

Original

# Comparación de métodos de escalamiento de actividades de la vida diaria en personas mayores

Alba Ayala<sup>a,b,\*</sup>, Rogelio Pujol<sup>c</sup>, Maria João Forjaz<sup>b,d</sup> y Antonio Abellán<sup>a</sup><sup>a</sup> Instituto de Economía, Geografía y Demografía, Centro de Ciencias Humanas y Sociales, Consejo Superior de Investigaciones Científicas, Madrid, España<sup>b</sup> Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), España<sup>c</sup> Instituto Nacional de Estadística, Madrid, España<sup>d</sup> Departamento de Epidemiología y Bioestadística, Escuela Nacional de Sanidad, Instituto de Salud Carlos III, Madrid, España

## INFORMACIÓN DEL ARTÍCULO

### Historia del artículo:

Recibido el 6 de marzo de 2018

Aceptado el 20 de julio de 2018

On-line el 2 de noviembre de 2018

### Palabras clave:

Actividades de la vida diaria

Guttman

Rasch

Métodos de escalamiento

Población mayor

### Keywords:

Activities of daily living

Guttman

Rasch

Scaling methods

Older people

## RESUMEN

**Objetivo:** Comparar las propiedades psicométricas de escalas de medición de las actividades de la vida diaria según distintos métodos y comprobar si los más complejos tienen más capacidad discriminatoria. **Método:** Muestra de personas mayores de la Encuesta sobre discapacidad, autonomía personal y situaciones de dependencia. Se utilizaron 14 ítems que medían actividades de la vida diaria. Cinco métodos de escalamiento: Suma y Rasch (los dos con ítems en forma dicotómica o politómica) y Guttman (forma dicotómica). Se evaluaron su capacidad discriminatoria (precisión relativa [PR]) y el área bajo la curva (AUC).

**Resultados:** Todos los métodos mostraron altas correlaciones de Pearson entre ellos (0,765-0,993). Tenían similar poder discriminatorio al comparar categorías extremas de individuos no limitados respecto a gravemente limitados (PR: 0,93-1,00). El procedimiento Suma politómico mostró la mayor AUC (0,934; intervalo de confianza del 95% [IC95%]: 0,928-0,939) y el Guttman la menor (0,853; IC95%: 0,845-0,861).

**Conclusiones:** Se observa una mayor fiabilidad en ítems politómicos que dicotómicos. Los métodos sencillos (Suma) y los complejos (Rasch) son opciones igualmente válidas. El método Guttman presentó peor capacidad discriminatoria.

© 2018 SESPAS. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Comparison of scaling methods for activities of daily living in older people

### ABSTRACT

**Objective:** To compare the psychometric properties of scales to measure activities of daily living, constructed with different scaling methods, and to check whether the most complex scales have higher discriminatory capacity.

**Method:** Sample of elderly people from the Spanish Survey on Disability, Personal Autonomy and Dependency. We used 14 items that measured activities of daily living. Five scaling methods were applied: Sum and Rasch (both for dichotomous and polytomous items) and Guttman (dichotomous). We evaluated the discriminatory capacity (relative precision [RP]) and area under the curve (AUC).

**Results:** All methods showed high Pearson correlations among them (0.765-0.993). They had similar discriminatory power when comparing extreme categories of individuals with no disability with severely limited (RP: 0.93-1.00). The polytomous Sum procedure showed the highest AUC (0.934; 95% confidence interval [95%CI]: 0.928-0.939) and Guttman the lowest (0.853; 95%CI: 0.845-0.861).

**Conclusions:** Polytomous items have greater reliability than the dichotomous ones. Simplest methods (Sum) and most complex (Rasch) are equally valid. Guttman method presented worse discriminatory capacity.

© 2018 SESPAS. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introducción

El proceso de envejecimiento suele estar asociado a un deterioro en la capacidad funcional, que en muchas ocasiones puede producir problemas en la realización de las actividades cotidianas y conducir a una situación de dependencia en la que se

requiere la ayuda de otra persona para desempeñarlas. El desarrollo de medidas de valoración es fundamental en la asistencia geriátrica y en la organización de los cuidados de larga duración<sup>1,2</sup>. La capacidad funcional puede medirse a través de la combinación de actividades de la vida diaria (ítems), básicas (AVD) e instrumentales (AIVD)<sup>3</sup>. Para ello suelen utilizarse diversos métodos de escalamiento, desde los más sencillos con la aplicación de la teoría clásica de los test (suma de varios ítems o de sus categorías) hasta los más complejos como los métodos Guttman o Rasch.

\* Autor para correspondencia.

Correo electrónico: [arwen.alba@gmail.com](mailto:arwen.alba@gmail.com) (A. Ayala).

Uno de los métodos más sencillos consiste en obtener la puntuación de un individuo mediante la suma algebraica de sus respuestas sobre desempeño de las actividades cotidianas; sin embargo, una desventaja es que para poder puntuar a un individuo es necesario conocer sus respuestas a todos los ítems. Los métodos Guttman y Rasch, aparte de ofrecer una puntuación, tienen la ventaja de ordenar ítems y personas con respecto a una dimensión acumulativa subyacente<sup>4-6</sup>. El método Guttman tiene la ventaja de ser relativamente fácil de llevar a cabo, pues no se necesita conocer las respuestas a todos los ítems para asignar una puntuación; sin embargo, es determinista, pues asume que la probabilidad de respuesta favorable a un ítem es 0 o 1, es decir, no toma un valor continuo entre esos valores. El modelo Rasch supera la rigidez de Guttman, tiene la ventaja de ser robusto a valores perdidos y permite obtener una escala lineal de intervalo<sup>5,6</sup>, pero tiene el inconveniente de su mayor complejidad y tiempo para su desarrollo en comparación con los otros dos métodos. Además de considerar las ventajas y los inconvenientes de los citados métodos, sus cualidades psicométricas nos darían también un criterio para elegir entre ellos.

Por todo lo anterior, el objetivo de este estudio fue comparar estos métodos de escalamiento de actividades de la vida diaria de personas mayores, analizar sus propiedades psicométricas y comprobar si los más complejos proporcionan una mayor capacidad discriminativa.

## Método

### Participantes

Los datos proceden de la *Encuesta sobre discapacidad, autonomía personal y situaciones de dependencia*, realizada por el Instituto Nacional de Estadística<sup>7</sup>. El análisis se realizó con las respuestas de personas de 65 y más años; cuando no pudieron contestar, se utilizó la respuesta de un *proxy*.

### Instrumentos

Se tuvieron en cuenta 14 ítems sobre AVD y AIVD que demostraron buenas propiedades psicométricas en un estudio anterior<sup>8</sup>: levantarse/acostarse, mantener la postura corporal, andar o moverse dentro la vivienda, andar o moverse fuera de la vivienda, desplazarse en medios de transporte, lavarse/ ducharse, peinarse/cortarse las uñas, control de la necesidad de orinar, control de la necesidad de defecar, vestirse, comer, hacer la compra, preparar la comida y realizar las tareas de casa. Se preguntó si tenía dificultad para realizar cada una de estas actividades (forma dicotómica: 0, no discapacidad; 1, discapacidad), y además sobre el nivel de esa dificultad (forma politómica: 0, no discapacidad; 1, moderado; 2, importante; 3, no poder realizar la actividad). Se excluyeron los individuos que no indicaron el nivel de dificultad y aquellos con dificultad para otra actividad no seleccionada.

Por último, se recogieron datos de una variable específica de control: «¿se ha visto limitado para realizar las actividades que la gente habitualmente hace, debido a un problema de salud o discapacidad?». Las opciones de respuesta fueron: gravemente limitado, limitado pero no gravemente, y no limitado.

### Análisis estadístico

Se desarrollaron cinco métodos de escalamiento: Guttman, Suma y Rasch; se utilizaron ítems en forma dicotómica y los dos últimos también en forma politómica. Se analizaron sus propiedades psicométricas. Se comprobó la unidimensionalidad del conjunto de ítems, es decir, si todos medían el mismo constructo de discapacidad, mediante análisis factorial exploratorio a partir de la

matriz de correlaciones tetracóricas para ítems dicotómicos y de correlaciones policóricas para ítems politómicos. El número de factores se determinó según el número de autovalores mayores que 1 y el porcentaje de varianza explicada por el factor<sup>9</sup>. Posteriormente se estudió la consistencia interna calculando el coeficiente 20 Kuder-Richardson (KR20, ítems dicotómicos) y el alfa de Cronbach (ítems politómicos). Los valores por encima de 0,9 muestran una consistencia interna excelente<sup>10</sup>.

### 1) Método Rasch

En el análisis para ítems politómicos se utilizó el modelo de crédito parcial<sup>11</sup>. Tanto para los ítems en forma dicotómica como politómica se evaluaron las propiedades del modelo y se hicieron modificaciones iterativamente hasta llegar al modelo final. La primera propiedad evaluada es el ajuste de los datos. Se consideró que hay buen ajuste cuando la probabilidad del estadístico  $\chi^2$  cuadrado, con corrección de Bonferroni por número de ítems, era no significativa. Además, los valores de los estadísticos de los residuos, tanto de los ítems como de las personas, no deberían exceder los valores  $\pm 2,5$ <sup>12</sup>. Para evitar un ajuste condicionado por un conjunto de datos excesivamente grande se seleccionó una muestra aleatoria de 250 personas<sup>13</sup>. Si los umbrales de los ítems no seguían una secuencia ordenada, las categorías de respuesta adyacentes fueron agrupadas<sup>6</sup>. La fiabilidad se evaluó a través del *Person Separation Index* (PSI), considerando como bueno un valor  $>0,70$ <sup>12</sup>.

Una correlación de residuos  $>0,30$  significa que los ítems tienen dependencia local. Una manera de solucionar este problema es juntar los ítems causantes de esa dependencia en superítems<sup>14</sup>. Por ejemplo, si se juntan dos ítems (en forma dicotómica) en un superítem, la suma de la respuesta de no discapacidad (valor = 0) o discapacidad (valor = 1) da lugar a tres categorías del nuevo superítem: no discapacidad en ninguno de los dos ítems (valor = 0), discapacidad en uno solo (valor = 1) o discapacidad en los dos (valor = 2). En el caso de un superítem formado con tres ítems daría lugar a cuatro categorías: no discapacidad en ninguno de los tres (valor = 0), discapacidad en uno solo (valor = 1), en dos (valor = 2) o en los tres (valor = 3). Si en la forma politómica se construye un superítem se procede de la misma manera, pero en este caso se suman los valores de las diferentes respuestas sobre el nivel de dificultad; así, si una persona tiene dificultad moderada (valor = 1) en un ítem y dificultad importante (valor = 2) en el otro, el superítem formado por los dos ítems sencillos tendría como puntuación 3. Obviamente, con este procedimiento se pueden alcanzar muchas más categorías.

El método Rasch debe cumplir la propiedad de unidimensionalidad, medida con un análisis de componentes principales de residuos, separando los ítems en dos conjuntos según su carga factorial (valor positivo y negativo). La diferencia en las estimaciones para cada persona en estos dos conjuntos se analizó mediante un test T independiente, suponiendo que la escala era unidimensional si menos del 5% de los test T se encontraban fuera del intervalo de  $-1,96$  a  $1,96$ . Otro criterio para probar la unidimensionalidad fue que, para el número observado de test significativos, el intervalo de confianza (IC) de la binomial al 95% debería superponerse al valor esperado del 5%<sup>12</sup>.

Finalmente, con las puntuaciones Rasch se obtuvo una medida lineal de intervalo (escala logit), en la que los valores negativos denotan mayor habilidad en la realización de actividades de la vida diaria y los positivos significan mayor discapacidad.

### 2) Método Guttman

Los ítems fueron escalados según el procedimiento Guttman de menor a mayor problema para realizar la actividad, asignando valores correlativos desde 1 tras el ordenamiento de los ítems<sup>15</sup>. La

**Tabla 1**

Distribución porcentual de los encuestados según la codificación de la actividad de la vida diaria, N = 39.430

	Categorías de respuesta									
	0	1	2	3	4	5	6	7	8	9
<i>Forma dicotómica (tener o no tener discapacidad)</i>										
Levantarse + mantenerse	89,7	3,3	6,9							
Andar dentro	91,4	8,6								
Andar fuera + autobús	84,6	3,1	12,4							
Lavarse + peinarse	86,2	3,7	10,0							
Orinar + defecar	92,3	2,6	5,1							
Vestirse	90,6	9,4								
Comer	96,1	3,9								
Compra + comida + tareas casa	83,4	3,1	3,3	10,3						
<i>Forma politómica (puntuación según nivel de dificultad)</i>										
Levantarse + mantenerse	89,7	1,4	2,5	1,3	2,4	0,8	1,8			
Andar dentro	91,4	2,1	3,5	3,0						
Andar fuera + autobús	84,6	1,1	2,4	1,6	3,4	1,9	4,9			
Lavarse + peinarse	86,2	1,3	2,3	1,8	2,5	1,1	4,8			
Orinar + defecar	92,3	1,3	1,5	0,7	1,3	0,4	2,5			
Vestirse	90,6	2,1	2,9	4,4						
Comer	96,1	1,0	1,4	1,5						
Compra + comida + tareas	83,4	0,8	1,5	1,9	1,0	0,9	2,1	0,6	0,7	7,3

El signo + indica la presencia de un superítem formado por dos o tres ítems sencillos. En la forma dicotómica se suman las dos posibles categorías de los ítems sencillos (tener o no tener discapacidad). En la forma politómica se suman las tres posibles categorías (con dificultad moderada, con dificultad importante, no poder realizar la actividad).

escala Guttman no utiliza respuestas politómicas, sino dicotómicas; en caso de existencia de superítems, estos se dicotomizaron, considerando valor 0 (sin discapacidad) y valor 1 (discapacidad en el resto de categorías). Se calcularon los índices de reproducibilidad y escalabilidad que indican la capacidad/validez para reproducir las respuestas de los individuos a partir de su puntuación total y conformar una escala acumulativa; son aceptables índices por encima de 0,9 y 0,6, respectivamente<sup>16</sup>.

### 3) Método Suma

Se asignó a cada individuo una puntuación igual a la suma de sus respuestas a cada ítem tanto en la versión del método en forma dicotómica (rango de puntuaciones: 0-14) como en la politómica (rango de puntuaciones: 0-42).

### 4) Comparación de métodos

Se transformaron todas las puntuaciones de los cinco métodos en escalas lineales de 0-100 para hacerlas comparables entre sí; de este modo, una mayor puntuación indica una mayor discapacidad. Se calcularon correlaciones de Pearson entre escalas para estudiar su intensidad de relación, siendo alta con correlaciones por encima de 0,7<sup>17</sup>. Además, se realizaron diferentes pruebas no paramétricas de Mann-Whitney para comparar los rangos de puntuaciones medias respecto de las categorías de la variable específica de control. A continuación, con los valores Z resultantes de cada una de estas pruebas se calcularon los correspondientes valores de precisión relativa (PR), tomando como referencia la escala Rasch de ítems politómicos<sup>18</sup>; el valor de PR se obtiene para cada par de escalamientos y cada par de categorías de la variable específica de control, como el cociente de los respectivos valores Z obtenidos en las pruebas no paramétricas. Estos valores de PR informaron sobre cuánto más preciso es el escalamiento de referencia respecto a cada uno de los otros escalamientos. Además, se calculó el tamaño del efecto asociado al valor Z, que puede ser pequeño (alrededor de 0,1), mediano (alrededor de 0,3) y grande (alrededor de 0,5)<sup>19</sup>.

Por último, para saber en qué medida las diferentes escalas discriminarían bien a una persona mayor gravemente limitada de una con limitación pero no grave o sin limitación, se realizó una prueba diagnóstica de cada escalamiento mediante el uso de curvas ROC (*receiver operating characteristics*). Estas curvas comparan la probabilidad de clasificar correctamente a un individuo gravemente

limitado (sensibilidad de la prueba) con la probabilidad de clasificar como gravemente limitado a un individuo que no lo es (1 – especificidad). El análisis Rasch se realizó con el programa RUMM2030<sup>20</sup>, y el resto de los análisis con SPSS versión 22 y Stata 12.0.

## Resultados

El tamaño final efectivo de la muestra fue de 39.430 personas, de las que el 55,9% eran mujeres. Las distribuciones porcentuales de las personas encuestadas según categorías de los ítems se recogen en la tabla 1. El análisis factorial exploratorio indicó la existencia de un único factor, tanto en el conjunto de ítems dicotómicos (autovalor igual a 7,05 con una variabilidad total explicada del 95,24%) como en el de ítems politómicos (6,96% y 96,74%, respectivamente). Las cargas factoriales presentaron un rango de 0,671 a 0,877 en el análisis de ítems dicotómicos y un rango de 0,666 a 0,902 para ítems politómicos. Además, se obtuvo un coeficiente KR20 igual a 0,91 (ítems dicotómicos) y un alfa de Cronbach de 0,92 (ítems politómicos).

En los primeros análisis Rasch con 14 ítems se observaron problemas de dependencia local. La reducción por adición de los ítems dio lugar a una lista final de ocho ítems, tres sencillos (andar dentro, vestirse, comer) y cinco superítems (levantarse + mantenerse, andar fuera + autobús, lavarse + peinarse, orinar + defecar, compra + comida + tareas). Estos ocho elementos definitivos fueron los utilizados también para los métodos Guttman y Suma, con el fin de comparar escalas con el mismo número de elementos.

El modelo Rasch final presentó buen ajuste de los datos al modelo ( $\chi^2(40) = 54,954$ ;  $p = 0,058$ ), con un valor medio en los estadísticos de ajuste de los residuos de  $-0,424$  (desviación típica [DT] = 0,882) para los ítems y de  $-0,265$  (DT = 0,619) para las personas. Ningún estadístico de los residuos para ítems y personas de manera individual excedía de los valores  $\pm 2,5$ . Se obtuvo un valor de PSI de 0,762. El análisis final no mostró problemas de dependencia local entre ítems, y se confirmó la unidimensionalidad de la escala (3,20% de valores extremos, IC binomial 95% = 0,005;  $p = 0,059$ ). En el caso de los ítems politómicos se observó un buen ajuste ( $\chi^2(32) = 39,343$ ;  $p = 0,174$ ), con un valor medio de  $-0,175$  (DT = 0,993) para los ítems y de  $-0,248$  (DT = 0,726) para las personas. Los residuos para ítems y personas de manera individual se encontraron dentro de los valores  $\pm 2,5$ , con una buena fiabilidad (PSI = 0,877). Se confirmaron la independencia

**Tabla 2**  
Orden de los ítems de menor a mayor carga de discapacidad en distintos métodos de escalamiento

Rasch dicotómico	Rasch politómico	Guttman
Andar fuera + autobús	Compra + comida + tareas casa	Lavarse + peinarse
Compra + comida + tareas casa	Andar fuera + autobús	Compra + comida + tareas casa
Lavarse + peinarse	Lavarse + peinarse	Andar fuera + autobús
Vestirse	Vestirse	Vestirse
Levantarse + mantenerse	Orinar + defecar	Levantarse + mantenerse
Andar dentro	Levantarse + mantenerse	Andar dentro
Orinar + defecar	Andar dentro	Orinar + defecar
Comer	Comer	Comer

**Tabla 3**  
Estadísticos descriptivos y correlaciones de Pearson entre los diferentes métodos de escalamiento (N = 39.430)

	Rasch dicotómico	Rasch politómico	Guttman	Suma dicotómico	Suma politómico
Media	9,0	7,4	5,4	10,3	8,0
Desviación típica	22,0	17,9	17,7	24,9	20,8
<b>Correlaciones</b>					
Rasch politómico	0,979				
Guttman	0,800	0,765			
Suma dicotómico	0,993	0,968	0,805		
Suma politómico	0,979	0,984	0,789	0,976	

Todas las correlaciones son significativas con un valor 0,01.

**Tabla 4**  
Valores Z del test de diferencias Mann Whitney y precisión relativa (referencia: escalamiento Rasch politómico)

Escala	Valores Z test Mann Whitney (tamaño del efecto)			Precisión relativa (ratio de valores Z) Base: escala Rasch		
	No limitado vs. no gravemente	No limitado vs. gravemente	No gravemente vs. gravemente	No limitado vs. no gravemente	No limitado vs. gravemente	No gravemente vs. gravemente
Rasch politómico	105,77 (0,56)	153,31 (0,84)	58,64 (0,59)	1,000	1,000	1,000
Rasch dicotómico	105,86 (0,56)	153,28 (0,84)	57,89 (0,59)	1,001	1,000	0,987
Guttman	80,27 (0,43)	142,60 (0,78)	45,22 (0,46)	0,759	0,930	0,771
Suma dicotómica	105,86 (0,56)	153,28 (0,84)	57,89 (0,59)	1,001	1,000	0,987
Suma politómica	105,75 (0,56)	153,35 (0,84)	59,12 (0,60)	1,000	1,000	1,008

local entre ítems (correlaciones <0,3) y la unidimensionalidad de la escala (4% de valores extremos; IC binomial 95% = 0,013; p = 0,067).

En el procedimiento de escalamiento Guttman, el coeficiente de reproducibilidad fue de 0,96 y el de escalabilidad fue de 0,87, por encima de los umbrales aceptables. El orden de los ítems de menor a mayor carga de discapacidad según los diferentes métodos de escalamiento se presenta en la [tabla 2](#).

Los cinco escalamientos tuvieron coeficientes altos de correlación ([tabla 3](#)), aunque más bajos en el método Guttman. Las correlaciones entre los métodos Suma y Rasch fueron las más altas.

La [tabla 4](#) muestra los estadísticos del test de diferencias de Mann Whitney y la precisión relativa. Todos los métodos de escalamiento mostraron similar poder discriminatorio en la variable específica de control al comparar categorías extremas de individuos, aunque no en las intermedias, en las que el método Guttman resultó ser menos preciso que el Rasch politómico. Para el resto de los procedimientos se obtuvieron PR muy cercanas o iguales a 1; los procedimientos Suma y Rasch dicotómico presentaron similar o igual poder discriminatorio que el Rasch politómico. Los tamaños del efecto asociado a los valores Z fueron grandes (alrededor y por encima de 0,5).

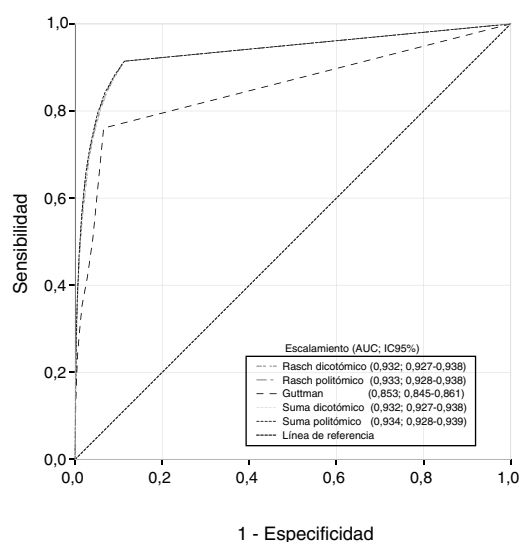
El método Suma politómico es el que mostró una mayor área bajo la curva ROC ([fig. 1](#)). Así, la probabilidad de que una persona gravemente limitada puntúe más alto que una no gravemente limitada es de 0,934. Se observó que los intervalos de confianza asintóticos de los procedimientos Rasch y Suma (ítems politómicos y dicotómicos) se solapaban. El método de escalamiento Guttman resultó el menos discriminante (área bajo la curva = 0,853); además, su intervalo de confianza no se solapó con el del resto de los métodos.

## Discusión

Este estudio tuvo como objetivo comparar las propiedades psicométricas de cinco métodos de escalamiento que combinan actividades de la vida diaria. Primero se comprobó que los ítems de AVD y AIVD forman un único constructo unidimensional y muestran una buena consistencia interna. Por tanto, estos ítems son adecuados para la aplicación de métodos de escalamiento para medir la discapacidad en las personas mayores. Los cinco métodos utilizados presentaron buenas propiedades psicométricas.

Los índices de consistencia interna mostraron unos valores mínimos comúnmente aceptados, y adecuados para comparaciones individuales<sup>21</sup>. El método Rasch politómico presentó una mayor fiabilidad que el de ítems dicotómicos, lo que significa que es apropiado para la comparación entre individuos, mientras que la escala de ítems dicotómicos solo puede ser comparada entre grupos de población (PSI >0,70)<sup>12</sup>.

Existen diferencias importantes en el grado de complejidad de los métodos, que pueden ordenarse del más complejo al más simple: Rasch (en forma politómica y dicotómica), Guttman y finalmente Suma (en ambas formas). En términos de poder discriminante, los cinco son similares para diferenciar las personas gravemente limitadas de las no limitadas (casos extremos). Si se pretende discriminar entre otras categorías de limitación (no limitados frente a no gravemente limitados; gravemente limitados frente a no gravemente limitados), son más convenientes los procedimientos Rasch y Suma que el Guttman. En consecuencia, los investigadores y los clínicos que traten con personas mayores sin discapacidad o con discapacidad leve o moderada pueden



**Figura 1.** Curvas ROC (*receiver operating characteristics*) para los métodos de escalamiento Suma, Guttman y Rasch. AUC: área bajo la curva ROC; IC95%: intervalo de confianza del 95%. Criterio de referencia: si se ha visto limitado para realizar las actividades que la gente habitualmente hace, debido a un problema de salud o discapacidad en el último año, o prevé que va a durar más de 1 año (gravemente limitado vs. no limitado o limitado pero no gravemente). Cuanto más alto sea el valor de AUC, mejor discrimina el procedimiento entre una persona gravemente limitada y otra que no lo está.

considerar los métodos Rasch y Suma como opciones igualmente válidas en esos casos. Si no se dispone de capacidad computacional, se propone utilizar el método Suma por ser menos complejo en términos de análisis y tiempo de ejecución.

Nuestro estudio sugiere que el método Guttman presenta peores resultados en cuanto a poder discriminante. Esto difiere de los resultados encontrados en otro estudio que no muestra diferencias entre los procedimientos<sup>22</sup>, pero en él solo se utilizaron AIVD y en el nuestro se incluyeron también AVD, lo que nos permite colocar a todos los ítems en un continuo más amplio<sup>23</sup>. En otro estudio se señala que los instrumentos de medida que incluyen AVD y AIVD son adecuados, y que una medida de solo AVD es sesgada y no adecuada para comparar la discapacidad<sup>24</sup>. Por ello, se recomienda utilizar las AVD y las AIVD para la evaluación geriátrica<sup>8,25,26</sup>.

Los métodos Rasch y Guttman permitieron ordenar las actividades en una jerarquía de discapacidad, según el grado de habilidad o capacidad del individuo para ejecutarlas. Aunque hay diferencias, los procedimientos mostraron una ordenación lógica y similar. Los problemas de discapacidad suelen comenzar en las actividades instrumentales, más complejas y que requieren un mayor esfuerzo y habilidad para su ejecución, y mayor conocimiento para toma de decisiones, pero son de carga de cuidado más leve (p. ej., comprar o realizar tareas domésticas); suelen terminar en aquellas actividades básicas más sencillas de ejecutar (p. ej., lavarse, comer), pero que implican mayor carga de cuidado<sup>8,27,28</sup>.

Las curvas ROC y la precisión relativa mostraron diferencias importantes entre el método Guttman y los otros. Nuestro estudio sugiere también que estrategias complejas de escalamiento, como Rasch, proporcionan igual validez discriminativa del grado de discapacidad de las personas mayores que los métodos más sencillos, por ejemplo Suma, aunque otros estudios apoyan las ventajas del modelo Rasch<sup>29</sup>.

Deben tenerse en consideración varias limitaciones. Las estrategias de escalamiento utilizadas son quizás las más comunes, pero no las más exhaustivas<sup>30</sup>. Además, las razones de fondo por las que los métodos más complejos no producen puntuaciones sustancialmente más válidas ni discriminantes no han sido estudiadas. Otra

limitación es que individuos con diferentes combinaciones de ítems pueden llegar a puntuar igual.

Una fortaleza del estudio es la comparación de medidas de asistencia geriátrica que combinan AVD y AIVD en una muestra de personas mayores representativa del ámbito nacional.

Por último, las principales conclusiones son:

- El hallazgo más relevante es que la complejidad en los procedimientos no aporta mayor poder discriminante para evaluar el grado de discapacidad en las personas mayores.
- Se recomienda la utilización de ítems politómicos que midan el nivel de dificultad de las actividades de la vida diaria.
- En la práctica clínica, si se prefieren métodos más sencillos, se puede recurrir al Suma.
- Nuestros resultados no apoyan la utilización del método Guttman frente al resto de los procedimientos.

### ¿Qué se sabe sobre el tema?

El desarrollo de medidas de valoración de la capacidad funcional es fundamental y está muy extendido en la asistencia geriátrica y en la organización de los cuidados de larga duración. Existen diversos métodos de escalamiento, desde los más simples hasta los más complejos, y se han utilizado en muestras de población variadas. El tipo de escalamiento influye en cómo recoger los resultados en la evaluación sobre la discapacidad.

### ¿Qué añade el estudio realizado a la literatura?

Nuestro estudio encuentra que métodos sencillos de escalamiento proveen similar poder discriminante que los complejos para evaluar el grado de discapacidad en las personas mayores. La utilización de herramientas más simples que requieran menos tiempo tiene implicación en la práctica clínica. Se utiliza una muestra representativa nacional.

### Editor responsable del artículo

David Cantarero.

### Declaración de transparencia

La autora principal (garante responsable del manuscrito) afirma que este manuscrito es un reporte honesto, preciso y transparente del estudio que se remite a GACETA SANITARIA, que no se han omitido aspectos importantes del estudio, y que las discrepancias del estudio según lo previsto (y, si son relevantes, registradas) se han explicado.

### Contribuciones de autoría

A. Abellán y M.J. Forjaz realizaron la concepción y el diseño del trabajo. A. Ayala y R. Pujol desarrollaron los análisis estadísticos y la interpretación de los datos. A. Ayala redactó el primer borrador del artículo e incorporó los cambios del resto de autores. Todas las personas firmantes aportaron su revisión crítica con importantes contribuciones intelectuales y aprobaron la versión final para su publicación.

## Financiación

El estudio fue parcialmente financiado por ENVACES, Plan Nacional I+D+I, Ref. CSO2015-64115-R, y el Programa Envejecimiento Activo, Calidad de Vida y Género, Comunidad de Madrid (ENCAGE-CM), Ref: S2015/HUM-3367.

## Conflicto de intereses

Ninguno.

## Bibliografía

- Kane RA, Kane RL, Carrillo E. Evaluación de las necesidades en los ancianos: guía práctica sobre los instrumentos de medición. Barcelona: SG; 1993.
- Martínez Rodríguez T. La atención centrada en la persona en los servicios gerontológicos. Modelos de atención y evaluación. Estudios de la Fundación Pílares para la autonomía personal; 2016.
- Gross AL, Jones RN, Inouye SK. Development of an expanded measure of physical functioning for older persons in epidemiologic research. *Res Aging*. 2015;37:671–94.
- Guttman L. A basis for scaling qualitative data. *Am Sociol Rev*. 1944;9:139–50.
- Andrich D. Rating scales and Rasch measurement. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11:571–85.
- Bond T, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. New York: Routledge; 2015.
- Instituto Nacional de Estadística. Encuesta sobre discapacidad, autonomía personal y situaciones de dependencia, 2008. Madrid: Instituto Nacional de Estadística; 2008. Disponible en: <http://www.ine.es/metodologia/t15/t1530418.pdf>
- Forjaz MJ, Ayala A, Abellán A. Hierarchical nature of activities of daily living in the Spanish Disability Survey. *Rheumatol Int*. 2015;35:1581–9.
- Osborne JW. Best practices in exploratory factor analysis. CreateSpace Independent Publishing Platform; 2014.
- Hara N, Matsudaira K, Masuda K, et al. Psychometric assessment of the Japanese version of the Zurich Claudication Questionnaire (ZCQ): reliability and validity. *PLoS One*. 2016;11:e0160183.
- Masters GN. A Rasch model for partial credit scoring. *Psychometrika*. 1982;47:149–74.
- Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum*. 2007;57:1358–62.
- Hagell P, Westergren A. Sample size and statistical conclusions from tests of fit to the Rasch model according to the Rasch Unidimensional Measurement Model (RUMM) program in health outcome measurement. *J Appl Meas*. 2016;17:416–31.
- Marais I, Andrich D. Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *J Appl Meas*. 2008;9:200–15.
- Mclver J, Carmines EG. Unidimensional scaling. 24. California: Sage Publications; 1981.
- Maranell G. Scaling: a sourcebook for behavioral scientists. New York: Routledge; 2017.
- Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J*. 2012;24:69–71.
- Wijers IG, Ayala A, Rodríguez-Blázquez C, et al. Rasch analysis and construct validity of the disease burden morbidity assessment in older adults. *The Gerontologist*. 2017;17:1102–8.
- Coolican H. Research methods and statistics in psychology. 6th ed. New York: Psychology Press; 2014.
- Andrich D, Sheridan BED, Luo G. RUMM2030: Rasch unidimensional models for measurement. Perth, Western Australia: RUMM Laboratory; 2009.
- Srimongkon P, Aslani P, Chen TF. A systematic review of measures of medication adherence in consumers with unipolar depression. *Res Soc Adm Pharm*. 2017.
- Vittengl JR, White CN, McGovern RJ, et al. Comparative validity of seven scoring systems for the instrumental activities of daily living scale in rural elders. *Aging Ment Health*. 2006;10:40–7.
- Buz J, Cortés-Rodríguez M. Measurement of the severity of disability in community-dwelling adults and older adults: interval-level measures for accurate comparisons in large survey data sets. *BMJ Open*. 2016;6:e011842.
- LaPlante MP. The classic measure of disability in activities of daily living is biased by age but an expanded IADL/ADL measure is not. *J Gerontol B Psychol Sci Soc Sci*. 2010;65:720–32.
- Morris JN, Berg K, Fries BE, et al. Scaling functional status within the interRAI suite of assessment instruments. *BMC Geriatr*. 2013;13:128.
- Roehrig B, Hoeffken K, Pientka L, et al. How many and which items of activities of daily living (ADL) and instrumental activities of daily living (IADL) are necessary for screening. *Crit Rev Oncol Hematol*. 2007;62:164–71.
- Verbrugge LM, Yang L-S, Juarez L. Severity, timing, and structure of disability. *Soz-Präventivmedizin*. 2004;49:110–21.
- González MDP. El proceso de discapacidad: un análisis de la Encuesta sobre discapacidades, deficiencias y estado de salud. Madrid: Fundación Pfizer; 2004. p. 210.
- Sheehan TJ, DeChello LM, García R, et al. Measuring disability: application of the Rasch model to activities of daily living (ADL/IADL). *J Outcome Meas*. 2002;5:839–63.
- Fricke J, Unsworth C. Time use and importance of instrumental activities of daily living. *Aust Occup Ther J*. 2001;48:118–31.