



Revista de Psicología

ISSN: 0716-8039

revista.psicologia@facso.cl

Universidad de Chile

Chile

Junco, Juan Ignacio; Prieto, Gerardo

Análisis del test neuropsicológico Addenbrooke's Cognitive Examination mediante el Modelo de Rasch

Revista de Psicología, vol. 23, núm. 1, 2014, pp. 40-52

Universidad de Chile

Santiago, Chile

Disponible en: <http://www.redalyc.org/articulo.oa?id=26432004006>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Análisis del test neuropsicológico *Addenbrooke's Cognitive Examination* mediante el Modelo de Rasch

Analysis of the Addenbrooke's Cognitive Examination using the Rasch Model

Juan Ignacio Junco y Gerardo Prieto

Universidad de Salamanca, España

Resumen

El *Addenbrooke's Cognitive Examination* (ACE) es un test cognitivo breve diseñado para detectar y diferenciar entre demencia tipo Alzheimer y demencia frontotemporal, recientemente validado en España. En el presente estudio se evaluaron las propiedades psicométricas de la versión española del ACE mediante el modelo de Rasch y se comparó la eficacia discriminativa de esta prueba con el *Mini-Mental State Examination* (MMSE). La prueba se aplicó a un grupo de 103 participantes, 53 de los cuales tenían establecido un diagnóstico de demencia. Los resultados indican que, en general, el ACE posee adecuadas propiedades psicométricas en cuanto ajuste y precisión de las medidas, aunque el rendimiento en la prueba se ve influido por algunas características sociodemográficas de los participantes (sexo y nivel de escolarización). Más aún, en comparación con el MMSE, los ítems del ACE permiten identificar un conjunto más amplio de síntomas de deterioro cognitivo y presenta mayor poder de discriminación. En conclusión, la versión española del ACE parece ser una herramienta adecuada para el diagnóstico de demencia, aunque son necesarias más investigaciones para seguir profundizando en sus características psicométricas.

Palabras clave: *Addenbrooke's Cognitive Examination*, ACE versión española, modelo de Rasch, demencia, evaluación cognitiva.

Abstract

The *Addenbrooke's Cognitive Examination* (ACE) is a brief cognitive test, recently validated in Spain, designed to distinguish between Alzheimer's disease and frontotemporal dementia. In this study we evaluated the psychometric properties of the Spanish version of ACE using the Rasch model and the discriminative performance of the test was compared with the *Mini-Mental State Examination* (MMSE). The test was applied to a group of 103 participants, 53 of whom had previously received a diagnosis of dementia. Results indicate that, in general, ACE has adequate psychometric properties for test items and the overall measurement, although test performance is influenced by some sociodemographic characteristics of participants (gender and educational level). Moreover, compared with the MMSE, ACE items identify a broader set of cognitive impairment symptoms and it has greater discrimination power. In conclusion, the Spanish version of the ACE seems to be a suitable tool for the diagnosis of dementia, although further research is needed to deepen its psychometric characteristics.

Keywords: *Addenbrooke's Cognitive Examination*, ACE Spanish version, Rasch model, dementia, cognitive assessment.

Agradecemos al Dr. García-Caballero por facilitarnos la versión española del ACE y la colaboración del Centro de Referencia Estatal de Alzheimer de Salamanca por permitirnos recoger datos para la presente investigación.

Contacto: G. Prieto. Facultad de Psicología. Universidad de Salamanca (España). Avenida de la Merced, 109-131. 37005 Salamanca. España. gprieto@usal.es

Cómo citar este artículo:

Junco, J. I. y Prieto, G. (2014). Análisis del test neuropsicológico *Addenbrooke's Cognitive Examination* mediante el Modelo de Rasch. *Revista de Psicología*, 23(1), 40-52. doi: 10.5354/0719-0581.2014.32873

Introducción

El desarrollo económico global ha propiciado el aumento de la esperanza de vida de los ciudadanos en los países desarrollados, debido a lo cual ha cobrado un mayor interés tomar las medidas oportunas para mejorar la calidad de vida de las personas mayores. En este sentido, la detección precoz de las principales enfermedades que afectan a este grupo poblacional se ha convertido en un objetivo prioritario para mejorar el acceso a los tratamientos y establecer las medidas preventivas oportunas (Bermejo-Pareja et al., 2009; Contador, Fernández-Calvo, Ramos, Tapias-Merino y Bermejo-Pareja, 2010; Iturra-Mena, 2007; Ladera y Perea, 2012).

En España, las personas mayores de 65 años representan en torno al 16% de la población total (Pedro-Cuesta et al., 2009) y la prevalencia de la demencia se sitúa en torno al 6% de la población anciana (Pedro-Cuesta et al., 2009; World Health Organization [WHO], 2012). Debido a que la mayoría de estos casos son atendidos en primer lugar por los servicios de atención primaria, se hace primordial contar con instrumentos de cribado breves que permitan identificar adecuadamente a los adultos mayores con síntomas de demencia y que puedan ser integrados con normalidad en la práctica clínica del médico generalista (Bermejo-Pareja et al., 2009; Gifford y Cummings, 1999; Sarasola, Calcagno, Sabe, Cabañero y Manes, 2004; Zunzunegui Pastor et al., 2003).

Si bien se han desarrollado diversos test de cribado o tamizaje, ninguno se ha convertido en la medida estándar (Gifford y Cummings, 1999). Actualmente, uno de los instrumentos más utilizados internacionalmente es el *Mini-Mental State Examination* (MMSE; Folstein, Folstein y McHugh, 1975) que evalúa seis dominios cognitivos: orientación temporal, orientación espacial, registro de información, atención y cálculo, recuerdo y lenguaje. El empleo del MMSE ha sido criticado por su baja sensibilidad en la detección de pacientes con demencia leve (Contador et al., 2010; Tierney, Szalai, Dunn, Geslani y McDowell, 2000), por su incapacidad de detectar síntomas frontales (Gregory, Orrell, Sahakian y Hodges, 1997) y porque el desempeño en la prueba puede verse influido por la edad y el nivel cultural del examinado (Tombaugh y McIntyre, 1992).

Para superar estas dificultades, el grupo de Mathuranath desarrolló el test breve *Addenbrooke's Cognitive Examination* (ACE; Mathuranath, Nestor, Berrios, Rakowicz y Hodges, 2000), el cual integra los ítems del MMSE y amplía la valoración de las capacidades mnésicas, frontales, de lenguaje y visuoespaciales. Su investigación pone de manifiesto que la prueba es capaz de diferenciar los diferentes subtipos de demencia, especialmente entre demencia tipo Alzheimer y de-

mencia frontotemporal (Mathuranath, et al., 2000). Por otra parte, se han realizado adaptaciones de la prueba en otras culturas, confirmandose su utilidad para detectar demencias (Bier et al., 2005; Mathuranath et al., 2004; Sarasola et al., 2004).

En España, García-Caballero et al. (2006) llevaron a cabo un primer estudio de validación del ACE adaptando la prueba a la población española rural de bajo nivel educativo, con una muestra de 167 personas (95 sujetos como grupo clínico y 72 sujetos como controles). En su estudio encontraron que es un instrumento útil para el diagnóstico de demencia y para diferenciar entre demencia tipo Alzheimer y demencia frontotemporal. No obstante, el análisis de García-Caballero et al. (2006) fue llevado a cabo bajo los supuestos de la Teoría Clásica de los Test (TCT), un modelo ampliamente conocido y utilizado que no está exento de limitaciones al estudiar las propiedades psicométricas de un instrumento. Los modelos psicométricos más avanzados se inscriben en la denominada Teoría de Respuesta al Ítem (Embretson y Hershberger, 1999). En este trabajo optamos por el modelo de Rasch (1960), dado que permite estimar, en la misma dimensión latente, medidas intervalares de las personas y los ítems. La medición conjunta en la misma dimensión de las personas y los ítems facilita analizar su interacción, determinando la probabilidad de que una persona domine un ítem en función de la magnitud de la diferencia entre su nivel en la variable medida y la dificultad de ítem. Esta propiedad permite construir normas referidas a la variable latente, además de las tradicionales normas referidas al grupo. Además, el modelo no parte de la suposición de que el error estándar de medida tenga la misma magnitud en toda la variable (como en la TCT). Por el contrario, permite estimar la precisión de cada medida de la persona y el ítem, cuestión de sumo interés para estimar la fiabilidad de los puntos de corte que se emplean en la clasificación diagnóstica (e.g., demencia y no demencia).

Debido a estas ventajas (Masters, 2005; Prieto y Delgado, 2003; Wilson, 2005), se ha extendido el uso del modelo de Rasch en el campo de las ciencias de la salud (Chachamovich, Fleck y Power, 2010), particularmente en el análisis de las pruebas neuropsicológicas (Conrad y Smith, 2004; Prieto, Contador, Tapias-Merino, Mitchell y Bermejo-Pareja, 2012; Prieto, Delgado, Perea y Ladera, 2010; Tennant et al., 2004).

El objetivo del presente estudio fue evaluar las propiedades psicométricas de la versión española del ACE en una población de ancianos, utilizando el modelo de Rasch (1960). En primer lugar, se analizó el ajuste de los datos al modelo y la fiabilidad de las medidas de los ítems y de las personas. En segundo lugar, se analizó el rendimiento en la prueba aso-

ciado a las características sociodemográficas de la muestra. Por último, se verificó la eficacia discriminativa del ACE en comparación con el MMSE.

Método

Participantes

La muestra se compone de un total de 103 personas (64 hombres y 39 mujeres), entre los 41 y los 95 años de edad. Como criterios de inclusión en el grupo clínico, los participantes debían contar con un diagnóstico previo de demencia, con un nivel de demencia entre leve y moderado (GDS entre 3 y 4 puntos o MMSE entre 10 y 26 puntos) y habilidades básicas de lectoescritura. De esta manera, del total de 103 personas, 55 presentaban el diagnóstico clínico de demencia: 23 personas tenían un diagnóstico de demencia tipo Alzheimer (22.3%), 12 personas tenían diagnóstico de demencia mixta (11.7%), siete personas tenían un diagnóstico de demencia vascular (6.8%), seis personas tenían un diagnóstico de demencia frontotemporal (5.8%), tres personas tenían un diagnóstico de demencia por cuerpos de Lewy (2.9%) y dos personas tenían un diagnóstico de demencia asociada a esclerosis múltiple (1.9%). Debido al pequeño tamaño de la submuestra clínica y a su heterogeneidad, todos los participantes del mismo fueron categorizados bajo una misma categoría (demencia) al efecto de analizar la utilidad diagnóstica del ACE en base a su eficiencia para determinar la discriminación con el grupo de control. Respecto al grupo de control, los criterios de inclusión fueron que los participantes tuvieran habilidades de lectoescritura básicas y que no tuvieran un diagnóstico previo de demencia. Se procuró emparejar a los participantes en cuanto a edad y sexo con el grupo clínico.

Además de las variables evaluadas por el ACE, se recogieron los datos sociodemográficos de todos los participantes (edad, sexo, nivel educativo, profesión y diagnóstico). Las características clínicas y sociodemográficas de la muestra se presentan en la tabla 1.

Instrumento

Para el desarrollo de esta investigación se utilizó la versión española del ACE (García-Caballero et al., 2006), que incluye adaptaciones en los ítems de aprendizaje verbal, recuerdo diferido, memoria semántica, repetición y lectura, y la substitución de los ítems del MMSE por la correspondiente versión española de la prueba (Lobo, Marcos, Día y De la Cámara, 1995). Los ítems del MMSE incluidos en el ACE aparecen marcados con un asterisco en la primera columna de la tabla 2. El ACE evalúa 6 dominios cognitivos, siendo la puntuación máxima equivalente a 100 puntos: orientación (10 ítems), atención (8 ítems), memoria (35 ítems), fluencia verbal (14 ítems), lenguaje (28 ítems) y habilidades visuoespaciales (5 ítems). Se asigna un punto a cada respuesta correcta. Según los autores de la versión española, la sensibilidad y especificidad de la prueba se sitúa en 90% y 86%, respectivamente, para un punto de corte de 68 puntos.

Procedimiento

Para la adquisición de los datos del grupo clínico se solicitó la colaboración de diferentes centros de día y centros residenciales; mientras que los datos del grupo control se obtuvieron por muestreo de conveniencia. Todos los participantes fueron entrevistados individualmente.

Análisis de datos

Los datos de la muestra se analizaron bajo los supuestos del modelo dicotómico de Rasch. Como ya se ha indicado con anterioridad, este modelo permite la cuantificación de los parámetros para las personas y los ítems en la misma escala y una representación conjunta en la misma dimensión. Los valores de los parámetros se expresan en una unidad de medida especial denominada *Logit* (Log-odd), que se expresa en la fórmula (1), explicada más adelante. Aunque la escala *logit* puede adoptar valores entre más y menos infinito, la gran mayoría de los casos se sitúa en el rango ± 5 . La localización del punto 0 de la escala es arbitraria, situándose en dicho punto la dificultad media de los ítems (Prieto y Velasco, 2011).

Tabla 1

Características sociodemográficas de la muestra

	Demencia (<i>n</i> = 53)	No demencia (<i>n</i> = 50)
Edad (media y DT)	74.62 (11.13)	72.26 (11.63)
Sexo		
Mujeres	25 (47.2%)	14 (28%)
Hombres	28 (52.8%)	36 (72%)
Educación		
≤ 14 años de escolarización	32 (60.4%)	16 (32%)
> 14 años de escolarización	21 (39.6%)	34 (68%)

Nota: DT = Desviación típica.

La fundamentación del modelo dicotómico de Rasch puede ser resumida en los siguientes puntos:

1. El atributo que se desea medir puede representarse en una única dimensión donde se sitúan conjuntamente ítems y personas;
2. El nivel de la persona en el atributo y la dificultad del ítem son los únicos determinantes de la probabilidad de que la respuesta sea correcta. Si el control de la situación es adecuado, esta expectativa es razonable y puede ser representada matemáticamente a través del modelo.

Para modelar esta relación, Rasch parte de la siguiente fórmula:

$$\ln(P_{is} / 1 - P_{is}) = (\theta_s - \beta_i) \quad (1)$$

La ecuación establece que el cociente entre la probabilidad de una respuesta correcta y la probabilidad de una respuesta incorrecta a un ítem dado ($P_{is} / 1 - P_{is}$), es una función de la diferencia en el atributo entre el nivel de la persona (θ_s) y el nivel del ítem (β_i) (Prieto y Delgado, 2003). Así, el modelo de Rasch se basa en el supuesto de que la probabilidad de una respuesta depende básicamente de la diferencia en el atributo entre la persona y el ítem, esto es la propiedad de unidimensionalidad. Es decir, personas *con el mismo nivel* han de tener la misma probabilidad de una respuesta a un ítem, independientemente de su pertenencia a distintos grupos sociales, culturales, patológicos, etc.

La formulación más conocida del modelo de Rasch (2), se deriva de la predicción de la probabilidad de una respuesta al ítem a partir de la diferencia en el atributo entre el nivel de la persona (θ_s) y el nivel del ítem (β_i) y con una base de logaritmos naturales (e) (Prieto y Delgado, 2003):

$$P_{is} = e^{(\theta_s - \beta_i)} / 1 + e^{(\theta_s - \beta_i)} \quad (2)$$

Las ventajas de este modelo solo pueden ser obtenidas si, cumpliéndose los supuestos del procedimiento, los datos empíricos se ajustan al modelo. Para comprobar el ajuste, se calculan los estadísticos basados en residuos (diferencias entre las respuestas observadas y las esperadas), empleándose la media no ponderada de los residuos estandarizados al cuadrado (*Outfit*) y la media de estos residuos ponderados con su varianza (*Infit*). Cuando los valores de estos indicadores oscilan entre 0.5 y 1.5 se considera que los datos tienen un ajuste adecuado al modelo, mientras

que los valores mayores a 2.0 indican un desajuste severo (Linacre, 2011). Por tanto, cuando estos estadísticos se mantienen por debajo de 2, podemos concluir que la prueba mide el atributo con una aceptable calidad psicométrica.

Adicionalmente al empleo de un análisis *omnibus* del ajuste al modelo como el realizado con los estadísticos *Infit* y *Outfit*, se ha generalizado en los últimos años la utilización de contrastes específicos del supuesto más importante del modelo: la unidimensionalidad. En el contexto del modelo de Rasch, la unidimensionalidad suele analizarse mediante el Análisis de Componentes Principales (ACP) de los residuos (Chow y Wang, 2010). Aunque el logro de una unidimensionalidad perfecta es muy difícil en la práctica (Zickar y Broadfoot, 2009), el ACP de los residuos permite evaluar si la dimensión extraída con el modelo de Rasch explica una cantidad suficiente de la varianza de las observaciones y si en la varianza no explicada (residual) existe algún factor o dimensión adicional con suficiente entidad para amenazar la validez de las puntuaciones obtenidas. Suele afirmarse que los datos son esencialmente unidimensionales si la medida Rasch da cuenta de un porcentaje de la varianza moderadamente alto (al menos un 20% en opinión de Reckase, 1979) y si además el porcentaje de la varianza residual correspondiente al autovalor del primer componente de los residuos es bajo (Chou y Wang, 2010; Smith y Miao, 1994).

La precisión de los parámetros se cuantifica mediante el *error estándar*, que es la desviación típica de los errores al estimar el parámetro de la persona o del ítem y permite evaluar la precisión de un test en cada nivel del rasgo. Además es posible cuantificar la fiabilidad de forma similar a la TCT mediante los estadísticos denominados *Person Separation Reliability* (PSR, fiabilidad de las medidas de los sujetos), e *Item Separation Reliability*, (ISR, fiabilidad de las medidas de los ítems) (Schumacker, 2004). Estos estadísticos indican la proporción de la varianza observada en un grupo que se corresponde con la varianza verdadera.

Los datos se analizaron con el programa WINSTEPS (Wright y Linacre, 1998). Además de estimar los estadísticos habituales de Rasch (parámetros de las personas e ítems, estadísticos de ajuste, etc.), se utilizó este programa para evaluar la unidimensionalidad de los datos mediante ACP de los residuales para analizar la asociación de las puntuaciones de los participantes con sus características sociodemográficas y para comparar la eficiencia diagnóstica del ACE con la del MMSE (fijando la dificultad de los ítems del MMSE en la obtenida al calibrar el ACE).

Resultados

Análisis del ajuste de los ítems y las personas

En lo que se refiere a la inspección de la unidimensionalidad de las medidas, el ACP de los residuos mostró que el porcentaje de la varianza de las observaciones explicada por el componente extraído por el modelo de Rasch es moderadamente alto (45.4%) y superior al valor mínimamente aceptable según el criterio de Reckase (1979). Además, el porcentaje de la varianza de los residuos explicado por el primer componente es muy bajo (4.8%). En consecuencia, parece razonable la consideración de que las medidas son fundamentalmente unidimensionales. Este hallazgo es semejante al obtenido en

los análisis mediante el modelo de Rasch de otras pruebas de rastreo de la demencia cuyo contenido es muy similar al del ACE, tales como el MMSE (Prieto et al., 2012; Tennant et al., 2004) y el test *Montreal Cognitive Assessment* (MoCA; Freitas, Prieto, Simões y Santana, 2014). En dichos estudios se observó que bastaba una dimensión básica para estimar de forma precisa el nivel de las personas y de los ítems. Esta unidimensionalidad fundamental justificaría representar el rendimiento de los evaluados mediante una única puntuación y utilizar un punto de corte para clasificarlos en las categorías diagnósticas de *normal* o *demencia probable*.

En la tabla 2 se pueden observar los estadísticos de ajuste de cada uno de los ítems del cuestionario.

Tabla 2

Ítems del cuestionario con sus estadísticos principales

Ítems	Dificultad	Error estándar	Infit	Outfit	P _d	P _{nd}	P _{nd} - P _d
Orientación							
Temporal							
1. Año*	-0.97	0.33	1.00	0.52	.79	.99	.20
2. Estación*	-0.56	0.31	1.56	7.06	.71	.99	.28
3. Mes*	-0.03	0.29	0.91	1.83	.59	.98	.39
4. Día de la semana*	0.84	0.28	1.12	1.10	.38	.96	.58**
5. Fecha*	0.30	0.28	1.09	2.17	.51	.98	.47
Espacial							
6. País*	-2.12	0.45	0.83	0.29	.92	1.00	.08
7. Región/Provincia*	-1.76	0.40	0.89	0.41	.89	1.00	.11
8. Ciudad*	-1.93	0.43	0.98	0.46	.91	1.00	.09
9. Hospital*	-0.20	0.29	0.88	1.71	.63	.99	.36
10. Piso*	0.14	0.28	0.98	1.38	.55	.98	.43
Atención							
11. Peseta*	-3.37	0.74	1.10	0.74	.98	1.00	.02
12. Caballo*	-2.12	0.45	1.23	0.96	.92	1.00	.08
13. Manzana*	-2.60	0.54	1.23	1.35	.95	1.00	.05
14. Cálculo 1*	-0.97	0.33	1.52	1.21	.79	.99	.20
15. Cálculo 2*	0.06	0.29	1.24	0.82	.57	.98	.41
16. Cálculo 3*	-0.03	0.29	1.11	1.99	.59	.98	.39
17. Cálculo 4*	0.77	0.28	1.11	1.00	.39	.97	.58**
18. Cálculo 5*	0.92	0.28	1.14	1.61	.36	.96	.60**
19. O*	-0.37	0.30	1.27	2.30	.67	.99	.32
20. D*	0.54	0.28	1.02	1.77	.45	.97	.52**
21. N*	1.53	0.28	1.11	1.17	.23	.93	.70**
22. U*	1.68	0.28	0.70	0.53	.21	.92	.71**
23. M*	1.37	0.27	0.76	0.58	.26	.94	.68**
Memoria							
Recuerdo diferido 1							
24. Peseta*	1.75	0.28	1.02	0.94	.19	.91	.72**
25. Caballo*	1.30	0.27	1.12	0.94	.27	.94	.67**
26. Manzana*	2.37	0.28	0.89	0.68	.12	.85	.73**
Curva de aprendizaje							
27. José (1º ensayo)	-0.37	0.30	1.03	0.59	.67	.99	.32
28. Álvarez (1º ensayo)	0.38	0.28	0.87	0.56	.49	.98	.49
29. Gran (1º ensayo)	1.07	0.27	1.06	1.00	.32	.95	.63**
30. Vía (1º ensayo)	1.07	0.27	1.06	1.00	.32	.95	.63**
31. 36 (1º ensayo)	2.14	0.28	1.64	1.74	.14	.88	.74**
32. Primero (1º ensayo)	0.84	0.28	0.91	0.71	.38	.96	.58**
33. Valencia (1º ensayo)	0.61	0.28	0.91	0.66	.43	.97	.54**
34. José (2º ensayo)	-0.86	0.33	0.74	0.37	.77	.99	.22
35. Álvarez (2º ensayo)	-0.20	0.29	0.74	0.45	.63	.99	.36
36. Gran (2º ensayo)	-0.20	0.29	0.78	0.44	.63	.99	.36
37. Vía (2º ensayo)	-0.20	0.29	0.78	0.44	.63	.99	.36
38. 36 (2º ensayo)	0.77	0.28	0.90	1.01	.39	.97	.58**

Tabla 2 (continuación)

Ítems	Dificultad	Error estándar	Infit	Outfit	P _d	P _{nd}	P _{nd} - P _d
39. Primero (2º ensayo)	0.22	0.28	0.82	0.49	.53	.98	.45
40. Valencia (2º ensayo)	-0.28	0.30	1.03	0.60	.65	.99	.34
41. José (3º ensayo)	-1.76	0.40	1.18	0.54	.89	1.00	.11
42. Álvarez (3º ensayo)	-0.75	0.32	1.25	0.70	.75	.99	.24
43. Gran (3º ensayo)	0.06	0.29	0.74	0.45	.57	.98	.41
44. Vía (3º ensayo)	0.06	0.29	0.74	0.45	.57	.98	.41
45. 36 (3º ensayo)	-0.11	0.29	0.88	0.52	.61	.99	.38
46. Primero (3º ensayo)	0.06	0.29	0.71	0.41	.57	.98	.41
47. Valencia (3º ensayo)	-1.32	0.36	0.91	0.44	.84	1.00	.16
Recuerdo diferido 2							
48. José	2.06	0.28	1.07	1.05	.15	.88	.73**
49. Álvarez	2.29	0.28	0.94	0.79	.12	.86	.74**
50. Gran	1.75	0.28	0.56	0.39	.19	.91	.72**
51. Vía	1.75	0.28	0.56	0.39	.19	.91	.72**
52. 36	2.52	0.28	0.83	0.68	.10	.83	.73**
53. Primero	1.98	0.28	0.82	0.60	.16	.89	.73**
54. Valencia	1.53	0.28	0.92	0.72	.23	.93	.70**
Memoria retrógrada							
55. Presidente actual	0.14	0.28	0.91	0.60	.55	.98	.43
56. Presidente anterior	0.92	0.28	0.78	0.51	.36	.96	.60**
57. Líder de la oposición	1.53	0.28	0.78	0.58	.23	.93	.70**
58. Nombre del Papa	0.84	0.28	1.17	1.14	.38	.96	.58**
Fluidez Verbal							
59. Categoría	2.29	0.28	1.51	1.60	.12	.86	.74**
60. Letra P	2.29	0.28	1.48	1.73	.12	.86	.74**
Lenguaje							
Denominación							
61. Reloj*	-5.33	1.84	-	-	1.00	1.00	.00
62. Lápiz*	-2.93	0.61	0.87	0.19	.96	1.00	.04
63. Jirafa	-0.11	0.29	0.88	0.58	.61	.99	.38
64. Cometa	1.00	0.28	1.07	1.13	.34	.96	.62**
65. Cerdo	-1.32	0.36	0.85	0.76	.84	1.00	.16
66. Canguro	1.37	0.27	0.76	0.60	.26	.94	.68**
67. Corona	-0.20	0.29	1.35	2.46	.63	.99	.36
68. Barril	0.22	0.28	0.98	0.62	.53	.98	.45
69. Camello	-0.20	0.29	1.01	1.46	.63	.99	.36
70. Cabra	-0.97	0.33	0.92	0.72	.79	.99	.20
71. Helicóptero	1.00	0.28	1.01	1.03	.34	.96	.62**
72. Molino	-0.37	0.30	0.88	0.56	.67	.99	.32
Comprensión							
73. Señale puerta	-2.93	0.61	1.08	0.48	.96	1.00	.04
74. Señale techo	-2.93	0.61	1.07	0.75	.96	1.00	.04
75. Cierre ojos*	-0.97	0.33	1.13	0.85	.79	.99	.20
76. Papel con mano derecha*	-1.08	0.34	1.27	3.06	.80	.99	.19
77. Doblar por la mitad*	-1.60	0.39	1.35	3.05	.87	1.00	.13
78. Poner sobre mesa*	-2.12	0.45	0.88	0.48	.92	1.00	.08
79. Señale techo y puerta	-1.08	0.34	1.14	0.96	.80	.99	.19
80. Señale mesa y puerta	0.22	0.28	1.25	1.17	.53	.98	.45
Repetición							
81. Prueba	-2.93	0.61	1.15	0.93	.96	1.00	.04
82. Traer	-1.76	0.40	1.44	9.90	.89	1.00	.11
83. Impresionante	-1.93	0.43	1.01	0.62	.91	1.00	.09
84. Frase 1*	-1.76	0.40	1.23	0.72	.89	1.00	.11
85. Frase 2	1.15	0.27	0.88	0.69	.31	.95	.64**
Lectoescritura							
86. Lectura 1	-2.93	0.61	1.09	0.98	.96	1.00	.04
87. Lectura 2	0.92	0.28	0.78	0.61	.36	.96	.60**
88. Escritura*	0.84	0.28	0.98	0.86	.38	.96	.58**
Habilidades visuoespaciales							
A la copia							
89. Pentágonos *	0.06	0.29	0.87	0.52	.57	.98	.41
90. Cubo	1.45	0.28	0.62	0.49	.25	.93	.68**
A la orden							
91. Esfera reloj	-1.20	0.35	1.18	4.04	.82	.99	.17
92. Números reloj	1.53	0.28	0.98	0.92	.23	.93	.70**
93. Agujas reloj	1.83	0.28	0.93	0.88	.18	.91	.73**

Nota: Pd = Probabilidad de acertar el ítem las personas con un nivel medio en el Grupo con Demencia (Media de $\theta = 0.33$); Pnd = Probabilidad de acertar el ítem las personas con un nivel medio en el Grupo sin Demencia (Media de $\theta = 4.09$).

* Ítems del MMSE.

** Ítems más discriminativos entre ambos grupos.

Los resultados de los valores promedio y las desviaciones típicas de ajuste de los ítems ponen de manifiesto un buen ajuste global del ACE al modelo de Rasch ($Infit = 1.01$; $DT = 0.22$; $Outfit = 1.12$, $DT = 1.29$), variando los valores $Infit$ de los ítems entre 0.56 y 1.64. No obstante, si se toma en consideración el estadístico más sensible al desajuste ($Outfit$), es posible observar que 9 de los 93 ítems desajustan severamente con el modelo ($Outfit > 2$), lo cual es indicativo de un severo desajuste de las respuestas dadas a los mismos y pone en entredicho la capacidad de esos ítems para medir adecuadamente el nivel de las personas en una dimensión cuyos polos podrían ser identificados como *rendimiento cognitivo normal* y *deterioro cognitivo severo*. No se eliminaron los ítems desajustados, dado que solo sería aconsejable proceder a la eliminación si el desajuste se replicase en futuros estudios (Smith, 2004).

Por su parte, el 91.75% de los participantes presenta valores de $Infit$ y/o $Outfit$ por debajo de 2.0, y un promedio de ajuste aceptable ($Infit = 0.99$, $DT = 0.13$; $Outfit = 1.12$, $DT = 0.77$). Por tanto, el ajuste de los ítems y las personas al modelo es básicamente adecuado.

Análisis de los valores de ítems y personas

En la tabla 2 se muestran asimismo las estimaciones del parámetro de dificultad de cada ítem y su error estándar asociado. Se ha de notar que el ítem 61 (Denominación del Reloj) es un ítem que fue resuelto por todos los participantes, por lo que su eficiencia discriminativa es nula y su parámetro de dificultad es asignado por defecto con una alta imprecisión.

Excluyendo este ítem extremo (ítem 61), el promedio de los errores estándar y su desviación típica son relativamente bajos (0.33 y 0.10, respectivamente), por lo que los parámetros de los ítems fueron estimados con una fiabilidad aceptable ($ISR = .94$). El ítem 61 no se empleó para la estimación de los parámetros de las personas por tratarse de un ítem extremo (resuelto correctamente por todos los sujetos).

En la figura 1 se muestra el escalamiento conjunto de personas e ítems. Aun excluyendo el ítem 61 resuelto correctamente por todos los participantes, los valores de los ítems en la escala muestran una gran variabilidad ($DT = 1.46$), oscilando entre 2.52 y -3.37 *logits*. Esto sugiere que los ítems del ACE pueden identificar un amplio rango de síntomas de deterioro cognitivo.

La mayor dificultad corresponde a los ítems de la curva de aprendizaje, de recuerdo diferido, de fluidez verbal y el dibujo de las agujas del reloj. Por su parte, los ítems más fáciles

son los de fijación, comprensión de órdenes simples, el primer ítem de lectura, la repetición de la palabra “prueba” y los ítems de denominación del MMSE.

Mediante la ecuación 2 del modelo de Rasch, se ha calculado la probabilidad de acertar cada uno de los ítems en función del rendimiento promedio de cada grupo. Así, en la columna Pd de la tabla 2 encontramos la probabilidad que una persona con la puntuación promedio del grupo de demencia ($M = 0.33$) tiene de emitir una respuesta correcta a cada ítem, mientras que en la columna Pnd se encuentra esta probabilidad para una persona que tenga el rendimiento promedio del grupo control (i.e., $M = 4.09$). Al comparar ambas probabilidades (columna Pnd - Pd), comprobamos qué ítems son más eficientes para detectar la demencia: el ítem correspondiente al recuerdo del día de la semana (ítem 4), los dos últimos ítems de cálculo (ítems 17 y 18), los cuatro últimos ítems de deletreo inverso (ítems del 20 al 23), el recuerdo diferido de las tres palabras (ítems del 24 al 26), algunos elementos de la curva de aprendizaje (ítems del 29 al 33 e ítem 38), recuerdo diferido de la dirección (ítems del 48 al 54), algunos elementos de memoria retrógrada (ítems del 56 al 58), los ítems de fluidez verbal (59 y 60), algunos elementos de denominación (ítems 64, 66 y 71), la repetición de la segunda frase (ítem 85), el segundo apartado de lectura (ítem 87) y escritura (ítem 88), el dibujo del cubo (ítem 90) y el dibujo de los números y las agujas del reloj (ítems 92 y 93).

En la tabla 3 se muestran los estadísticos descriptivos de las personas, tanto para el total de la muestra como para cada grupo.

Las estimaciones globales de los parámetros de las personas son muy fiables en la muestra total ($PSR = .95$) y en el grupo clínico ($PSR = .92$). Como era esperable, la fiabilidad en el grupo de control es menor ($PSR = .51$) debido a la diferencia entre su nivel de rendimiento y la dificultad de los ítems. El nivel medio de rendimiento es de 2.16 *logits* ($DT = 2.17$), fluctuando este valor entre 6.55 y -1.60 *logits*. Esta gran variabilidad se debe a la presencia de personas con y sin demencia. Los resultados revelan que existe una diferencia de tamaño grande y significativa ($t(94) = 17.29$; $p < .001$; d de Cohen = 3.56) entre las medias de ambos grupos (clínico y control) en el rendimiento en la prueba. En la figura 1, podemos observar cómo la media de rendimiento del grupo control es claramente superior a la media de rendimiento del grupo de personas con demencia. Este dato evidencia la utilidad diagnóstica del ACE para la detección del deterioro cognitivo característico de la demencia.

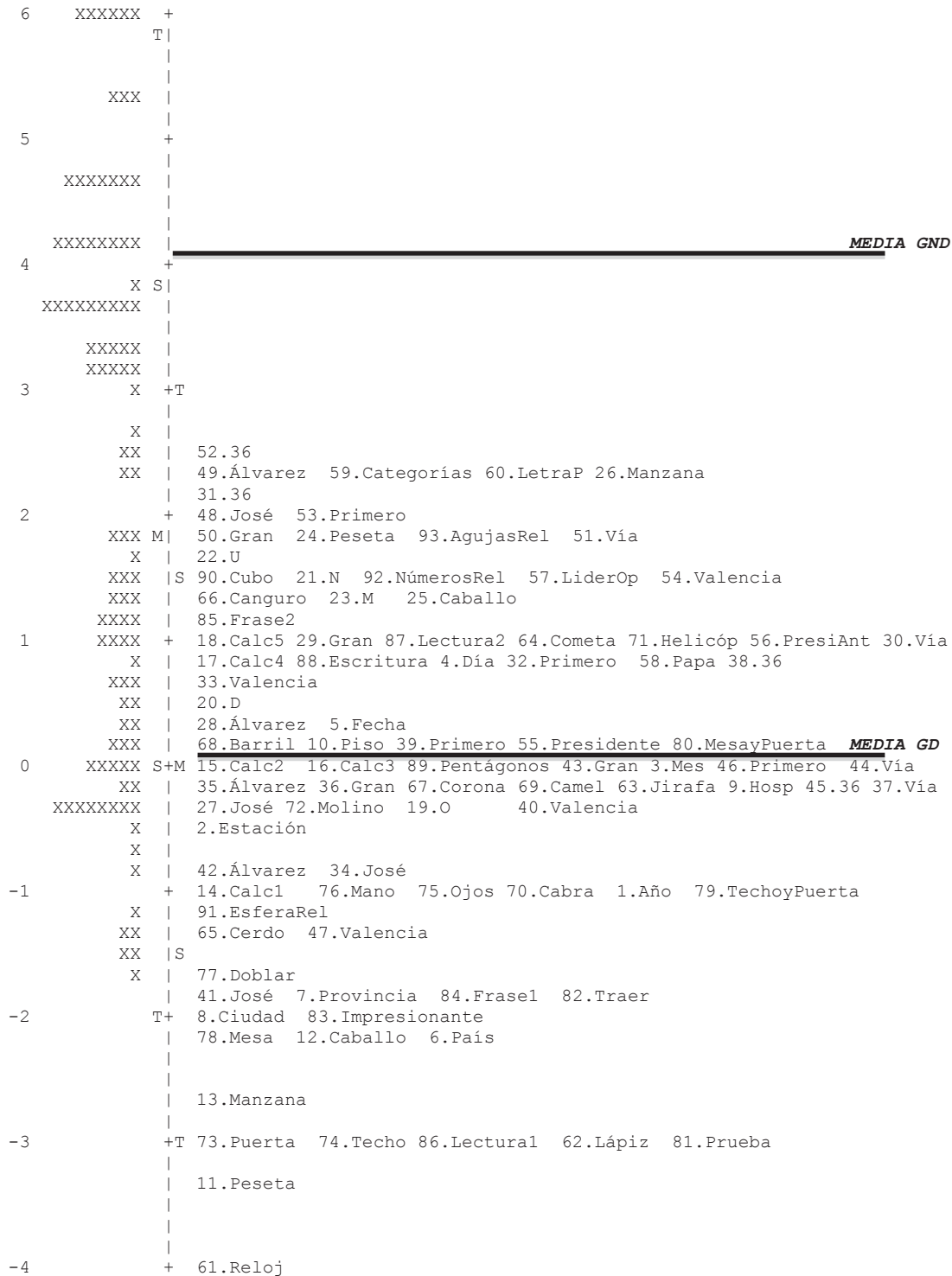


Figura 1. Distribución de los ítems del cuestionario (aparece su número y contenido en la parte derecha) y de los participantes (aparecen representados por las "x" en la parte izquierda) a lo largo del continuo de deterioro cognitivo. Las líneas horizontales representan las medias del grupo de personas con demencia (GD) y del grupo de personas sin demencia (GND).

Tabla 3*Valores de las personas en el ACE*

Estadísticos	X(t)	L(t)	EE(t)	X(nd)	L(nd)	SE(nd)	X(d)	L(d)	EE(d)
Media	69.2	2.16	0.47	88.2	4.09	0.70	51.2	0.33	0.26
DT	21.5	2.17	0.38	4.7	1.19	0.45	14.7	0.97	0.01
Máximo	93	6.55	1.84	93	6.55	1.84	79	2.33	0.32
Mínimo	23	-1.60	0.25	65	1.20	0.26	23	-1.60	0.25
PSR	-	.95	-	-	.51	-	-	.92	-

Nota: X = número de aciertos; L = puntuación logit; EE = error estándar de L; t = total de la muestra; nd = No Demencia; d = Demencia; DT = Desviación Típica; PSR = Person Separation Reliability.

Rendimiento en función de las características sociodemográficas: edad, sexo y nivel de escolarización

Para comprobar si el rendimiento en la escala se veía influenciado por las características sociodemográficas de la muestra, se realizaron comparaciones de los promedios de las puntuaciones *logits* en función de la edad (dicotomizada a partir de la mediana = 74 años), sexo y años de escolarización. En la tabla 4 se presentan los resultados. Exceptuando la edad, el resto de factores mostró diferencias significativas en las puntuaciones a un 95% de confianza. Así, los hombres obtienen mayores puntuaciones en el ACE que las mujeres ($t(72) = -3.03$; $p < .01$), y las personas con un nivel más alto de escolarización presentan un mayor rendimiento en ACE, en contraste con las personas de menor nivel educativo ($t(98) = -3.13$; $p < .05$).

Comparación del ACE y el MMSE: rango de medida, precisión y poder discriminativo

La valoración de la utilidad diagnóstica del ACE como herramienta de screening en demencias impone efectuar una comparación con el MMSE, utilizado en muchos países como el criterio predominante para un cribado de la enfermedad. Dado que el ACE incluye los ítems del MMSE, ha sido posible comparar ambas pruebas en tres aspectos: el rango del constructo que muestrean los ítems, la precisión de las medidas y la eficiencia de ambas pruebas para discriminar entre el grupo clínico y el de control. Para llevar a cabo esta comparación métrica, se calibraron las personas fijando la dificultad de los ítems del MMSE en los parámetros obtenidos en la calibración del ACE. Este procedimiento permite una equiparación de ambas escalas en la misma métrica

Tabla 4*Media, desviación típica y puntuación en el estadístico t de Student en función de la edad, el sexo y la educación*

	Media (logits)	DT (logits)	t
Edad			
≤ 74 años	2.47	2.39	1.48
> 74 años	1.84	1.87	
Sexo			
Mujer	1.33	2.23	-3.03**
Hombre	2.66	1.97	
Educación			
≤ 14 años	1.48	1.74	-3.13**
> 14 años	2.75	2.33	

Nota: * $p < .05$; ** $p < .01$.

Tabla 5*Estadísticos descriptivos de los ítems en el ACE y el MMSE*

Estadísticos	Dificultad (Media)	DT	Máximo	Mínimo	Infit (Media)	Outfit (Media)	ISR
ACE	0.00	1.46	2.52	-3.37	1.01	1.12	.94
MMSE	-0.38	1.46	2.37	-3.37	1.01	1.05	.93

Nota: DT = desviación típica de los parámetros de dificultad; ISR = Item Separation Reliability.

Tabla 6*Estadísticos descriptivos de las personas en el ACE y el MMSE*

Estadísticos	Total de la muestra	Demencia	No Demencia
ACE			
Media (logits)	2.16	0.33	4.09
DT	2.17	0.97	1.19
Máximo	6.55	2.33	6.55
Mínimo	-1.60	-1.60	1.20
PSR	.92	.92	.50
MMSE			
Media (logits)	2.06	0.38	3.83
DT	2.01	1.05	1.01
Máximo	5.22	2.69	5.22
Mínimo	-1.24	-1.24	1.50
PSR	.79	.80	.00

Nota: DT = desviación típica; PSR = Person Separation Reliability.

(Prieto y Velasco, 2011). En las tablas 5 y 6 aparecen los estadísticos de la muestra de ítems y personas en ambas pruebas.

En primer lugar, podemos observar que la dificultad promedio de los ítems del MMSE se sitúa por debajo de la dificultad del ACE (en -0.38 logits). Además, la localización de los ítems del MMSE se sitúa entre 2.37 y -3.37 logits, mientras que los valores de los ítems del ACE oscilan entre 2.52 y -3.37 . En consecuencia, puede deducirse que los ítems del MMSE muestrean un rango del constructo más limitado.

Respecto a la precisión de las medidas, ambas pruebas mantienen adecuados índices de fiabilidad de los ítems, disminuyendo en apenas una décima en el MMSE ($ISR = .93$).

No obstante, la fiabilidad de las estimaciones de los parámetros de las personas en el MMSE disminuye a 0.79 para el total de la escala y en el grupo clínico se sitúa en 0.80 . Para el grupo control, la estimación de los parámetros de persona no es fiable ($ISR = 0$) debido a que el nivel de rendimiento de las personas de este grupo es muy superior a la dificultad de los ítems del MMSE. Observamos también que la consistencia interna del MMSE (α de Cronbach = $.92$) es levemente menor a la del ACE (α de Cronbach = $.98$).

Por último, en cuanto a eficacia discriminativa los resultados revelan que el MMSE también permite objetivar diferencias significativas entre el grupo control y el grupo de demencia, aunque con un menor tamaño del efecto ($t(100) = 16.86$; $p < .001$; d de Cohen = 3.37) respecto al ACE ($t(94) = 17.29$; $p < .001$; d de Cohen = 3.56).

Discusión y conclusiones

El presente trabajo tenía como objetivos valorar la calidad de las medidas de la versión española del ACE por medio del modelo de Rasch y comprobar su utilidad diagnóstica como herramienta de screening en demencias. Hasta donde conocemos, este es el primer estudio que se hace en España de esta prueba siguiendo un modelo de la Teoría de Respuesta al Ítem. Gracias al modelo de Rasch, pudimos cuantificar la dificultad de los ítems y el nivel de competencia de las personas en una misma escala.

Para verificar la calidad de las medidas, analizamos el grado de ajuste de las respuestas a los supuestos del modelo y la precisión de la estimación de los parámetros, tanto para ítems como para personas. Las medias de los residuos permiten inferir un adecuado ajuste al modelo de los ítems y de las personas a nivel grupal. No obstante, algunos ítems presentan un desajuste severo, lo cual puede deberse a diversos factores como el funcionamiento diferencial de los ítems, su excesiva variación en el parámetro de discriminación de su curva característica, falta de motivación o cooperación de algunos participantes, errores al anotar la respuesta, el escaso tamaño muestral, etc. (c.f., Karabatsos, 2000). En el futuro sería conveniente realizar nuevos estudios con muestras más amplias para comprobar si el desajuste de esos ítems se mantiene.

El examen de la unidimensionalidad de la prueba mediante el ACP de los residuos (Chou y Wang, 2010) muestra que el ACE pareciera ser una prueba fundamentalmente unidi-

mensional, dado que el porcentaje de la varianza observada explicado por la medida Rasch es moderadamente alto y el primer componente extraído explica un porcentaje muy bajo de la varianza residual. Esta unidimensionalidad básica se ha observado en varios estudios psicométricos de tests de cribado de la demencia muy empleados internacionalmente como el MMSE y el MoCA. Pese a que estas pruebas estén integradas por ítems que proceden de facetas de contenido diversas (orientación, atención, memoria, etc.), dichos subconjuntos de ítems no se comportan empíricamente como factores distintos que requieran la construcción de subescalas (Prieto et al., 2012; Tennant et al., 2004; Freitas et al., 2014). Esta unidimensionalidad básica justificaría la representación del rendimiento cognitivo de los evaluados mediante una única puntuación y la utilización de un punto de corte para clasificarlos en las categorías diagnósticas como *normal* o *demen- cia probable*. Sin embargo, se requiere de mayores estudios para comprobar si estos resultados se sostienen en muestras de mayor tamaño.

Respecto a la estimación de los parámetros, los resultados revelan que los ítems se estimaron con una adecuada fiabilidad. Igualmente, las estimaciones de los parámetros de las personas son muy fiables, especialmente en el grupo con demencia. Como era de esperar en un test clínico, la precisión de las medidas de los sujetos de control no alcanza tan alto nivel de precisión, ya que el nivel de dificultad de los ítems es bajo en este subgrupo. Debido al escaso número de personas en los subgrupos clínico y de control, será conveniente llevar a cabo estudios de confirmación de estas características métricas con muestras de mayor tamaño.

Asimismo, comprobamos las diferencias en el rendimiento en la prueba asociadas a las características sociodemográficas de las personas de la muestra. En este punto, encontramos algunas discrepancias con algunas de las investigaciones anteriores (Mathuranath et al., 2004; Sarasola et al., 2004). Si bien la puntuación total en la prueba parece no depender de la edad de los participantes, el sexo de los participantes sí influye en la puntuación total y encontramos que el nivel de escolarización de los participantes influye en el rendimiento en la prueba, lo cual se observa también en otras investigaciones (García-Caballero et al., 2006; Mathuranath et al., 2004; Sarasola et al., 2004). Se considera necesario por tanto, realizar nuevas investigaciones con muestras de mayor tamaño, que permitan evaluar si estos hallazgos son diferencias poblacionales o son resultado de funcionamiento diferencial de los ítems (DIF) y/o funcionamiento diferencial del test (DTF) asociado al sexo, nivel educativo y/o el diagnóstico.

Una vez evaluada la calidad métrica del ACE, el siguiente paso fue analizar su eficacia discriminativa y compararla con

una medida más tradicional como es el MMSE. A la luz de los resultados, podemos concluir que ambas pruebas son efectivas para la detección de demencia. No obstante, un análisis más detallado de los estadísticos de cada instrumento, permite observar que la dificultad media de los ítems del MMSE y especialmente su rango son menores que en el ACE. Estos datos sugieren que la eficacia del MMSE para detectar a las personas que se encuentran en las fases iniciales de la demencia sería menor que la del ACE. Es decir, el mayor rango de los ítems del ACE a lo largo del continuo evaluado permiten muestrear más adecuadamente la variabilidad de los síntomas de deterioro cognitivo.

Por otro lado, la diferencia entre las probabilidades de responder correctamente a los ítems entre las personas con nivel medio en el ACE en el grupo clínico y el de control, nos permitió identificar los ítems más eficientes para diferenciar entre el grupo de control y el grupo con demencia. Así, encontramos que los ítems con mayor capacidad para discriminar entre ambos grupos son, en su mayoría, aquellos relacionados con las áreas que se amplían en el ACE respecto al MMSE (memoria, la fluidez verbal y el lenguaje). Estos resultados son coherentes con investigaciones anteriores (Bier et al., 2005; García-Caballero et al., 2006; Mathuranath et al., 2004; Sarasola et al., 2004).

Como señala Mathuranath et al. (2004), el ACE consume más tiempo en su aplicación que otras pruebas de screening, pero ofrece mayor información respecto al rendimiento cognitivo de las personas. Las ampliaciones que incorpora esta prueba mejoran los aspectos que en el MMSE habían recibido críticas (Feher et al., 1992). En este sentido, la versión validada por el equipo de García-Caballero (2006) demuestra su utilidad como herramienta de cribado en el ámbito de las demencias.

En el futuro sería conveniente contar con una muestra más amplia para corroborar los resultados obtenidos en este estudio, incorporando más casos de demencia frontotemporal y otras patologías no relacionadas con la demencia, para seguir profundizando en el funcionamiento de este instrumento y comprobar su capacidad discriminativa en el diagnóstico de demencias (Dudas, Berrios y Hodges, 2005). Además, sería conveniente validar la nueva versión de la prueba, la ACE-R (Mioshi, Dawson, Mitchell, Arnold y Hodges, 2006), adaptándola a las características socioculturales de la población española. Por último, cabe señalar la necesidad de seguir utilizando modelos tipo Rasch para el análisis de medidas neuropsicológicas, en tanto que constituyen una herramienta poderosa para la selección de ítems adaptados al rendimiento de las personas y para comparar la calidad de instrumentos que evalúan el mismo constructo.

Referencias

- Bermejo-Pareja, F., Benito-Leon, J., Vega, S., Olazarán, J., De Toledo, M., Díaz-Guzmán, J., ... Román, G. C. (2009). Consistency of clinical diagnosis of dementia in NEDICES: A population-based longitudinal study in Spain. *Journal of geriatric psychiatry and neurology*, 22(4), 246-255. doi: 10.1177/0891988709335794
- Bier, J. C., Donckels, V., Van Eyll, E., Claes, T., Slama, H., Fery, P. y Vokaer, M. (2005). The French Addenbrooke's cognitive examination is effective in detecting dementia in a French-speaking population. *Dementia and geriatric cognitive disorders*, 19, 15-17. doi:10.1159/000080965
- Chachamovich, E., Fleck, M. P. y Power, M. (2010). Is Geriatric Depression Scale-15 a suitable instrument for measuring depression in Brazil? Results of a Rasch analysis. *Psychology, Health & Medicine*, 15, 596-606. doi:10.1080/13548506.2010.487108
- Chou, Y. y Wang, W. (2010). Checking dimensionality in Item Response Models with principal component analysis on standardized residuals. *Educational and Psychological Measurement*, 70, 717-731. doi: 10.1177/0013164410379322
- Conrad, K. J. y Smith, E. V., Jr. (2004). International conference on objective measurement: Applications of Rasch analysis in health care. *Medical care*, 42(1 Suppl.), 11-6. doi:10.1097/01.mlr.0000103527.52821.1c
- Contador, I., Fernández-Calvo, B., Ramos, F., Tapias-Merino, E. y Bermejo-Pareja, F. (2010). El cribado de la demencia en atención primaria. Revisión crítica. *Revista Neurología*, 51, 677-686.
- Dudas, R. B., Berrios, G. E. y Hodges, J. R. (2005). The Addenbrooke's Cognitive Examination (ACE) in the differential diagnosis of early dementias versus affective disorder. *American Journal of Geriatric Psychiatry*, 13(3), 218-226. doi: 10.1097/00019442-200503000-00007
- Embretson, S. E. y Hershberger, S. L. (1999). *The new rules of measurement: What every psychologist and educator should know*. Mahwah, NJ: LEA.
- Feher, E. P., Mahurin, R. K., Doody, R. S., Cooke, N., Sims, J. y Pirozzolo, F. J. (1992). Establishing the limits of the Mini-Mental State. Examination of «subtests». *Archives of neurology*, 49, 87-92. doi: 10.1001/archneur.1992.00530250091022
- Folstein, M. F., Folstein, S. E. y McHugh, P. R. (1975). «Mini-mental state». A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189-198.
- Freitas, S., Prieto, G., Simões, M. R. y Santana, I. (2014). Psychometric properties of the Montreal Cognitive Assessment (MoCA): An analysis using the Rasch model. *The Clinical Neuropsychologist*, 28, 65-83. doi: 10.1080/13854046.2013.870231
- García-Caballero, A., García-Lado, I., González-Hermida, J., Recimil, M. J., Area, R., Manes, F., ... Berrios, G. E. (2006). Validation of the Spanish version of the Addenbrooke's Cognitive Examination in a rural community in Spain. *International Journal of Geriatric Psychiatry*, 21, 239-245. doi:10.1002/gps.1450
- Gifford, D. R. y Cummings, J. L. (1999). Evaluating dementia screening tests methodologic standards to rate their performance. *Neurology*, 52(2), 224-224. doi:10.1212/WNL.52.2.224
- Gregory, C. A., Orrell, M., Sahakian, B. y Hodges, J. R. (1997). Can frontotemporal dementia and Alzheimer's disease be differentiated using a brief battery of tests? *International journal of geriatric psychiatry*, 12, 375-383. doi: 10.1002/(SICI)1099-1166(199703)12:3%3C375::AID-GPS518%3E3.0.CO;2-%23
- Iturra-Mena, A. M. (2007). Adaptación y validación preliminar de un test para el screening de demencia en Chile: el Eurotest. *Revista chilena de neuro-psiquiatría*, 45, 296-304. doi:10.4067/S0717-92272007000400005
- Karabatsos, G. (2000). A critique of Rasch residual fit statistics. *Journal of applied measurement*, 1, 152-176.
- Ladera, V. y Perea, M. V. (2012). Rastreo cognitivo en demencia. *Revista Neuropsicología, Neuropsiquiatría y Neurociencias*, 12, 133-150.
- Linacre, J. M. (2011). *A user's guide to WINSTEPS & MINISTEPS: Rasch Model computer programs*. Chicago, IL: Winsteps.com.
- Lobo, A., Marcos, G., Día, J. L. y De la Cámara, C. (1995). The prevalence of dementia and depression in the elderly community in a southern European population. *Archives of General Psychiatry*, 52, 497-506. doi: 10.1001/archpsyc.1995.03950180083011
- Masters, G. (2005). Objective Measurement. En S. Alagumalai, D. D. Curtis y N. Hungi (Eds.), *Applied Rasch Measurement: A book of exemplars* (pp. 15-26). Dordrecht: Springer.
- Mathuranath, P. S., Hodges, J. R., Mathew, R., Cherian, P. J., George, A. y Bak, T. H. (2004). Adaptation of the ACE for a Malayalam speaking population in southern India. *International journal of geriatric psychiatry*, 19, 1188-1194. doi:10.1002/gps.1239
- Mathuranath, P. S., Nestor, P. J., Berrios, G. E., Rakowicz, W. y Hodges, J. R. (2000). A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology*, 55, 1613-1620. doi: 0.1212/WNL.55.11.1613
- Mioshi, E., Dawson, K., Mitchell, J., Arnold, R. y Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): A brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry*, 21, 1078-1085. doi:10.1002/gps.1610
- Pedro-Cuesta, J. de, Virués-Ortega, J., Vega, S., Seijo-Martínez, M., Saz, P., Rodríguez, F., ... del Barrio, J. L. (2009). Prevalence of dementia and major dementia subtypes in Spanish populations: A reanalysis of dementia prevalence surveys, 1990-2008. *BMC Neurology*, 9, 55. doi:10.1186/1471-2377-9-55
- Prieto, G., Contador, I., Tapias-Merino, E., Mitchell, A. J. y Bermejo-Pareja, F. (2012). The Mini-Mental-37 Test for dementia screening in the Spanish population: An analysis using the Rasch Model. *The Clinical Neuropsychologist*, 26, 1006-1018. doi: 10.1080/13854046.2012.704945

- Prieto, G., y Delgado, A. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15, 94-100.
- Prieto, G., Delgado, A. R., Perea, M. V. y Ladera, V. (2010). Scoring neuropsychological tests using the Rasch model: An illustrative example with the Rey-Osterrieth Complex Figure. *The Clinical neuropsychologist*, 24, 45-56. doi:10.1080/13854040903074645
- Prieto, G. y Velasco, A. D. (2011). Uso del modelo de Rasch para poner en la misma escala las puntuaciones de distintos tests. *Actualidades en Psicología*, 19, 5-23.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: The Danish Institute for Educational Research.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 207-230. doi: 10.2307/1164671
- Sarasola, D., Calcagno, M., Sabe, L., Caballero, A. y Manes, F. (2004). Utilidad del Addenbrooke's Cognitive Examination en español para el diagnóstico de demencia y para la diferenciación entre la enfermedad de Alzheimer y la demencia frontotemporal. *Revista Argentina de Neuropsicología*, 4, 1-11.
- Schumacker, R. E. (2004). Rasch measurement: The dichotomous model. En E. V. Smith y R. M. Smith, Jr. (Eds.), *Introduction to Rasch measurement* (pp. 226-257). Maple Grove, Mn: JAM Press.
- Smith, R. M. (2004). Fit Analysis in Latent Trait Measurement Models. En E. V. Smith y R. M. Smith, Jr. (Eds.), *Introduction to Rasch measurement* (pp. 73-92). Maple Grove, Mn: JAM Press.
- Smith, R. M y Miao, C. Y. (1994). Assessing unidimensionality for Rasch measurement. En M. Wilson (Ed.), *Objective Measurement: Theory into Practice. Volume 2* (pp. 316-328). Greenwich: Ablex.
- Tennant, A., Penta, M., Tesio, L., Grimby, G., Thonnard, J. L., Slade, A., ... Phillips, S. (2004). Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: The PRO-ESOR project. *Medical care*, 42, 137-148. doi:10.1097/01.mlr.0000103529.63132.77
- Tierney, M. C., Szalai, J. P., Dunn, E., Geslani, D. y McDowell, I. (2000). Prediction of probable Alzheimer disease in patients with symptoms suggestive of memory impairment. Value of the Mini-Mental State Examination. *Archives of family medicine*, 9, 527-532. doi: 10.1001/archfam.9.6.527
- Tombaugh, T. N. y McIntyre, N. J. (1992). The Mini-Mental State Examination: A comprehensive review. *Journal of the American Geriatrics Society*, 40, 922-935.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: LEA.
- World Health Organization. (2012). *Dementia: A public health priority*. Ginebra: Autor
- Wright, B. y Linacre, J. M. (1998). *A Rasch computer program*. Chicago: MESA Press.
- Zickar, M. J. y Broadfoot, A. A. (2009). The partial revival of a dead horse? Comparing classical test theory and item response theory. En C. E. Lance y R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends* (pp. 37-61). New York, NY: Routledge.
- Zunzunegui Pastor, M. V., del Ser, T., Rodríguez Laso, A., García Yébenes, M. J., Domingo, J. y Otero Puime, A. (2003). Demencia no detectada y utilización de los servicios sanitarios: implicaciones para la atención primaria. *Atención Primaria*, 31, 581-586. doi:10.1016/S0212-6567(03)79220-8

Fecha de recepción: 4 de abril de 2014
 Fecha de aceptación: 15 de mayo de 2014