

AVALIAÇÃO DA ESCALA DE AUTO-ESTIMA DE ROSENBERG MEDIANTE O MODELO DE RASCH

*Sónia Quintão*¹

*Ana R.. Delgado*²

*Gerardo Prieto*²

Resumo: O objectivo deste estudo consistiu em avaliar as características psicométricas da versão portuguesa da escala de Rosenberg mediante o modelo de Escalas de Classificação, uma extensão para itens politómicos do modelo de Rasch. Foi recolhida uma amostra de 510 estudantes universitários, 223 do sexo masculino e 287 do sexo feminino, com uma média etária de 21.7 anos ($DP = 3.80$). As análises indicam que o ajuste dos dados ao modelo é aceitável. Conclui-se que tanto o formato de resposta da versão portuguesa (tipo *Likert* com quatro categorias) como a fiabilidade dos dados resultam adequados. No entanto, de um ponto de vista substantivo e, em concordância com estudos prévios, constata-se que a escala não permite distinguir as pessoas com níveis médio e alto de auto-estima.

Palavras-chave: Auto-estima; Modelo de Rasch; Rosenberg; Portugal.

Evaluation of Rosenberg self-esteem scale using rasch model (Abstract): The objective of this study consisted on evaluating the psychometrical characteristics of the Portuguese version of the Rosenberg Scale by means of the Rating Scale model, an extension for the polytomous items of the Rasch Model. A sample of 510 college students was collected, 223 males and 287 females, with an average of 21.7 years ($SD = 3.80$). The analysis indicates that the adjustment of the data to the model is acceptable. We conclude that the response format from the Portuguese version (*Likert* type with four categories) as well as the reliability of the data, results adequate. However, from a substantive point of view, and in agreement with previous studies, it appears that the scale does not distinguish people with medium and high levels of self-esteem.

Key-words: Self-esteem; Rasch model; Rosenberg; Portugal

¹ Faculdade de Ciências Médicas da Universidade Nova de Lisboa (Portugal); CEDOC.

² Universidad de Salamanca (Espanha)

Avaliação da Escala de Auto-Estima de Rosenberg mediante o Modelo de Rasch

Rosenberg (1965) define auto-estima como uma avaliação que o indivíduo faz sobre si mesmo, que se expressa numa atitude positiva ou negativa em função de si, em que este aprova ou não o resultado. Uma auto-estima elevada implica que o indivíduo sinta que tem valor e inclui a dimensão dos sentimentos positivos que tem sobre si. Neste contexto, uma baixa auto-estima é originada por uma diminuição do indivíduo perante si próprio, através de uma auto-avaliação negativa. De acordo com este autor, a auto-estima global é baseada não só na avaliação da qualidade dos seus constituintes, como o auto-conceito, mas concomitantemente na avaliação das qualidades que são importantes para o indivíduo.

A escala de auto-estima de Rosenberg (1965) é a medida de auto-estima mais comumente utilizada em investigação (Ghaderi, 2005; Kwan, John, Kenny, Bond, & Robins, 2004) e está traduzida e validada para Portugal (Santos & Maia, 2003). A brevidade de aplicação, a linguagem acessível dos itens e os resultados positivos evidenciados no estudo original de validação, permitiram que esta escala se transformasse no padrão de referência na avaliação da auto-estima (Santos & Maia, 2003), embora exista um estudo (Gray-Little, Williams & Hancock, 1997) em que se sugeria que a escala de Rosenberg só é adequada para medir com precisão pessoas situadas nos níveis baixos do constructo de auto-estima.

A Teoria Clássica dos Testes (TCT) foi o principal modelo psicométrico empregue na construção e análise de testes. No entanto, foram assinadas várias limitações à TCT, como utilizar distintas escalas para medir os sujeitos e os itens (o que impede a análise das suas interações para construir interpretações sobre a variável medida), pressupor falsamente que o erro de medida é invariante ao longo da variável, etc. (Embretson, 1996). As suas limitações levaram a uma proposta de modelos alternativos, dos quais o mais harmonioso é o modelo de Rasch, que permite a medição conjunta de pessoas e itens numa mesma dimensão ou constructo (Prieto & Delgado, 2003; Stone, 2003). Assim, é um dos principais modelos que proporciona a objectividade específica para a construção de escalas, permitindo a independência das estimações com respeito à distribuição do atributo pelas pessoas em que é medido (Bond & Fox, 2001). Este modelo tem recebido muita atenção como revelam as publicações em distintas línguas, especificamente na língua portuguesa (Ziviani & Primi, 2002).

O modelo de Rasch, proposto por Georg Rasch (1960), fundamenta-se em dois pressupostos principais: o atributo que se pretende medir pode representar-se numa dimensão onde só se situariam conjuntamente as pes-

soas e os itens; o nível da auto-estima do inquirido e a dificuldade do item são os únicos determinantes da probabilidade de uma determinada resposta. Se o controlo da situação é adequado, esta expectativa é razoável e pode ser representada matematicamente através do modelo.

Rasch apresenta uma fórmula para modelar esta relação em itens dicotómicos: $\ln(P_{is} / 1 - P_{is}) = (B_s - D_i)$. Esta equação indica que o quociente entre a probabilidade de uma determinada resposta (por exemplo, concordo) e a probabilidade da outra resposta possível (discordo) num item ($P_{is} / 1 - P_{is}$), é função da diferença no atributo entre o nível da pessoa (B_s) e o nível do item (D_i). Assim, quando uma pessoa responde a um item equivalente ao seu nível no atributo, terá a mesma probabilidade de uma resposta de concordar ou discordar ($P_{is} / 1 - P_{is} = 0.50/0.50$). Neste caso, o logaritmo natural de ($P_{is} / 1 - P_{is}$), reflecte que a dificuldade do item é equivalente ao nível do sujeito no atributo medido ($B_s - D_i=0$) (Bond & Fox, 2001; Prieto & Delgado, 2003; Stone, 2003).

Existem extensões do modelo de Rasch para itens politómicos. Uma das formas de apresentar este tipo de dados, mais utilizadas na avaliação em saúde mental, são as escalas tipo *Likert*. O modelo estabelece a localização de cada ponto do item desde os itens mais baixos até os mais altos. Assim, cada item tem uma localização estimada. O modelo também estima os valores das categorias da escala para dar uma estrutura de qualificação à escala, que é partilhada por todos os itens que a compõem. Com dados dicotómicos, cada item tem uma estimação de localização e um erro de estimação associado. Com dados politómicos, não só cada item tem uma localização estimada como adicionalmente a escala tem uma série de pontos de transição (incremento na probabilidade de escolha de uma categoria de resposta em função do nível da auto-estima do inquirido) entre as categorias sucessivas. Podemos apresentar uma extensão do modelo básico de Rasch para escalas com categorias ordenadas segundo o Modelo de Escalas de Classificação (Rating Scale Model, Andrich, 1978): $\ln(P_{nik} / P_{ni(k-1)}) = B_n - D_i - F_k$. Onde: P_{nik} é a probabilidade de que a pessoa n , perante o item i seja observada, ou responda, na categoria k ; $P_{ni(k-1)}$ é a probabilidade de que a observação ou a resposta esteja na categoria $k-1$; B_n é a habilidade, atitude, etc., da pessoa n ; D_i é a dificuldade do item i ; F_k é o ponto de transição entre a categoria k e a categoria $k-1$, i.e. o k ésimo passo de calibração, de onde as categorias se numeram 0, m .

Com base no exposto, foi objectivo do presente estudo a análise das características psicométricas da tradução portuguesa da escala de Auto-Estima de Rosenberg mediante o Modelo de Escalas de Classificação (MEC, Wright & Masters, 1982). Era esperado que a escala apresentasse características razoáveis ao nível da consistência interna, mas que não permitisse distinguir níveis médios e altos de auto-estima, tal como sugerido por

Gray-Little, Williams e Hancock (1997) que referiram que a escala de Rosenberg só é adequada para medir com precisão pessoas situadas nos níveis baixos do constructo de auto-estima, devido aos itens serem muito orientados para os extremos do constructo.

Método

Participantes

Participaram no estudo 510 estudantes universitários, 223 do sexo masculino e 287 do sexo feminino, com uma média etária de 21.7 anos ($DP = 3.80$). A maior parte dos participantes eram solteiros (95.24%), estudantes do primeiro e segundo ano da universidade (52.39%), futuros licenciados em Letras, Direito e História (35.97%) e em Economia, Gestão, Contabilidade e Marketing (29.25%), sem profissão (90.34%).

Medidas

Rosenberg Self-esteem Scale (RSES; Rosenberg, 1965) é uma medida de auto-avaliação da auto-estima global. A RSES foi originalmente construída como uma escala de tipo *Guttman* embora, na maioria dos casos, os investigadores optem por um formato tipo *Likert* (Santos & Maia, 2003). A RSES original apresentava boas propriedades psicométricas: a sua consistência interna era elevada, com um valor de α de *Cronbach* igual a .92 (Rosenberg, 1965).

Um estudo internacional sobre as características psicométricas da escala, em 53 países, mostrou que a escala apresenta uma consistência interna pelo menos adequada em 50 desses países (α de *Cronbach* superior a .70), sendo o valor mais elevado de .90, com excepção da República Democrática do Congo, Etiópia e Tanzânia (Schmitt & Allik, 2005).

Este questionário, na versão portuguesa, é constituído por 10 itens, com um formato de resposta tipo *Likert* de quatro pontos (1. Concordo fortemente; 4. Discordo fortemente), 5 de orientação positiva e 5 de orientação negativa. A medida permite um resultado total, que varia de 10 a 40, crescendo em função do nível de auto-estima.

A escala traduzida apresenta, de uma forma genérica, qualidades psicométricas comparáveis às relatadas originalmente por Rosenberg (1965) e a outras versões de outras línguas (Schmitt & Allik, 2005; Shapurian, Hojat, & Nayerahmadi, 1987). Apresenta uma consistência interna adequada, com um valor de α de *Cronbach* igual a .86, e boa estabilidade temporal, dado que com um intervalo de duas semanas entre avaliações, o coeficiente de correlação de *Pearson* era igual a .90 (Santos & Maia, 2003).

Procedimento

A aplicação do questionário foi realizada em universidades (bibliotecas e salas de estudo). Pediu-se a colaboração desinteressada dos participantes de forma oral e por escrito, com informação prévia do objectivo do estudo e do tipo de participação pretendida, da confidencialidade e anonimato, da possibilidade de desistência durante a elaboração da avaliação, assim como da inexistência de respostas correctas ou incorrectas.

A recolha da amostra decorreu em diversas universidades da cidade de Lisboa, tendo a aplicação da prova demorado em média cinco minutos.

Codificação e análise dos dados

Uma vez recolhidos os dados, estes foram recodificados de maneira a que as categorias correspondentes aos itens que mediam em sentido inverso passarem a estar no sentido da variável medida, isto é, a maior valor corresponder maior auto-estima.

Os dados foram analisados com o programa *Winsteps* (Linacre & Wright, 2000).

Em primeiro lugar, realizou-se uma análise da funcionalidade das categorias de resposta, de acordo com os seguintes critérios estatísticos convencionais propostos por Linacre (2002, 2010):

1. Frequência suficiente e distribuição regular das categorias: as frequências baixas não são úteis para estimar as calibrações dos passos (F_k). O primeiro critério exige um mínimo de 10 observações dentro de cada categoria da escala. Baixas contagens dentro de uma categoria pode levar a estimativas imprecisas ou instabilidade nas calibrações dos passos. A situação ideal é uma distribuição uniforme, mas também é adequada uma distribuição unimodal ou bimodal com as frequências maiores nas categorias dos extremos.

2. As categorias não devem apresentar um desajuste elevado com o modelo.

Para pôr à prova o ajuste das pessoas, itens e categorias, usaram-se as médias dos resíduos (diferenças entre as respostas dadas e as esperadas) *Outfit* (média dos resíduos estandardizados elevados ao quadrado) e *Infit* (média dos resíduos estandardizados elevados ao quadrado, ponderados com a função de informação) que são as estatísticas que indicam o grau de ajuste. *Outfit* é muito sensível aos padrões de resposta inesperados (*outliers*), que têm menos influência no *Infit* (Wright & Mok, 2004). O valor esperado para as duas estatísticas é a unidade (Wright, 1996). Valores altos indicam padrões anormais com respeito ao modelo, significando que há maior probabilidade de erro do que a probabilidade sistemática no item e valores baixos indicam possibilidade de existirem padrões aproximados aos modelos deterministas de resposta como o de *Guttman* (significando que pode existir

probabilidade sistemática nas respostas ao item que não é explicada pela variável latente de interesse, que no presente estudo é a auto-estima). Os valores de *Infit* ou *Outfit*, superiores a 1.5 e inferiores a 2, indicam que o desajuste dos dados empíricos é moderadamente alto, mas não enfraquecem gravemente as medidas. No entanto, os valores superiores a 2 revelam um alto desajuste, assim como valores inferiores a .50 indicam possibilidade de existirem padrões aproximados aos modelos deterministas de resposta (Linacre, 2010).

3. Os pontos de transição (passos) entre as categorias sucessivas (Fk) devem aumentar monotonicamente (com ordenação crescente) o que significa que todas as categorias de resposta são funcionais, uma vez que cada categoria é a mais provável em algum ponto da variável.

No MEC, a probabilidade de seleccionar uma categoria depende do nível de auto-estima da pessoa. É desejável que exista algum intervalo no contínuo no qual cada categoria seja a mais provável de ser seleccionada (Linacre, 2002). Se isto suceder, os pontos de transição (Fk) estão ordenados monotonicamente. Quer dizer, $F1 < F2 < F3$, etc. Isto significa que no caso de quatro categorias tipo Likert, que as pessoas com um valor superior a $F3$ terão uma probabilidade superior de responder à categoria 4.

O funcionamento diferencial dos itens (FDI) é uma evidência da falta de validade das medidas. Produz-se quando existem factores alheios ao constructo, que se pretende medir, que afectam ilegitimamente as respostas. Em presença de FDI pode suceder que sujeitos com o mesmo nível no constructo, mas pertencentes a diferentes grupos sociodemográficos (por exemplo, homens e mulheres), tenham uma probabilidade distinta de uma resposta (Draba, 1977). De acordo com os pressupostos dos modelos tipo Rasch, a probabilidade da resposta a um item só dependerá do parâmetro da pessoa no atributo latente. Isto implica, por exemplo, que a probabilidade de 0.50 de estar de acordo com um item corresponde a um valor na variável que há-de ser a mesma para as pessoas de distintos grupos sociais. Ao contrário, a probabilidade da resposta está determinada por factores alheios ao constructo medido. Em consequência, as medidas obtidas com os itens careceriam da validade requerida.

Actualmente a análise de FDI é um passo obrigatório nos programas de validação de um teste (AERA, APA, NCME, 1999). Em consequência, levou-se a cabo uma análise do funcionamento diferencial dos itens da RSES, relacionada com o sexo, uma vez que a literatura aponta para a existência de diferenças entre sexos, apresentando os homens maior auto-estima (Hendricks, et al., 2001; Saigal, Lambert, Russ, & Hault, 2002; Santos & Maia, 2003).

O procedimento implementado em Winsteps consiste em estimar em cada item a diferença entre o parâmetro de dificuldade em cada grupo e na

amostra total. O contraste leva-se a cabo mediante a fórmula proposta por Wrigt e Panchapakesan (1969): $t = B_f - B_r / (SE_f^2 + SE_r^2)^{1/2}$. Onde $B_f - B_r$ são os parâmetros de localização ajustados por sexo, e SE_f^2 e SE_r^2 são os seus SE (erros padrão) ao quadrado.

Um valor absoluto superior a 2 da estatística de contraste pode-se considerar uma evidência de FDI estatisticamente significativa. No entanto, segundo Wright e Douglas (1975), os valores de FDI que enfraquecem as medidas correspondem a diferenças ($B_f - B_r$) superiores a 0.5 *logit*.

Resultados

Em primeiro lugar, foi analisada a qualidade psicométrica das categorias de resposta, de acordo com os critérios de Linacre (2002). Tal como se pode observar na Tabela 1, o sistema de quatro categorias, analisado com o Rating Scale Model, é adequado.

Pode observar-se que as categorias cumprem os critérios propostos por Linacre (2010): as frequências e a sua distribuição são adequadas, não há um desajuste elevado em nenhuma categoria (*Outfit* < 2) e os passos (*step*) entre as categorias sucessivas estão ordenados de forma crescente, ver tabela 1.

Tabela 1. Estatísticas correspondentes às categorias

Categoria	Frequências	B	<i>Infit</i>	<i>Outfit</i>	Passo
1 = DF	109 (2%)	-0,81	1,42	1,69	-
2 = D	582 (12%)	-0,06	0,88	0,90	-2,29
3 = C	2591 (52%)	1,64	0,91	0,92	-,63
4 = CF	1648 (33%)	3,35	1,00	0,99	2,92

Uma vez comprovada a adequação das categorias, foi analisado o ajuste dos itens e das pessoas, foram estimados os seus parâmetros e avaliada a sua fiabilidade. Na tabela 2 são apresentadas as estatísticas de ajuste (*Infit* e *Outfit*), a localização (*Di*) e o erro padrão de medida dos itens (*SE*), ver tabela 2.

Pode-se observar que não existem itens com valores de *infit* e/ou *outfit* superiores a 1.5 o que revela que os itens não se desajustam de forma severa (Linacre & Wright, 2000).

Tabela 2. Características psicométricas dos itens

Item (1)	<i>Infit</i>	<i>Outfit</i>	Di	SE
1 (Globalmente, estou satisfeito(a) comigo próprio(a))	0,82	0,83	0,28	0,08
2 (Por vezes penso que não sou bom/boa em nada)	1,06	1,13	1,06	0,08
3 (Sinto que tenho algumas qualidades)	1,00	1,11	-0,74	0,09
4 (Sou capaz de fazer as coisas tão bem como a maioria das pessoas)	0,94	1,00	-0,43	0,09
5 (Sinto que não tenho muito de me orgulhar)	0,92	0,93	-0,14	0,90
6 (Por vezes sinto-me, de facto, um(a) inútil)	1,21	1,20	0,07	0,90
7 (Sinto-me uma pessoa de valor, pelo menos quanto a generalidade das pessoas)	0,74	0,76	-0,16	0,90
8 (Gostaria de ter mais respeito por mim próprio(a))	1,33	1,41	1,32	0,08
9 (Bem vistas as coisas, inclino-me a sentir que sou um(a) falhado(a))	0,92	0,78	-1,25	0,10
10 (Adopto uma atitude positiva para comigo)	0,88	0,87	0,00	0,09
Media	0,98	1,00	0,00	0,09
DP	0,17	0,20	0,73	0,00

(1) As categorias foram adaptadas de modo a que todos os itens vão em sentido positivo.

Di – Localização; SE – Erro padrão.

Na coluna Di, referente à dificuldade ou localização dos itens, podemos ver que o item 8 é o indicador de maior nível de auto-estima e o item 9 é o de menor nível.

Os erros padrão da medida dos itens oscilam entre .08 a .10, o que indica que a fiabilidade dos itens é elevada. A fiabilidade global das estimações dos itens pode ser avaliada mediante a estatística denominada Item Separation Reliability que indica a proporção da variância observada dos itens não explicada pelo erro de medida (Smith & Smith, 2004). Como se pode observar na tabela 3 a fiabilidade dos itens é muito alta (.99), ver tabela 3.

O ajuste das pessoas ao modelo não é razoável, pois apesar de o valor da média e do desvio-padrão serem 1.02 e .71 (*infit*), e 1.00 e .74 (*outfit*), o número de pessoas com *infit* e/ou *outfit* superior a 1.5 é elevado: 87 (18%). Inferior a .5 estão 107 sujeitos (22%), o que pode dever-se à existência de

vários itens similares ou muito correlacionados que inflacionam as fiabilidades.

Tabela 3. Resumo dos resultados da RSES.

Estatísticas	Max.	Min.	Média	DP	Valor
<i>Infit</i> dos itens	1,33	0,74	0,98	0,17	—
<i>Outfit</i> dos itens	1,41	0,76	1,00	0,20	—
% itens desajustados (1)					0
<i>Infit</i> das pessoas	6,27	0,15	1,02	0,71	—
<i>Outfit</i> das pessoas	6,34	0,14	1,00	0,74	—
% pessoas desajustadas					18
Parâmetro itens	1,32	-1,25	0,00	0,73	—
Fiabilidade Itens	—	—	—	—	0,99
Parâmetro pessoas	5,34	-2,74	1,95	1,56	—
Fiabilidade das pessoas	—	—	—	—	0,79
Medida dos homens	-2,74	5,34	2,29	1,82	—
Medida das mulheres	-2,20	5,34	1,95	1,67	—
r_{bp} com género(2)	—	—	—	—	-0,10

(1) *Infit* e/ou *outfit* > 1.5 (2) $p < .05$

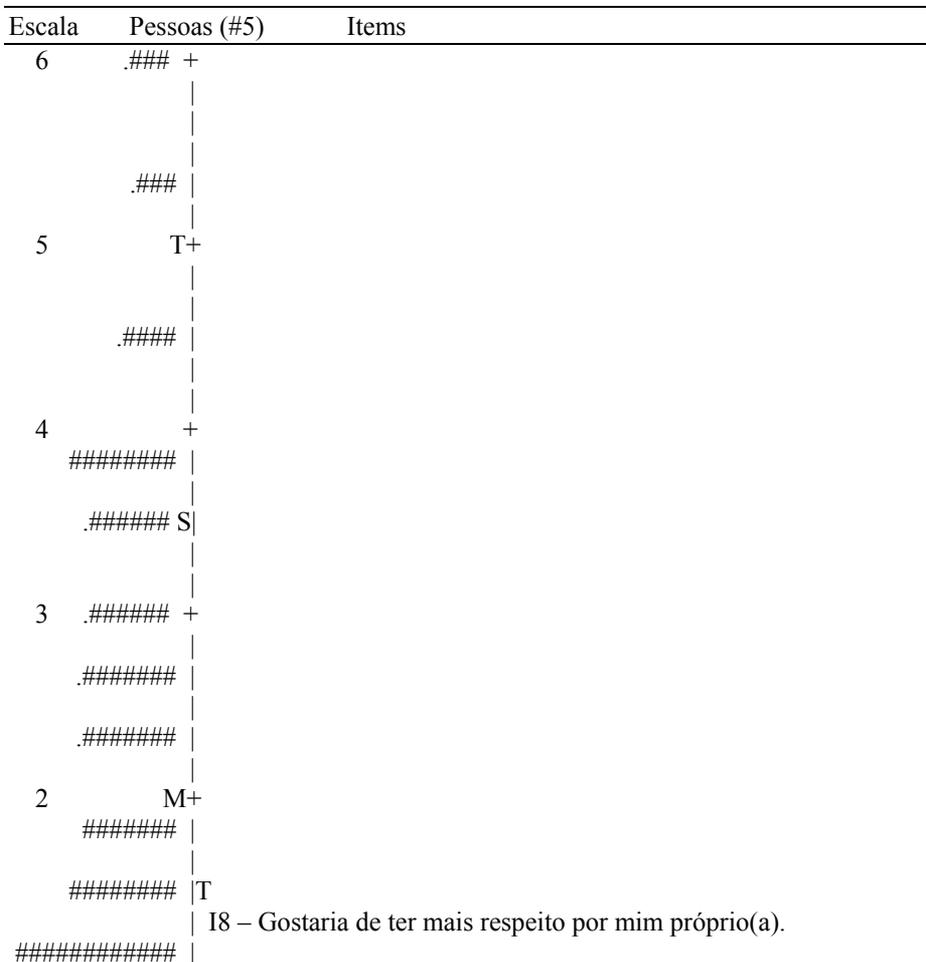
Os parâmetros das pessoas vão de 5.34 a -2.74 com uma média de 1.95 e desvio-padrão de 1.56. A fiabilidade global das pessoas (.79) não é excelente, embora razoável (percentagem da variância das pessoas não explicada pelo erro). Este valor é similar ao α de Cronbach da teoria clássica.

A correlação entre sexo e a medida dos itens, embora seja significativa ($p = .03 < .05$) é muito reduzida ($r = -.10$). Não aparece nenhuma outra correlação significativa com outras variáveis demográficas.

A Tabela 4 mostra uma representação conjunta pessoa-item, onde se pode observar a posição dos itens e das pessoas na variável. Este tipo de representação facilita a comunicação dos resultados e das suas implicações. A coluna da “Escala” representa os valores estimados da variável latente

auto-estima, sendo que quanto maior forem estes valores, maior é o nível de auto-estima. Dado que a média das pessoas (1.95) é muito superior à dos itens, pode-se considerar que o nível de auto-estima dos sujeitos analisados é muito alto. Além do mais, a maior parte das pessoas encontram-se acima do ponto da variável em que os itens se situam. Em consequência, porque a média dos itens é alta (i.e., dificuldade baixa) o que corresponde a indivíduos com valores de logit baixos (i.e., baixos valores na variável latente) tenderem a concordar fortemente com os itens da escala), observa-se que os itens da escala não permitem medir com elevada precisão as pessoas com níveis médio e alto de auto-estima, ver tabela 4.

Tabela 4. Representação Conjunta de Pessoas e Itens



1	+	I2 – Por vezes penso que não sou bom/boa em nada.
	#####	
		S
	#####	
		S I1 – Globalmente, estou satisfeito(a) comigo próprio(a).(1)
	####	
0	##	+M I10 – Adopto uma atitude positiva para comigo(1) I6 – Por vezes
		sinto-me, de facto, um(a) inútil.
		I5 – Sinto que não tenho muito de me orgulhar. I7 – Sinto-me uma
		pessoa de valor, pelo menos quanto a generalidade das pessoas.(1)
	#	
	#	I4 – Sou capaz de fazer as coisas tão bem como a maioria das
		pessoas.(1)
		S I3 – Sinto que tenho algumas qualidades.(1)
-1	#	+ T
	#	I9 – Bem vistas as coisas, inclino-me a sentir que sou um(a) falhado(a).

(1) *Itens invertidos de maneira a que as maiores pontuações indiquem maior auto-estima.*

Na tabela 5 apresentam-se os resultados da análise de Funcionamento Diferencial dos Itens associado ao sexo, ver tabela 5.

Tabela 5. Funcionamento Diferencial dos Itens associados ao sexo.

Item	DM –DH	SE	t
1	-0,08	0,17	-0,44
2	0,44	0,16	2,71
3	0,02	0,18	0,13
4	0,34	0,18	1,89
5	-0,42	0,18	-2,39
6	0,33	0,17	1,92
7	-0,34	0,18	-1,97
8	-0,28	0,16	-1,75
9	-0,06	0,19	-0,29
10	0,02	0,17	0,14

De acordo com a distribuição normal, $t_{\geq|2|}$ indica DIF. Neste caso, só os itens 2 e 5 mostram DIF, mas os valores não são muito altos, não enfraquecem a medida, pois não têm diferenças superiores a 0.5 *logit*. O item 2 indica um maior nível de auto-estima nas mulheres ($d_{i2}=.44$) e o item 5 nos homens ($d_{i5}=-.42$).

Discussão

O objectivo deste estudo foi realizar uma análise das características psicométricas da escala de auto-estima de Rosenberg (RSES) numa amostra de alunos universitários portugueses. Os dados foram analisados mediante o modelo de Escalas de Classificação (Wright & Masters, 1982), uma extensão para itens politómicos do modelo de Rasch.

Neste estudo pretendia-se determinar o melhor sistema de categorias de resposta. Os resultados mostraram que o sistema original de quatro categorias tem boa qualidade psicométrica.

Com relação aos itens da RSES, a fiabilidade é muito alta. O ajuste das pessoas ao modelo está no limite do razoável. A fiabilidade global das pessoas não é excelente, embora resulte adequada. Sendo este valor similar ao α de *Cronbach* da teoria clássica, pode-se assim afirmar que está de acordo com os resultados da fiabilidade clássica da RSES apresentados por outros autores (Schmitt & Allik, 2005), ainda que mais reduzido do que aparece em outros estudos (Rosenberg, 1965; Santos & Maia, 2003; Shapurian, Hojat, & Nayerahmadi, 1987; Schmitt & Allik, 2005).

Alguns estudos prévios com a escala original colocaram a possibilidade de que a escala não permitisse distinguir as pessoas com nível médio das pessoas com nível elevado de auto-estima (Gray-Little, Williams & Hancock, 1997). Os resultados indicam que a média das pessoas é superior à dos itens, o que revela que a maior parte das pessoas se encontra num nível superior ao do nível da variável em que os itens se situam, ou seja com mais auto-estima. Em consequência, observa-se que os itens da escala não permitem medir com alta precisão as pessoas com níveis médio e alto de auto-estima. Isto sucede porque a escala só tem itens adequados para o leque inferior do constructo. Este resultado está de acordo com o afirmado no estudo de Gray-Little et al. (1997) que sugeriam que a escala de Rosenberg só é adequada para medir com precisão pessoas situadas nos níveis baixos do constructo de auto-estima.

Só os itens 2 e 5 mostram FDI, mas os valores não são muito altos, não enfraquecem a medida, logo a comparação entre homens e mulheres resulta adequada. O item 2 indica um maior nível de auto-estima nas mulheres e o item 5 nos homens. Tanto no item 2 (Por vezes penso que não sou

bom em nada) como no item 5 (Sinto que não tenho muito de que me orgulhar), não é fácil encontrar uma explicação clara para as diferenças de resposta entre sexos. Podemos assim concluir que a medida pode ser utilizada igualmente nos dois géneros.

Tal como se esperava, os homens apresentaram um valor médio de auto-estima mais elevado que as mulheres, embora o tamanho do efeito seja baixo, seguindo critérios convencionais (Cohen, 1998). Este dado está de acordo com o publicado por outros autores (Hendricks, et al., 2001; Saigal, Lambert, Russ, & Houlst, 2002; Santos & Maia, 2003; Turner, Pickering, & Johnson, 1998). Não foram encontradas correlações significativas com outras variáveis demográficas, o que pode dever-se ao facto de a população universitária ser uma população com características muito mais homogéneas que as da população geral.

Finalmente, tal como se havia discutido em estudos anteriores, os dados recolhidos com a escala de auto-estima de Rosenberg mostraram uma boa fiabilidade.

Em conclusão, os aspetos mais positivos da escala são que as categorias funcionam adequadamente, a fiabilidade média dos sujeitos é aceitável e a fiabilidade dos itens é elevada.

Por outro lado, confirma-se que os itens da escala não permitem medir com elevada precisão as pessoas com níveis médio e elevado de auto-estima. Isto acontece porque o nível das pessoas é elevado e a escala só tem itens para auto-estima reduzida. Este é o aspeto mais criticável da escala RSES. Estes resultados podem estar de qualquer forma algo mascarados pelo efeito da desejabilidade social, embora se tenha passado o questionário de forma anónima e pelo facto de que os participantes serem estudantes universitários, não sendo portanto representativos da população geral. Esta poderia ser também a justificação para que as diferenças sexuais em auto-estima tenham resultado menores de que o defendido em estudos prévios.

Uma limitação deste estudo prende-se por a amostra só ter sido recolhida em Lisboa, não sendo representativa da população universitária portuguesa.

Em qualquer caso, parece necessário construir uma nova escala de auto-estima que permita avaliar adequadamente um leque mais amplo do constructo.

Referências

- AERA, APA, NCME (1999). *The standards for educational and psychological testing*. Washington, DC: AERA.
- Andrich, D. A. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 561-573.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model*. Mahwah, NJ: LEA.
- Cohen, J. (1998). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: LEA.
- Draba, R. E. (1977). The identification and interpretation of item bias.
- Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, *4*, 341-349.
- Ghaderi, A. (2005). Psychometric properties of the self-concept questionnaire. *European Journal of Psychological Assessment*, *21*, 139-146.
- Gray-Little, B., Williams, V.S.L., & Hancock, T.D. (1997). An item response theory analysis of the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *23*, 443-451.
- Hendricks, C. S., Tavakoli, A., Hendricks, D. L., Harter, N. R., Campbell, K. P., L'Ecuver, R. I., Geddings, A. A., Hackett, D., & Mathis, D. (2001). Self-esteem matters: Racial & gender differences among rural southern adolescents. *Journal of Nacional Black Nurses' Association*, *12*, 15-22.
- Kwan, V. S. Y., John, O. P., Kenny, D., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: An interpersonal approach. *Psychological Review*, *111*, 94-110.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, *3*, 85-106.
- Linacre, J. M. (2010). *A user's guide to winsteps ministep. Rasch-model computer programs*. Retirado em 5, Abril, 2010, de www.winsteps.com.
- Linacre, J. M., & Wright, B. D. (2000). *WINSTEPS: Multiple-choice, rating scale, and partial credit Rasch analysis (Computer software)*. Chicago: MESA Press.
- Prieto, G., & Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, *15*, 94-100.
- Rasch, G. (1960). *Probabilistic Models for some intelligence and attainment tests*. Copenhagen: Institute for Educational Reseca.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Saigal, S., Lambert, M., Russ, C., & Hoult, L. (2002). Self-esteem of adolescents who were born prematurely. *Pediatrics*, *109*, 429-433.
- Santos, P. J., & Maia, J. (2003). Análise factorial confirmatória e validação preliminar de uma versão portuguesa da escala de autoestima de Rosenberg. *Psicologia: Teoria, Investigação e Prática*, *8*, 253-268.
- Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the rosenberg self-esteem scale in 53 nations: exploring the universal and culture-specific

- features of global self-esteem. *Journal of Personality & Social Psychology*, 89, 623-642.
- Shapurian, R., Hojat, M., & Nayerahmadi, H. (1987). Psychometric characteristics and dimensionality of Pearsian version of Rosenberg Self-Esteem Scale. *Perceptual and Motor Skills*, 65, 27-34.
- Smith, E. V., & Smith, R. M. (Eds) (2004). *Introduction to Rasch measurement*. Maple Grove, Mn: JAM Press.
- Stone, M. H. (2003). Substantive scale construction. *Journal of Applied Measurement*, 4, 282-297.
- Turner, L. A., Pickering, S., & Johnson, R. B. (1998). The relationship of attributional beliefs to self-esteem. *Adolescente*, 33, 477-484.
- Wright, B. D. (1996) Reasonable mean-square fit values. In Wright, B.D., & Linacre JM, (Eds.) *Rasch measurement transactions. Part 2*. Chicago: MESA Press.
- Wright, B. D., & Douglas, G. A. (1975). A better procedures for sample-free item analysis. *Research Memorandum*. Statistical Laboratory. Department of Education. University of Chicago.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.
- Wright, B. D., & Mok, M. M. C. (2004). An overview of the family of Rasch measurement models. In Smith, E. V., & Smith, R. M. (Eds.) *Introduction to Rasch Measurement*. Maple Grove: JAM Press.
- Wright, B. D., & Panchapakesan, N. (1969). A procedure for simple-free item analysis. *Educational and Psychological Measurement*, 29, 23-48.
- Ziviani, C., & Primi, R. (2002). Teoria da resposta ao item o modelo Rasch de mensuração: Uma análise de provas de psicologia. In R. Primi (Org.). *Temas em Avaliação Psicológica* (pp. 131-151). Campinas: IBAP.