

Comparação de dois procedimentos de mensuração de respostas avaliativas: “*rating*” e “*ranking*”*

Maria Salomé Pinho**

Resumo: Na maioria dos estudos realizados em psicologia cognitiva, em que se pretende mensurar as respostas avaliativas dos participantes, o uso de escalas de pontos parece ser irresistível. No presente trabalho são analisados os resultados de um estudo preliminar, no qual os participantes classificaram numa escala de fama com cinco pontos (*rating*), completamente etiquetada e ordenaram (*ranking*), em termos de fama relativa, o mesmo conjunto de itens. Na tarefa de *rating* foram ainda consideradas duas condições: emissão de juízos absolutos e de juízos relativos. As vantagens e os inconvenientes de cada um destes procedimentos são também ponderados, bem como a relevância da técnica de *rating* “o mais – o menos”.

Na discussão dos resultados, uma possível explicação para as diferenças relativamente aos resultados alcançados por outros investigadores, a respeito do problema da diferenciação colocado pelo procedimento de *rating*, é apresentada.

Palavras-chave: *rating*, *ranking*, juízos absolutos e juízos relativos.

Abstract: In the majority of studies accomplished in cognitive psychology in which we intend measuring the participants evaluative responses the application of rating scales seems to be irresistible. In this paper we analyse the results of a preliminary study in which the participants rated in a five-point fame scale fully labeled and ranked the same set of items. In the rating procedure we included two conditions: making absolute judgements and relative judgements. The benefits and the costs of each of the procedures are weighted as well as the importance of the most – least method. In the discussion of the results, a possible explanation for the differences found when we compare the results obtained by other researchers concerning the differentiation problem in rating is proposed.

Key words: rating, ranking, absolute judgement and relative judgement.

Na maioria dos estudos em que se pretende mensurar respostas avaliativas dos sujeitos recorre-se ao uso de escalas de pontos (*rating scales*) (cf., por ex., Fife-Schaw, 2000; McCarty & Shrum, 2000; Preston & Colman, 2000).

Será que os procedimentos de *rating* e de *ranking* conduzem a resultados equivalentes?

Subjacente à adopção de um ou outro procedimento encontram-se, por vezes em exclusivo, considerações de natureza prática, como a possibilidade de contemplar um menor número de itens e a sua apresentação simultânea, exigidas pelo procedimento de *ranking*. Quando são exequíveis ambos os procedimentos, o

* Este texto retoma os resultados apresentados em formato *poster* no V Simpósio Nacional de Investigação em Psicologia, Lisboa (16 a 18 de Setembro) 2003, inserindo-os num contexto mais extenso. Trabalho realizado no âmbito do projecto “Estudo das relações entre memória, metamemória e desempenho académico em estudantes do ensino superior” financiado pela Fundação Calouste Gulbenkian.

** Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra [salome@fpce.uc.pt]

privilégio concedido às escalas de pontos baseia-se em argumentos relacionados com o tipo de informação obtido e as técnicas estatísticas permitidas. Assim, o procedimento de *rating* fornece informação sobre a escala de avaliação do respondente, i. e., a informação apurada não está circunscrita ao contexto criado com os itens apresentados e as propriedades estatísticas da pontuação de *rating* são menos restritivas, podendo usar-se a estatística paramétrica (McCarty & Shrum, 2000; Russell & Gray, 1994). Os problemas relacionados com a variação restrita da variável/atributo, não diferenciação dos itens e o evitamento de um dos extremos da escala, ou mesmo os dois, são aqui considerados pouco relevantes.

Por outro lado, a ordenação dos itens (*ranking*) obriga a uma discriminação mais exigente (não se pode atribuir, a mais do que um item, a mesma posição), contudo esta pode nem sempre corresponder a diferenças reais de valoração (cf. McCarty & Shrum, 2000; Sayadi *et al.*, 2002). O facto da pontuação total resultante do *ranking* ser ipsativa impossibilita análises comparativas interindividuais (McCarty & Shrum, 2000). As escalas com sete pontos, apresentadas visualmente, são consideradas, por diversos investigadores, globalmente melhores no que diz respeito à fiabilidade, validade (de critério e convergente) e poder discriminativo (Preston & Colman, 2000) ou à percentagem de respondentes indecisos e à sua capacidade de diferenciação entre os valores da escala (Cox cit. por Schwarz *et al.*, 1991). Em termos das preferências dos respondentes, Preston e Colman (2000) observaram, num dos seus estudos, que estas recaíam, em primeiro lugar, nas escalas com dez pontos e, logo a seguir, nas escalas com sete e nove pontos. Estes mesmos investigadores sugerem que a popularidade das escalas com cinco categorias de resposta, embora

abrangidas pelo chamado “número mágico sete mais ou menos dois”, como foi baptizado o resultado da análise de G. Miller sobre os limites da capacidade de processamento da informação, não decorre das suas propriedades psicométricas. Mas, nas escalas com um número ímpar de pontos, à categoria intermédia poderá estar associada uma propensão de resposta (*response set*) e a dificuldade em nomear com exactidão este ponto (i. e., torna-se aqui complicado dispor de uma descrição verbal, ou etiqueta, que preserve a distância numérica) podendo ambos os aspectos contribuir para a obtenção de médias diferentes quando se utilizam escalas com ou sem etiquetas e ainda escalas com outras etiquetas (Chang, 1997).

Num estudo com escalas de pontos em número par, tanto com etiquetas em todos eles como apenas nos extremos, Chang (1997) observou, através de análises baseadas na teoria da generalizabilidade, que os respondentes se baseiam maioritariamente na informação numérica para a atribuição de pontuações de *rating*. Assim, a existência eventual de uma discrepância entre distâncias iguais dos algarismos/números e etiquetas que não preservem essa igualdade não teria consequências negativas, pelo menos nas escalas de pontos em número par. Aqui, a preocupação em dispor de etiquetas verbais cuja intensidade conotativa mantenha distâncias iguais deixaria de ser relevante.

Para promover a diferenciação entre os itens submetidos a apreciação e evitar a acumulação das respostas em torno de um ponto privilegiado da escala, McCarty e Shrum (2000) propuseram uma modificação no procedimento usual de *rating*. De acordo com esta alteração, os respondentes deveriam começar por seleccionar os itens que consideram ser “o mais” e “o menos”

(âncoras) no atributo a ser julgado e proceder à sua classificação na escala de pontos fornecida. Mais precisamente, o primeiro item a ser considerado é “o mais” ao qual se segue a escolha de um valor na escala, depois o método é o mesmo para o item designado “o menos” e seguidamente adopta-se o modo habitual de classificação na escala de pontos para os restantes itens. O procedimento de *rating* assim concebido foi designado método “o mais – o menos”.

Vários investigadores como, por exemplo, Russell e Gray (1994) referem que os procedimentos de *ranking* e de *rating* estão associados a tarefas distintas. No primeiro procedimento estariam implicados juízos de natureza relativa, i.e., tratar-se-ia de atribuir posições relativas aos itens apresentados a julgamento considerando, portanto, comparações que se restringem ao conjunto de itens actualmente presente. Já o procedimento de *rating* daria ensejo ao julgamento de tipo absoluto, pois esta situação seria mais favorável a uma avaliação, de certo modo, independente dos itens apresentados, podendo os itens considerados âncoras não pertencer ao conjunto proposto para ser julgado (comparação com itens similares acessíveis na memória).

Contudo, esta distinção cujo critério assenta no seguimento preferencial de uma orientação absoluta (*absolute set*), no caso do procedimento de *rating*, não tem sido testada experimentalmente. Note-se, a este propósito, que embora as exigências de ambos os procedimentos possam ser diferentes é, no entanto, conhecido que se forem considerados os valores médios de juízos de *rating* e de *ranking*, numa amostra suficientemente grande de indivíduos, haverá, por certo, elevada concordância, desde que as pontuações de *rating* e de *ranking* apresentem uma tendência

moderada para variarem monotonamente com o atributo a ser julgado. Compatível com uma correlação alta entre médias de pontuações de *rating* e de *ranking*, poder-se-á encontrar, nas condições acima mencionadas, um número apreciável de respondentes relativamente ao qual as referidas correlações sejam baixas ou não positivas (Russell & Gray, 1994).

A comparação em termos de convergência de juízos obtidos por *rating* e por *ranking* tem sido, sobretudo, estudada com planos experimentais intersujeitos (*ibid*). Mas, para que esta análise comparativa contemple igualmente diferenças individuais, e não apenas a média do grupo, importa alargar o planeamento a um estudo intra-sujeitos. É ainda controversa a resposta à questão sobre qual dos dois procedimentos de mensuração de respostas avaliativas apresenta melhores propriedades psicométricas. Autores como, por exemplo, Alwin, Krosnick, Miethe e Reynolds (cit. por McCarty & Shrum, 2000) referem a superioridade do procedimento de *ranking* no que diz respeito à validade e fiabilidade enquanto Munson, McIntyre, Rankin e Grube (*ibid*) afirmam que não existem diferenças relevantes entre ambos. O argumento principal que sustenta a primeira posição remete para o nível de diferenciação dos itens avaliados que seria, frequentemente, inferior quando se opta pelo procedimento de *rating*. Se se excluíssem da amostra estudada os casos de não diferenciação, então os resultados obtidos com ambos os procedimentos seriam equivalentes. Neste sentido, importa, então, fomentar a diferenciação na pontuação de *rating*. Uma das formas seria o método “o mais – o menos” proposto por McCarty e Shrum (2000), referido anteriormente e uma outra possibilidade consistiria num misto dos dois procedimentos, no qual os participantes começam por

ordenar os itens e depois classificam os mesmos numa escala de pontos. Este procedimento provocaria um “efeito de ancoragem” na pontuação de *rating*, uma vez que no *ranking* inicial há exposição simultânea a todos os itens que posteriormente serão alvo de classificação na escala de pontos (os itens extremos poderão servir de âncoras na escala de *rating*). O presente trabalho secunda o de Russell e Gray (1994) respeitante ao estudo da concordância entre pontuações de *rating* e de *ranking*, como medida de respostas avaliativas, tanto ao nível do indivíduo como do grupo. Manipula, ainda, as instruções referentes ao procedimento de *rating* no sentido de tornar directamente comparáveis os resultados de *rating* e de *ranking*, na suposição de que estes estariam relacionados com tarefas distintas (juízos absolutos e juízos relativos, respectivamente). Contudo, forçando o procedimento de *rating* ao uso de juízos relativos, mediante instruções adequadas, seriam as pontuações atribuídas, também ao nível individual, elevadamente concordantes com o *ranking* dos mesmos itens? Poderia obter tal *ranking* uma correspondência menor, embora elevada, quando as instruções de *rating* dispusessem os sujeitos para realizarem somente juízos absolutos?

Neste contexto, a hipótese colocada seria a seguinte:

- a pontuação de *rating*, no caso de serem dadas instruções conducentes à emissão de juízos relativos, apresentaria, em resultado de uma melhor discriminação entre os itens, uma extensão mais ampla (um desvio padrão mais elevado) e como corolário ter-se-ia:
- o grau de concordância entre pontuações de *rating* e *ranking* seria mais elevado para esse procedimento específico (*rating* com indução de juízos relativos).

Método

Variáveis independentes: instruções de *rating* e ordem de *rating*

Variável dependente: avaliação da fama (pontuação de *rating*)

Amostra: n = 188 estudantes dos primeiros e segundos anos das licenciaturas em Psicologia e em Ciências da Educação da FPCE-UC¹, sendo 174 do género feminino e 14 do género masculino, com idades compreendidas entre os 18 e os 38 anos (média = 19.77 e d.p. = 2.00).

Material: Lista com 15 nomes de cantores famosos, ordenados alfabeticamente; folhas com uma escala de cinco pontos completamente etiquetada (adjacente a cada um dos nomes) para registo das pontuações de *rating* e folhas com quinze linhas, com indicação das posições desde a 1^a até à 15^a, para inscrição da ordenação decrescente realizada pelos respondentes.

Procedimento: Cada um dos 188 sujeitos participou individualmente na classificação dos 15 nomes de cantores famosos na escala de pontos apresentada para esse efeito e procedeu também à sua ordenação. Estes 188 sujeitos foram distribuídos por quatro grupos com as seguintes tarefas cuja ordem foi contrabalanceada: 47 sujeitos pertenceram à condição de passar primeiro pelo procedimento de *ranking* e em seguida pelo procedimento de *rating* com instruções para efectuarem juízos absolutos (G1), para outros 47 o procedimento geral foi o mesmo, embora submetidos a instruções de julgamento relativo (G2), os 94 restantes realizaram em segundo lugar a tarefa de ordenação tendo metade previamente procedido à classificação dos nomes de cantores famosos na escala de pontos de acordo com instruções

¹ Agradece-se a todos os estudantes a sua participação neste estudo.

de juízos absolutos (G3) enquanto a outra seguiu instruções de julgamento relativo (G4). Em todos os casos, antes de se passar à aplicação do segundo procedimento, a folha de registo das respostas já preenchida foi recolhida. As instruções foram apresentadas por escrito, nas próprias folhas de resposta, tendo sido também explicadas oralmente antes da realização de cada uma das tarefas.

Resultados

Uma vez que foi estabelecido um plano factorial intersujeitos 2 (tipo de instruções de *rating*) x 2 (ordem de *rating*), os resultados médios de cada sujeito nas pontuações de *rating* para os quinze nomes famosos e as médias dos respectivos desvios padrões foram estudadas através da ANOVA univariada com dois factores fixos.

No Quadro 1 são apresentadas as médias e os desvios padrão das médias das pon-

no que diz respeito ao factor instruções de *rating*, é também significativa: $F(1,184) = 4.65$, $p = 0.03$; o mesmo não acontecendo com a interacção entre os factores ordem das instruções e instruções de *rating*: $F(1,184) = 0.68$, $p = 0.08$. Sublinhe-se que sob as instruções de indução de juízos relativos, os respondentes tendem a apresentar médias de *rating* mais elevadas (3.37) do que quando as instruções se referem à indução de juízos absolutos (3.23). Contudo, as médias de *rating* são mais elevadas quando os respondentes realizaram em primeiro lugar o procedimento de *rating* (3.42) comparativamente com a situação em que tal procedimento é antecedido pelo *ranking* (3.19). As médias dos desvios padrão das pontuações de *rating* para os quinze nomes famosos são indicadas no Quadro 2 da página seguinte.

Tanto o factor instruções (considere-se a diferença entre os desvios padrões de *rating* nos grupos juízos absolutos (1.24) e juízos

Quadro 1 - Efeitos dos factores instruções de *rating* (indução de juízos absolutos e de juízos relativos) e ordem (primeiro *ranking*, primeiro *rating*) nas médias das pontuações de *rating* dos sujeitos para os quinze nomes de cantores famosos (entre parênteses indicam-se os d.p.s)

Instruções de <i>rating</i>	Ordem das instruções		
	1º <i>ranking</i>	1º <i>rating</i>	
indução de juízos absolutos	3.18 (0.32)	3.28 (0.60)	3.23
indução de juízos relativos	3.20 (0.45)	3.55 (0.47)	3.37
	3.19	3.42	3.30

tuações de *rating* para os quinze nomes famosos.

Como foi referido houve contrabalançamento da ordem das instruções. É significativa a diferença entre as médias de *rating* nos grupos 1º *ranking* (3.19) e 1º *rating* (3.42): $F(1,184) = 10.46$, $p = 0.001$. A diferença entre as médias de *rating* nos grupos indução de juízos absolutos (3.23) e de juízos relativos (3.37),

relativos (1.27)), como o factor ordem (veja-se a diferença entre as médias de dispersão de *rating* nos grupos 1º *ranking* (1.27) e 1º *rating* (1.24)), como ainda a interacção entre ambos não estão associados a efeitos estatisticamente significativos quando se consideram as médias dos desvios padrões das pontuações de *rating*: $F(1,184) = 1.19$, $p = 0.28$; $F(1,184) = 0.60$, $p = 0.44$ e $F(1,184) = 1.43$, $p = 0.23$, respectivamente.

Quadro 2 – Efeitos dos factores instruções de *rating* (indução de juízos absolutos e de juízos relativos) e ordem (primeiro *ranking*, primeiro *rating*) nos desvios padrões das pontuações de *rating* dos sujeitos para os quinze nomes de cantores famosos (indicam-se entre parênteses os d.p.s)

Instruções de <i>rating</i>	Ordem das instruções		
	1º <i>ranking</i>	1º <i>rating</i>	
indução de juízos absolutos	1.27 (0.10)	1.20 (0.27)	1.24
indução de juízos relativos	1.26 (0.19)	1.28 (0.22)	1.27
	1.27	1.24	1.26

Procedeu-se também à análise das correlações intra-sujeitos entre as pontuações de *rating* e de *ranking* para os quinze nomes de cantores famosos calculando-se, portanto, 188 coeficientes de correlação de Spearman. Note-se que a escala para *rating* apresenta menos pontos do que o número de itens para classificar o que determina, por um lado, que o valor máximo do desvio padrão é mais pequeno quando comparado com o desvio padrão do procedimento de *ranking* e, por outro, o valor máximo da correlação de Spearman entre pontuações de *rating* e *ranking* é, neste caso, de 0.98 (cf. Russell & Gray, 1994).

No Quadro 3.1 são referidos alguns indicadores das distribuições das correlações

intra-sujeitos para os quatro grupos considerados; os respectivos diagramas de caule e folhas são apresentados a seguir em 3.2. Os valores das correlações de Spearman variam entre 0.57 e 0.98 para G1, entre 0.41 (valor estatisticamente não significativo) e 0.98 em G2, 0.60 e 0.98 para G3 e 0.68 e 0.98 em G4. À excepção do valor de correlação já assinalado, todos os restantes são estatisticamente significativos (teste bilateral, $p = 0.01$). Para os grupos G2 e G3, 75% dos respondentes, ou seja, cerca de 35 em 47, apresentam correlações de valor igual a 0.84 e superiores; para os grupos G1 e G4 cerca de 35 em 47 respondentes alcançam valores de correlação igual a 0.83 e 0.86,

Quadro 3.1 - Estatística das distribuições das correlações intra-sujeitos entre os valores resultantes dos procedimentos de *ranking* e de *rating* para os sujeitos pertencentes aos quatro grupos considerados (G1, G2, G3 e G4) e respectivos diagramas de caule e folhas (3.2)

[G1 – instruções de indução de juízos absolutos e procedimento de *ranking* em 1º lugar
G2 – instruções de indução de juízos relativos e procedimento de *ranking* em 1º lugar
G3 – instruções de indução de juízos absolutos e procedimento de *rating* em 1º lugar
G4 – instruções de indução de juízos relativos e procedimento de *rating* em 1º lugar]

Estadística	G1	G2	G3	G4
N	47	47	47	47
Média	0.881	0.874	0.883	0.887
Erro padrão da média	0.013	0.015	0.012	0.009
Mediana	0.913	0.905	0.914	0.904
Desvio padrão	0.091	0.100	0.080	0.064
Assimetria	-1.520	-2.495	-1.599	-1.159
Mínimo	0.571	0.413	0.598	0.681
Máximo	0.982	0.977	0.984	0.979
1º Quartil	0.834	0.842	0.840	0.856
3º Quartil	0.943	0.937	0.938	0.928

respectivamente e valores superiores. No que diz respeito aos valores mais baixos de correlação de Spearman, apenas um caso pertencente ao grupo G2, como se referiu acima, apresenta um valor baixo não significativo (0.41); este grupo inclui dois casos com valores de correlação

inferiores ou iguais a 0.66; existem três casos cujos valores de correlação são inferiores a 0.69 no grupo G3, dois respondentes com valores iguais ou abaixo de 0.67 no grupo G1 e dois sujeitos com valores inferiores ou iguais a 0.76 no grupo G4.

3.2 - Diagramas de caule e folhas:

G1			
Frequência	caule	e	folhas
2.00	Extremos	(= \leq 0.67)	Largura do caule: .100
1.00	6	.	8
3.00	7	.	244
1.00	7	.	6
7.00	8	.	1223333
6.00	8	.	578999
18.00	9	.	000112222223334444
9.00	9	.	555566778
G2			
Frequência	caule	e	folhas
2.00	Extremos	(= \leq 0.66)	Largura do caule: .100
2.00	7	.	44
4.00	7	.	7789
5.00	8	.	00444
10.00	8	.	5667788999
15.00	9	.	000122222233344
9.00	9	.	566666677
G3			
Frequência	caule	e	folhas
3.00	Extremos	(= \leq 0.69)	Largura do caule: .100
3.00	7	.	678
7.00	8	.	1223344
9.00	8	.	567788889
19.00	9	.	0112222222333344444
6.00	9	.	556668
G4			
Frequência	caule	e	folhas
2.00	Extremos	(= \leq 0.76)	Largura do caule: .100
4.00	7	.	6789
4.00	8	.	0334
11.00	8	.	55777778899
20.00	9	.	00000011112222234444
6.00	9	.	556667

Os valores da média das correlações situam-se entre 0.88 (grupos G1, G2 e G3) e 0.89 (grupo G4) e os das medianas entre 0.90 (grupo G4) e 0.91 (grupos G1, G2 e G3) apresentando as distribuições dos quatro grupos uma assimetria negativa elevada, verificando-se a maior concentração dos valores em torno de 0.92.

Relativamente à hipótese formulada, de que a pontuação de *rating* resultante da condição “instruções indutoras de juízos relativos” (grupos G2 e G4) teria uma extensão mais ampla (desvio padrão superior), procedeu-se à comparação das médias dos desvios padrão das pontuações de *rating* nos 3º e 1º quartis das distribuições das correlações de Spearman (entre os procedimentos de *rating* e de *ranking*), para os quatro grupos considerados (ver Quadro 4), através de testes *t* simultâneos.

4 (variâncias não homogêneas) para o grupo G1; $t(20) = 1.76$ e $p = 0.09 \times 4$ para o grupo G2; $t(20) = 2.13$ e $p = 0.05 \times 4$ para o grupo G3; finalmente, $t(20) = 0.49$ e $p = 0.63 \times 4$ para o grupo G4. O enunciado do corolário (ver anteriormente), segundo o qual o grau de concordância entre pontuações de *rating* e *ranking* seria mais elevado na situação em que as instruções do procedimento de *rating* induzissem juízos relativos, foi testado através de uma ANOVA univariada com dois factores fixos (tipo de instruções e ordem de *rating*). A escolha da ANOVA apoia-se nos seguintes argumentos: as distribuições embora não sejam simétricas apresentam uma forma similar (todas são negativamente assimétricas; ver Quadro 3), a maior variância (0.010) não é mais do que

Quadro 4 - Médias dos desvios padrão das pontuações de *rating* respeitantes aos 3º e 1º quartis das distribuições das correlações de Spearman (entre parênteses indicam-se os d.p.s)

	G1	G2	G3	G4
3º Quartil	1.27 (0.14)	1.31 (0.20)	1.35 (0.22)	1.30 (0.17)
1º Quartil	1.16 (0.29)	1.17 (0.18)	1.09 (0.349)	1.25 (0.27)

Quadro 5 - Efeitos dos factores instruções de *rating* (indução de juízos absolutos e de juízos relativos) e ordem (primeiro *ranking*, primeiro *rating*) nos valores das correlações de Spearman entre os procedimentos de *ranking* e de *rating* para os quatro grupos de sujeitos (entre parênteses indicam-se os d.p.s)

Instruções de <i>rating</i>	Ordem das instruções		
	1º <i>ranking</i>	1º <i>rating</i>	
indução de juízos absolutos	0.88 (0.09)	0.88 (0.08)	0.88
indução de juízos relativos	0.87 (0.10)	0.89 (0.06)	0.88
	0.88	0.89	0.88

Em todos os grupos as diferenças entre as médias dos desvios padrão (0.12; 0.14; 0.26 e 0.05 respectivamente nos grupos G1, G2, G3 e G4) não são estatisticamente significativas (atendeu-se à correcção de Bonferroni): $t(14.5) = 1.19$ e $p = 0.25 \times$

quatro vezes superior à menor variância (0.008) e o tamanho das amostras é igual ($n = 47$) (Howell, 1992). As médias dos valores das correlações de Spearman, para os quatro grupos, estão indicados no Quadro 5.

Nenhuma das diferenças entre as correlações de Spearman é estatisticamente significativa: para o factor instruções (não existe diferença entre as médias das correlações de Spearman nos grupos com indução de juízos absolutos (0.88) e de juízos relativos (0.88)) $F(1,184)$ é, obviamente, igual a zero; para o factor ordem (diferença entre as correlações de Spearman nos grupos 1º *ranking* (0.88) e 1º *rating* (0.89)) $F(1,184) = 0.46$ e $p = 0.5$. Refira-se que o valor de potência observado para $\alpha = 0.05$ é aqui sempre baixo (muito próximo de zero). Por simples inspeção visual do Quadro 5 (construído com base no *output* que precede, no SPSS, o resultado final da ANOVA) chegar-se-ia a estes resultados.

Se se observar o Quadro 3 verifica-se também que as medianas dos valores da correlação de Spearman são quase exactamente iguais (0.905 para G2 e 0.904 para G4; em ambos os grupos as instruções seriam indutoras de juízos relativos) e muito próximos dos restantes dois grupos (0.91 para G1 e também 0.91 para G3). Foram ainda calculadas as médias das pontuações de *rating* e *ranking* atribuídas pelos respondentes a cada um dos quinze nomes de cantores famosos (ver Quadro 6). Em seguida, o grau de concordância entre as médias das pontuações de *rating* e *ranking*, para cada um dos quatro grupos considerados, foi avaliado com base no valor do coeficiente de correlação de Pearson (ver ainda Quadro 6).

Os valores das correlações de Pearson para os quatro grupos variam entre os extremos 0.87 ($r^2 = 76\%$) e 0.99 ($r^2 = 98\%$). Observa-se, portanto, um nível elevado de concordância (associação positiva forte) entre as pontuações de *ranking* e de *rating* (indução de juízos absolutos e de juízos relativos), qualquer que seja a ordem dos procedimentos, para os quinze nomes de cantores famosos.

Discussão dos resultados

Perante os resultados obtidos tentar-se-á perceber, em primeiro lugar, porque é que a previsão contida na hipótese formulada e no seu corolário não se verificou neste estudo, diferentemente do sucedido na investigação realizada por Russell e Gray (1994). Se se examinar mais em profundidade a primeira ANOVA efectuada (relativa às médias das pontuações de *rating* considerando os factores instruções de *rating* e ordem), cujos resultados são estatisticamente significativos, há a assinalar que o valor para os resíduos é muito elevado, sendo r^2 ajustado apenas igual a 0.075. Nesta situação, o modelo não explica uma grande proporção da variação da variável dependente (médias das pontuações de *rating*) devendo, então, procurar-se outros factores mais eficientes. A este propósito, Howell (1992) refere que o investigador deverá interrogar-se se a diferença encontrada possui ou não algum

Quadro 6 - Médias das pontuações de *rating* e *ranking* respectivas correlações de Pearson para os quinze nomes de cantores famosos considerando os quatro grupos de respondentes

	G1	G2	G3	G4
média <i>ranking</i> de nomes famosos	8.00 (3.59)	7.54 (4.43)	8.00 (3.54)	8.04 (3.70)
média <i>rating</i> de nomes famosos	3.18 (1.06)	3.24 (1.04)	3.28 (0.97)	3.39 (1.19)
cor. de Pearson	0.99**	0.91**	0.98**	0.87**

**p = 0.01 (teste bilateral)

significado prático, pois valores consideravelmente baixos de r^2 podem indicar que se está perante um efeito, apesar de estatisticamente significativo, trivial ou, pelo contrário, que na situação experimental existem outros factores que inflacionam a variabilidade global.

Da revisão da literatura levada a cabo, destaca-se que a elucidação do problema da diferenciação colocado pelo procedimento de *rating* passa também por factores relacionados com as diferenças individuais. Entre as variáveis estudadas encontra-se a motivação do respondente e a sua “sofisticação cognitiva” (Krosnick & Alwin, 1988). Neste contexto, um menor envolvimento na tarefa significaria menos esforço na distinção dos itens apresentados para avaliação e, conseqüentemente, pouca diferenciação nas pontuações de *rating*. A variável “sofisticação cognitiva”, operacionalizada em termos do grau de educação formal/escolar obtido, estaria associado à anterior, uma vez que a reduzida motivação para avaliar os itens propostos na escala de pontos surge frequentemente nos respondentes com menor instrução académica (o encadeamento seria o seguinte: menos motivação, menos investimento cognitivo, menor diferenciação). A maior dispersão na pontuação de *rating* encontrar-se-ia nos respondentes com um nível mais elevado de formação académica (*ibid*). Os resultados de uma investigação realizada por McCarty e Shrum (2000), na qual confrontaram dois procedimentos de *rating* (o usual e o método “o mais – o menos”), são consistentes com a ideia de que os respondentes com mais formação académica apresentam uma maior diferenciação das pontuações quando comparados com outros cujo grau de escolaridade é inferior (no que diz respeito ao método “o mais – o menos”

tal variável, compreensivelmente, não exerce influência na diferenciação).

A amostra no presente trabalho é constituída unicamente por estudantes universitários o que, à luz da explicação sugerida acima, poderá levar a supor que se trata de respondentes cuja “sofisticação cognitiva” é elevada e, nessa medida, efectuem uma melhor diferenciação na pontuação de *rating*. Deste modo, o facto da classificação dos itens na escala de pontos ser antecedida pelo procedimento de *ranking* desses mesmos itens ou de se procurar induzir a formação de juízos relativos no procedimento de *rating* não teria impacto no desempenho discriminativo destes respondentes. A validade das pontuações de *rating* não seria fomentada por nenhuma destas orientações. O grau elevado de concordância entre os resultados de ambos os procedimentos (*rating* e *ranking*) nas condições ou grupos analisados seria praticamente seguro.

Neste enquadramento ficaria por esclarecer como é que os sujeitos com instrução académica mais elevada negociam um eventual conflito quando as instruções de *rating* solicitam a emissão de juízos absolutos: ignoram pura e simplesmente as instruções dadas ou procuram segui-las, mas a medida utilizada pelo investigador, na avaliação dos efeitos daí decorrentes, não possui sensibilidade suficiente para revelar diferenças?

Referências bibliográficas

- Chang, L. (1997). Dependability of anchoring labels of Likert-type scales. *Educational and Psychological Measurement*, 57 (5), 800-807.
- Fife-Schaw, C. (2000). Questionnaire design. In G. M. Breakwell, S.

- Hammond & Fife-Schaw (Eds.), *Research methods in psychology* (2nd edition, pp. 158-174). London: Sage.
- Howell, D. C. (1992). *Statistical methods for psychology* (3rd edition). Belmont, CA: Duxbury.
- Krosnick, J. A. & Alwin, D. F. (1988). A test of the form-resistant correlation hypothesis: Ratings, rankings, and measurement of values. *Public Opinion Quarterly*, 52, 526-538.
- McCarty, J. A. & Shrum, L. J. (2000). The measurement of personal values in survey research: A test of alternative procedures. *Public Opinion Quarterly*, 64, 271-298.
- Preston, C. C. & Coleman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, 1-15.
- Russell, P. A. & Gray, C. D. (1994). Ranking or rating? Some data and their implications for the measurement of evaluative response. *British Journal of Psychology*, 85, 79-92.
- Sayadi, S., Gonzalez, M. C. & Calatrava, J. (2002). Ranking versus scale rating in conjoint analysis: Evaluating landscapes in mountainous regions in southeastern Spain. Acedido em Maio 16, 2003, em <http://neptune.c3edu.uvsq.fr/eee/pages2002isee/paper/p89.pdf>
- Schwarz, N., Knäuper, B., Hippler, H.-J., Noelle-Neumann, E. & Clark, L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly*, 55, 570-582.
- Weng, L.-J. & Cheng, C.-P. (2000). Effects of response order on Likert-type scales. *Educational and Psychological Measurement*, 60 (6), 908-924.

