



## El Modelo de Rasch Aplicado a las Ciencias Psicológicas

[O Modelo de Rasch Aplicado às Ciências Psicológicas]

LUÍS ALBERTO COELHO REBELO MAIA<sup>1</sup>

### Resumo

Um aspecto de importância crucial quando um investigador se coloca perante a questão do projecto metodológico e da respectiva aplicação dos testes psicológicos incide nas teorias de medição que suportam a sua orientação em aspectos como a atribuição da pontuação (cotação), as características dos itens, etc., permitindo realizar as análises dos dados e mesmo originar outros métodos de interesse. Relacionada com este aspecto, a cotação dos itens de um teste psicológico nem sempre corresponde a uma informação o mais adequada possível em relação ao sujeito, devido, por exemplo, à situação de teste, bem como a outras variáveis.

Em 1960, o matemático Georg Rasch desenvolveu um modelo estatístico que se propunha a resolver algumas das principais críticas apresentadas à Teoria Clássica dos Testes (TCT), conseguindo ganhar um importante campo de aplicação na avaliação no âmbito das ciências psicológicas e da educação. O modelo de Rasch é bem conhecido pela divulgação da Teoria de Resposta ao Item (TRI), cujo objectivo é apresentar considerações conjuntas sobre as respostas dos sujeitos e dos itens, examinando a probabilidade de se observar uma resposta correcta, assim como a relação entre os sujeitos e os itens. Assim, este modelo permite ter uma maior acuidade no acesso adequado a um dado constructo, uma vez que a sua análise (item a item, mas também sujeito a sujeito) permite uma estimação sobre se determinado sujeito é ou não capaz de acertar a um determinado item.

Assim, apresentamos um estudo baseado na aplicação de um teste para avaliar o défice cognitivo (*Mini Mental State Examination*) em 92 indivíduos portugueses internados numa instituição de apoio a idosos. Num primeiro nível, a conclusão que podemos tirar é a de que o modelo de Rasch permite uma análise detalhada (profunda) de parâmetros que geralmente não são considerados na TCT. Com esta metodologia podemos evidenciar que, nesta amostra específica de 92 idosos, o MMSE se adapta de forma muito forte, uma vez que os dados que sobressaem indicam uma distribuição normalizada do desempenho destes sujeitos quando avaliados pelo MMSE.

Palavras-chave: Modelo de Rasch, défice cognitivo, MMSE.

---

<sup>1</sup> Professor Auxiliar do Departamento de Psicologia e Educação da UBI. Doutorado em Neuropsicologia Clínica pela UBI e Universidad de Salamanca – Departamento de Psicobiología, Psicología Básica y Metodología de las ciencias del comportamiento / Instituto de Neurociencias de Castilla y León.  
E-mail: [hmaia@ubi.pt](mailto:hmaia@ubi.pt)

## Resumen

Un aspecto de crucial relevancia cuando un investigador se plantea ante la cuestión de su proyecto metodológico y respectiva implementación de pruebas son las teorías de medición que auxilian su orientación en aspectos como la asignación de puntajes, características de los ítems, *etc.*, permitiendo realizar análisis de los datos y mismo hacer derivar otros métodos de interés. Relacionado con este aspecto, la cotación de los ítems de un teste psicológico ni siempre corresponden a una información lo más adecuada posible, relativo al sujeto, debido, por ejemplo, a la situación del teste, bien como otras variables.

En 1960 el matemático Georg Rasch desarrolló un modelo estadístico que se proponía resolver algunas de las principales críticas presentadas a la *Teoría Clásica de los Testes* (TCT), por ese motivo está consiguiendo ganar un importante campo de aplicación en la evaluación dentro de las ciencias psicológicas y educativas. Lo modelo de Rasch es más conocido a través de la divulgación de la *Teoría de Respuesta a los Ítems* (TRI) cuyo objetivo general es presentar consideraciones conjuntas acerca de las respuestas de los sujetos a determinados ítems, tratando la probabilidad de observarse una respuesta correcta, así como la estimación de la relación entre los sujetos y los ítems. Así, con este modelo se puede tener más acuidade a la hora de acceder de fuerma adecuada a un dado constructo, una vez que la estimación (ítem a ítem, pero también sujeto a sujeto) permiten una estimación de que un sujeto particular sea capaz de acertar un determinado ítem o no.

Así, presentaremos un estudio realizado tras la aplicación de un teste de evaluación de deterioro cognitivo (*Mini Mental State Examination*) en 92 sujetos ingresados en una institución de apoyo a ancianos. Un primer nivel de conclusión que podemos plantear es que el Modelo de Rasch permite un análisis profundo de parámetros usualmente no considerados en la TCT. Con esta metodología podemos certificar que, en esta muestra original de 92 ancianos, el MMSE se adapta de forma muy fuerte, una vez que los datos se superponen a los objetos indicando una distribución normalizada del desempeño de estos sujetos cuando evaluados por el MMSE.

Palabras-clave: Modelo de Rasch, Deterioro Cognitivo, MMSE.

## Introducción

De acuerdo con Dussaildant (2003, p.92) “*Las teorías de medición sirven como marco teórico en el diseño e implementación de pruebas. Estas teorías emplean la metodología para la asignación de puntajes, proveen mecanismos para determinar las características de las preguntas o ítems, y a partir de ellas se derivan métodos para realizar otros análisis de interés*”. Así, un aspecto de crucial relevancia cuando un investigador se plantea ante la cuestión de su diseño metodológico y respectiva implementación de pruebas son las teorías de medición que auxilian su orientación en aspectos como la asignación de puntajes, características de los ítems, *etc.*, permitiendo realizar análisis de los datos y mismo hacer derivar otros métodos de interés (Stenner, Burdick & Stone, 2008; Dussaildant, 2003). Relacionado con este aspecto está el hecho que, como refieren Feddag (2008), Chi-Wen (2007) y Hornke (2000) la cotación de los ítems de un teste psicológico ni siempre corresponde a una información lo más adecuada posible, relativo al sujeto, debido, por ejemplo, a la situación del teste, bien como otras variables).

Prieto y Delgado (2000) refieren que el propósito de la Psicometría debe ser la formulación de modelos procedimentales que permitan la obtención de medidas psicológicas. Estas medidas psicológicas, suelen llamarse *constructos*, que de acuerdo con Pinto (1990), remite para una idea construida por el investigador para explicar los fenómenos ocurridos e observados en una situación particular (p.168).

En 1960 el matemático *Georg Rasch* desarrolló un modelo estadístico que se proponía resolver algunas de las principales críticas presentadas a la *Teoría Clásica de los Testes* (TCT). Lo modelo de Rasch es más conocido tras la divulgación de la *Teoría de Respuesta a los Ítems* (TRI) cuyo objetivo general es presentar consideraciones conjuntas relacionadas con las respuestas de los sujetos a determinados ítems, tratando la probabilidad de observarse una respuesta correcta, así como la estimación de la relación entre los sujetos y los ítems (Feddag, 2008; Stenner, Burdick & Stone, 2008; Embretson, 2006; Juhe, 1999, Rasch, 1960, 1980).

Así, Prieto y Delgado (2003) refuerzan que con este modelo se puede tener más acuidade a la hora de acceder a un dado constructo, una vez que la estimación (ítem a ítem, pero también sujeto a sujeto) permiten una estimación de que un sujeto particular sea capaz de acertar un determinado ítem o no. Aunque nos apoyemos fundamentalmente en Prieto y Delgado (2003) para la descripción de las características principales del modelo de Rasch, se

pueden consultar obras de Feddag (2008), Stenner, Burdick & Stone (2008), Embretson (2006), Bond y Fox (2001), Rost (2000) y Wright y Stone (1979), para una abordaje profundizada al tema.

La función logística utilizada por Rasch para modelar la relación en los ítems (su respectiva dificultad y relativa probabilidad de ser resuelto correcta o incorrectamente) fue:

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

(In Prieto & Delgado, 2003, p.94)

Prieto y Delgado (2003, p.94) recapitulan así la función presentada (Cf. también Battisti, Nicolini y Salini, p. 2): *“La ecuación del cociente entre la probabilidad de una respuesta correcta y la probabilidad de una respuesta incorrecta a un ítem ( $P_{is} / 1 - P_{is}$ ), é una función de la diferencia en el atributo entre el nivel de la persona ( $\theta_s$ ) y el nivel del ítem ( $\beta_i$ ). (...) cuando una persona respunde a un ítem equivalente a su umbral de competencia, tendrá la misma probabilidad de una respuesta correcta y de una respuesta incorrecta ( $P_{is} / 1 - P_{is} = 0,50/0,50$ ). En este caso, el logaritmo natural de  $P_{is} / 1 - P_{is}$ , refleja que la dificultad del ítem es equivalente al nivel de competencia de la persona ( $\theta_s - \beta_i = 0$ ).”* (Prieto & Delgado, 2003, p.94).

Así, de acuerdo con el expuesto, si un sujeto presenta una mayor competencia para contestar a un determinado atributo mensurado por un ítem ( $\theta_s - \beta_i > 0$ ) se verificará que existe una mayor probabilidad del sujeto responder correctamente al mismo ítem, que responderlo incorrectamente. Lo contrario, es fácilmente verificado por la ecuación ( $\theta_s - \beta_i < 0$ ), en que la probabilidad de una persona responder correctamente a un ítem es inferior a de responderlo incorrectamente.

De acuerdo con Prieto y Delgado (2003; Cf. también Stenner, Burdick & Stone, 2008; Embretson, 2006; García-Cueto, Muñiz, & Lozano, 2003; Verstralen, Bechger & Maris, 2001) el modelo de Rasch, se hizo más conocido por el público general tras su difusión en publicaciones relacionadas con la TRI. Según los autores *“deriva de la predicción de la probabilidad de responder correctamente al ítem a partir de la diferencia en el atributo entre el nivel de la persona ( $\theta_s$ ) y el nivel del ítem ( $\beta_i$ ). En este caso,  $P_{is} = e^{(\theta_s - \beta_i)} / 1 + e^{(\theta_s - \beta_i)}$ , donde  $e$  es la base de los logaritmos naturales (2,7183)”* (Prieto & Delgado, 2003, p.94).

Basandonos en lo referido del *ajuste de los datos al modelo*<sup>2</sup>, Green y Frantom (2002) refieren que, en este modelo, un parámetro estadístico conocido como *fit* sirve para mantener un mecanismo interior identificando las contestaciones incorrectas a los ítems, mientras permita la exclusión o re-valoración de personas cuyas contestaciones no tienen sentido, es decir, no encajen, según nuestra comprensión de la estructura medida – *constructo*<sup>3</sup>.

Para Green y Frantom (2002) debe cuestionarse la validez de las cuestiones cuando la lógica del *constructo* no prevalece, como cuando una persona está de acuerdo con un ítem que remite para una idea suicida pero no con un ítem relativo a sentimientos de tristeza. Así, según los autores, *fit* proporciona un índice del grado según las contestaciones se conforman a un modelo lógico, bien como una indicación de la validez de la medida para un individuo específico. Semejantemente, *fit* permite la valoración de la validez de la medida global proporcionando una herramienta para identificar ítems que se adecuan más rica o más pobremente al modelo. El *fit* de un ítem en particular es un índice de cuanto el ítem funcione en la reflexión del rasgo. Los ítems buenos, indicadores de *fit*, son más útiles a la hora de medición o valoración de un determinado rasgo que los ítems con un *fit* pobre. Con estas consideraciones en mente, se busca entonces el desarrollo de testes que proporcionen el mejor *ajuste de los datos* posible.

Seguidamente presentaremos un estudio realizado tras la aplicación de un teste de evaluación de deterioro cognitivo (*Mini Mental State Examination*, de Folstein y colaboradores, 1975, en su versión adaptada y traducida para el portugués, por Guerreiro, en 1993). No sólo serán presentadas consideraciones estadísticas descriptivas e inductivas, como también se procederá a un análisis profundo de acuerdo con el modelo e Rasch.

## **Método**

### ***Participantes y metodología de aplicación del teste***

Fueron analizados los datos de 92 ancianos portugueses ingresados en dos instituciones de acogimiento social perteneciente a la misma entidad cuidadora. La

---

<sup>2</sup> La idea que si los datos empíricos se ajustan al modelo utilizado, los sujetos con menor capacidad solo deberían responder correctamente ítems fáciles o de mediana dificultad, así como sujetos con mayor capacidad deberían ser capaces de responder correctamente ítems fáciles, medios y difíciles (Cf. Prieto & Delgado, 2003).

<sup>3</sup> Así, por ejemplo, los autores refieren que nuestra comprensión de una dada condición, por ejemplo la depresión, debe reflejarse como un *constructo*, en un conjunto patrón de contestaciones de los participantes. Se esperaría entonces que una persona que estea más deprimida conteste de determinada forma, o esté más fuertemente de acuerdo con los ítems en un estudio de depresión, que alguien que estea menos deprimido.

selección fue consecutiva, es decir, todos los sujetos ingresados en las dos instituciones tenían la misma probabilidad de ser evaluados. Solo no fueron evaluados los enfermos con un elevado grado de discapacidad (estados vegetativos-comatosos, estados de demencia severa, *etc.*). Así, 92 ancianos fueron evaluados después de haber sido pedido el respectivo consentimiento informado.

Como se comprueba en la Tabla I la edad media es de 80,48 años con una desviación típica elevada de 12,24 (la edad mínima es de 44 y la edad máxima es de 101).

		Edad del sujeto
N	Válidos	90
	Missing	2
	Medias	80,48
	Desviación Típica	12,236
	Mínima	44
	Máxima	101

Tabla I – Estadística descriptiva.

Género sexual		Frecuencia	Porcentaje
Válidos	Femenino	60	65,2
	Masculino	32	34,8
	Total	92	100,0

Tabla II – Estadística descriptiva.

La Tabla II muestra que de los 92 sujetos 60 son mujeres (65,2%) y 32 (34,8%) son varones.

		Género		
		Femenino	Masculino	Total (%)
Edad en 3 categorías	Asta 65 años	6	8	14 (15.22)
	>65-75	8	4	12 (13.04)
	> 75	45	19	64 (69.57)
Total		59	31	90 (97.83)

Tabla III – Estadística descriptiva.

Considerando la variable *género* teniendo en cuenta varias categorías de edad (Tabla III), verificamos que 14 sujetos se encuentran bajo la edad de 65 años. Esto significa que, de 92 sujetos la mayoría absoluta (82.61%) esta claramente encima de la línea de los 65 años. Estos datos ganan mayor relevancia cuando se comprueba que, de estos últimos, 64 sujetos (69.57%) son considerados *ancianos avanzados* (en la clasificación portuguesa, sujetos con más que 75 años de edad).

### **Instrumento**

Fue utilizado el *Mini Mental State Examination* (M. Folstein, S. Folstein & Mchugh, 1975) en su versión portuguesa (adaptada por Guerreiro, 1993). Este teste es constituido por

30 ítems de evocación libre dividido en 6 bloques que reportan para diversas dimensiones (constructos) que se cree están relacionados con la orientación o, al revés, deterioración cognitiva del sujeto (Ejemplo de parte do MMSE – versión portuguesa, Figura I).

Cada uno de los seis bloques de cuestiones está respectivamente relacionado con: Orientación, Repetición, Atención y Cálculo, Evocación, Lenguaje y Habilidad Constructiva. Cada cuestión está compuesta por un enunciado que pide una respuesta que se considera correcta o incorrecta (como un todo tenemos entonces 30 ítems de respuesta cerrada e así, los hemos tratado como datos dicotómicos).

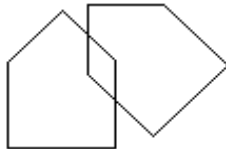
De acuerdo con Guerreiro (1991) el MMSE es un teste muy breve, creado en los Estados Unidos de América (EUA), siendo el teste de *screening* de deterioro cognitivo más utilizado en estudios epidemiológicos (como referido por nosotros en estudios previos, este test solamente debe ser utilizado como um test de *screening* neuropsicológico para deterioro cognitivo – Maia, 2006).

---

**6. HABILIDADE CONSTRUTIVA** (1 ponto pela cópia correcta)

Deve copiar um desenho. Dois pentágonos parcialmente sobrepostos; cada um deve ficar com cinco lados, dois dos quais intersectados. Não valorizar, tremor ou rotação.

**DESENHO**



**CÓPIA**

(Máximo 30 pontos)

Total:

Pontos de corte  
(População Portuguesa)

Considera-se com defeito cognitivo

- § Analfabetos ≤ 15
  - § 1 a 11 anos de escolaridade ≤ 22
  - § Com escolaridade superior a 11 anos ≤ 27
- 

Figura I – MMSE – Versión Portuguesa; Fonte: Adaptado de Guerreiro (1993)

Así, Guerreiro (1993) refiere que este teste debe ser más utilizado como un teste breve de *screening* para áreas que están normalmente relacionadas con funciones ejecutivas como orientación, memoria, lenguaje y capacidad constructiva. Un aspecto que fue considerado muy importante en su adaptación portuguesa fue la escolaridad de los sujetos. En su estudio con 137 sujetos controles y 151 enfermos con diferentes tipos de demencias, Guerreiro comprobó que la escolaridad se correlacionaba muy sencillamente (*Índice de Correlación de*

*Pearson* –  $r = 0.64$ ), siendo la diferencia de resultados entre los dos grupos, cuando se considera la escolaridad, altamente significativa ( $t_s = 11,11$ ;  $\rho = 0.000$ ). Las variables sexo y edad no se encontraban correlacionadas con los resultados globales en el teste.

Segun el método de bipartición, para averiguar la garantía y validez del teste, Guerreiro (1993) verificó un valor de 0.82 en lo que se relaciona con la correlación entre las dos partes del teste con el *Coefficiente Alfa de Cronbach*, bien como el *coeficiente de garantía* con un *valor de Corrección de Gutman* de 0.88. Con un estudio de *validad de criterio concurrente* (los resultados obtenidos por los mismos sujetos, cuando son comparados con los resultados de los mismos sujetos, en el mismo espacio temporal, en otra medida de valor reconocido) con las *Matrices Progresivas de Raven*, se verifica una fuerte asociación entre las dos medidas ( $r = 0.70$ ).

Guerreiro (1993) recupera un conjunto de estudios relativo a las características psicométricas y diagnósticas del MMSE: partiendo del barémo de 23/24 (de 0 a 30) presentado originalmente por Folstein *et al.* (1975) como el punto limite indicador de deterioro cognitivo; Anthony, Leresche y Unaiza (1982) encontraron niveles de falsos positivos próximos de los 5% en un grupo de enfermos con escolaridad inferior a 9 años.

Todavía, solo con la utilización del barémo 23/24 originalmente introducido por Folstein *et al.* (1975) se ha verificado un conjunto variado de estudios con números de falsos positivos muy elevados (por ejemplo, Bertolucci, Brucki & Campacci, 1994 – en un estudio con sujetos de variadas categorías académicas, verificaron que en los 107 sujetos analfabetos del estudio, la utilización del barémo 23/24 crearía “artificialmente” cerca de 75% de falsos positivos). Así, se refuerza la idea que las potencialidades diagnósticas del teste sólo se afirman en máxima exponencia cuando se tiene en cuenta aspectos como: el nivel de escolaridad formal y la edad del sujeto [Cossa, Sala & Musicco (1997), Malloy, Cummings & Coffey (1997), Bertolucci *et al.*, (1994), Guerreiro, Silva & Botelho (1994), O'Connor, Pollit, & Hyde (1989), Cavanaugh & Wettstein (1983), Chandler & Gerndt (1988)].

Para el tratamiento de la información con el Modelo de Rasch, y con base en otros trabajos (Prieto & Delgado, 2003), las respuestas fueron codificadas dicotómicamente siendo los datos analizados utilizando el programa informático *Quest* (Adams & Khoo, 1996).

## **Resultados**

Relativamente a la variable *Género* (Tabla IV), los resultados de los sujetos no se diferencian significativamente, considerando la Estadística  $\chi^2$  (Chi-square).



Teste $\chi^2$ (Chi-square) para la variable sexo y categorización de edad	Valor	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3,842 <sup>(a)</sup>	2	.146
Likelihood Ratio	3,663	2	.160
Linear-by-Linear Association	3,376	1	.066
N de casos válidos	90		

Tabla IV – Teste  $\chi^2$  (Chi-square) para la variable sexo y categorización de edad; <sup>(a)</sup> 2 células (33,3%) presentan un N esperado inferior a 5. Lo mínimo esperado es 4,13.

El análisis de la Tabla V presenta los resultados medios en el MMSE: media de 13,32, con una desviación típica de 8,936 siendo la puntuación mínima conseguida de 0 puntos y máxima de 30 puntos.

		Resultado total del MMSE
N	Válidos	92
	Missing	0
	Medias	13,32
	Desviación Típica	8,936
	Mínima	0
	Máxima	30

Tabla V – Resultados en el MMSE.

Teste $\chi^2$ (Chi-square) para resultados en el MMSE y la variable sexo	Valor	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	29,968 <sup>(a)</sup>	30	.467
Likelihood Ratio	38,199	30	.145
Linear-by-Linear Association	3,550	1	.060
N de casos válidos	92		

Tabla VI – Teste  $\chi^2$  (Chi-square) para resultados en el MMSE y la variable sexo; <sup>(a)</sup> 62 células (100,0%) presenta un N esperado inferior a 5. Lo mínimo esperado es 0,35.

También no se verifica cualquier diferencia estadísticamente significativa en lo que se refiere a los resultados de mujeres y varones, considerando los resultados globales en el MMSE (Tabla VI). Considerando los baremos específicos para la población portuguesa con el MMSE, la media de resultados obtenidos por estos sujetos fue de 13,32 con una desviación típica de 8,936.

La Tabla VII muestra que el índice de *Consistencia Interna*, calculado el *Coficiente Alpha de Cronbach* es de .95.

El análisis de las correlaciones de los ítems con el resultado total en el MMSE, como las correlaciones ítem-ítem permite verificar que todos los ítems presentan niveles de correlación elevada con el resultado final, a un elevado nivel de significancia estadística ( $\rho < .01$ ) con excepción del Ítem 29.

---

**Reliability Coefficients**

Número de Casos = 92

Número de Ítems = 30

Media en el MMSE = 13,32

Desviación Típica = 8,936

**Alpha = .9458**

---

Tabla VII – *Reliability Analysis – Scale (Alpha)*.

Según el Modelo de Rasch, no solo se puede analizar el poder discriminativo de cada ítem (Figura XII), en este caso, el ítem 29 presenta un índice de discriminación de .46, como analizar cuantos sujetos lo respondieron correctamente (Embretson, 2006). En este caso, solo 7 sujetos acertaron el ítem 29.

**Fonte:** Adaptado de Guerreiro (1997).

---

```
-----
Análisis del Minimental. Covilha
-----
Item Estimates (Thresholds)
13:50:58
all on all (N = 92 L = 30 Probability Level= .50)
-----
Summary of item Estimates
=====
Mean                0.00
SD                  2.07
SD (adjusted)      2.04
Reliability of estimate .97

Fit Statistics
=====
Infit Mean Square      Outfit Mean Square
Mean                .98                Mean                1.78
SD                  .27                SD                  3.19

Infit t                Outfit t
Mean                -.20                Mean                .37
SD                  1.38                SD                  1.13

0 items with zero scores
0 items with perfect scores
-----

Case Estimates
all on all (N = 92 L = 30 Probability Level= .50)
-----
Summary of case Estimates
=====
Mean                -.35
SD                  2.38
SD (adjusted)      2.29
Reliability of estimate .93

Fit Statistics
=====
Infit Mean Square      Outfit Mean Square
Mean                .97                Mean                1.78
SD                  .29                SD                  4.96

Infit t                Outfit t
Mean                -.07                Mean                .68
SD                  .89                SD                  1.10

7 cases with zero scores
0 cases with perfect scores
-----
```

Figura III – Análisis del MMSE; estimación de los ítems y de los casos de acuerdo con el Programa *Quest*.

El análisis de la Figura III demuestra varios parámetros dignos de nota. La fiabilidad de estimación de los ítems (*Reliability of Estimate*) es muy buena, con un índice de 0.98. A su vez, los indicadores de *Infit*, para la estimación de los ítems, son casi perfectos, con un valor de 0.97. Lo mismo no ocurre para la estimación del *Outfit* de los ítems, con un valor muy elevado de 1.74. Por otro lado, la estimación de los casos presenta un índice de fiabilidad igualmente elevado, con un valor de 0.93. Además en el análisis de la Figura III se puede verificar que, en lo que concierne a los sujetos, el *Infit* es muy bueno (0.97) siendo el *Outfit*, de 1.78, muy elevado.

Análisis del Minimental. Covilha								
Item Estimates (Thresholds) In input Order							18-May-2004 13:51:02	
all on all (N = 92 L = 30 Probability Level= .50)								
ITEM NAME	SCORE	MAXSCR	THRSH	INFT	OUTFT	INFT	OUTFT	
			1	MNSQ	MNSQ	t	t	
1	item 1	2	92	5.47	1.47	12.35	.9	2.8
				.78				
2	item 2	42	92	-.37	.88	1.06	-.8	.3
				.30				
3	item 3	17	92	2.15	.94	.68	-.2	.1
				.36				
4	item 4	47	92	-.80	.89	.84	-.7	-.1
				.30				
5	item 5	37	92	.07	1.33	2.31	1.9	2.1
				.31				
6	item 6	49	92	-.98	.97	.87	-.2	0.0
				.30				
7	item 7	37	92	.07	1.05	1.00	.4	.2
				.31				
8	item 8	58	92	-1.78	1.15	1.13	.9	.4
				.31				
9	item 9	58	92	-1.78	1.15	.87	.9	.1
				.31				
10	item 10	46	92	-.71	.77	.49	-1.6	-1.0
				.30				
11	item 11	57	92	-1.69	.75	.47	-1.7	-.6
				.31				
12	item 12	52	92	-1.24	.83	.58	-1.1	-.6
				.30				
13	item 13	55	92	-1.50	.97	1.03	-.1	.3
				.31				
14	item 14	33	92	.44	.80	.62	-1.2	-.6
				.31				
15	item 15	25	92	1.23	.58	.31	-2.6	-1.1
				.33				
Análisis del Minimental. Covilha (continuación)								
16	item 16	17	92	.69	.69	-.7	-.4	
				.32				
Item Estimates (Thresholds) In input Order							18-May-2004 13:51:02	
all on all (N = 92 L = 30 Probability Level= .50)								
ITEM NAME	SCORE	MAXSCR	THRSH	INFT	OUTFT	INFT	OUTFT	
			1	MNSQ	MNSQ	t	t	
18	item 18	22	92	1.55	1.34	1.3	0.0	
				.34				
19	item 19	26	92	1.12	1.14	1.27	.8	.6
				.33				
20	item 20	28	92	.92	.95	.88	-.2	0.0
				.32				
21	item 21	17	92	2.15	1.20	14.13	1.0	4.0
				.36				
22	item 22	72	92	-3.34	.76	.44	-1.1	.2
				.38				
23	item 23	76	92	-3.97	.82	.30	-.6	.4
				.43				
24	item 24	49	92	-.98	1.05	1.42	.4	.8
				.30				
25	item 25	55	92	-1.50	1.26	2.75	1.6	1.9
				.31				
26	item 26	51	92	-1.15	.77	.62	-1.7	-.5
				.30				
27	item 27	52	92	-1.24	.77	.53	-1.7	-.7
				.30				
28	item 28	56	92	-1.60	1.60	2.05	3.4	1.3
				.31				
29	item 29	7	92	3.78	.67	.19	-1.2	.4
				.48				
30	item 30	10	92	3.19	1.60	2.14	2.2	1.1
				.43				
Mean								
				0.00	.98	1.78	-.2	.4
SD								
				2.07	.27	3.19	1.4	1.1

Figura IV.

En su representación gráfica (Figura V) los ítems 1, 28 y 30 se presentan con un *Infit* superior a 1.3.

```

Análisis del Minimental. Covilha
-----
Item Fit                                     18-May-2004 13:51:14
all on all (N = 92 L = 30 Probability Level= .50)
-----
INFIT
MNSQ      .53      .63      .77      1.00      1.30      1.60      1.90
-----+-----+-----+-----+-----+-----+-----+
1 item 1      .          .          .          .          .          .          .
2 item 2      .          .          .          *          .          .          .
3 item 3      .          .          .          *          .          .          .
4 item 4      .          .          .          *          .          .          .
5 item 5      .          .          .          .          .          .          *
6 item 6      .          .          .          *          .          .          .
7 item 7      .          .          .          .          *          .          .
8 item 8      .          .          .          .          .          *          .
9 item 9      .          .          .          .          .          *          .
10 item 10     *          .          .          .          .          .          .
11 item 11     .          *          .          .          .          .          .
12 item 12     .          .          *          .          .          .          .
13 item 13     .          .          .          *          .          .          .
14 item 14     .          .          *          .          .          .          .
15 item 15     *          .          .          .          .          .          .
16 item 16     .          .          .          *          .          .          .
17 item 17     .          *          .          .          .          .          .
18 item 18     .          *          .          .          .          .          .
19 item 19     .          .          .          .          *          .          .
20 item 20     .          .          .          *          .          .          .
21 item 21     .          .          .          .          .          *          .
22 item 22     *          .          .          .          .          .          .
23 item 23     .          *          .          .          .          .          .
24 item 24     .          .          .          .          *          .          .
25 item 25     .          .          .          .          .          *          .
26 item 26     .          *          .          .          .          .          .
27 item 27     .          *          .          .          .          .          .
28 item 28     .          .          .          .          .          .          *
29 item 29     .          *          .          .          .          .          .
30 item 30     .          .          .          .          .          .          *
=====

```

Figura V.

Se puede aún verificar que no existe ningún ítem que no haya sido resuelto por todos los 92 sujetos, existiendo, no obstante, siete (7) sujetos con un resultado total de “0” en el teste (es decir, 7.6% de los sujetos presentan niveles de deterioro cognitivo severo).

El análisis de la Figura IV permite comprobar que de los treinta ítems, solo los ítems 1, 28 y 30 presentan un *Infit* claramente elevado (respectivamente 1.48, 1.60 y 1.60).

Una vez que estamos utilizando el método de estimación conjunta y no tenemos datos debidamente calibrados, podemos ilustrar nuestra cuestión con la reproducción de los resultados de un sujeto tomado de un estudio de Prieto y Delgado (2003, p. 98) – Figura VI.

Por el análisis de la figura VI se puede verificar el mapa de ejecución del alumno 259 a un teste de matemática con 30 ítems. Una vez que los ítems ya estaban calibrados (es decir, ya se conocía su grado de dificultad en una determinada población – fácil/difícil) es así posible determinar la probabilidad de acierto ítem a ítem, demostrando cuales son los sujetos que presentan mejor “*fit*”, es decir, se ajustan al modelo.

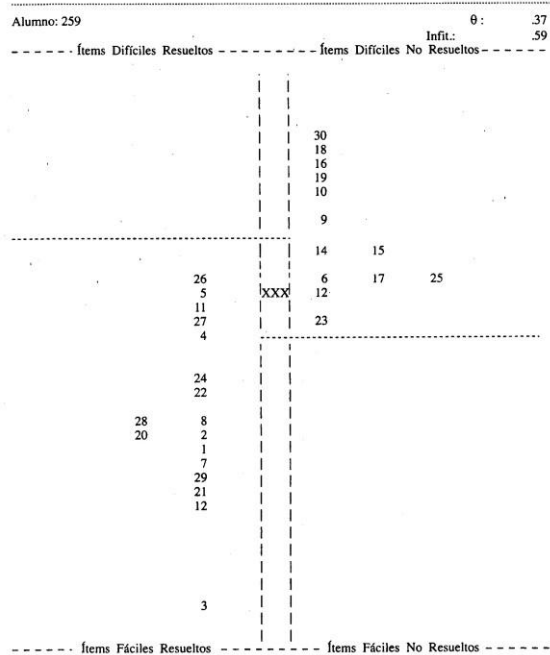


Figura VI – Mapa de los resultados de un alumno en un teste de matemática; Fuente: Prieto y Delgado (2003, p. 98).

La Figura VI muestra un alumno que presenta una ejecución típica de un modelo de Rasch, para un alumno con un resultado mediano (ver discusión de la Figura en el apartado Discusión). Relativamente a los sujetos (Figura VII), 7 sujetos (7,6 %) presentan un *Infit* superior a 1,3 (desde 1,40 hasta 1,78). Además de lo referido relativamente a la posibilidad de representación individual de la relación de cada sujeto con los ítems (Figura VI) y de la estimación de los parámetros para cada sujeto (Figura VII), se puede también presentar la representación gráfica de los sujetos relativamente a su *Infit* (Figura VIII).

Análisis del Minimental. Covilha

Case Estimates In input Order 18-May-2004 13:51:05  
 all on all (N = 92 L = 30 Probability Level= .50)

NAME	SCORE	MAXSCR	ESTIMATE	ERROR	INFIT MNSQ	OUTFIT MNSQ	INFT t	OUTFT t
1 001	13	30	-.59	.46	1.28	19.75	1.24	5.70
2 002	0	30	Case has zero score					
3 003	21	30	1.28	.52	1.10	.77	.44	-.01
4 004	8	30	-1.70	.49	.88	1.00	-.49	.54
5 005	9	30	-1.46	.48	1.12	1.31	.63	.67
6 006	8	30	-1.70	.49	.94	.66	-.23	.29
7 007	25	30	2.58	.64	.81	2.36	-.39	1.16
8 008	1	30	-4.74	1.09	.61	.09	-.28	2.28
9 009	4	30	-2.88	.62	.70	.37	-.69	.68
10 010	3	30	-3.31	.69	.61	.29	-.80	.93
11 011	26	30	3.02	.69	1.18	2.23	.53	1.12
12 012	0	30	Case has zero score					
13 013	27	30	3.56	.78	.42	.12	-1.26	.24
14 014	2	30	-3.87	.82	1.61	2.56	1.08	1.84
15 015	20	30	1.01	.51	1.25	1.31	.93	.62
16 016	24	30	2.20	.59	1.78	2.68	1.95	1.34
17 017	5	30	-2.53	.57	1.12	.69	.47	.67
18 018	24	30	2.20	.59	.59	.28	-1.28	-.39
19 019	28	30	4.28	.92	.69	.15	-.33	.83
20 020	26	30	3.02	.69	.51	.18	-1.22	-.04
21 021	13	30	-.59	.46	.94	1.11	-.18	.43
22 022	4	30	-2.88	.62	.70	.37	-.69	.68
23 023	1	30	-4.74	1.09	1.28	.69	.59	2.46
24 024	20	30	1.01	.51	1.20	1.23	.76	.53
25 025	0	30	Case has zero score					
26 026	5	30	-2.53	.57	.93	.88	-.10	.76
27 027	15	30	-.16	.47	1.14	.95	.63	.20
28 028	19	30	.76	.50	1.38	1.35	1.35	.67
29 029	8	30	-1.70	.49	1.13	.85	.63	.44
30 030	14	30	-.37	.46	.69	.52	-1.45	-.38
31 031	12	30	-.80	.46	.99	.69	0.00	.02
32 032	9	30	-1.46	.48	.97	.67	-.08	.21
33 033	8	30	-1.70	.49	1.04	.82	.26	.42
34 034	18	30	.52	.49	.85	.74	-.51	-.16
35 035	23	30	1.86	.56	.69	.43	-.98	-.32
36 036	27	30	3.56	.78	1.48	.99	.98	.84
37 037	17	30	.29	.48	1.07	.86	.34	.04
38 038	17	30	.29	.48	1.10	.89	.46	.09
39 039	23	30	1.86	.56	1.11	2.46	.43	1.33
40 040	11	30	-1.02	.47	.58	.40	-2.40	-.31
41 041	19	30	.76	.50	1.08	1.07	.38	.33
42 042	8	30	-1.70	.49	.78	.47	-1.01	.12
43 043	6	30	-2.22	.54	1.13	.73	.53	.55
44 044	26	30	3.02	.69	.75	.79	-.45	.52
45 045	15	30	-.16	.47	1.00	.85	.07	.07
46 046	2	30	-3.87	.82	1.52	1.53	.96	1.70
47 047	3	30	-3.31	.69	1.47	.70	1.02	1.12
48 048	27	30	3.56	.78	.89	.27	-.05	.42
49 049	19	30	.76	.50	.74	.56	-.94	-.48
50 050	21	30	1.28	.52	.51	.29	-2.02	-.91
51 051	22	30	1.56	.54	1.04	.94	.23	.25
52 052	25	30	2.58	.64	.96	.74	.04	.32
53 053	16	30	.06	.47	1.04	1.11	.26	.39

Análisis del Minimental. Covilha (continuación)

Case Estimates In input Order 18-May-2004 13:51:05  
 all on all (N = 92 L = 30 Probability Level= .50)

NAME	SCORE	MAXSCR	ESTIMATE	ERROR	INFIT MNSQ	OUTFIT MNSQ	INFT t	OUTFT t
54 054	2	30	-3.87	.82	1.62	2.68	1.10	1.85
55 055	15	30	-.16	.47	.96	.74	-.07	-.08
56 056	28	30	4.28	.92	.95	.34	.14	.97
57 057	15	30	-.16	.47	1.00	.79	.10	-.01
58 058	0	30	Case has zero score					
59 059	14	30	-.37	.46	.84	1.77	-.68	1.01
60 060	4	30	-2.88	.62	1.08	.53	.32	.78
61 061	0	30	Case has zero score					
62 062	21	30	1.28	.52	1.14	1.06	.57	.35
63 063	10	30	-1.24	.47	.77	.50	-1.17	-.05
64 064	7	30	-1.95	.51	1.20	2.25	.82	1.12
65 065	26	30	3.02	.69	.51	.18	-1.22	-.04
66 066	1	30	-4.74	1.09	1.28	.69	.59	2.46
67 067	11	30	-1.02	.47	1.11	1.39	.59	.69
68 068	29	30	5.38	1.21	.38	.05	-.74	1.85
69 069	23	30	1.86	.56	.99	1.53	.08	.78
70 070	19	30	.76	.50	.85	.57	-.51	-.45
71 071	7	30	-1.95	.51	1.03	.81	.22	.50
72 072	8	30	-1.70	.49	1.43	1.22	1.79	.66
73 073	15	30	-.16	.47	.67	.50	-1.46	-.50
74 074	12	30	-.80	.46	1.21	1.50	1.00	.77
75 075	1	30	-4.74	1.09	1.40	38.15	.70	3.45
76 076	14	30	-.37	.46	.78	.60	-.95	-.25
77 077	10	30	-1.24	.47	1.11	.96	.61	.39
78 078	10	30	-1.24	.47	.66	.43	-1.89	-.15
79 079	0	30	Case has zero score					
80 080	15	30	-.16	.47	.64	.50	-1.65	-.49
81 081	21	30	1.28	.52	.65	.37	-1.33	-.71
82 082	9	30	-1.46	.48	.76	.48	-1.20	.02
83 083	13	30	-.59	.46	1.28	19.75	1.24	5.70
84 084	0	30	Case has zero score					
85 085	21	30	1.28	.52	1.10	.77	.44	-.01
86 086	8	30	-1.70	.49	.88	1.00	-.49	.54
87 087	9	30	-1.46	.48	1.12	1.31	.63	.67
88 088	8	30	-1.70	.49	.94	.66	-.23	.29
89 089	25	30	2.58	.64	.81	2.36	-.39	1.16
90 090	1	30	-4.74	1.09	.61	.09	-.28	2.28
91 091	4	30	-2.88	.62	.70	.37	-.69	.68
92 092	3	30	-3.31	.69	.61	.29	-.80	.93
Mean			-.35		.97	1.78	-.07	.68
SD			2.38		.29	4.96	.89	1.10

Figura VII – Estimación de los sujetos según el program *Quest*.

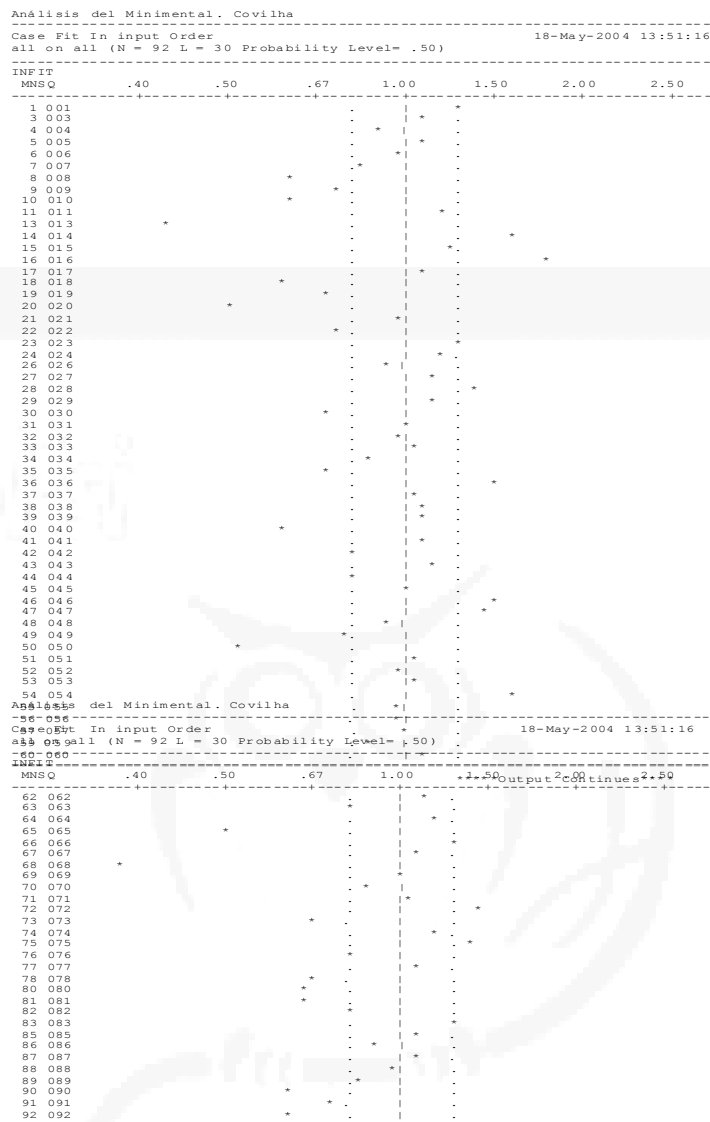


Figura VIII – Estimación gráfica del *Infit* de los sujetos según el program *Quest*.

En la Figura IX se comprueba una representación gráfica del escalamiento conjunto, donde se verifica la representación de las personas y de los ítems en un rango de valores entre -5 y 6 *logit*.

En la Figura X son presentadas las principales características de cada ítem en términos de su capacidad discriminativa, bien como el porcentaje de sujetos que los resuelve correctamente. Tomemos como ejemplo el ítem 1. Por los datos presentados, se comprueba que apenas 2 sujetos (2.2%, en un total de 92) responden correctamente al ítem 1. Su índice de discriminación es de 0.00.

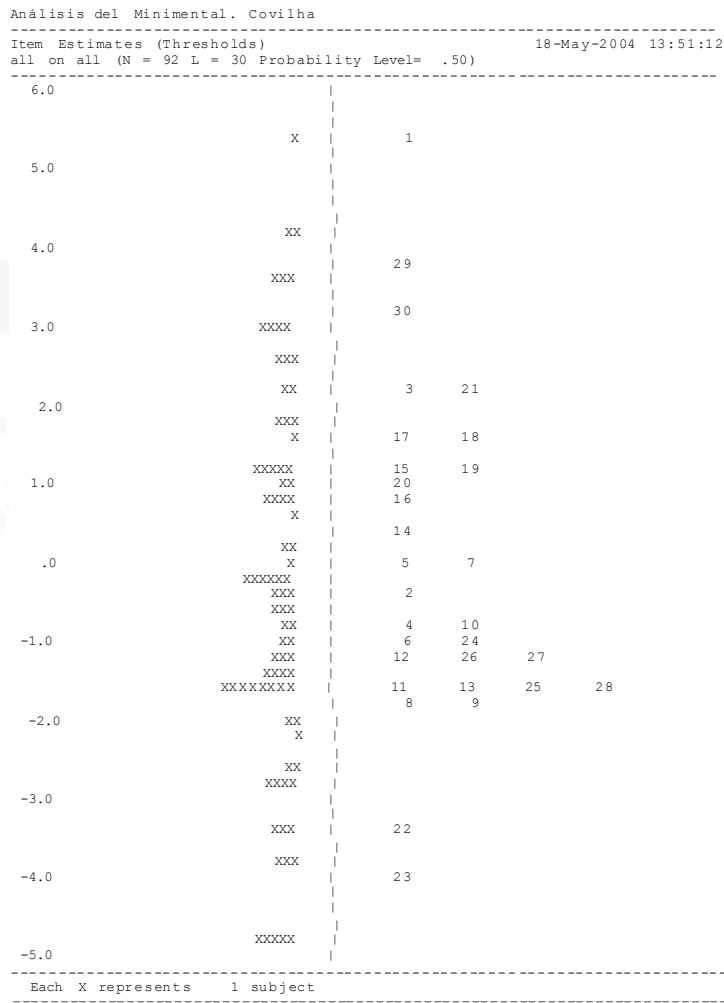


Figura IX – Representación gráfica del escalamiento conjunto.

### ***Revaloración excluyendo los sujetos que no se adaptan al Modelo***

Una cuestión que se puede plantear relacionado con los datos presentados anteriormente es *¿Cómo se comportarían los datos se excluimos los sujetos con respuestas aberrantes, es decir, con un infit superior a 1,3?*

La preocupación es simple. Una vez que estamos a proceder a una *estimación conjunta de parámetros*, o sea, estamos a calibrar los valores de los datos bien como también de las personas, y como los sujetos con *infit* superior a 1,3 pueden estar a sesgar la calibración de los datos, se estuviéramos a desarrollar un nuevo teste, lo mejor sería quitar los sujetos en cuestión. Así, después de eliminados los sujetos con un *infit* superior a 1,3 se presentan de seguida los datos reanalizados. Este esfuerzo correspondería a una tentativa de calibración de los datos, para que el teste desarrollado, o mejor, cada uno de sus ítems, lograsen evaluar la forma más adecuada posible del constructo pretendido. El análisis de la Figura XI



demuestra varios parámetros dignos de nota. La fiabilidad de estimación de los ítems (*Reliability of Estimate*) es muy elevada, con un índice de 0.97.

Análisis del Minimental. Covilha					Análisis del Minimental. Covilha (continuación)						
Item Analysis Results for Observed Responses all on all (N = 92 L = 30 Probability Level= .50)					Item Analysis Results for Observed Responses all on all (N = 92 L = 30 Probability Level= .50)						
Item	Categories	Count	Percent (%)	Pt-Biserial	Disc	Item	Categories	Count	Percent (%)	Pt-Biserial	Disc
Item 1: item 1	0	1*	missing		Disc = 0.00	Item 16: item 16	0	1*	missing		Disc = .71
Count	90	2	0			Count	63	29	0		
Percent (%)	97.8	2.2				Percent (%)	68.5	31.5			
Pt-Biserial	0.00	0.00				Pt-Biserial	-.71	.71			
p-value	.493	.493				p-value	.000	.000			
Item 2: item 2	0	1*	missing		Disc = .74	Item 17: item 17	0	1*	missing		Disc = .68
Count	50	42	0			Count	71	21	0		
Percent (%)	54.3	45.7				Percent (%)	77.2	22.8			
Pt-Biserial	-.74	.74				Pt-Biserial	-.68	.68			
p-value	.000	.000				p-value	.000	.000			
Item 3: item 3	0	1*	missing		Disc = .60	Item 18: item 18	0	1*	missing		Disc = .68
Count	75	17	0			Count	70	22	0		
Percent (%)	81.5	18.5				Percent (%)	76.1	23.9			
Pt-Biserial	-.59	.59				Pt-Biserial	-.68	.68			
p-value	.000	.000				p-value	.000	.000			
Item 4: item 4	0	1*	missing		Disc = .74	Item 19: item 19	0	1*	missing		Disc = .61
Count	45	47	0			Count	66	26	0		
Percent (%)	48.9	51.1				Percent (%)	71.7	28.3			
Pt-Biserial	-.74	.74				Pt-Biserial	-.60	.60			
p-value	.000	.000				p-value	.000	.000			
Item 5: item 5	0	1*	missing		Disc = .60	Item 20: item 20	0	1*	missing		Disc = .68
Count	55	37	0			Count	64	28	0		
Percent (%)	59.8	40.2				Percent (%)	69.6	30.4			
Pt-Biserial	-.59	.59				Pt-Biserial	-.68	.68			
p-value	.000	.000				p-value	.000	.000			
Item 6: item 6	0	1*	missing		Disc = .72	Item 21: item 21	0	1*	missing		Disc = .51
Count	43	49	0			Count	75	17	0		
Percent (%)	46.7	53.3				Percent (%)	81.5	18.5			
Pt-Biserial	-.71	.71				Pt-Biserial	-.51	.51			
p-value	.000	.000				p-value	.000	.000			
Item 7: item 7	0	1*	missing		Disc = .69	Item 22: item 22	0	1*	missing		Disc = .63
Count	55	37	0			Count	20	72	0		
Percent (%)	59.8	40.2				Percent (%)	21.7	78.3			
Pt-Biserial	-.69	.69				Pt-Biserial	-.63	.63			
p-value	.000	.000				p-value	.000	.000			
Item 8: item 8	0	1*	missing		Disc = .66	Item 23: item 23	0	1*	missing		Disc = .60
Count	34	58	0			Count	16	76	0		
Percent (%)	37.0	63.0				Percent (%)	17.4	82.6			
Pt-Biserial	-.66	.66				Pt-Biserial	-.60	.60			
p-value	.000	.000				p-value	.000	.000			
Item 9: item 9	0	1*	missing		Disc = .65	Item 24: item 24	0	1*	missing		Disc = .68
Count	34	58	0			Count	43	49	0		
Percent (%)	37.0	63.0				Percent (%)	46.7	53.3			
Pt-Biserial	-.64	.64				Pt-Biserial	-.68	.68			
p-value	.000	.000				p-value	.000	.000			
Item 10: item 10	0	1*	missing		Disc = .78	Item 25: item 25	0	1*	missing		Disc = .59
Count	46	46	0			Count	37	55	0		
Percent (%)	50.0	50.0				Percent (%)	40.2	59.8			
Pt-Biserial	-.78	.78				Pt-Biserial	-.59	.59			
p-value	.000	.000				p-value	.000	.000			
Item 11: item 11	0	1*	missing		Disc = .75	Item 26: item 26	0	1*	missing		Disc = .76
Count	35	57	0			Count	41	51	0		
Percent (%)	38.0	62.0				Percent (%)	44.6	55.4			
Pt-Biserial	-.75	.75				Pt-Biserial	-.76	.76			
p-value	.000	.000				p-value	.000	.000			
Item 12: item 12	0	1*	missing		Disc = .75	Item 27: item 27	0	1*	missing		Disc = .76
Count	40	52	0			Count	40	52	0		
Percent (%)	43.5	56.5				Percent (%)	43.5	56.5			
Pt-Biserial	-.75	.75				Pt-Biserial	-.76	.76			
p-value	.000	.000				p-value	.000	.000			
Item 13: item 13	0	1*	missing		Disc = .69	Item 28: item 28	0	1*	missing		Disc = .51
Count	37	55	0			Count	36	56	0		
Percent (%)	40.2	59.8				Percent (%)	39.1	60.9			
Pt-Biserial	-.68	.68				Pt-Biserial	-.51	.51			
p-value	.000	.000				p-value	.000	.000			
Item 14: item 14	0	1*	missing		Disc = .75	Item 29: item 29	0	1*	missing		Disc = .46
Count	59	33	0			Count	85	7	0		
Percent (%)	64.1	35.9				Percent (%)	92.4	7.6			
Pt-Biserial	-.74	.74				Pt-Biserial	-.46	.46			
p-value	.000	.000				p-value	.000	.000			
Item 15: item 15	0	1*	missing		Disc = .77	Item 30: item 30	0	1*	missing		Disc = .32
Count	67	25	0			Count	82	10	0		
Percent (%)	72.8	27.2				Percent (%)	89.1	10.9			
Pt-Biserial	-.76	.76				Pt-Biserial	-.32	.32			
p-value	.000	.000				p-value	.001	.001			

Figura X – Análisis individual de los parámetros discriminativos de los ítems.

A su vez, los indicadores de *Infit*, para la estimación de los ítems, son casi completos, con un valor medio de 1.00 (desviación típica de .29). Lo mismo no ocurre para la estimación del *Outfit* de los ítems, con un valor de 1.41. Por otro lado, la estimación de los casos presenta un índice de fiabilidad igualmente fuerte, con un valor de 0.93. Aunque en el análisis de la Figura XI se pueda verificar que, en lo que concierne a los sujetos, el *Infit* es elevado (0.95) siendo el *Outfit*, de 1.41.

Análisis del Minimental. Covilha				Análisis del Minimental. Covilha			
Item Estimates (Thresholds)				03-Jun-2004 10:09:01			
all on all (N = 84 L = 30 Probability Level= .50)				all on all (N = 84 L = 30 Probability Level= .50)			
Summary of item Estimates				Summary of case Estimates			
Mean	0.00			Mean	-.20		
SD	2.27			SD	2.39		
SD (adjusted)	2.23			SD (adjusted)	2.31		
Reliability of estimate	.97			Reliability of estimate	.93		
Fit Statistics				Fit Statistics			
Infit Mean Square		Outfit Mean Square		Infit Mean Square		Outfit Mean Square	
Mean	1.00	Mean	1.41	Mean	.95	Mean	1.41
SD	.29	SD	2.61	SD	.27	SD	3.35
Infit t		Outfit t		Infit t		Outfit t	
Mean	-.15	Mean	.34	Mean	-.05	Mean	.83
SD	1.37	SD	.88	SD	.84	SD	1.13
0 items with zero scores				7 cases with zero scores			
0 items with perfect scores				1 cases with perfect scores			

Figura XI – Re-estimación de los ítems y de los casos de acuerdo con el programa *Quest*.

En su representación de los parámetros (Figura XII) y gráfica (Figura XIII) los ítems 1, 28 y 30 continúan a presentar un *Infit* superior a 1.3, no se ajustando así al modelo.

En la Figura XIV se verifica una representación gráfica del escalamiento conjunto después de quitados los sujetos con mayor desajuste al modelo, donde se señala la representación de las personas y de los ítems en un rango de valores entre -6 y 6 *logit*.

Como se evidencia en la Figura XIV, una vez más, los objetos (datos y sujetos) presentan una superposición casi absoluta, demostrando que las distribuciones de los parámetros se presentan con distribución normal.

En la Figura XV se presenta la representación gráfica de los parámetros de los ítems. Así se verifica de forma evidente que los ítems 1, 28 y 30 están posicionados muy para la derecha del límite 1,30 para un *infit* aceptable (se recuerda que el *Infit* absoluto sería el 1.0).

Análisis del Minimental. Covilha

Item Estimates (Thresholds) In input Order 03-Jun-2004 09:00:05  
 all on all (N = 84 L = 30 Probability Level= .50)

	ITEM NAME	SCORE MAXSCR		THRSH	INFT	OUTFT	INFT	OUTFT
				1	MNSQ	MNSQ	t	t
1	item 1	2	83	5.54	1.55	14.81	1.0	3.1
2	item 2	40	83	-.46	.70	.49	-2.1	-.9
3	item 3	15	83	2.31	1.02	.80	.2	.3
4	item 4	44	83	-.83	.81	.58	-1.3	-.6
5	item 5	32	83	.31	1.27	1.22	1.5	.5
6	item 6	46	83	-1.02	.88	.66	-.8	-.3
7	item 7	34	83	.12	.98	.98	-.1	.1
8	item 8	54	83	-1.82	1.17	1.42	1.0	.7
9	item 9	55	83	-1.92	1.11	.76	.7	.1
10	item 10	42	83	-.65	.78	.50	-1.5	-.8
11	item 11	54	83	-1.82	.82	.52	-1.1	-.3
12	item 12	49	83	-1.31	.90	.65	-.6	-.3
13	item 13	52	83	-1.61	1.06	1.29	.4	.6
14	item 14	29	83	.62	.77	.60	-1.3	-.5
15	item 15	22	83	1.40	.59	.31	-2.4	-.8
16	item 16	26	83	.94	.92	.81	-.4	0.0
17	item 17	18	83	1.90	.74	.59	-1.3	0.0
18	item 18	19	83	1.77	.77	.97	-1.1	.3
19	item 19	26	83	.94	1.04	1.11	.3	.4
20	item 20	27	83	.83	.89	.77	-.5	-.1
21	item 21	14	83	2.46	1.22	.87	1.0	.4
22	item 22	69	83	-4.02	.83	.57	-.4	.8
23	item 23	72	83	-4.91	1.07	.38	.3	1.4
24	item 24	45	83	-.93	1.10	1.68	.7	1.1
25	item 25	51	83	-1.51	1.19	2.71	1.2	1.7
26	item 26	47	83	-1.12	.79	.68	-1.5	-.3
27	item 27	49	83	-1.31	.82	.60	-1.2	-.4
28	item 28	51	83	-1.51	1.68	2.29	3.7	1.4
29	item 29	5	83	4.29	.59	.13	-1.3	1.0
30	item 30	9	83	3.33	1.80	2.68	2.5	1.3
Mean				0.00	1.00	1.41	-.1	.3
SD				2.27	.29	2.61	1.4	.9

Figura XII – Estimación de los ítems según el program Quest.

Análisis del Minimental. Covilha

Case Estimates In input Order 03-Jun-2004 09:00:08  
all on all (N = 84 L = 30 Probability Level= .50)

NAME	SCORE	MAXSCR	ESTIMATE	ERROR	INFIT MNSQ	OUTFIT MNSQ	INFT t	OUTFT t	
1 001	13	30	-.58	.47	1.32	21.27	1.37	5.37	
2 002	0	30	Case has zero score						
3 003	21	30	1.37	.53	1.10	.75	.41	.16	
4 004	8	30	-1.74	.50	.89	.90	-.43	.54	
5 005	9	30	-1.50	.49	1.14	1.21	.72	.65	
6 006	8	30	-1.74	.50	.97	.71	-.07	.41	
7 007	25	30	2.72	.65	.82	2.72	-.35	1.28	
8 008	1	30	-5.18	1.16	.51	.06	-.52	3.24	
9 009	4	30	-3.02	.66	.77	.40	-.41	.90	
10 010	3	30	-3.51	.75	.63	.33	-.60	1.24	
11 011	26	30	3.19	.71	1.27	2.52	.69	1.33	
12 012	0	30	Case has zero score						
13 013	27	30	3.75	.80	.41	.11	-1.26	.88	
14 015	20	30	1.09	.52	1.26	1.30	.92	.62	
15 017	5	30	-2.63	.60	1.27	.78	.80	.84	
16 018	24	30	2.33	.61	.64	.31	-1.06	.10	
17 019	28	30	4.49	.93	.79	.16	-.16	1.63	
18 020	26	30	3.19	.71	.49	.17	-1.25	.49	
19 021	13	30	-.58	.47	.98	1.22	-.02	.54	
20 022	4	30	-3.02	.66	.77	.40	-.41	.90	
21 023	1	30	-5.18	1.16	1.64	1.05	-.96	3.42	
22 024	20	30	1.09	.52	1.31	1.40	1.08	.69	
23 025	0	30	Case has zero score						
24 026	5	30	-2.63	.60	.98	1.07	.07	.95	
25 027	15	30	-.14	.48	1.22	1.00	.92	.30	
26 029	8	30	-1.74	.50	1.24	1.05	1.07	.62	
27 030	14	30	-.36	.47	.73	.53	-1.21	-.26	
28 031	12	30	-.81	.47	1.04	.72	.24	.11	
29 032	9	30	-1.50	.49	1.01	.68	.13	.30	
30 033	8	30	-1.74	.50	1.10	.91	.52	.54	
31 034	18	30	.57	.50	.89	.74	-.31	-.01	
32 035	23	30	1.98	.58	.68	.39	-1.00	0.00	
33 037	17	30	.33	.49	1.16	.90	.65	.18	
34 038	17	30	.33	.49	1.13	.94	.56	.23	
35 039	23	30	1.98	.58	1.16	2.77	.57	1.30	
36 040	11	30	-1.03	.47	.60	.39	-2.25	-.21	
37 041	19	30	.82	.51	1.16	1.11	-.64	.43	
38 042	8	30	-1.74	.50	.82	.47	-.80	.22	
39 043	6	30	-2.30	.55	1.24	1.03	.84	.80	
40 044	26	30	3.19	.71	.79	.88	-.35	.91	
41 045	15	30	-.14	.48	1.07	.91	.35	.21	
42 048	27	30	3.75	.80	.98	.27	.13	1.00	
43 049	19	30	.82	.51	.81	.57	-.64	-.20	
44 050	21	30	1.37	.53	.56	.31	-1.71	-.44	
45 051	22	30	1.66	.55	1.07	.98	.31	.63	
46 052	25	30	2.72	.65	1.04	.75	.22	.44	
47 053	16	30	-.09	.48	1.11	1.23	.48	.54	
48 055	15	30	-.14	.48	.99	.74	.05	0.00	
49 056	28	30	4.49	.93	1.04	.37	.27	1.72	
50 057	15	30	-.14	.48	1.05	.80	.27	.08	
51 058	0	30	Case has zero score						
52 059	14	30	-.36	.47	.87	1.97	-.49	1.10	
53 060	4	30	-3.02	.66	1.26	.67	.68	1.02	
54 061	0	30	Case has zero score						
55 062	21	30	1.37	.53	1.24	1.15	.86	.52	
56 063	10	30	-1.26	.48	.81	.51	-.95	.04	
57 064	7	30	-2.01	.52	1.28	2.76	1.08	1.27	
58 065	26	30	3.19	.71	.52	.18	-1.13	.50	
59 066	1	30	-5.18	1.16	1.64	1.05	.96	3.42	
60 067	11	30	-1.03	.47	1.17	1.92	.86	1.00	
61 068	29	30	5.58	1.19	.45	.05	-.67	3.08	
62 069	23	30	1.98	.58	.99	1.55	.07	.82	
63 070	19	30	.82	.51	.82	.56	-.59	-.22	
64 071	7	30	-2.01	.52	1.10	.84	.46	.59	
65 073	15	30	-.14	.48	.67	.48	-1.43	-.39	
66 074	12	30	-.81	.47	1.23	1.75	1.09	.92	
67 076	14	30	-.36	.47	.83	.63	-.70	-.11	
68 077	10	30	-1.26	.48	1.17	1.04	.87	.50	
69 078	0	30	-1.26	.48	.66	.41	-1.90	-.07	
70 079	0	30	Case has zero score						
71 080	15	30	-.14	.48	.63	.48	-1.61	-.40	
72 081	21	30	1.37	.53	.67	.37	-1.18	-.33	
73 082	9	30	-1.50	.49	.79	.47	-1.03	.11	
74 083	13	30	-.58	.47	1.32	21.27	1.37	5.37	
75 084	0	30	Case has zero score						
76 085	21	30	1.37	.53	1.10	.75	.41	.16	
77 086	8	30	-1.74	.50	.89	.90	-.43	.54	
78 087	9	30	-1.50	.49	1.14	1.21	.72	.65	
79 088	8	30	-1.74	.50	.97	.71	-.07	.41	
80 089	25	30	2.72	.65	.82	2.72	-.35	1.28	
81 090	1	30	-5.18	1.16	.51	.06	-.52	3.24	
82 091	4	30	-3.02	.66	.77	.40	-.41	.90	
83 092	3	30	-3.51	.75	.63	.33	-.60	1.24	
Mean			-.20		.95	1.41	-.05	.83	
SD			2.39		.27	3.35	.84	1.13	

Figura XIII – Estimación de los sujetos según el program *Quest*.

En la figura XVI se presenta la representación gráfica de los parámetros de los sujetos. Así se verifica que solo los sujetos 23 y 66 están posicionados muy para la derecha del limite 1,30 para un *infit* aceptable (se recuerda que el *Infit* absoluto sería el 1.0).

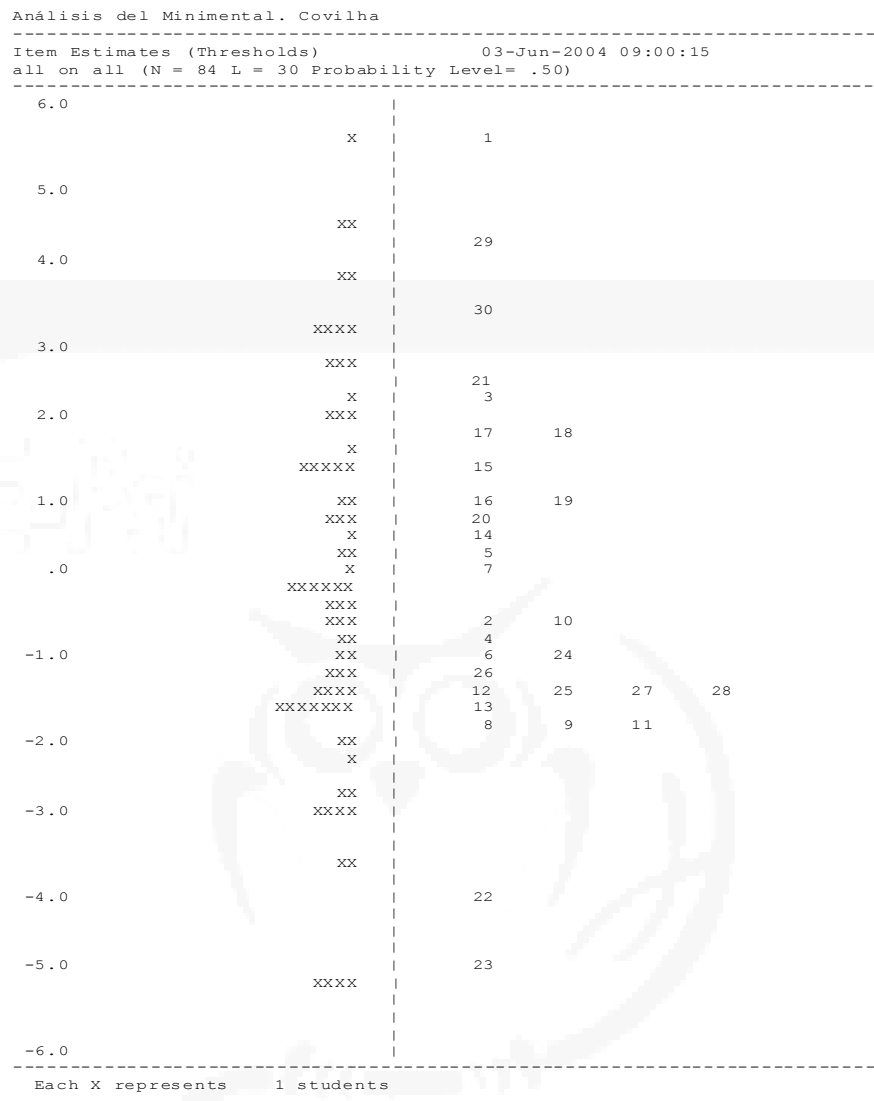


Figura XIV – Representación gráfica del escalamiento conjunto después de la re-estimación de parámetros.

### Discusión

Recordamos que utilizando los baremos específicos para la población portuguesa con el MMSE, los 92 sujetos originales presentan una media de 13,32 con una desviación típica elevada de 8,936. Estos resultados demuestran que los resultados de los sujetos varían de forma considerable ( $13,32 \pm 8,936$ ). De hecho, los 92 sujetos evaluados presentan una gran variación en el nivel de deterioro cognitivo no solo cuando son evaluados con un teste objetivo como el MMSE, sino también teniendo en cuenta indicadores de observación y de información recogida junto a los cuidadores.

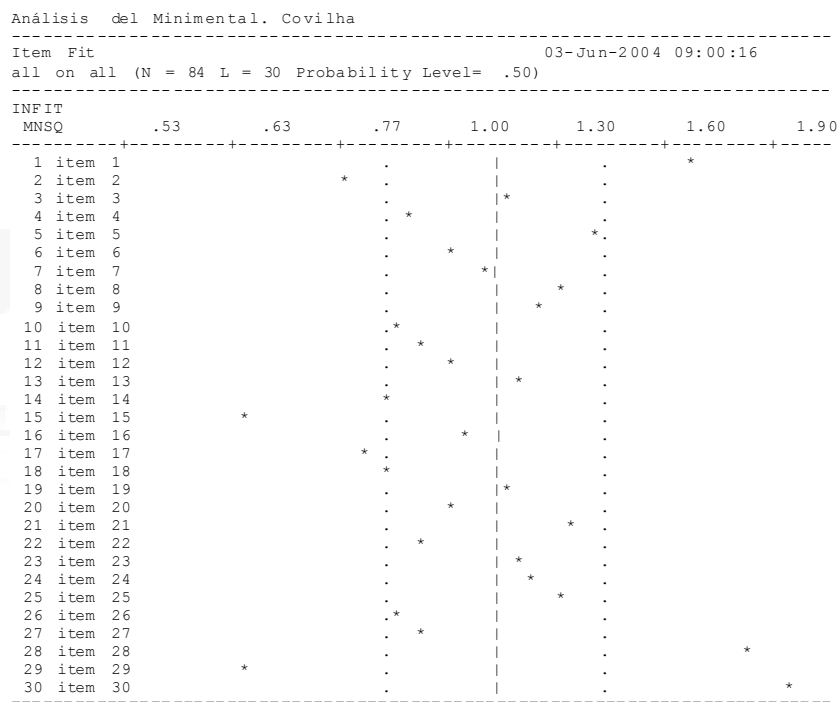


Figura XV – Re-estimación gráfica del *Infit* de los ítems según el program *Quest*.

El análisis de fiabilidad permite estudiar las propiedades de medición del teste y de los ítems que los constituyen. El procedimiento de *Análisis de Fiabilidad* calcula varias medidas de fiabilidad del teste proporcionando también información sobre las relaciones entre los ítems individuales en el teste. Pueden usarse los coeficientes de correlación intra-clase para analizar las estimaciones de fiabilidad.

La pregunta, relacionada con nuestro trabajo es: *¿El MMSE mide el deterioro cognitivo de los sujetos de una manera útil?* El análisis de fiabilidad usado puede determinar hasta que punto se relacionan los ítems en el MMSE, puede presentar un índice global de repetición o de consistencia interna del teste en conjunto, o bien puede identificar ítems problemáticos que deben excluirse del teste (con base en SPSS – *Statistical Package for Social Science* – Version 13 – *Topics*). Con base en las características de nuestro estudio nos parece que el indicador más apropiado es el *Índice Alpha de Cronbach*, que presenta indicadores de consistencia interna, basado en la correlación media inter-ítems.

En la Tabla VII se mostró que el índice de *Consistencia Interna*, calculado el *Coefficiente Alpha de Cronbach* fué de .95, demostrando que el teste, en esta muestra de

sujetos, se presenta como bastante fiable a la hora de acceder al conjunto de constructos que se pretenden medir con el teste (Cronbach, 1951; Almeida & Freire, 1997).

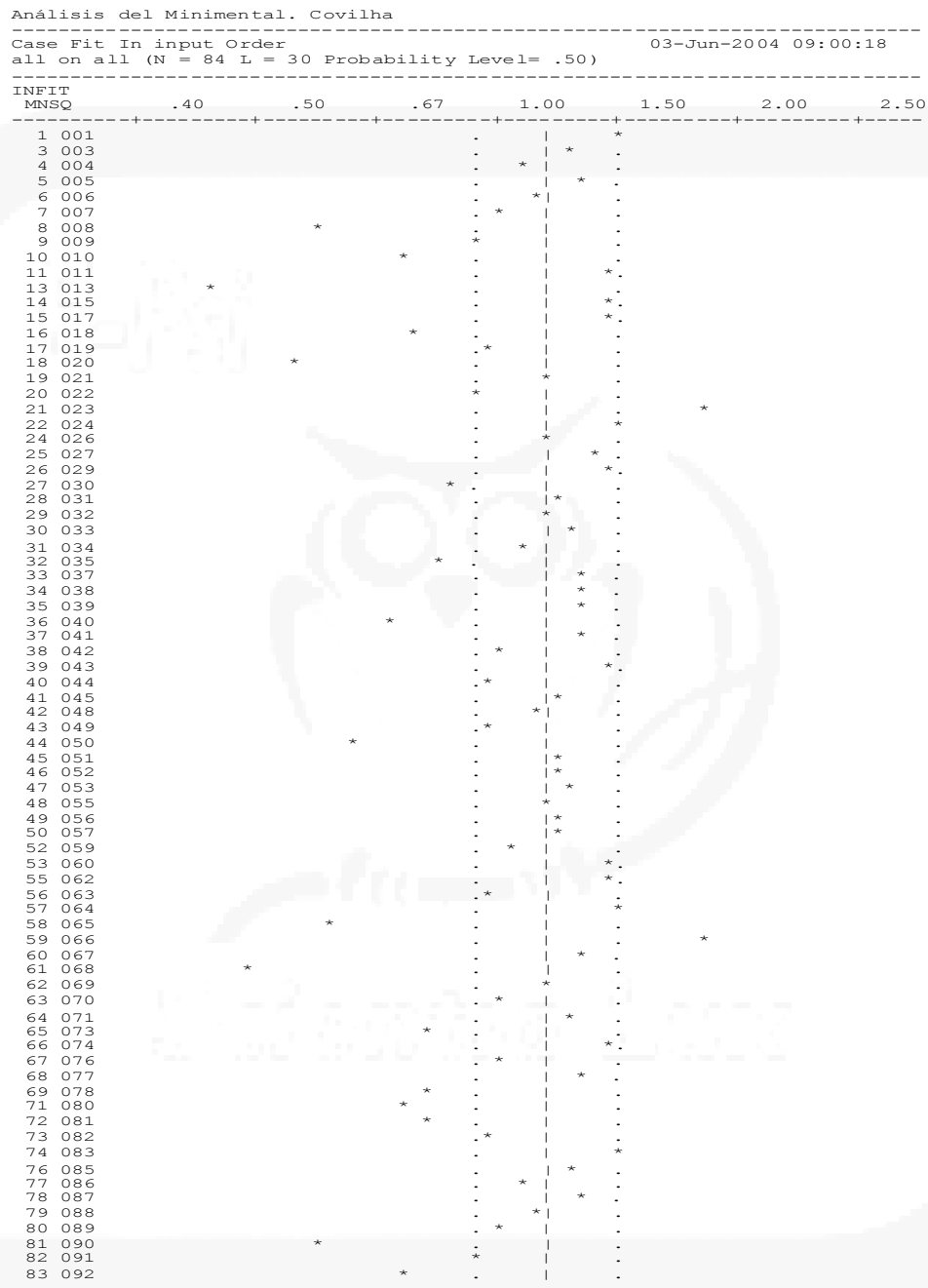


Figura XVI – Re-estimación gráfica del *Infit* de los sujetos según el program *Quest*.

El análisis de las correlaciones de los ítems con el resultado total en el MMSE, como las correlaciones ítem-ítem permite verificar el dato curioso que todos los ítems presentan niveles de correlación elevada con el resultado final, a un elevado nivel de significancia estadística ( $p < .01$ ) con excepción del Ítem 29. Según las TCT estos datos no son de gran

utilidad una vez que la única indicación que dan al análisis de los datos es que, los ítems, con excepción del 29, se correlacionan fuertemente con el resultado global. Así, dos grandes consideraciones se pueden plantear: según las TCT los datos presentados no permiten un análisis más profundizado que la presentada, así como también no permite analizar particularmente los resultados de los sujetos en el ítem 29.

En la Figura III podemos verificar que, en lo que concierne a los 92 sujetos originales, el *Infit* es muy bueno (0.97) siendo el *Outfit*, de 1.78, muy elevado. Esta conjugación demuestra que, aunque los ítems del teste se adapten muy adecuadamente al modelo, algunos sujetos presentan resultados muy inconsistentes, y por eso, muy desadaptados relativamente al modelo (nombradamente los sujetos 75 con un *Outfit* de 38.15, el 83 con un *Outfit* de 19.75, así como varios sujetos con *Outfit* superior a 2 – en el total, 15 sujetos – 16 % – presentan *Outfit* igual o superior a 1.39).

Por el análisis de la Figura IV se permite comprobar que de los 30 ítems, solo los ítems 1, 28 y 30 presentan un *Infit* claramente elevado (respectivamente 1.48, 1.60 y 1.60). El análisis minucioso de los ítems en cuestión permite comprender mejor el porque de este desajuste. Relativamente a la Cuestión “1 – ¿En que año estamos?” puede plantearse la hipótesis que, en esta muestra de ancianos, esta cuestión se configure como de dificultad elevadísima (¡2 sujetos la responden correctamente! – Figura IV). La edad media de los sujetos es bastante elevada – 80.48, con un error estándar de 12.24 (como se puede verificar por la Tabla I). De hecho, en la literatura, ha sido referido que la orientación temporal es uno de los aspectos más afectados en ancianos de edad avanzada (> 75 años) (American Psychiatric Association – DSM-IV, 1994; Cf. también algunos trabajos del autor relativos a la temática de la memoria y/o sus afecciones en ancianos: Matteo, Cabaco, Capataz & Maia, 2003; Maia & de Mendonça, 2002a; Maia & de Mendonça, 2002b; Maia, 2001).

Si añadimos a este hecho la demostración que estos ancianos están en situación de ingreso institucional se puede sustentar que estos resultados se presentan como normales, o al menos, esperables, en tal muestra. El ítem 28, con una instrucción escrita “¡Cierre su ojos!”, pero con la posibilidad de ser presentada grafémica ó fonéticamente, y el hecho de 36 sujetos (39.1%) no lo teneren resuelto adecuadamente refleja por un lado los índices elevados de iliteracia o baja escolaridad en esta muestra de sujetos bien como la dificultad en comprender y seguir adecuadamente instrucciones verbales.

Los bajos resultados en la cuestión 30 (habilidad constructiva – “reproducir gráficamente dos figuras superpuestas”) parecen reflejar los niveles elevados de



(dis)praxias, dificultades de conjugación óculo-motora como la falta de práctica en tales tareas, sobretodo en sujetos con niveles etáricos tan avanzados.

Por estas razones, si estuviéramos en una fase de construcción del teste *Mini Mental State Examination*, o al menos su adaptación para la población anciana ingresada de larga duración, estos ítems deberían ser excluidos, no solo para que el teste se ajustase al modelo, sino que, de hecho, estos ítems se plantean como de dificultad muy elevada, luego, potencialmente desajustada a este tipo de población. Por ejemplo, podrían ser sustituidos por otros ítems que pudiesen acceder a los mismos parámetros del constructo que se quisiera medir, sin hacerlo con estos 3 ítems.

Con esta metodología del Modelo de Rasch, se podría entonces adaptar mejor el teste a la medición de un conjunto de parámetros que, al final, busca evaluar varias dimensiones de un mismo constructo (o se quisieramos, en este caso, un conjunto de constructos, una vez que, un teste de evaluación del nivel de orientación o deterioro cognitivo evalúa un conjunto variado de aspectos *constructuales*).

Cuanto a la representación gráfica con los 92 sujetos originales (no re-calibrados) (Figura V) los ítems 1, 28 y 30 se presentan con un *Infit* superior a 1.3, demostrando que estos ítems, en esta muestra de sujetos, no se ajustan al modelo.

Relativamente a los sujetos representados en la Figura VII, 7 sujetos (7,6 %) presentan un *Infit* superior a 1,3 (desde 1,40 hasta 1,78). Estos datos indican que la forma como los sujetos están a contestar los ítems, utilizando el método de *estimación conjunta*, hace que no se adapten al modelo. Según el mismo raciocinio, cuándo intentamos comprender cual sería el motivo que hace que los ítems 1, 28 y 30 sean los que más se desajustaban al modelo, aquí se puede también intentar comprender lo que tenga contribuido para ese desajuste. Para eso se puede recurrir a la presentación gráfica de los resultados particulares de cada sujeto en un gráfico de estimación conjunta donde se demuestra el resultado específico del sujeto en cada ítem, y se ese ítem es considerado fácil o difícil (relativamente a una *calibración previa*). Una vez que estamos utilizando el método de estimación conjunta y no tenemos datos debidamente calibrados, podemos ilustrar nuestra cuestión con la reproducción de los resultados de un sujeto tomado de un estudio de Prieto y Delgado (2003, p. 98) – Figura VI.

Por el análisis de la figura VI se puede verificar el mapa de ejecución del alumno 259 a un teste de matemática con 30 ítems. Una vez que los ítems ya estaban calibrados (es decir, ya se conocía su grado de dificultad en una determinada población – fácil/difícil) es así posible determinar la probabilidad de acierto ítem a ítem, demostrando cuales son los sujetos que presentan mejor “*fit*”, es decir, se ajustan al modelo.

La Figura VI muestra un alumno que presenta una ejecución típica de un modelo de Rasch, para un alumno con un resultado mediano. Así, en el mapa, se ve representado en nivel medio del sujeto (XXX), bien como de los ítems en un continuo.

El intervalo  $\theta \pm SE(\theta)$  es presentado en las dos líneas de puntos. Así, según los autores, los ítems en que el alumno presenta una mayor probabilidad de resolución correcta se presentan por bajo de la línea de puntos en el lado izquierdo; los ítems que el alumno presenta una menor probabilidad de resolución adecuada se presentan por cima de la línea del lado derecho. En este caso, este alumno presenta un desempeño normativo a su nivel de performance (situado en el punto XXX).

El alumno en cuestión acierta todos los ítems fáciles (ítems por bajo de la línea de puntos del lado izquierdo) y no resuelve los ítems difíciles de la prueba (ítems por cima de la línea de puntos del lado derecho). Se verifica que el *Infit* de este alumno es de .59, demostrando un ajuste al modelo, pero permitiendo verificar que es un alumno mediano (o sea, regla general, resuelve los ítems más fáciles de su nivel de desempeño, pero no resuelve los que están por cima de su desempeño medio).

Si nuestros sujetos con un *infit* superior a 1.3 pudiesen ser representados según este modelo de diagnóstico individual, probablemente se demostrarían patrones aberrantes o al menos, que no se ajustarían al modelo. Por ejemplo, cuando analizamos particularmente los sujetos con *infits* superiores o iguales a 1.3 verificamos que la elevación del *infit* en algunos sujetos parece deberse a su performance insuficiente en el teste (ejemplos de los sujetos 75, 54, 46 y 14, con resultados de 1, 2, 2 y 2 respectivamente – se recuerda que el máximo posible de acierto en el MMSE es de 30 puntos; además, la media de ítems correctamente resueltos en esta muestra de 92 sujetos es de 13 puntos). Todavía, los sujetos 36, 28 y 16 presentan un resultado claramente superior a la media, 27, 19 y 24 respectivamente, pero *infits* de 1,48, 1,38 y 1,78. Estos resultados, si presentados en un mapa de ejecución, como el presentado por el alumno de la figura VI, presentarían, probablemente, patrones aberrantes.

Por ejemplo, el sujeto 36, obteniendo un resultado de 27 en el MMSE, presenta un *infit* de 1,48. Es decir, según las normas psicométricas de interpretación del MMSE este sujeto es considerado “*normal*”, de acuerdo con los baremos portugueses, la verdad es que su patrón de respuestas no es completamente lo esperable de acuerdo con la estimación de su nivel de desempeño (el XXX en la Figura VI). Por lo dicho, este modelo es también muy importante para el esclarecimiento de características individuales, al nivel del diagnóstico de aspectos particulares, en la relación sujeto a sujeto, bien como ítem a ítem, así como ítem-sujeto (que ni siempre es posible en la TCT).

Se ha verificado en la Figura IX que los objetos (datos y sujetos) mostraron una superposición casi absoluta, demostrando que las distribuciones de los parámetros se presentan de forma normal, y en la Figura X se puede verificar las principales características de cada ítem en lo que toca a su capacidad discriminativa, bien como el porcentaje de sujetos que los resuelve correctamente. Hemos tomado como ejemplo el ítem 1. Por los datos presentados, se comprueba que apenas 2 sujetos (2.2%, en un total de 92) responden correctamente al ítem 1. Su índice de discriminación es de 0.00. Se demuestra así que este ítem se presenta como poco discriminativo del resultado global, en esta muestra específica de sujetos. Debiendo ser eliminado o ajustado, se estuviéramos creando un teste nuevo.

Analizando los ítems 28 y 30, según los mismos parámetros (“*parámetros discriminativos de los ítems*” – Figura X), se verifica respectivamente niveles de acierto de 60.9 % (n = 56) y 10.9 % (n = 10). Los respectivos niveles de discriminación son también .51 y .32. Del análisis global de los tres ítems con un desajuste más evidente, el ítem 28 es el que se presenta como el más susceptible de no ser sustituido, pero si adaptado, una vez que los niveles de error pueden ser explicados por la baja escolaridad de la muestra. Así, una metodología de presentación sistemática del ítem en una nueva formulación (por ejemplo, ser presentada siempre de forma verbal) o su sustitución por una instrucción diferente como “*ponga su mano derecha por encima de su cabeza*”, una vez que este ítem pretende acceder no solo a habilidades de lenguaje escrita receptiva sino también a la capacidad de comprensión de instrucciones y su capacidad para seguirlas adecuadamente.

Se debe llevar en consideración que, la estimación de los casos presenta un índice de fiabilidad igualmente fuerte, con un valor de 0.93. Aunque en el análisis de la figura XI se pueda verificar que, en lo que concierne a los sujetos, el *Infit* es elevado (0.95) siendo el *Outfit*, de 1.41. Esta conjugación refuerza lo que referimos anteriormente acerca de los ítems se adaptaren al modelo, existiendo algunos sujetos en la muestra original de 92 sujetos que contribuían para un cierto desajuste al modelo. Con esta revaloración con 84 sujetos los ítems continúan a presentar una fuerte adaptación con excepción hecha a los ya referidos 1, 28 y 30. Algunos sujetos presentan resultados muy inconsistentes, y por eso, muy desadaptados relativamente al modelo (nombradamente los sujetos 75 con un *Outfit* de 38.15, el 83 con un *Outfit* de 19.75, así como varios sujetos con *Outfit* superior a 2 – en el total, 15 sujetos – 16 % – presentan *Outfit* igual o superior a 1.39).

Hemos referido que en su representación de los parámetros (Figura XII) y gráfica (Figura XIII) los ítems 1, 28 y 30 continúan a presentar un *Infit* superior a 1.3, no se ajustando así al modelo. Por lo expuesto se fortalece la interpretación de la dificultad de

estés ítems en esta población de sujetos, mismo cuando quitados los siete sujetos de la población original de 92 ancianos. Si consideramos que los siete sujetos retirados podrían estar a contribuir de forma considerable para el desajuste de los ítems 1, 28 y 30, y el hecho de después de quitarlos (a los sujetos) los ítems se mantienen con indicadores de desajuste (elevada dificultad), podríamos reforzar nuestra interpretación inicial que estés ítems se presentan como desajustados a la hora de evaluar sujetos ancianos, ingresados en instituciones de acogimiento, donde se comprueba ya un claro deterioro cognitivo.

Relativamente a los 84 sujetos que contribuyeron para la reestimación de parámetros apenas los sujetos 23 y 66 presentan un *infit* elevado (1.64). Cuando evaluamos el desempeño específico de los dos sujetos, verificamos que los sujetos solo responden correctamente 1 ítem.

Como hemos referido para la Figura XVI se presenta la representación gráfica de los parámetros de los sujetos donde se verifica que solo los sujetos 23 y 66 están posicionados muy para la derecha del límite 1,30 para un *infit* aceptable (se recuerda que el *Infit* absoluto sería el 1.0). Los dos sujetos presentan un *infit* de 1,66. Así, se sugeriría la evaluación específica del patrón de respuestas de estos dos sujetos (por ejemplo, tras la representación conjunta de los ítems fáciles y difíciles resueltos y no resueltos, bien como sus respectivos niveles medios de desempeño).

## **Conclusion**

En este artículo se presentaron datos e análisis que permitieron el abordaje del tema de la psicometría desde el punto de vista de los modelos de medición objetivos, específicamente, lo Modelo de Rasch.

Un primer nivel de conclusión que podemos plantear es que el Modelo de Rasch permite un análisis profundo de ciertos parámetros usualmente no considerados en la TCT. Con esta metodología podemos certificar que, en esta muestra original de 92 ancianos, el MMSE se adapta muy fuertemente, una vez que los datos se superponen a los objetos indicando una distribución normalizada del desempeño de estos sujetos cuando evaluados por el MMSE.

Al mismo tiempo, el análisis particularizado ítem a ítem, bien como sujeto a sujeto permite acceder a un nivel de análisis esencial a la hora de desarrollar testes para la supuesta medición de constructos, bien como testar el ajuste de un teste ya desarrollado a una dada población. Por ejemplo, en la última parte de este trabajo, procedemos a la reestimación de

parámetros con base en la exclusión de sujetos particulares que solo este modelo permite identificar de forma tan sencilla.

Esta metodología permite aún analizar los resultados particulares de un determinado sujeto, verificando y comprendiendo el desempeño específico en cada ítem. Este es un nivel de análisis diagnóstico solo posible en este tipo de metodología. La representación gráfica de los resultados individuales de los sujetos bien como los valores de *Infit* y *Outfit* en un gráfico permite también una rápida identificación de los objetos (ítems o sujetos) que se quieren evaluar de forma particular.

Como conclusión final, se puede plantear la idea que, con estos modelos (de Rasch), la psicometría puede resolver algunas de las limitaciones en lo que se refiere a la individualidad de los objetos. Aquí, los objetos no son solo considerados en un *pool* generalista de *medias* y *desviación típica*, pero si, se considera su especificidad. Así se puede prever la probabilidad de un ítem ser adecuadamente resuelto (correctamente) o no. Se pueden clasificar los ítems como de *baja*, *media* o de *alta dificultad*, con base en criterios objetivos, basados en su normalización en la población de sujetos evaluados, más do que quedarse por un criterio (por veces arbitrario) de aquello que el “*creador*” del teste piensa ser un teste de *baja*, *media* o *alta dificultad* (un psicometrista, un profesor, *etc.*).

Finalmente nos gustaría referir que cuando utilizamos estos modelos tenemos en consideración que la asunción de uní dimensionalidad, según Embretson (2006) y Verhelst (2001), el corazón del modelo de Rasch y de muchos otros modelos TRI, se encuentra asegurado. No obstante, como nos recuerda Verhelst (2001) en un artículo de revisión, relativamente poca atención se ha prestado a la comprobación estadística de esta asunción, aspecto este que es especialmente sensible a la violación del axioma de la uní dimensionalidad. Autores como Glas (1989), Molenaar (1983), Van den Wollenberg (1979, 1982), Martin-Löf (1973) bien como el propio Verhelst (2001) son ejemplos de investigadores que buscan contribuir para el fortalecimiento de este modelo.

Por lo expuesto, pensamos que los resultados presentados en este artículo contribuyen para la consubstanciación de su utilización, particularmente en la utilización de datos clínicos que puedan ser tratados de forma dicotómica<sup>4</sup>.

Reflejo de lo expuesto anteriormente es el hecho de se estar verificando un incremento en la utilización del Modelo de Rasch, o de las TRI en estudios variados como sean, por

---

<sup>4</sup> Con el término “*dicotómico*”, no queremos decir que una cuestión, ítem, indicador... lo que sea, solo pueda presentar dos formas, pero si que, aunque puedan ser presentadas varias posibilidades de respuesta, apenas una sea, indudablemente la correcta).

ejemplo *The rasch model to measure service quality* (Battisti, Nicolini & Salini, 2003); *“Quantum Measurement Technique” a Powerful Tool for Measuring Quality of Life* (Pimentel, Álvarez, Román, & Melon, 2003); *Uso del Modelo de Rasch para poner en la misma escala las puntuaciones de distintos testes* (Prieto & Velasco, en sumisión); *Análisis de un teste mediante el modelo de Rasch* (Prieto & Delgado, 2003); *Efecto de la reducción de alternativas en las escalas tipo Likert desde la perspectiva de la TRI* (García-Cueto et al., 2003); *Survey development and validation with the Rasch Model* (Green & Frantom, 2002); *The Combined Use of Classical Test Theory and Item Response Theory* (Verstralen et al., 2001); *Application of the Rasch Model for Testing Piaget's Theory of Cognitive Development* (Spiel, Gittler, Surch & Glück, 2000); *Applying the Mixed Rasch Model to Personality Questionnaires* (Rost, Carstensen & Davier, 2000); *Survey development and validation with the Rasch Model* (Green & Frantom, 2002); *The Application of the Rasch Model to a Marketing-Scale* (Salzbeger, 1999), entre otros.

El campo de la psicometría, y en nuestra opinión, el campo de la psicometría relacionada con la práctica clínica, permite que estos modelos de medición objetiva figuren como de gran utilidad y se espera una gran evolución en los próximos años en lo que concierne a su difusión junto a los psicólogos y psicometristas en general (Feddag, 2008; Chi-Wen, 2007).

## Referencias

- Adams, R.J. & Khoo, S. (1996). Quest. The interactive test analysis system. Victoria: ACER.
- Almeida, L.S. & Freire, T. (1997). *Metodologia da Investigação em Psicologia e Educação* (1 ed.). Coimbra: APPORT.
- American Psychiatric Association (1994). *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*. Washington, D.C.: American Psychiatric Association.
- Anthony, J. C., Leresche, L., & Unaiza, N. (1982). Limits of the "Mini Mental State" as a screening test for dementia and delirium among hospital patients. *Psychological Medicine*, 12, 397-408.
- Battisti, F. D., Nicolini, G., & Salini, S. (2003). *The Rasch model to measure service quality - Working Paper n. 27*. Milano: Dipartimento di Economia Politica e Aziendale - Università degli Studi di Milano.
- Bertolucci, M.L., Brucki, S.M., & Campacci, S.R. (1994). The Mini Mental State Examination in a general population: Impact of educational status. *Arquivos de Neuropsiquiatria*, 52(1), 1-7.
- Bond, T. G., & Fox, C. M. (2001). *Appling the Rasch Model*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Cavanaugh, S.A., & Wettstein, R.M. (1983). The relationship between severity of depression, cognitive dysfunction, and age in medical patients. *American Journal of Psychiatry*, 140, 495-496.
- Chandler, J.D. & Gerndt, J. (1988). Cognitive screening tests for organic mental disorders in psychiatric patients. *Journal of Nervous and Mental Disorders*, 176, 675-681.
- Chi-Wen (2007). *Using the Rasch model to validate the Peabody Developmental Motor Scales—second edition in infants and pre-school children*. Masters (Research) thesis, James Cook University.
- Cossa, F.M., Sala, S.D., & Musicco, M. (1997). Comparison of two scoring systems of the Mini-Mental State Examination as a screening test for dementia. *Journal of Clinical Epidemiology*, 50(8), 961-965.
- Cronbach, L.J. (1951). Coefficiente alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Dussailant, F. (2003). *Técnicas de medición en pruebas de admisión a las universidades*. Santiago: Centro de Estudios Públicos.

- Embretson, S.E. (2006). The continued search for nonarbitrary metrics in psychology. *American Psychologist*, 61(1), 50-55.
- Feddag, M.L. (2008). Statistical Inference for the Multidimensional Mixed Rasch Model. *Communications in Statistics – Simulation and Computation*. 37 (9 November), 1732 – 1749.
- Folstein, M. F., Folstein, S. E., & Mchugh, P. R. (1975). Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry Research*, 12, 189-198.
- Glas, C.A.W. (1989). Contributions to estimating and testing Rasch models. (Doctoral thesis). Enschede: University of Twente.
- Green, K. E., & Frantom, C. G. (2002). *Survey development and validation with the Rasch Model*. Paper presented at the International Conference on Questionnaire Development, Evaluation, and Testing, November 2002, University of Missouri – Columbia, Charleston, SC.
- Guerreiro, M., Silva, A.P., & Botelho, M.A. (1994). Adaptação à população portuguesa na tradução do "Mini Mental State Examination" (MMSE). *Revista Portuguesa de Neurologia*, 1, 9.
- Guerreiro, M. (1993). *Contributo da Neuropsicologia para o estudo das demências*. Dissertação de Doutoramento, não publicada. Lisboa: Universidade de Lisboa.
- Hornke, L. F. (2000). Item Response Times in Computerized Adaptive Testing. *Psicológica*, 21, 175-189.
- Juhel, J. (1999). Etude de la dimensionnalité d'un test de raisonnement à l'aide des Modèles de Réponse à l'Item. *Psychologie et Psychométrie*, 20 (2/3), 85-111.
- Maia, L. A. C. R. (2001). As doenças demenciais e o consumo de café. *Revista da Faculdade de Medicina de Lisboa*, 6(2), 93-103.
- Maia, L., & de Mendonça, A. (2002a). Does caffeine intake protect from Alzheimer Disease? *European Journal of Neurology*, 9.
- Maia, L., & de Mendonça, A. (2002b). Does caffeine intake protect from Alzheimer's Disease? *Journal D'Actualités D'Alzheimer (reprinted form European Journal of Neurology)*, 9, 1377-1382.
- Maia, L. (2006). *Avaliação Neuropsicológica na Esclerose Múltipla*. Viseu: Editora Psico & Soma.



- Malloy, P.F., Cummings, J.L., & Coffey, C.E. (1997). Cognitive screening instruments in neuroPsych – A report of the committee on research of the Am Neuropsychiatric Association. *Journal of Neuropsychiatry and Clinical Neuroscience*, 9, 189-197.
- Martin-Löf, P. (1973). Statistiska modeller [Statistical models.] Anteckningar från seminarier lasåret 1969-1970, utarbetade av Rolf Sundberg. Obetydligt ändrat nytryck, October 1973. Stockholm: Institutet för Försäkringsmatematik och Matematisk Statistisk vid Stockholms Universitet.
- Matteo, C., Cabaco, A., Capataz, I., & Maia, L. (2003). Memory Distortion & False memory creation - a review on recent research. *Revista Psicologia e Educação*, 2(1), 75-97.
- Molenaar, I.W. (1983). Some improved diagnostics for failure of the Rasch model. *Psychometrika*, 48, 49-72.
- O'Connor, D.W., Pollit, P.A., & Hyde, J.B. (1989). The reliability and validity of the Mini Mental State Examination in a Br community survey. *Journal of Psychiatry and Research*, 23, 87-96.
- Pimentel, Álvarez, Román, & Melon. (2003). “Quantum Measurement Technique” a Powerful Tool for Measuring Quality of Life. *Quality of Life News Letter*, 5-6.
- Pinto, A. C. (1990). *Metodologia da Investigação Psicológica*. Edições Jornal de Psicologia: Porto.
- Prieto, G. & Delgado, A.R. (2003). Análisis de un teste mediant el modelo de Rasch. *Psicothema*, 15(1), 94-100.
- Prieto, G. & Delgado, A.R. (2000). Utilidad y representación en la psicometría actual. *Metodologia de las Ciencias del Comportamiento*, 2(2), 111-127.
- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
- Rost, J. (2000). The Growing Family of Rasch Models - Chapter 2. In A. Boomsma, M. A. J. v. Duijn & T. A. B. Snijders (Eds.), *Lecture Notes in Statistics 157 – Essays on Item Response Theory* (pp. 25-42). Groningen: Springer-Verlag.
- Rost, J., Carstensen, C., & Davier, M. v. (2000). Applying the Mixed Rasch Model to Personality Questionnaires. In A. Boomsma, M. A. J. v. Duijn & T. A. B. Snijders (Eds.), *Lecture Notes in Statistics 157 – Essays on Item Response Theory* (pp. 324-332). Groningen: Springer-Verlag.
- Salzberger, T. (1999). *The Application of the Rasch Model to a Marketing-Scale*. Paper presented at the 11th European Meeting of the Psychometric Society, July 19-22, Lüneburg.

- Spiel, C., Gittler, G., Sirsch, U., & Glück, J. (2000). Application of the Rasch Model for Testing Piaget's Theory of Cognitive Development. In A. Boomsma, M. A. J. v. Duijn & T. A. B. Snijders (Eds.), *Lecture Notes in Statistics 157 – Essays on Item Response Theory* (pp. 111-117). Groningen: Springer-Verlag.
- SPSS (2002). SPSS Statistical Package for Social Science (Version 11.5).
- Stenner, A. J., Burdick, D. S., & Stone, M. H. (2008). Formative and reflective models: Can a Rasch analysis tell the difference? *Rasch Measurement Transactions*, 22(1), 1152-1153, [www.rasch.org/rmt/rmt221.pdf](http://www.rasch.org/rmt/rmt221.pdf)
- Van den Wollenberg, A.L. (1979). The Rasch model and time limit tests. Doctoral thesis. Nijmegen: University of Nijmegen.
- Van den Wollenberg, A.L. (1982). Two new test statistics for the Rasch model. *Psychometrika*, 47, 123-139.
- Verhelst, N. (2001). Testing the unidimensionality assumption of the Rasch model. *Methods of Psychological Research Online*, 6(3), 231-271.
- Verstralen, H., Bechger, T., & Maris, G. (2001). *The Combined Use of Classical Test Theory and Item Response Theory*. Retrieved September 14, 2001, [http://www.cito.nl/pok/poc/eind\\_fr.htm](http://www.cito.nl/pok/poc/eind_fr.htm)
- Wright, B. D., & Stone, M. H. (1979). *Best test design. Rasch measurement*. Chicago: Mesa Press.

**Como citar este artigo:**

Maia, L. (2012). El Modelo de Rasch Aplicado a las Ciencias Psicológicas. *Revista E-Psi*, 2 (1), 1-34.