en 0 y se han estrechado los límites de acuerdo.

los datos de la PzMedia y la PzMedia predicha por el TYMz y calculada aplicando la ecuación de regresión, se pudo centrar la diferencia dentro de los límites del IC para una diferencia media de 0 y los límites de acuerdo se estrechan hasta valores de $\pm 0,87$ puntos z (fig. 2). Si consideramos como clínicamente significativa una diferencia entre valores z superior a 0,59 puntos z (ICFm), el porcentaje de desacuerdo alcanza el 30% de los casos, como se observa en la figura de supervivencia-acuerdo (fig. 3). Para conseguir el 95% de acuerdo hay que fijar los límites de lo clínicamente significativo en una diferencia entre ambas medidas de más de 0,87 puntos z, es decir 0,28 puntos z más que el ICFm.

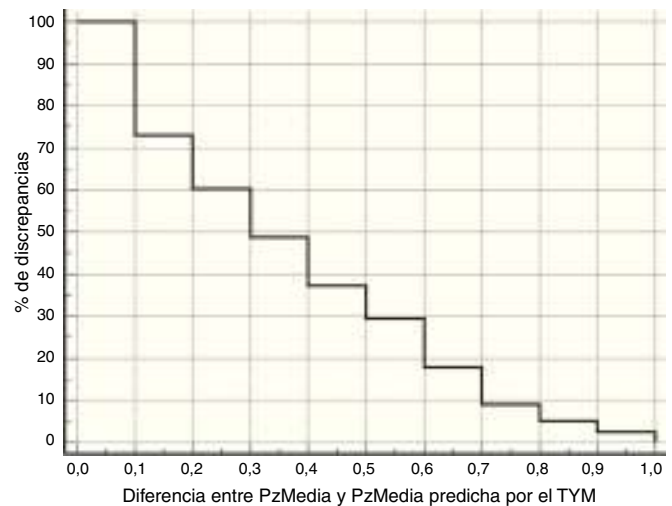


Figura 3 Gráfico de supervivencia-acuerdo para mostrar el porcentaje de desacuerdos entre la PzMedia y la PzMedia Predicha por el TYM. Se obtiene un 95% de acuerdo solo cuando se consideran como valor no idéntico una diferencia mayor de 0,87 puntos z.

Discusión

Los valores del TYM se correlacionan sobre todo con las pruebas de memoria verbal y lenguaje y mucho menos con los test que valoran el funcionamiento ejecutivo, de forma que probablemente detectará más fácilmente bajos rendimientos mnésicos y lingüísticos que en otros dominios cognitivos. En este sentido, sería una prueba más adecuada para el cribado de enfermedad de Alzheimer (EA) o una demencia de expresión inicial lingüística que, por ejemplo, la variante conductual de degeneración lobar frontotemporal o determinadas formas de EA de presentación peculiar (atrofia cortical posterior). La normalización del TYM realizada en este trabajo permite definir con mayor precisión los puntos de corte más idóneos para cada grupo etario y educacional que la regla general dada en el trabajo inicial de validación de nuestra versión en español⁴. Sin embargo, pensamos que los datos normativos dados aquí han de ser aplicados con suma cautela, ya que no parece existir una buena concordancia con la valoración psicométrica más amplia.

La idea de que las pruebas cognitivas breves encarnan la valoración neurocognitiva más extensa y pueden predecir los resultados de esta es antigua^{2,3,18-22}. Una hipótesis subyacente, e implícita, es que ambos reflejan de alguna manera una única dimensión: el rendimiento cognitivo general como reflejo del estado cognitivo global. Este concepto se adopta en trabajos científicos en los que se hace uso literal de estas expresiones¹⁴. Creemos que la idea de «cognición global» es más bien elusiva, pero la aceptaremos como concepto operativo. A las pretendidas utilidades de detección o cribado se han ido añadiendo las de graduador de defecto, indicador del cambio o reflejo fiel del funcionamiento cognitivo de las personas. Sin embargo, las pruebas cognitivas breves tienen desventajas obvias, como son la ocurrencia de efectos techo y suelo, distribuciones sesgadas, representación parcial e incompleta de los dominios neurocognitivos o calibración deficiente²³. Estas propiedades métricas, que comparten la mayoría de las pruebas breves, limitan sustancialmente su potencialidad para medir el funcionamiento cognitivo de amplio espectro y los cambios que se producen longitudinal, evolutivamente. En realidad, la mayoría de los estudios en los que se intenta relacionar una o varias pruebas breves con la psicometría utilizan técnicas analíticas basadas en establecer la correlación entre variables, que no valoran adecuadamente la concordancia de los valores obtenidos con 2 métodos con los que se intenta medir la misma magnitud²⁴.

Por lo que respecta a este estudio, se puede afirmar que, aunque el TYM explica cerca de la mitad de la varianza observada en la PzMedia, no es suficiente para que pueda cumplir una función vicariante de la valoración neurocognitiva más extensa. Las distancias que separan las puntuaciones *z* obtenidas con el TYM y con la valoración neurocognitiva tienen una dispersión lo bastante grande para que no se deba equiparar el valor obtenido con ambas técnicas de evaluación. No se trata solo de la deficiente calibración del TYM, en particular en los valores extremos, sino sobre todo de la dispersión de sus valores para cada valor PzMedia. En cerca del 30% de los sujetos de este estudio la diferencia entre los hallazgos observados con ambos instrumentos de medida supera el cambio mínimo detectable

y ello ocurre, además, y esto es lo realmente importante desde el punto de vista clínico, sin que se pueda predecir el sentido de la desviación. La conclusión es que aunque el TYM es una prueba cognitiva adecuada de cribado de deterioro cognitivo y demencia, no deben adscribirse propiedades que no posee. Tampoco se ha demostrado, por el momento, su sensibilidad al cambio ni se ha comprobado su relación con los hallazgos de la neuropatología, la neuroimagen u otros biomarcadores, como posibles criterios de validación externa.

Podemos preguntarnos por qué emplear un índice de rendimiento cognitivo global como el aplicado en este estudio cuando algunos autores han observado que ciertas pruebas breves pueden predecir mejor el estado funcional de los sujetos que el examen neurocognitivo más extenso²⁵. Así, se ha sometido a escrutinio la validez ecológica de los test neuropsicológicos, habiéndose encontrado que los mismos no alcanzan a explicar el 30% de la varianza observada en las medidas de las actividades funcionales diarias^{25,26}. Hay que reconocer, sin embargo, que los instrumentos con que contamos para cuantificar estas últimas tienen sus propios problemas²⁷: los datos derivados de los informadores tienen una confiabilidad muy variable, con frecuencia inconstante, las medidas pueden enfocarse a un heterogéneo grupo de actividades muy dispares (básicas, instrumentales, avanzadas, tareas caseras, actividades sociales, aficiones y pasatiempos...) que no se incluyen uniformemente como parte integrante de las escalas, las actividades de toda la vida de un sujeto a menudo no son comparables con las de otros sujetos con estilos de vida muy diferentes pero, por lo demás, invariantes con respecto a los factores sociodemográficos habitualmente contemplados, y muchos instrumentos carecen de las exigencias métricas mínimas, de modo que la proliferación de escalas no es menor que la de los test breves cognitivos, indicando muy claramente que no existe una verdaderamente satisfactoria²⁸. Algunos de los problemas inherentes a la construcción de escalas ordinales y test cognitivos podrían resolverse utilizando técnicas de construcción modernas^{27,29,30} (teoría de respuesta al ítem, análisis según el modelo Rasch...). Nosotros tenemos serias dudas de que cualquier menoscabo cognitivo no produzca necesariamente déficit funcional, aunque no dispongamos de instrumentos para detectarlos. De esta manera, el concepto de deterioro cognitivo ligero sería una categoría artificiosa derivada de nuestra propia impercepción de la realidad.

Siendo clásicos, y conservadores, es probable que la mayoría de los profesionales sanitarios piensen, como nosotros, que la evaluación formal neurocognitiva, polinómica y relativamente extensa represente mejor que cualquier prueba breve el verdadero estado cognitivo de un sujeto y, previsiblemente, exprese de forma más confiable el funcionamiento cerebral que es al final nuestro objetivo principal. Además, su validez prospectiva es claramente superior a la de los test breves, justamente allí donde más se necesita la precisión, cuando nos invade la incertidumbre y los rendimientos cognitivos anómalos parecen mucho más obvios a los pacientes o allegados que a nosotros mismos²³. Recientemente, un concienzudo análisis del ADAS-Cog ha dejado al descubierto algunas debilidades notables^{31,32}, a pesar de tratarse de una prueba ampliamente utilizada

para evaluar resultados en ensayos clínicos de la enfermedad de Alzheimer³³⁻³⁵. Pensamos que seguramente lo mismo puede aplicarse a muchas otras pruebas cognitivas breves. Los requisitos necesarios para la construcción de pruebas breves no siempre se cumplen, aunque han sido enumerados de forma pormenorizada³⁶. Por supuesto, la valoración neurocognitiva detallada aporta además una información cualitativa de grano fino, incomparablemente más rica y valiosa que la de una prueba breve.

Hemos de concluir que los resultados de este estudio sugieren que las puntuaciones del TYM (y posiblemente de otras pruebas cognitivas breves de cribado) no deberían utilizarse para graduar el estado cognitivo actual de un sujeto, dejar constancia de sus vicisitudes evolutivas cognitivas, determinar la ocurrencia de cambios en ensayos clínicos o predecir los hallazgos derivados de un examen psicométrico extenso.

Financiación

Este estudio no ha sido financiado por ninguna institución pública o privada.

Conflicto de intereses

Los autores no declaran tener conflicto de intereses.

Anexo. Material adicional

Se puede consultar material adicional a este artículo en su versión electrónica disponible en <http://dx.doi.org/10.1016/j.nrl.2015.03.015>.

Bibliografía

- Folstein MF, Folstein SE, McHugh PR. Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12:189–98.
- Farber JF, Schmitt FA, Logue PE. Predicting intellectual level from the Mini-Mental State Examination. *J Am Geriatr Soc.* 1988;36:509–10.
- Nakamura H, Nakanishi M, Furukawa TA, Hamanaka T, Tokudome S. Validity of brief intelligence tests for patients with Alzheimer's disease. *Psychiatry Clin Neurosci.* 2000;54:435–9.
- Ferrero-Arias J, Turrión-Rojo MÁ. Validación de una versión española del Test Your Memory. *Neurología.* 2014, <http://dx.doi.org/10.1016/j.nrl.2013.12.009>.
- Hughes CP, Berg L, Danziger WL, Cohen LA, Martin RL. A new clinical scale for the staging of dementia. *Br J Psychiatr.* 1982;140:566–72.
- Morris JC. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology.* 1993;43:2412–4.
- Lucas JA, Ivnik RJ, Willis FB, Ferman TJ, Smith GE, Parfitt FC, et al. Mayo's Older African Americans Normative Studies: Normative data for commonly used clinical neuropsychological measures. *Clin Neuropsychol.* 2005;19:162–83.
- Peña-Casanova J, Blesa R, Aguilar M, Gramunt-Fombuena N, Gómez-Ansón B, Oliva R, NEURONORMA Study Team. Spanish Multicenter Normative Studies (NEURONORMA Project): Methods and sample characteristics. *Arch Clin Neuropsychol.* 2009;24:307–19.
- Peña-Casanova J, Quiñones-Ubeda S, Quintana-Aparicio M, Aguilar M, Badenes D, Molinuevo JL, et al., NEURONORMA Study Team. Spanish Multicenter Normative Studies (NEURONORMA Project): Norms for verbal span, visuospatial span, letter and number sequencing, trail making test, and symbol digit modalities test. *Arch Clin Neuropsychol.* 2009;24:321–41.
- Peña-Casanova J, Quiñones-Ubeda S, Gramunt-Fombuena N, Quintana-Aparicio M, Aguilar M, Badenes D, et al., NEURONORMA Study Team. Spanish Multicenter Normative Studies (NEURONORMA Project): Norms for verbal fluency tests. *Arch Clin Neuropsychol.* 2009;24:395–411.
- Peña-Casanova J, Gramunt-Fombuena N, Quiñones-Ubeda S, Sánchez-Benavides G, Aguilar M, Badenes D, et al., NEURONORMA Study Team. Spanish Multicenter Normative Studies (NEURONORMA Project): Norms for the Rey-Osterrieth complex figure (copy and memory), and free and cued selective reminding test. *Arch Clin Neuropsychol.* 2009;24:371–93.
- Peña-Casanova J, Quiñones-Ubeda S, Gramunt-Fombuena N, Aguilar M, Casas L, Molinuevo JL, et al., NEURONORMA Study Team. Spanish Multicenter Normative Studies (NEURONORMA Project): Norms for Boston naming test and token test. *Arch Clin Neuropsychol.* 2009;24:343–54.
- Jones RN, Rudolph JL, Inouye SK, Yang FM, Fong TG, Milberg WP, et al. Development of a unidimensional composite measure of neuropsychological functioning in older cardiac surgery patients with good measurement precision. *J Clin Exp Neuropsychol.* 2010;32:1041–9.
- Gross AL, Jones RN, Fong TG, Tommet D, Inouye SK. Calibration and validation of an innovative approach for estimating general cognitive performance. *Neuroepidemiology.* 2014;42:144–53.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307–10.
- Luiz RR, Costa AJ, Kale PL, Werneck GL. Assessment of agreement of a quantitative variable: A new graphical approach. *J Clin Epidemiol.* 2003;56:963–7.
- Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol.* 1991;59:12–9.
- Ihl R, Frölich L, Dierks T, Martin EM, Maurer K. Differential validity of psychometric tests in dementia of the Alzheimer type. *Psychiatry Res.* 1992;44:93–106.
- Giordani B, Boivin MJ, Hall AL, Foster NL, Lehtinen SJ, Bluemlein LA, et al. The utility and generality of Mini-Mental State Examination scores in Alzheimer's disease. *Neurology.* 1990;40:1894–6.
- Brooks JO 3rd, Yesavage JA, Taylor J, Friedman L, Tanke ED, Luby V, et al. Cognitive decline in Alzheimer's disease: Elaborating on the nature of the longitudinal factor structure of the Mini-Mental State Examination. *Int Psychogeriatr.* 1993;5:135–46.
- Dick JP, Guiloff RJ, Stewart A, Blackstock J, Bielawska C, Paul EA, et al. Mini-mental state examination in neurological patients. *J Neurol Neurosurg Psychiatry.* 1984;47:496–9.
- Sugarman MA, Axelrod BN. Utility of the Montreal Cognitive Assessment and Mini-Mental State Examination in predicting general intellectual abilities. *Cogn Behav Neurol.* 2014;27:148–54.
- Wouters H, Appels B, van der Flier WM, van Campen J, Klein M, Zwiderman AH, et al. Improving the accuracy and precision of cognitive testing in mild dementia. *J Int Neuropsychol Soc.* 2012;18:314–22.

24. Bland JM, Altman DG. Agreed statistics: Measurement method comparison. *Anesthesiology*. 2012;116:182–5.
25. Royall DR, Lauterbach EC, Kaufer D, Malloy P, Coburn KL, Black KJ, et al., Committee on Research of the American Neuropsychiatric Association. The cognitive correlates of functional status: A review from the Committee on Research of the American Neuropsychiatric Association. *J Neuropsychiatry Clin Neurosci*. 2007;19:249–65.
26. Van der Elst W, van Boxtel MP, van Breukelen GJ, Jolles J. A large-scale cross-sectional and longitudinal study into the ecological validity of neuropsychological test measures in neurologically intact people. *Arch Clin Neuropsychol*. 2008;23:787–800.
27. Lindeboom R, Vermeulen M, Holman R, de Haan RJ. Activities of daily living instruments: Optimizing scales for neurologic assessments. *Neurology*. 2003;60:738–42.
28. Fieo RA, Austin EJ, Starr JM, Deary IJ. Calibrating ADL-IADL scales to improve measurement accuracy and to extend the disability construct into the preclinical range: A systematic review. *BMC Geriatr*. 2011;11:42.
29. McGrory S, Doherty JM, Austin EJ, Starr JM, Shenkin SD. Item response theory analysis of cognitive tests in people with dementia: A systematic review. *BMC Psychiatry*. 2014;14:47.
30. Crane PK, Narasimhalu K, Gibbons LE, Mungas DM, Haneuse S, Larson EB, et al. Item response theory facilitated calibrating cognitive tests and reduced bias in estimated rates of decline. *J Clin Epidemiol*. 2008;61:1018–27.
31. Hobart J, Cano S, Posner H, Selnes O, Stern Y, Thomas R, et al. Alzheimer's Disease Neuroimaging Initiative. Putting the Alzheimer's cognitive test to the test I: Traditional psychometric methods. *Alzheimers Dement*. 2013;9(1 Suppl):S4–9.
32. Hobart J, Cano S, Posner H, Selnes O, Stern Y, Thomas R, et al. Alzheimer's Disease Neuroimaging Initiative. Putting the Alzheimer's cognitive test to the test II: Rasch measurement theory. *Alzheimers Dement*. 2013;9(1 Suppl):S10–20.
33. Salloway S, Sperling R, Fox NC, Blennow K, Klunk W, Raskind M, et al. Bapineuzumab 301 and 302 Clinical Trial Investigators. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. *N Engl J Med*. 2014;370:322–33.
34. Doody RS, Raman R, Farlow M, Iwatsubo T, Vellas B, Joffe S, et al. Alzheimer's Disease Cooperative Study Steering Committee, Siemers E, Sethuraman G, Mohs R; Semagacestat Study Group. A phase 3 trial of semagacestat for treatment of Alzheimer's disease. *N Engl J Med*. 2013;369:341–50.
35. Doody RS, Thomas RG, Farlow M, Iwatsubo T, Vellas B, Joffe S, et al. Alzheimer's Disease Cooperative Study Steering Committee; Solanezumab Study Group. Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. *N Engl J Med*. 2014;370:311–21.
36. Peña-Casanova J, Monllau A, Gramunt Fombuena N. La psicometría de las demencias a debate. *Neurología*. 2007;22:301–11.