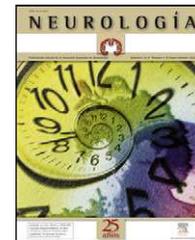


# NEUROLOGÍA

www.elsevier.es/neurologia



ORIGINAL

## Funcionamiento diferencial de los ítems del test Mini-mental en función de la patología

G. Prieto, A.R. Delgado\*, M.V. Perea y V. Ladera

Departamento de Psicología Básica, Psicobiología y Metodología, Universidad de Salamanca, Salamanca, España

Recibido el 27 de julio de 2010; aceptado el 7 de enero de 2011

Accesible en línea el 16 de marzo de 2011

### PALABRAS CLAVE

Prueba de cribado;  
Test Mini-mental;  
Psicometría;  
Funcionamiento  
diferencial de los  
ítems;  
Enfermedad de  
Parkinson;  
Demencia tipo  
Alzheimer

### Resumen

**Introducción:** Sería imposible comparar la estatura de los varones y las mujeres si el metro no tuviese las mismas propiedades en ambas poblaciones. De forma similar, no se debería comparar el deterioro cognitivo de sujetos con distintas patologías si el test empleado no tuviese las mismas propiedades métricas en los grupos analizados. La ausencia de funcionamiento diferencial de los ítems (DIF) es una condición de la invarianza métrica entre poblaciones.

**Material y métodos:** Este artículo analiza el test de cribado de la demencia más utilizado internacionalmente, el Mini-mental State Examination (MMSE), mediante un modelo psicométrico avanzado, el modelo de Rasch, con el objetivo de poner a prueba la invarianza de las medidas obtenidas en distintos grupos: pacientes con enfermedad de Parkinson, pacientes con demencia tipo Alzheimer y sujetos normales. Para ello, se ha contrastado la hipótesis de que el MMSE no muestra DIF contra ninguno de estos grupos en una muestra total de 400 sujetos.

**Resultados:** Los resultados del análisis indican que existen diferencias significativas entre los grupos; sin embargo, la comparación cuantitativa sólo tiene sentido si no existe evidencia en contra de la invarianza métrica: dado el tipo de ítems que muestran DIF contra el grupo de pacientes con enfermedad de Parkinson, cabe pensar que el MMSE podría estar sobreestimando el nivel de deterioro cognitivo de estos pacientes.

**Conclusiones:** Pese a lo extendido del uso del MMSE, el funcionamiento de 11 de sus 30 ítems no es igual para los distintos grupos por lo que la comparación de las puntuaciones no estaría justificada.

© 2010 Sociedad Española de Neurología. Publicado por Elsevier España, S.L. Todos los derechos reservados.

\* Autor para correspondencia.

Correo electrónico: [adelgado@usal.es](mailto:adelgado@usal.es) (A.R. Delgado).

**KEYWORDS**

Screening test;  
Mini-mental State Examination;  
Testing;  
DIF;  
Parkinson's disease;  
Alzheimer's type dementia

**Differential functioning of mini-mental test items according to disease****Abstract**

**Introduction:** Comparing the height of males and females would be impossible if the measuring device did not have the same properties for both populations. In a similar way, the cognitive level of diverse groups of patients should not be compared if the test has different measurement properties for these groups. Lack of Differential Item Functioning (DIF) is a condition for measurement invariance between populations.

**Material and methods:** The most internationally used screening test for dementia, the MMSE (or Mini-mental State Examination), has been analysed using an advanced psychometric technique, the Rasch Model. The objective was to determine the invariance of mini-mental measurements from diverse groups: Parkinson's disease patients, Alzheimer's type dementia and normal subjects. The hypothesis was that the scores would not show DIF against any of these groups. The total sample was composed of 400 subjects.

**Results:** Significant differences between groups were found. However, the quantitative comparison only makes sense if no evidence against measurement invariance was found: given the kind of items showing DIF against Parkinson's disease patients, the MMSE seems to underestimate the cognitive level of these patients.

**Conclusions:** Despite the extended use of this test, 11 items out of 30 show DIF and consequently score comparisons between groups are not justified.

© 2010 Sociedad Española de Neurología. Published by Elsevier España, S.L. All rights reserved.

**Introducción**

El Mini-Mental State Examination (MMSE) es la prueba de cribado de la demencia más utilizado internacionalmente. Los 30 ítems de la prueba se clasifican en 6 dominios cognitivos: orientación temporal, orientación espacial, registro de información, atención y cálculo, memoria y lenguaje. La versión original<sup>1</sup> ha sido adaptada a distintos países y traducida a distintos idiomas, entre los que se encuentran el español<sup>2</sup>, portugués<sup>3</sup>, japonés<sup>4</sup>, chino<sup>5</sup>, turco<sup>6</sup>, italiano<sup>7</sup> y hebreo<sup>8</sup>.

Las pruebas de cribado se emplean en poblaciones diversas. Una condición esencial de la equivalencia métrica de las puntuaciones de una prueba en distintos grupos es la ausencia de funcionamiento diferencial de los ítems (DIF, acrónimo inglés que emplearemos por hallarse muy extendido). Un ítem presenta DIF asociado a la pertenencia a un grupo cuando sujetos con el mismo valor en la variable medida, pero procedentes de grupos distintos, tienen distinta probabilidad de resolver dicho ítem correctamente. Los primeros procedimientos para detectar el DIF se aplicaron a las puntuaciones obtenidas mediante la teoría clásica de los tests (TCT)<sup>9</sup>. Aunque la TCT ha sido el principal modelo psicométrico empleado en el análisis de las puntuaciones de las pruebas, sus limitaciones han llevado a la propuesta de modelos alternativos, de los cuales el más parsimonioso es el modelo de Rasch<sup>10</sup>, que permite, dado un buen ajuste, la medición conjunta de personas e ítems en una misma variable latente con propiedades de intervalo. Es importante hacer constar que este modelo no requiere muestras representativas, ya que una de sus propiedades, la objetividad específica, garantiza que, dado un ajuste suficiente de los datos al modelo, los parámetros de los ítems sean independientes de la muestra de sujetos y los parámetros de los sujetos, de la muestra de ítems. Sí es necesario contar con un número suficiente de sujetos en todos los niveles de la variable latente.

El modelo de Rasch, estadísticamente incluido en la teoría de la respuesta a los ítems (TRI), permite poner a prueba fácilmente la invarianza de las medidas que una prueba proporciona en distintos grupos, además de contar con otras propiedades que hacen su uso particularmente recomendable<sup>11,12</sup>. En especial, el modelo de Rasch resulta muy sencillo de aplicar en su versión más básica, cuando se trata de analizar datos dicotómicos tales como las respuestas a los 30 ítems del MMSE, es decir, cuando hay una sola respuesta correcta y los datos pueden codificarse binariamente (1/0). En estos casos, puede modelarse la probabilidad de que el sujeto  $n$  dé una respuesta correcta al ítem  $i$ ,  $p_{ni}$ , con la siguiente fórmula:

$$p_{ni} = \exp(\beta_n - \delta_i) / (1 + \exp[\beta_n - \delta_i])$$

$\beta_n$ : nivel del sujeto  $n$  en la variable latente

$\delta_i$ : valor del ítem  $i$  en la variable latente

El ajuste de los datos obtenidos en la aplicación de la prueba al modelo de Rasch se calcula mediante estadísticos basados en los residuos, *Infit* y *Outfit*, cuya distribución se asemeja a  $\chi^2$ . En cuanto al DIF, su existencia se pone a prueba calculando la diferencia estandarizada entre los estimadores del parámetro de dificultad del ítem en cada grupo controlando las posibles diferencias entre grupos en la variable latente y utilizando la corrección de Bonferroni<sup>13</sup>. Suele recomendarse la eliminación de los ítems con DIF por su falta de validez generalizada, especialmente cuando las puntuaciones en una prueba van a servir para tomar decisiones de importancia<sup>14</sup>. Sin embargo, dependiendo de los objetivos de la medición, también podrían servir como fuente de hipótesis sobre la causa de la diferencia observada.

El objetivo de este artículo es analizar el MMSE mediante el modelo de Rasch con el fin de poner a prueba la invarianza de las medidas obtenidas mediante el MMSE en distintos grupos: pacientes con enfermedad de Parkinson, pacientes con

demencia tipo Alzheimer y sujetos normales. Para ello, se someterá a contraste la hipótesis de que el MMSE no muestra DIF contra alguno de los grupos.

## Sujetos y métodos

El MMSE<sup>2</sup> se administró mediante entrevistas individuales a 400 sujetos españoles residentes en Castilla y León, con el castellano como lengua materna, de los cuales 121 presentaban la enfermedad de Parkinson (grupo EP), 110 demencia tipo Alzheimer (grupo DTA) y 169 eran sujetos normales (grupo N). El rango de niveles en el constructo obtenido por estos tres grupos es muy amplio y garantiza que el error de estimación no varíe mucho a lo largo de la variable latente. Por otra parte, es conveniente recordar que la propiedad de objetividad específica nos permite obtener estimadores de los parámetros en muestras no representativas. El grupo EP está formado por 59 mujeres y 62 varones con diagnóstico de EP sin demencia. La EP fue diagnosticada por un neurólogo sobre la base de la acinesia asociada con uno de los otros dos signos cardinales (temblor, rigidez) y la capacidad de respuesta a la terapia con levodopa. Todos los pacientes cumplen los criterios del UK Parkinson's Disease Brain Bank para el diagnóstico de EP idiopática<sup>15</sup>.

El grupo DTA se compone de 62 mujeres y 48 varones diagnosticados de DTA probable, de acuerdo con los criterios de la NINCDS-ADRDA<sup>16</sup>, confirmado por especialistas. El grado de severidad de la demencia según la escala clínica de demencia (CDR)<sup>17</sup> era leve en 59 casos, moderado en 42 y severo en 9. Los sujetos con historia conocida o con sospecha de episodios de isquemia cerebral transitoria, alcoholismo, traumatismo craneoencefálico o enfermedades como cáncer, disfunción tiroidea y depresión mayor fueron excluidos. En cuanto al grupo de comparación (N), se trata de 90 mujeres y 79 varones de edades comprendidas entre los 15 y los 65 años y sin ningún tipo de alteración neurológica, psicopatológica y/o neuropsicológica en su historia clínica. Se excluyó de la muestra inicial a 77 sujetos con puntuación perfecta ( $X = 30$ ) en el MMSE, ya que el modelo no estima sus parámetros y no aportan información al análisis. De éstos, 75 eran sujetos normales y 2 con diagnóstico de EP.

## Resultados

En principio, el ajuste de los datos al modelo resulta suficiente: únicamente un ítem y un 10% de los sujetos desajustan severamente y de éstos solo uno pertenece al grupo DTA, cinco al grupo de EP y el resto son sujetos normales. El coeficiente de consistencia interna  $\alpha$  de Cronbach es de 0,93 y la fiabilidad de las personas, estimada siguiendo el modelo de Rasch es 0,86, un nivel aceptable. La fiabilidad de los ítems es muy alta, de 0,99, en términos del modelo.

Las estimaciones del parámetro de dificultad de los ítems pueden verse en la [tabla 1](#), junto con los errores y los indicadores de ajuste *Infit* y *Outfit*. Hay que señalar que el único ítem severamente desajustado, el número 25 (orden 1), tiene una discriminación de 0,17 que lo separa del resto de los ítems, cuya discriminación varía entre el 0,37 de los

ítems 28 (reloj) y 29 (repetición-frase) y el 0,77 del ítem 10 (piso). Es también reseñable el hecho de que el indicador *Outfit* detecte la dependencia entre varios ítems, por ejemplo los que forman parte de la función de cálculo, mediante valores excesivamente bajos que señalan un patrón determinista en las respuestas. En cuanto al orden de dificultad de los ítems, puede apreciarse fácilmente en la [figura 1](#) en la que se escalan conjuntamente sujetos e ítems a lo largo de la variable latente (el cero de la escala corresponde, convencionalmente, a la media de dificultad de los ítems). En la parte superior de este mapa, se encuentran los tres últimos ítems de cálculo seguidos de los tres de memoria, los que resultan más difíciles. En la parte inferior pueden verse, empezando por los más fáciles: los ítems de denominación por vía visual «reloj» y «bolígrafo», un ítem de registro, «lápiz», un ítem de orientación espacial, «ciudad» y los dos ítems de registro restantes. Como cabía esperar, los ítems relacionados con la función de denominación visual señalan el extremo inferior de la variable y los ítems de memoria y cálculo, el superior.

La comparación entre las medias de los tres grupos en las puntuaciones Rasch indica que existen diferencias grandes y significativas entre los tres grupos,  $F(2,397) = 654,41$ ,  $p > 0,001$ , cuyas medias son, de mayor a menor: 3,56 (N), 2,20 (EP) y  $-0,37$  (DTA).

El escalamiento conjunto de sujetos e ítems representado en la [figura 1](#) y las comparaciones entre grupos basadas en dicho escalamiento se han llevado a cabo a partir de la muestra total, lo que implica partir del supuesto de que la medida es invariante en los tres grupos. De no ser así, ni las medidas Rasch ni la suma de puntos habitualmente empleada (y que es el estadístico suficiente a partir del que se han estimado las puntuaciones de los sujetos y los ítems) reflejarían adecuadamente la aptitud de los grupos. De ahí nuestro objetivo de poner a prueba la invarianza de las medidas obtenidas mediante el MMSE en pacientes con enfermedad de Parkinson, pacientes con demencia tipo Alzheimer y sujetos normales.

Los resultados del análisis del DIF indican que son varios los ítems que funcionan diferencialmente en uno o más grupos. La [figura 2](#) muestra los indicadores de dificultad estimados separadamente para cada grupo controlando las diferencias entre grupos en la variable latente. Tomemos el ítem de orientación espacial «piso», un ítem con DIF contra el grupo DTA: no se trata de que haya una diferencia entre los grupos en la variable latente, que las hay, es que esa comparación cuantitativa sólo podrá hacerse legítimamente si previamente se ha rechazado la hipótesis del DIF entre los grupos. Es decir, el ítem ha de tener la misma dificultad para las personas que tienen el mismo nivel en el constructo sea cuál sea el grupo al que pertenezcan. Si, por el contrario, la dificultad de ese ítem para el grupo DTA fuera mayor que para los otros dos grupos (como se ve en la [fig. 2](#)), la comparación cuantitativa entre ellos se vería comprometida.

La [tabla 2](#) resume los resultados del análisis del DIF, comenzando por las cinco comparaciones que perjudican al grupo DTA en 4 ítems, de los cuales 3 son de orientación espacial y sólo 1 de cálculo (en éste la comparación es con el grupo N, no con el EP). A continuación, pueden verse los ítems con DIF contra el grupo EP, 5 ítems y un total de seis comparaciones: 1 ítem de cálculo, la repetición de la frase y el dibujo resultan más difíciles para los pacientes

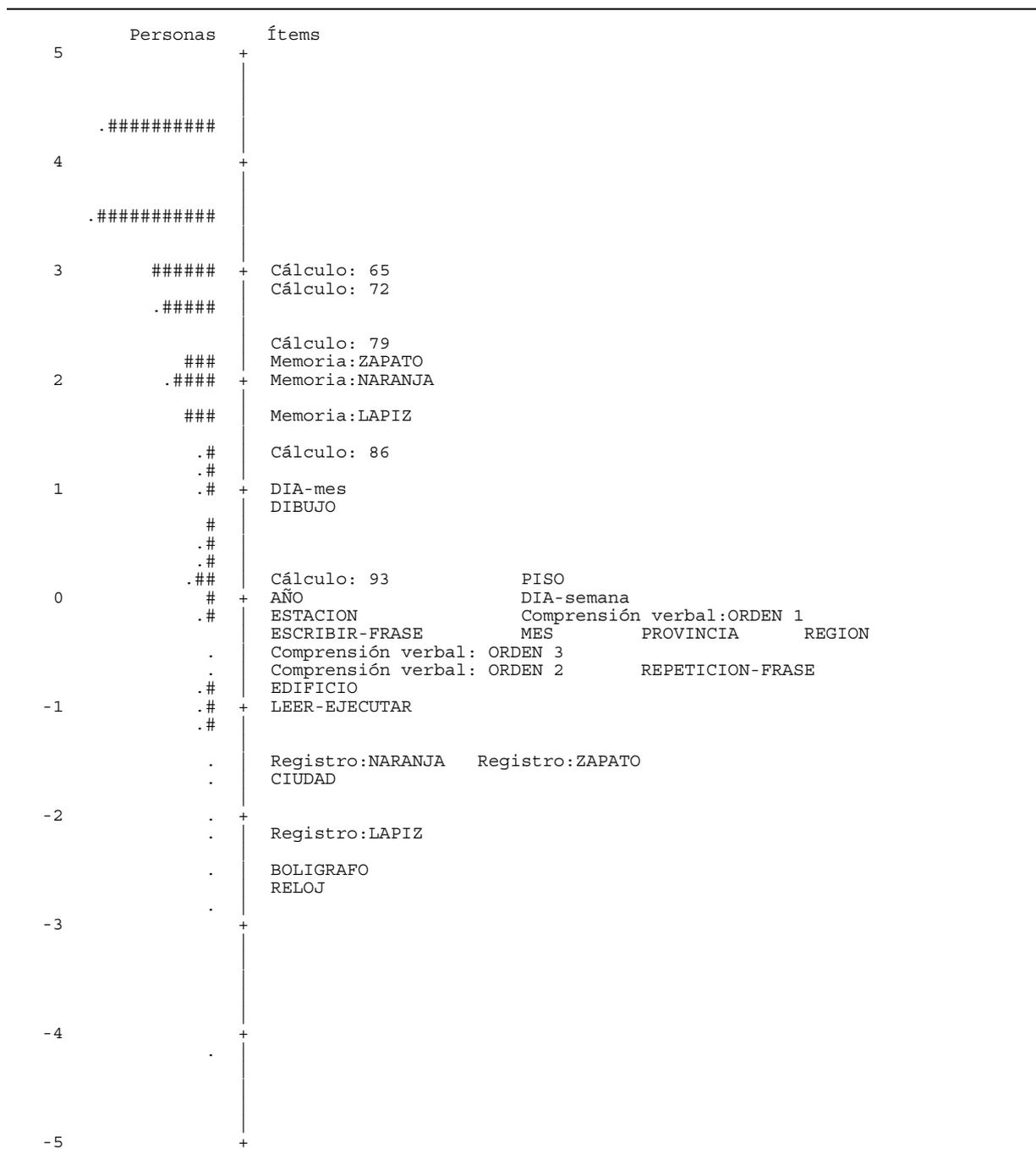
**Tabla 1** Dificultad, error de estimación e indicadores del ajuste de los ítems

Ítem	Dificultad	Error	Infit	Outfit	Etiqueta
1	-0,07	0,16	0,85	0,61	Orientación temporal: año
2	-0,18	0,16	1,01	1,24	Orientación temporal: estación
3	0,95	0,14	0,83	1,08	Orientación temporal: día-mes
4	-0,02	0,16	0,85	0,78	Orientación temporal: día-semana
5	-0,34	0,17	0,91	0,51	Orientación temporal: mes
6	-0,31	0,17	0,90	1,47	Orientación espacial: región
7	-0,34	0,17	0,74	0,93	Orientación espacial: provincia
8	-1,65	0,21	0,89	0,35	Orientación espacial: ciudad
9	-0,82	0,18	0,85	0,34	Orientación espacial: edificio
10	0,17	0,15	0,54	0,37	Orientación espacial: piso
11	-2,18	0,25	1,02	0,29	Registro: lápiz
12	-1,52	0,21	0,69	0,23	Registro: naranja
13	-1,52	0,21	0,69	0,24	Registro: zapato
14	0,10	0,16	1,26	1,59	Cálculo: 93
15	1,31	0,14	1,15	1,06	Cálculo: 86
16	2,40	0,13	0,72	0,64	Cálculo: 79
17	2,86	0,13	0,56	0,41	Cálculo: 72
18	2,99	0,13	0,61	0,45	Cálculo: 65
19	1,60	0,13	1,35	1,60	Memoria: lápiz
20	2,05	0,13	1,31	1,44	Memoria: naranja
21	2,13	0,13	1,06	1,01	Memoria: zapato
22	-2,44	0,27	0,75	0,17	Denominación visual: bolígrafo
23	-2,75	0,29	0,91	0,33	Denominación visual: reloj
24	-0,66	0,17	1,45	1,89	Repetición frase
25	-0,10	0,16	1,82	5,10	Comprensión verbal: orden 1
26	-0,63	0,17	1,10	1,27	Comprensión verbal: orden 2
27	-0,54	0,17	1,24	2,00	Comprensión verbal: orden 3
28	-1,02	0,19	1,02	0,93	Comprensión escrita: leer-ejecutar
29	-0,31	0,17	0,98	0,68	Escribir-frase
30	0,83	0,14	1,32	1,25	Dibujo

**Tabla 2** DIF: comparación, contraste, ítem, etiqueta y grupo perjudicado

Comparación	t	Ítem	Etiqueta	Difícil
EP-DTA	-4,26	6	Orientación espacial: región	DTA
EP-DTA	-4,09	7	Orientación espacial: provincia	DTA
EP-DTA	-6,01	10	Orientación espacial: piso	DTA
N-DTA	-3,62	10	Orientación espacial: PISO	DTA
N-DTA	-3,69	17	Cálculo: 72	DTA
EP-DTA	6,25	15	Cálculo: 86	EP
EP-DTA	4,46	24	Repetición frase	EP
EP-DTA	5,98	30	Dibujo	EP
N-EP	-4,06	17	Cálculo: 72	EP
N-EP	-3,72	18	Cálculo: 65	EP
N-EP	-4,76	30	Dibujo	EP
N-DTA	3,94	14	Cálculo: 93	N
N-DTA	3,87	19	Memoria: lápiz	N
N-DTA	8,56	25	Comprensión verbal: orden 1	N
N-EP	4,76	6	Orientación espacial: región	N
N-EP	6,09	19	Memoria: lápiz	N
N-EP	7,29	25	Comprensión verbal: orden 1	N

$p < 0,00055$  en todas las comparaciones ( $\alpha/n.$  de comparaciones).

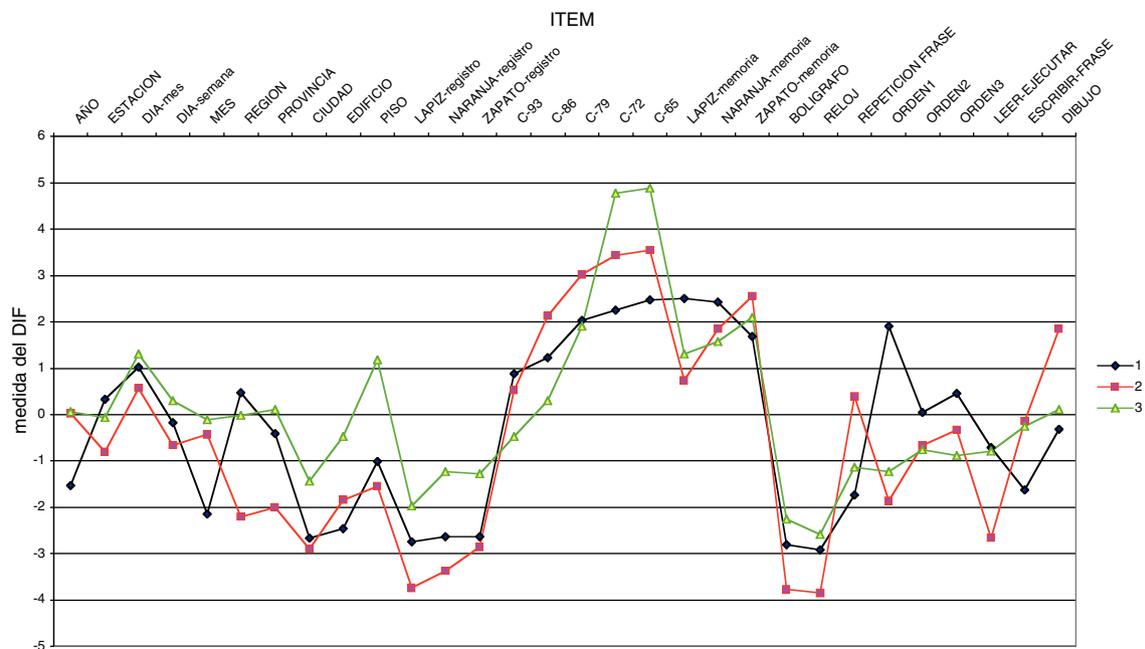


Nota: cada "#" representa 6 sujetos; cada "." de 1 a 5 sujetos.

**Figura 1** Mapa de la variable: personas e ítems escalados conjuntamente. Cada «#» representa 6 sujetos; cada «.» de 1 a 5 sujetos.

con EP que para los DTA aun cuando tengan el mismo nivel en la variable latente; 2 ítems de cálculo y el dibujo son más difíciles para los pacientes con EP que para los sujetos normales con el mismo nivel en la variable latente. Finalmente, seis comparaciones centradas en 4 ítems muestran DIF contra los sujetos del grupo normal, aunque uno de ellos desajusta severamente y tiene un bajo índice de discriminación. Todas las diferencias señaladas son estadísticamente significativas tras la corrección de Bonferroni, que corrige

el nivel  $\alpha$  para tener en cuenta el número de comparaciones realizadas, que en este estudio han sido noventa (30 ítems x 3 grupos). En total, 17 de las 90 comparaciones arrojan resultados estadísticamente significativos, que afectan a 11 ítems. Hay también un patrón claro en el tipo de ítems que perjudican a cada uno de los grupos con patología: los ítems de orientación espacial perjudican más al grupo DTA, los ítems de cálculo y dibujo perjudican más al grupo EP.



**Figura 2** Estimadores de dificultad de los ítems para cada grupo controlando el nivel en la variable latente. Los valores 1, 2 y 3 corresponden a N, EP y DTA, respectivamente.

**Discusión**

A diferencia de la TCT, todos los modelos de la TRI permiten cuantificar en la misma métrica el nivel de los ítems y de las personas. Los resultados de este estudio muestran que los ítems del MMSE presentan una alta variabilidad como indicadores de deterioro cognitivo, siendo los ítems más difíciles los de cálculo y memoria, y los más fáciles los de registro y denominación visual. Es decir, la respuesta incorrecta a estos últimos indica un mayor grado de deterioro que a los primeros. Las personas estudiadas presentan asimismo una alta variabilidad, existiendo diferencias grandes y significativas entre sujetos normales, EP y DTA. Aunque el interés diagnóstico de estos datos parezca indudable, se ha de tener en cuenta que, tanto las diferencias entre los ítems como entre los grupos de personas pueden estar distorsionadas por los ítems con funcionamiento diferencial. Las pruebas de cribado se emplean para clasificar a las personas en categorías «patológicas» en función de la comparación de la puntuación de la persona con puntos de corte determinados empíricamente. En el caso del MMSE, el procedimiento de clasificación sólo tendría una validez generalizable si las personas con la misma puntuación tuviesen el mismo nivel de deterioro cognitivo independientemente de su pertenencia a grupos afectados por distintas patologías. Es decir, el MMSE debería funcionar de forma similar (tener las mismas propiedades métricas) en distintos grupos de sujetos. En caso contrario, las medidas de distintos grupos no serían comparables. Este sería el caso de la imposible comparación de la estatura de varones y mujeres si el metro no tuviese las mismas propiedades (funcionase diferencialmente) en ambas poblaciones.

Uno de los procedimientos más usuales para analizar la invarianza métrica de un instrumento entre poblaciones es

el que se ha empleado en este trabajo: el análisis del funcionamiento diferencial de los ítems (DIF) entre sujetos normales y pacientes con DTA y EP. Los resultados indican que 11 de los 30 ítems del MMSE presentan algún tipo de DIF, lo que implica que una determinada puntuación en la prueba puede no ser indicadora del mismo grado de deterioro cognitivo en sujetos de DTA y EP. La comparación cuantitativa sólo tiene sentido si no existe evidencia en contra de la invarianza métrica. Dado el tipo de ítems que muestran DIF contra el grupo EP (repetición de una frase, expresión verbal de cálculos y dibujo), cabe pensar en factores que limitan la ejecución de esos ítems concretos, que no indican sólo deterioro cognitivo. Una explicación alternativa plausible podría ser la apatía, que ha comenzado a reconocerse como uno de los síntomas más relevantes en el diagnóstico diferencial<sup>18</sup>. De ser así, el MMSE podría estar sobreestimando el nivel de deterioro cognitivo de estos pacientes, por lo que no sería el instrumento más indicado para el cribado de la demencia en pacientes con la enfermedad de Parkinson o en personas en las que se sospeche su existencia. Pese a lo extendido del uso del MMSE, el funcionamiento de 11 de sus 30 ítems no es igual para los distintos grupos por lo que la comparación de las puntuaciones en distintos grupos no estaría justificada.

Aunque se ha propuesto el uso de otros tests basados en la TCT para el cribado de la demencia en los pacientes de EP<sup>19</sup>, esta solución dificulta aún más la comparación entre puntuaciones. La solución psicométricamente más satisfactoria pasa por emplear una prueba con validez generalizada, en la que no existan ítems que perjudiquen a los sujetos de un grupo por razones ajenas a la variable latente. El modelo de Rasch es la herramienta adecuada para construir y validar una prueba con tales características.

## Financiación

Este proyecto ha sido financiado por la Junta de Castilla y León (ref.: SA057A08).

## Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

## Bibliografía

1. 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.
2. Lobo A, Ezquerro J, Gómez F, Sala F, Seva A. El mini-examen cognoscitivo. Un test sencillo, práctico, para detectar alteraciones intelectivas en pacientes médicos. *Actas Luso Esp Neurol Psiquiatr Cienc Afines.* 1976;7:189–202.
3. Guerreiro M, Silva AP, Botelho M, Leitão O, Castro-Caldas A, Garcia C. Adaptação à população portuguesa da tradução do Mini Mental State Examination (MMSE). *Rev Port Neurol.* 1994;1:9.
4. Ishizaki J, Meguro K, Ambo H, Shimada M, Yamaguchi S, Hayasaka C, et al. A normative, community based study of Mini-Mental State in elderly adults: the effect of age and educational level. *J Gerontol B Psychol Sci Soc Sci.* 1998;53:359–63.
5. Sahadevan S, Lim P, Tan J, Chan S. Diagnosis performance of two mental status tests in the older Chinese: influence of education and age on cutoff values. *Int J Geriatr Psychiatry.* 2000;15:234–41.
6. Gungen C, Ertan T, Eker E, Yasar R, Engin F. Reliability and validity of the standardized Mini-Mental State Examination in the diagnosis of mild dementia in Turkish population. *Turk Psikiyatri Derg.* 2002;13:273–81.
7. Noale M, Limongi F, Minicuci N. Identification of factorial structure of MMSE based on elderly cognitive destiny: the Italian Longitudinal Study on Aging. *Dement Geriatr Cogn Disord.* 2006;21:233–41.
8. Werner P, Heinik J, Mendel A, Reicher B, Bleich A. Examining the reliability and validity of the hebrew version of the Mini-Mental State Examination. *Aging.* 1999;11:329–34.
9. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Ins.* 1959;22:719–48.
10. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research; 1960.
11. Prieto G, Delgado AR, Perea MV, Ladera V. Scoring Neuropsychological Tests Using the Rasch Model: an illustrative example with the rey-osterrieth complex figure. *Clin Psychol.* 2009;24:45–56.
12. Cadavid N, Delgado AR, Prieto G. Construcción de una escala de depresión con el modelo de Rasch. *Psicothema.* 2007;19:515–21.
13. Linacre JM. A user's guide to WINSTEPS & MINISTEPS: Rasch Model computer programs. Chicago: Winsteps.com; 2006.
14. Prieto G, Delgado AR. Fiabilidad y validez. *Pap Psicol.* 2010;31:67–74.
15. Gibbs WR, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *J Neurol Neurosurg Psychiatry.* 1988;51:745–52.
16. Mchann G, Drachan D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADR Work group under the auspice of Departamente of Health and Human Services Task Force on Alzheimer's disease. *Neurology.* 1984;34:939–44.
17. Hughes CP, Berg L, Danzinger WL. A new clinical scale for the staging of dementia. *Br J Psychiatry.* 1988;43:2412–3.
18. García-Ramos R, Villanueva C, Del Val J, Matías-Guío J. Apatía en la enfermedad de Parkinson. *Neurología.* 2010;25:40–50.
19. Parrao-Díaz T, Chaná-Cuevas P, Juri-Claverías C, Kunstmann C, Tapia-Núñez J. Evaluación del deterioro cognitivo en una población de pacientes con enfermedad de Parkinson mediante el test minimental Parkinson. *Rev Neurol.* 2005;40:339–44.