O emprego de escalas de magnitude na quantificação de apreciações psicológicas*

NELSON O' DE ALMEIDA**

1. Sinopse; 2. Introdução; 3. Metodologia; 4. Resultados; 5. Discussão; 6. Conclusões.

Apresentam-se, neste artigo, os resultados de um estudo sobre a possibilidade de generalizar a utilização de "escalas de magnitude" (Stevens, 1972), para expressar o julgamento de estímulos psicológicos. Nove tenentes (CAF), psicólogas do Serviço de Seleção de Pessoal da Marinha, submeteram-se ao treinamento de "calibração" proposto por Lodge e, em situação normal de trabalho, passaram a registrar seus julgamentos sobre os testes de personalidade que apreciavam, em termos de respostas clássicas de magnitude (números e tamanho de linhas), num paradigma de "modalidades cruzadas". A análise dos registros de cada psicóloga confirmou o julgamento em "magnitude" e os coeficientes expectados. Conclui-se que as escalas de magnitude propostas por Stevens, até agora empregadas em experimentos psicofísicos e sociológicos, poderão ser estendidas à mensuração das dinâmicas — e, por extensão, dos "traços" — usualmente identificadas pelos psicólogos ao interpretarem testes de personalidade.

1. Sinopse

Os últimos 20 anos têm assistido à publicação de grande número de experimentos empregando "escalas de magnitude" (de razões) na mensuração de fenômenos sensoriais, na área da psicofísica.

* O presente trabalho foi realizado no Serviço de Seleção de Pessoal da Marinha (SSPM), quando encarregado o Capitão-de-Fragata Guilherme Henrique Caspary Ribeiro, que o estimulou. Para ele contribuíram: 1.º ten. (Corpo Auxiliar Feminino) Psicólogas Vera Lúcia Gaia Campos e Ana Lúcia Martins Alves; 2.º ten. (Corpo Auxiliar Feminino) Psicólogas Angela Navarro Salles, Lucia Helena Ribeiro de Avelar, Maria Regina Josuá Abranches, Shirlene de Souza Carneiro, Tania Carvalho Richa, Thereza Christina Bastos Abraão e Vânia Carolina Moraes Alt.

A plotagem das avaliações e o cálculo das equações de regressão foram realizados pela 3.º sarg. (Quadro Feminino Geral) Tec. Cássia Inês da Silva Rezende, técnica em estatística da Força Aérea Brasileira. (Artigo apresentado à Redação em 30.4.85.))

** Psicólogo, mestre em psicologia industrial pela Ohio State University. (Endereço do autor: Rua Conde de Bonfim, 549 — apt.º 101 — 20.520 — Tijuca — Rio de Janeiro, RJ.)

Embora Thurstone, desde 1920, tenha buscado quantificar "atitudes" — e suas técnicas e derivações permanecem em uso até hoje —, coube a Stevens as marcantes contribuições de definir a "lei da resposta psicofísica", contraparte psicológica da Lei de Weber, mostrando a relação de potência entre estímulo e resposta e de publicar os expoentes a esperar, quando da mensuração de mais de 30 contínuos psicofísicos, abrangendo os cinco sentidos e, mesmo, certas sensações cinestésicas.

As medidas típicas de contínuos sociológicos — comparação aos pares, intervalos, escalas de Likert, Edwards, Coombs — têm-se relacionado apenas moderadamente com os comportamentos observados. Antes, entretanto, que se critique ou se abandone os constructos assim mensurados, faz-se mister analisar as características psicométricas oferecidas pelas técnicas citadas para concluir, com Lodge (1981), que tais escalas perdem informação significativa e que não se prestam a análises estatísticas mais potentes já que, no máximo, são pseudo-intervalares.

Se é que pode o ser humano médio utilizar-se, por exemplo, da série natural dos números inteiros para representar os julgamentos proporcionais (escala de razões) que faz sobre a estimulação recebida por quaisquer dos cinco sentidos, obedecendo sempre à "lei da potência", seria de se esperar que as metodologias, escalas e análises básicas utilizadas em psicofísica fossem estendidas à mensuração dos domínios sociológico e psicológico.

De fato, desde que Ekman (1962) buscou mensurar julgamentos morais, que Kuennapas & Wikstroem (1963) mensuraram o prestígio de diversas profissões e que Sellin & Wolfgang (1964) realizaram o célebre estudo sobre a severidade de diversos crimes, uma nova possibilidade de mensuração estava instituída. Chega-se a 1978 com Coleman & Rainwater reaplicando (15 anos após) Sellin & Wolfgang, como também o fez Figlio, no mesmo ano, utilizando-se de uma amostra de 54 mil pessoas.

Mensurou-se, pois, em escala de magnitude — uma escala de razões típicas, apenas sem ter o "zero" definido — desde o prestígio percebido em posições políticas, o poder nacional de diversos países, a qualidade estética de obras de arte, o "agradável" de peças musicais até, o grau de aceitação para comidas, bebidas e... relógios de pulso.

Em 1971, Dawson & Brinker, replicando nos EUA o estudo sobre o prestígio de diversas profissões, realizado na Suécia oito anos antes, apresentaram uma nova metodologia de validação, conhecida como das "modalidades cruzadas", o que significou a possibilidade prática de confirmar-se as qualidades das escalas de magnitude quando mensurando contínuos sociológicos.

Por outro lado, a apreciação de testes de personalidade tem sido realizada por profissionais treinados, visando identificar "traços" e "dinâmicas" determinados. Infelizmente, o resultado dessa apreciação é expresso em prosa ou em escalas nominais: a prosa (o laudo) tem-se mostrado de utilidade limitada em seleção, na medida em que, sendo basicamente qualitativa, exige abordagem clínica para as decisões finais; as categorias normalmente utilizadas, por seu lado, resumem-se a três ou quatro classes — com variância, portanto, reduzida. Seria, então, extremamente útil, sob o ponto de vista psicométrico, que se pudesse expressar o resultado dessas apreciações em escalas métricas que repre-

sentassem os "traços" avaliados ou compósitos ponderados representando as "dinâmicas" identificadas.

Desde que Guion (1965) virou, muitos acreditam, seus canhões contra as baixas validades obtidas nas seleções da época — a inexistência, mesmo, de qualquer comprovação de validade para certos testes de personalidade — e que Montmoullin (1974) publicou *A psicotécnica na berlinda* vêm os psicólogos industriais perdendo noites de sono na busca incessante de maiores validades e novos modelos que possam garantir melhores prognoses a serem apresentadas aos "inquisidores" que acusam os "ereges" por aplicarem a BTAG, o DAT ou a bateria da EDITES e, ainda mais, por se utilizarem de testes de personalidade em seleção, como se as teorias clássicas e toda uma vasta rede nomológica tivessem, apenas, o tamanho das validades até agora obtidas.

2. Introdução

O vasto número de experimentos com escalas de magnitude, como a pletora de domínios pesquisados levam a se perguntar por que não foram realizados experimentos sobre a mensuração de contínuos psicológicos clássicos.

Decidiu-se, pois, em boa hora, o Serviço de Seleção de Pessoal da Marinha a realizar um estudo sobre a possibilidade que não fora até então explorada: a utilização da metodologia, das escalas e das análises básicas da psicofísica na quantificação das "dinâmicas" de personalidade, como identificadas pelos psicólogos daquele serviço, para fins de seleção.

Esses profissionais, altamente treinados, buscam identificar "traços" e "dinâmicas" previamente determinados após exaustivas análises de trabalho — e a fixação de "padrões de excelência", por sua vez mensurados através de escalas de avaliação do desempenho — reportando os resultados obtidos através de laudos e numa escala de três categorias (Aconselhável (*A*), Aconselhável com restrições (*A/R*) e Desaconselhável (*D*)). Seria, pois, imensamente útil poder-se expressar os resultados das apreciações realizadas em uma escala de características psicométricas mais avançadas, de forma fidedigna e válida. Essa possibilidade torna-se presente quando as interpretações e julgamentos são reportados em escala de magnitude, com variância bastante ampla e com fidedignidade e validade bem mais fáceis de mensurar.

O modelo de Stevens mostra que $\hat{Y}\ aX^b$ ou $\Psi = aS^b$, onde

Ψ = magnitude percebida
a = constante (de escala)
S = magnitude do estímulo
b = característica do relacionamento.

Em termos psicofísicos pode-se verbalizar que "a magnitude da sensação cresce como uma função-potência da magnitude do estímulo" (Stevens, 1975, p. 230).

Em termos psicológicos traduzir-se-ia para "a magnitude subjetiva (apreciada) é uma função-potência da magnitude do estímulo" (Stevens, 1975, p. 266).

Na área da psicofísica é bastante simples demonstrar-se a "lei" que Cliff (1973) considerou como a mais testada das leis do julgamento humano, já que se pode comparar diretamente os "julgamentos" ou as "sensações" (variável dependente) com as quantidades métricas dos estímulos aplicados (variável independente). O estudo sistemático de diversos tipos de estimulação demonstrou o grau de generalização possível para a "lei" e permitiu a determinação dos expoentes (b) típicos para cada modo de resposta a utilizar. Os estímulos sociológicos, como os psicológicos, não apresentam, entretanto, as qualidades métricas inerentes aos estímulos psicofísicos: mas a estratégia de Dawson e Brinker, extremamente simples e eficiente, veio a propiciar a mensuração de "contínuos" sociológicos. Seu modelo mostra que se $\psi_1 = \phi_1^m$ e $\psi_2 = \phi_2^n$ e, ainda, se $\psi_1 = \psi_2$, então $\phi_1^m = \phi_2^n$ ou $\phi_1 = \phi_2^{n/m}$, o que, em coordenadas de escala logarítmica (log-log), é a representação de uma reta com "inclinação" (b) igual a n/m.

Assim, se duas sensações (ou julgamentos) são percebidas (ou considerados) como de igual magnitude, tem-se necessariamente que os estímulos se relacionam conforme um determinado expoente característico, *que não depende do estímulo* e, sim, dos modos de resposta utilizados. Portanto, como tem sido ricamente demonstrado, sempre que um estímulo sociológico é respondido através de um par qualquer de reações diferentes (alocação de números, traçado de linhas de diferentes tamanhos, aperto em um dinamômetro, escolha do volume de um som ou a intensidade de um foco de luz) pode-se determinar:

— se os julgamentos realizados sobre os estímulos estão sendo feitos em escala de magnitude;

— os "valores de escala" de tais julgamentos.

Por que não se estender, então, os princípios e as estratégias já consagrados na psicofísica e na sociologia à mensuração de contínuos psicológicos clássicos? Um tal experimento propiciaria a obtenção de índices de validade certamente mais elevados (maior possibilidade de variância na escala, que não mais limitaria a capacidade de discriminar, e permitiria a comprovação das validades aparentes admitidas pelos psicólogos (e por todos... ou quase todos), mas tão atacadas e de valores, em verdade, até agora tão baixos.

3. Metodologia

3.1 *Amostra*

A amostra estudada foi composta de nove tenentes (CAF), psicólogas servindo no Setor de Julgamento (SEJULG) do SSPM, que realizaram 552 apreciações baseadas em três instrumentos (HTP, Zulliger e um questionário estruturado) aplicados a 98 oficiais e a 87 praças ($N = 185$), e observando duas diferentes dinâmicas, assim resumidas:

Traços comuns: capacidade de adaptação, relacionamento humano, maturidade e estabilidade emocionais, capacidade de trabalho em equipe, iniciativa, senso de responsabilidade, capacidade de decisão e, ainda,

para o grupo de oficiais: capacidade de organização e capacidade de expressão oral;

para o grupo de praças: atenção ao detalhe e atenção concentrada.

As psicólogas eram formadas em universidades federais da área do Rio de Janeiro, com exceção de uma delas, formada em universidade particular. Todas estão cursando ou cursaram pós-graduação *stricto* ou *lato sensu*: o tempo médio após formatura varia de três a sete anos, com média de quatro anos; o tempo de trabalho nas funções atuais (SEJULG) varia de seis meses a três anos, com média de dois anos e sete meses; cada uma delas já realizou, em média, mais de 11 mil apreciações.

3.2 *Procedimentos*

1. A amostra foi introduzida às propostas de Stevens, conforme o sugerido por Lodge, o que compreendeu:

a) exposição (cerca de 30 minutos) sobre a teoria subjacente e sobre os resultados dos principais experimentos publicados;

b) exercício de "calibração", realizado através dos instrumentos descritos em (2);

c) troca de impressões e percepções sobre a teoria básica e sobre a proposta de pesquisa (*workshop*).

2. O exercício de calibração constou da aplicação de dois instrumentos, em forma de pequenos cadernos com nove estímulos cada, um único estímulo e resposta por página, apresentando na capa o "padrão" correspondente e assim constituídos:

a) nove números (N), em ordem aleatória, variando de 2 a 100, a serem respondidos através do traçado de uma linha horizontal, à mão livre e de comprimento correspondente à magnitude de N; padrão $N = 50$ e linha de 50mm;

b) nove linhas (TL), variando de 20 a 980mm, apresentadas em ordem aleatória, a serem respondidas pela alocação de um número proporcional ao tamanho da linha; padrão $TL = 50$mm e $N = 50$.

3. Dois dias após os exercícios de calibração, em entrevista informal com pequenos grupos (maior = 3, menor = 2), foram apresentados os resultados individuais e os obtidos por toda a amostra, tecendo-se comentários sobre o grau de sucesso de cada (obter a apreciação como esperado), além de ligeira orientação corretiva, quando necessária.

4. Imediatamente as psicólogas passaram a, além do organizacionalmente previsto (laudos sumários e escala de três categorias), alocar e anotar, para cada

apreciação realizada, um número (*N*), e a traçar uma linha de comprimento correspondente (*TL*): como "padrões" (ponto médio *das escalas*) adotou-se o número 500 (escala de *C* ➔ 1000) e uma linha de 50mm; esse procedimento permaneceu em vigor até que cada psicóloga houvesse realizado e anotado um número mínimo de julgamentos. Os resultados obtidos mostraram-se tão de acordo com o teoricamente esperado (apenas duas dentre as nove psicólogas ainda não obtinham resultados ideais) que se decidiu passar à coleta formal de dados.

5. Em novembro e dezembro de 1984 foram escolhidas duas "seleções" típicas, já realizadas mais de uma vez pelas psicólogas da amostra, para a coleta formal de dados e pesquisa; assim, cada psicóloga, *in labore* e além do organizacionalmente requerido, passou a registrar: número de código do *Pr*, número alocado (*N*), traçado de linha correspondente (*TL*) e indicação da categoria escolhida (*A, A/R, D*).

6. Os resultados dos exercícios de calibração e da coleta informal foram analisados ainda segundo Lodge, o que compreendeu:

a) mensuração das linhas em mm;

b) média aritmética, desvio-padrão e amplitude dos *N* e *TL* alocados pela mesma psicóloga;

c) cálculo das equações de regressão (potência) entre os *N* e os *TL* dados a um caso pela mesma psicóloga;

d) cálculo das médias geométricas de todas as apreciações recebidas pelo mesmo *Pr*, em termos de *N* e *TL*;

e) plotagem, em coordenadas log-log, dos *N* e das *TL* alocados pela mesma psicóloga;

f) levantamento e distribuição das diferenças, através das psicólogas, entre os *N* e os *TL* dados a um mesmo *Pr*;

g) levantamento e distribuição das diferenças entre os resultados alocados pelas diversas psicólogas e os resultados médios obtidos pelo *Pr;*

h) cômputo do índice e correlação médio entre o *N* e o *TL* recebidos pelo mesmo *Pr* (*r* de Pearson com a transformação de Fisher).

7. Os resultados coletados formalmente receberam o mesmo tratamento descrito.

4. **Resultados**

Os resultados obtidos, como explicitado em 3.2 (6), podem ser resumidos nas tabelas 1, 2 e 3, que condensam os mais importantes dados.

Tabela 1

Parâmetros da equação de regressão (potência), limites de confiança de 90% para "*b*", coeficiente de correlação de Pearson entre os *N* e os *TL* dados ao mesmo *Pr* e o número de protocolos julgados, por psicóloga

Psic.	a	b	b_{mx}	b_{mn}	$r_{N.TL}$	n
1	0,51*	0,83**	1,028	0,972	0,977	98
2	0,73*	0,75**	1,009	0,991	0,998	84
3	0,30*	0,89**	1,036	0,964	0,992	23
4	0,06	1,03	1,051	0,950	0,983	25
5	0,05	1,02	1,058	0,942	0,999	63
6	0,05	0,98	1,057	0,943	0,960	59
7	0,08	1,01	1,034	0,966	0,986	52
8	0,06	1,03	1,032	0,968	0,987	73
9	0,002	1,00	1,014	0,986	0,996	75
						552

* Significativamente diferente de Ø, @ 0,05.
** Fora dos limites de confiança de 90%.

Tabela 2

Distribuição das diferenças absolutas entre as apreciações dadas pelas psicólogas e a apreciação média de cada *Pr*

	Cl	Pm	d_N	d_{TL}
1	0- 44	22	241	230
2	45- 88	66	191	155
3	89-132	110	66	93
4	133-176	154	32	43
5	177-220	198	12	19
6	221-264	242	7	9
7	265-308	286	3	3

Tabela 3

Distribuição das médias das diferenças absolutas entre os valores de *N* e de *TL* atribuídos pela mesma psicóloga a cada caso apreciado

	Cl	Pm	$d_{\overline{N}.TL}$
1	0-30	15	6
2	31-60	45	2
3	61-90	75	1

Figura 1
Scree-Tests:
Diferenças em N

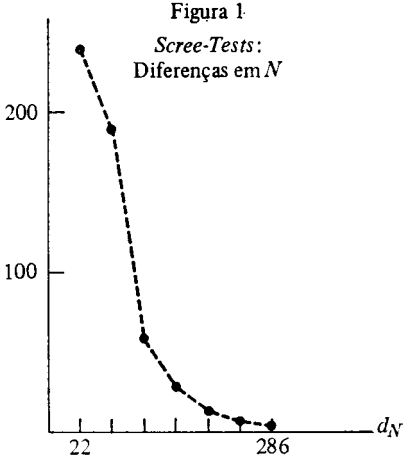


Figura 2
Diferenças em TL

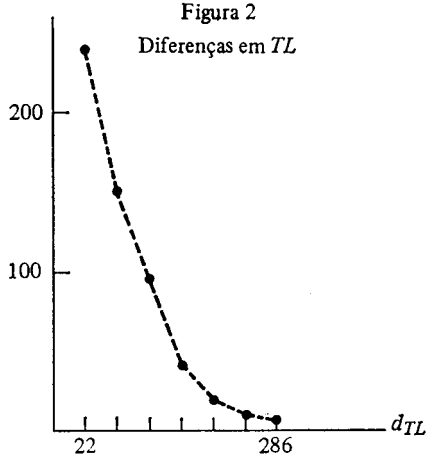
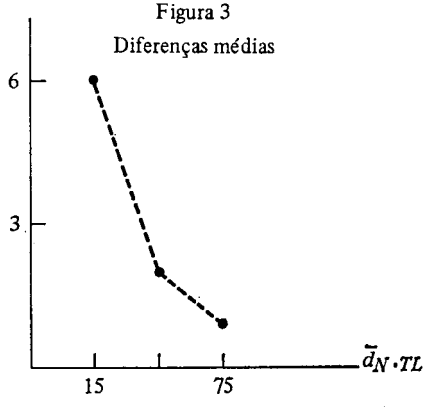


Figura 3
Diferenças médias



5. Discussão

5.1 *Geral*

Embora a amostra utilizada possa parecer pequena (apenas nove psicólogas), ela representa o universo de profissionais trabalhando no SEJULG. A amostra era, ainda, composta de psicólogas altamente treinadas. Ambos esses fatos podem diminuir a validade das conclusões e limitar a possibilidade de generalizarem-se os resultados alcançados. Assim, embora os ótimos resultados obtidos sejam reais para o SSPM, deve-se cercar de bastante cuidado ao generalizar ou ao condenar prontamente os resultados conseguidos para grupos mais heterogêneos de psicólogos.

Os comentários comuns confirmaram a atitude percebida de dúvida generalizada — se não mesmo de recusa quanto ao método e quanto ao resultado final do exercício. Atitude crítica semelhante tem sido observada em outros profissionais que tomam conhecimento, pela primeira vez, das propostas de Stevens, podendo ser considerada, portanto, como normal ao ambiente acadêmico do Rio de Janeiro.

Entretanto, nas entrevistas informais, quando já conhecidos os resultados apenas dos exercícios de calibração, as atitudes iniciais de recusa cambiavam, rapidamente, para atitudes de surpresa e, finalmente, para a de curiosidade ativa pelo método e suas potencialidades. Um tal evoluir seria de esperar e está mesmo implícito em Lodge, quando propõe, de início, que o leitor submeta-se logo ao exercício de calibração, como maneira pragmática — e eficiente — de convencê-lo e de melhor fazê-lo compreender os inúmeros experimentos confirmatórios que apresenta. O próprio autor deste artigo — hoje defensor das escalas de magnitude — teve reações semelhantes ao iniciar seus contatos com o método, por volta de 1982.

5.2 *Os exercícios*

Os resultados obtidos nos exercícios de calibração vieram confirmar sua utilidade — como treinamento tanto como evidência prática de apoio à melhor compreensão das escalas de magnitude. Pelas plotagens log-log ficou demonstrado que, já nos exercícios, 78% da amostra (7 em 9) estavam usando apreciações de magnitude (escala de razões!) ao ajuizarem números e tamanho de linhas. As únicas discrepâncias, de modo algum críticas, aconteceram justamente com as duas psicólogas que mais demonstraram ceticismo inicial, o que levaria a acreditar-se em alguma forma de bloqueio ou contaminação conducentes a esses resultados menos aceitáveis.

5.3 *A coleta livre*

Os resultados da coleta livre de dados, quando as psicólogas alocaram números e traçaram linhas "apreciando", cada uma, entre 20 e 100 protocolos ($\overline{X} = 48$), representaram etapa importante do estudo, não apenas como possibilidade de retreinamento, mas também como "aplicação-piloto", onde ficou clara a necessidade de padronizar-se as condições de registro. Como a etapa era dita "livre", algumas das psicólogas, naturalmente premidas pelo trabalho intenso, fizeram seus registros em folhas de tamanho diverso e que algumas vezes limitavam fisicamente o possível tamanho das linhas que desejariam traçar. Assim que os primeiros resultados foram coletados, passou-se a utilizar as folhas padronizadas de tamanho ofício. Pôde-se, ainda, identificar "erros de regressão" e recalcular os expoentes corrigidos: de fato, sete dentre as nove psicólogas

geraram expoentes dentro dos limites de 90% daquele previsto, confirmando julgamentos em "magnitude".

5.4 *A coleta formal*

Os resutlados da coleta formal de dados, realizada *in labore,* foram analisados como indicado em 3.2 (6). Acrescente-se que:

1. As folhas padronizadas permitiram o traçado completamente livre das linhas, o que veio confirmar-se pelo exame da amplitude média ($\bar{A}$) observada. Houve uma preferência natural pelo uso de apenas múltiplos de 10, tanto nos números quanto no tamanho de linhas, o que significa que a escala utilizada era, na realidade, de 0 ➔ 100; mas obteve-se uma liberdade de expressão de resultados radicalmente diferentes daqueles alcançados com a clássica escala de três categorias.

$N_{max} = 1000$	$N_{min} = 25$	$\bar{A} = 789$
$TL_{max} = 1200$	$TL_{min} = 25$	$\bar{A} = 825$

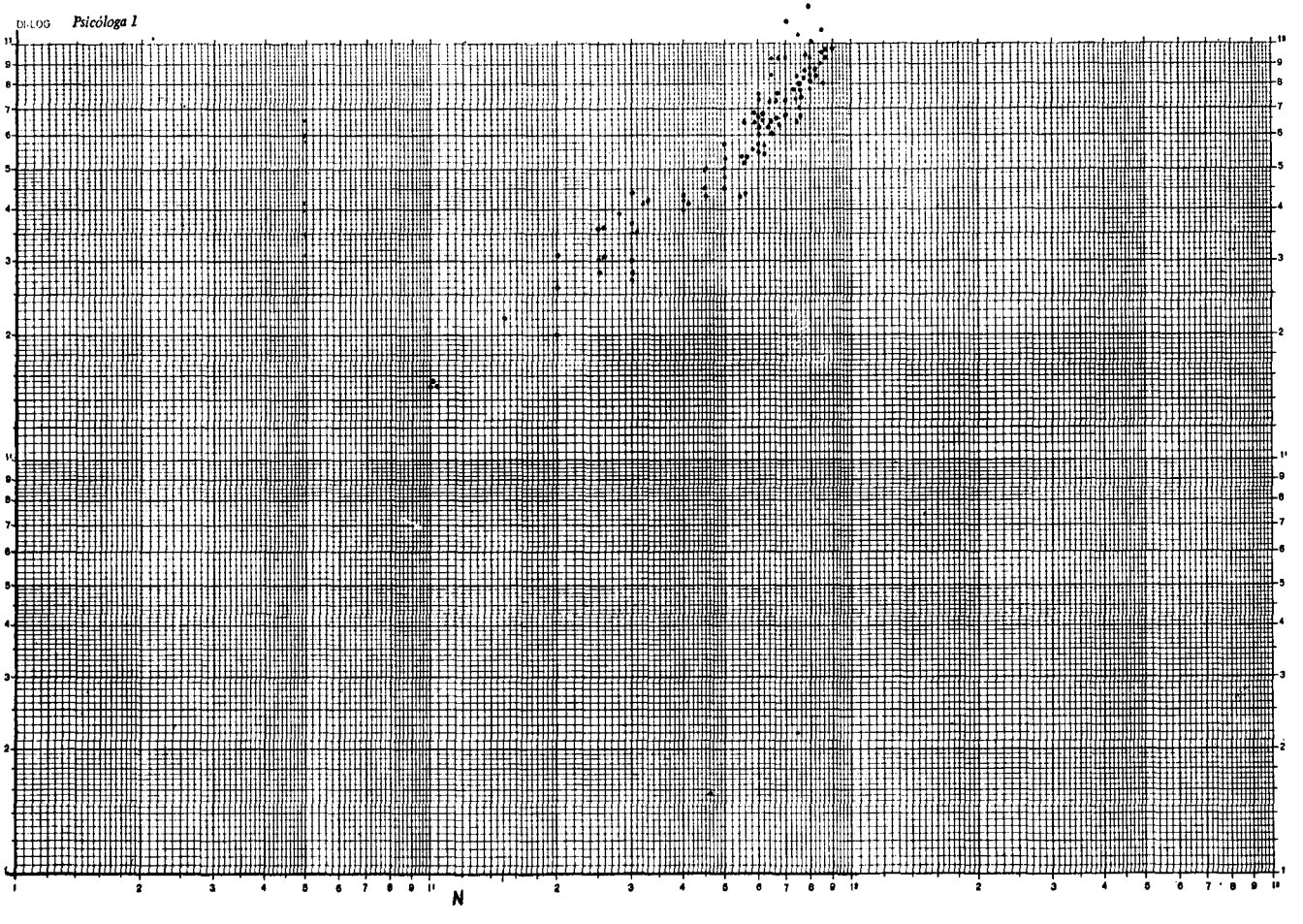
2. A precisão foi examinada sob dois aspectos: a diferença absoluta entre os *N* e os *TL* alocados ao mesmo *Pr* (precisão da mensuração — $\bar{d}_{N,TL}$) e as diferenças, também absolutas, entre as apreciações feitas pelas psicólogas e a avaliação média do *Pr* (precisão da avaliação — $\bar{d}_N$ *e* $\bar{d}_{TL}$). Deve-se considerar como excepcionalmente boa a precisão da mensuração ($d_{N.TL} = 34$; $A = 82$ a 0; $r_{N.TL} = 0{,}990$), mostrando que as psicólogas alocam números e traçam linhas de forma bastante coerente. A precisão das avaliações foi, também, muito boa ($\bar{d}_N = 64$; $A = 287$ a 0; $\bar{d}_{TL} = 70$; $A = 285$ a 3; imprecisão média menor que 7%). Tendo-se em conta que a escala de três categorias (*A, A/R, D*) apresentou, para os mesmos casos, 26% de classificações discordantes, pode-se afirmar que, na nova escala, as psicólogas conseguem fazer apreciações muito precisas. A distribuição das diferenças foi, como esperado, semelhante a um *J* à direita.

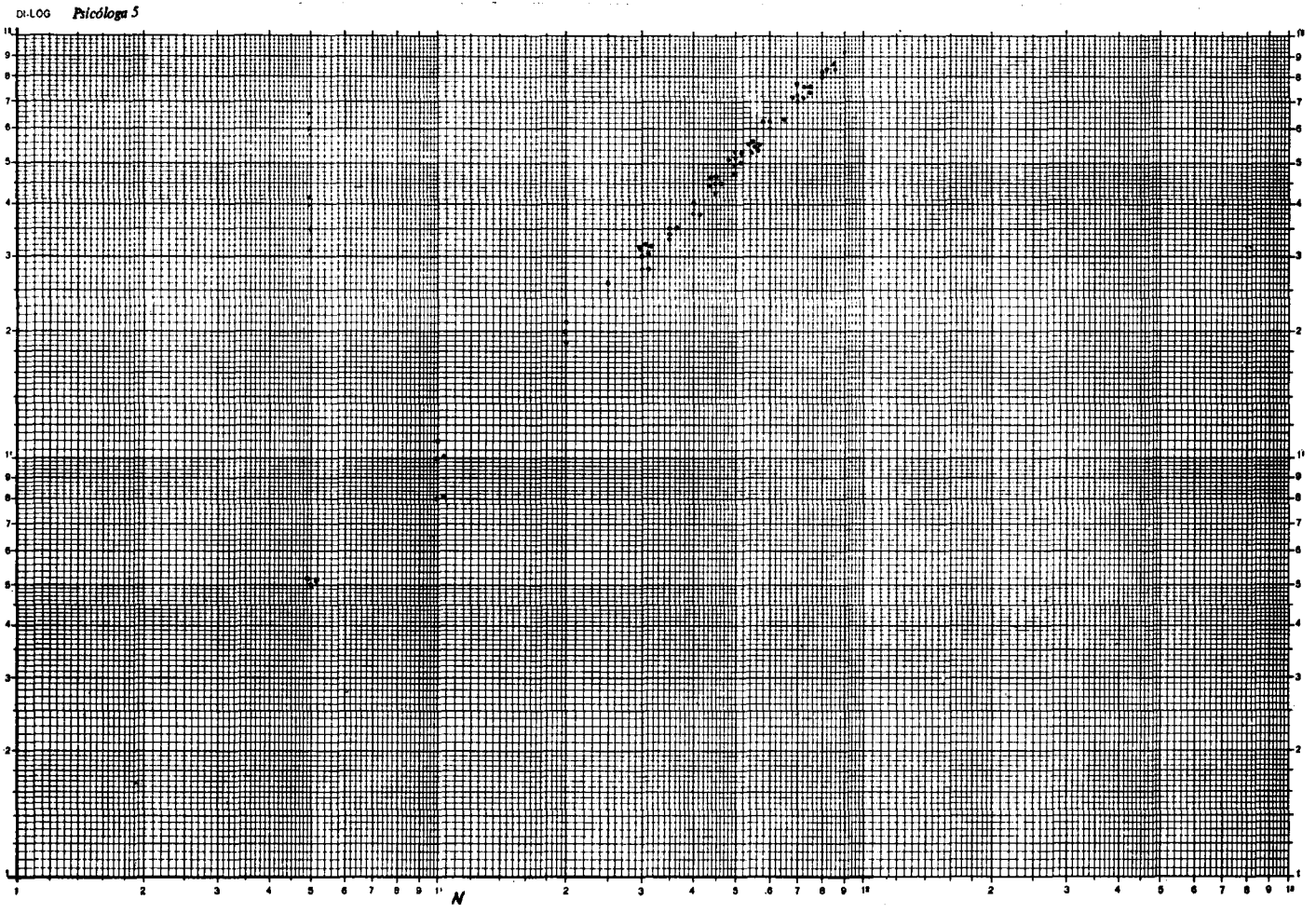
Tentou-se, ainda, obter uma estimativa grosseira de fidedignidade — embora Wherry (1952) tenha apontado as imensas dificuldades de interpretar esses índices quando referentes a apreciações — considerando como variância total aquela obtida de todos os 642 julgamentos realizados e como variância verdadeira a observada nas 185 avaliações médias dos *Pr*. Os índices obtidos foram $Fid_N = 0{,}872$ e $Fid_{TL} = 0{,}882$; parece ao autor, entretanto, que a percentagem de erro (ou a imprecisão) informaria melhor sobre o conceito em foco.

3. Mais importante que as análises anteriores — sendo, mesmo, o objetivo básico do presente artigo — seria a comprovação de que as psicólogas, ao apre-

Figuras 4A e 4B

Plotagem, em papel *log-log*, das duas apreciações extremas, realizadas pelas psicólogas 1 e 5, *N* (abscissas) *versus TL* (ordenadas).
Buscam-se distribuições lineares, homocedásticas, e uma inclinação (*b*) da reta de regressão de 45°




DI-LOG
Psicóloga 5
N

ciar testes de personalidade com objetivos definidos, podem expressar seus julgamentos em escala de magnitude e não mais apenas através de laudos qualitativos ou escalas limitadas de categorias. O método das modalidades cruzadas "valida" a escala de magnitude por qualquer uma de duas maneiras:

— o expoente (*b*) da equação de regressão estar dentro dos limites de confiança de 90% daquele esperado e/ou

— a plotagem log-log das avaliações apresentar-se como uma reta.

As plotagens mostram que as apreciações, embora com diferentes graus de variabilidade, podem ser tomadas como homocedásticas e lineares — como o esperado — exceção feita às psicólogas 1 e 2 que apresentam componentes não-lineares em suas "retas".

A análise dos expoentes obtidos mostra que seis dentre as nove psicólogas estão, realmente, utilizando escalas de magnitude. Três psicólogas, portanto, apresentaram expoentes fora dos limites previstos e as correções para viés em *TL,* propostas por Lodge, não chegaram a trazer esses expoentes para os valores requeridos. O exame detalhado da plotagem das três psicólogas em foco (1, 2 e 3) mostra que as duas primeiras não estão utilizando a mesma variabilidade ao longo de toda a escala (pequena variabilidade nos extremos, em *TL;* muita variabilidade no extremo inferior e pouca no superior, em *N*); este é o erro típico de "categorização", isto é, a escala representada é de categorias, embora possa ter um grande número delas. A psicóloga três, no entanto, não comete esse erro: seu expoente discrepante parece mais ser devido à alta variabilidade em ambos os tipos de resposta (heterocedasticidade). O fato de essa psicóloga ter sido a que menos protocolos apreciou ($n = 23$; $\overline{X}_n = 61$; $A = 98$ a 23), talvez possa significar que, com um maior número de protocolos, seu expoente viesse a coincidir com os parâmetros.

Note-se, mais, que nenhum esforço é feito para trazer as apreciações "para um mesmo padrão". Este continua tão explicitado (ou tão vago) quanto o era antes; mas agora a base métrica obtida permite abordagens bem mais ricas ao problema de padronização de interpretação.

6. Conclusões

Pelos resultados obtidos e pelas análises realizadas pode-se concluir que o modelo de Dawson e Brinker — método das modalidades cruzadas — é simples, fácil de utilizar, bem aceito e muito rico. Para psicólogos interessados — e motivados — basta, em princípio, o reconhecimento da rede nomológica e o exercício de calibração para se obter julgamentos em escala de magnitude (escala de razões). As atitudes frente ao processo tendem a mudar rapidamente do ceticismo para a aceitação curiosa e ativa. O método em nada restringe o julgamento dos profissionais, que podem expressar livremente suas apreciações e permite a comprovação clara e inequívoca de ter-se ou não atingido o nível (escala de razões) e, ainda, oferece meios para se obter "valores de escala" para esses julgamentos.

A amostra testada — nove psicólogas — após o exercício de calibração e ligeiro treinamento livre passou a expressar seus julgamentos em escala de magnitude, exceção feita a três profissionais (1, 2 e 3), que não atingiram o nível procurado, mas porque existam interveniências conhecidas, prefere-se considerar tais resultados como não-conclusivos e, ainda, sob estudo.

As apreciações foram muito precisas, tanto enquanto "mensurações", como enquanto "avaliações" e as discordâncias foram menores que as anteriormente observadas, quando do uso de escala de categorias. Os julgamentos foram expressados com bastante variabilidade, demonstrando perfeita liberdade de apreciação.

A validade dos exames de personalidade certamente será maior que a obtida até agora, não só pelo aumento de variabilidade na escala (e, portanto, de discriminação), mas também pela possibilidade de determinação mais precisa de pontos de corte, inclusive através de técnicas estatísticas que não poderiam ser utilizadas em escalas de categorias. As escalas agora obtidas (escalas métricas de razões, apenas sem ter o "zero" definido) permitirão o emprego de técnicas analíticas avançadas, até agora não utilizáveis ou empregadas sob severas restrições.

A mensuração das "dinâmicas", como foi realizada, é generalizável para a mensuração de "traços" significando a possibilidade de analisar-se, diretamente, a "estratégia" de cada profissional ao apreciar uma dinâmica qualquer, ou seja, a importância relativa que dá a cada traço. Poder-se-á melhorar a validade preditiva de cada aspecto e de todo o desempenho.

Os modelos estatísticos mais avançados de seleção e classificação de pessoal poderão ser utilizados com toda a objetividade, mesmo usando traços de temperamento e diagnoses bem mais explicitados poderão levar a decisões muito mais bem informadas.

Finalmente, os resultados de testes de personalidade devidamente quantificados — e em escalas de magnitude! — poderão ser trazidos para seleções quantitativamente justificadas e não, apenas, para as de cunho exclusivamente clínico.

Talvez se esteja iniciando uma resposta a Guion...

Abstract

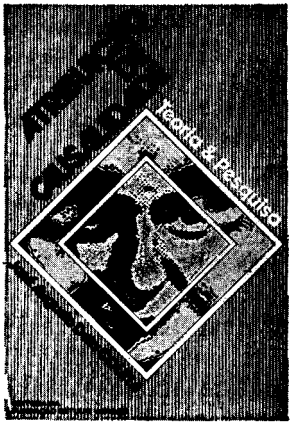
Discusses the extension of "magnitude scaling" (Stevens) to the measurement of psychological stimuli. Nine WAC lieutenants, psychologists working at the Brazilian Navy Personnel Selection Service, received the "calibration" training (Lodge) and, under normal working conditions, reported their judgements about sonality tests by means of classical magnitude responses (numbers and line lengh) in a cross-modality paradigm. Analysis of the results confirmed magnitude judgements and offered the expected exponents.

The conclusion is that magnitude scaling as proposed by Stevens and until now used only in psychophysics and in sociology, can be also used for quantifying the dynamics — and certainly the "traits" — usually identified by psychologists when analysing personality tests.

Referências bibliográficas

Cliff, N. Scaling. *Annual Review of Psychology, 24*:473-506, 1973.
Coleman, R. & Rainwater, L. *Social standing in America*. Basic Books, 1978.
Dawson, W. E. & Brinker, R. P. Validation of ratio scales of opinion by multimodality matching. *Perception & Psychophysics, 9*:413-7, 1971.
Ekman, G. Measurement of moral judgement: a comparison of scaling methods. *Perceptual & Motor Skills, 15*:3-9, 1962.
Figlio, R. M. *The national survey of crime severity*. Univ. of Pennsylvania, 1978. monograph.
Guilford, V. P. *Psychometric methods*. Wiley, 1954.
Guion, R. M. & Gothier, R. F. Validity of personality measures in personnel selection. *Personnel Psychology,* (18):135-64, 1965.
Kuennapas, T. & Wikstroem, I. Measurement of occupational preferences: a comparison of scaling methods. *Perceptual & Motor Skills, 17*:611-24, 1963.
Lodge, M. Magnitude scaling. *Quantitative applications in the social sciences*. Sage Publications, 1981.
Montmoullin, M. *A psicotécnica na berlinda*. Rio de Janeiro, Agir, 1974.
Sellin, J. T. & Wolfgang, M. E. *The measurement of delinquency*. Wiley, 1964.
Stevens, S. S. On the psychophysical law. *Psychophysical Review, 64*:153-81, 1957.
———. On predicting exponents for cross-modality matching. *Perception & Psychophysics, 6*:251-6, 1969.
———. *Psychophysics & Social scaling*. General Learning, 1972.
———. *Psychophisics: introduction to its perceptual, neural & social aspects*. Wiley, 1975.
Wherry, R. J. *Control of Bias in rating: a theory of rating*. US Army, 1952. Technical Report n. 922.

Os Porquês nas Relações Sociais



Com esta obra, o Prof. Dela Coleta, Doutor em Psicologia e psicólogo do ISOP — Instituto Superior de Estudos e Pesquisas Psicossociais, da FGV, fornece um acervo organizado de conhecimentos sobre o processo de atribuição de causalidade, tema palpitante da Psicologia Social atual, de grande interesse para estudiosos no assunto, sobretudo psicólogos de diversas áreas e estudantes de pós-graduação em psicologia.
Compreende 10 abordagens teóricas da atribuição de casualidade. Um livro para facilitar e incrementar os estudos e pesquisas nesta área no meio brasileiro.

Atribuição de Causalidade, teoria e pesquisa
José Augusto Dela Coleta
154 p.

Nas Livrarias da FGV *ou pelo Reembolso Postal*