

Validação preliminar da bateria de raciocínio diferencial — BRD para o Brasil¹

Luiz Pasquali e Robson M. de Araújo

Universidade de Brasília - LabPAM

Resumo

Trata-se da validação para o Brasil da Bateria de Raciocínio Diferencial, elaborada à luz da teoria diferencial da inteligência, baseada na bateria belga "Tests de Raisonnement Différentiel" de Meuris, adaptada à população portuguesa por Almeida e Campos em 1982. O teste consta de cinco provas, Raciocínio Numérico (NR – 30 itens), Raciocínio Abstrato (AR – 35 itens), Raciocínio Verbal (VR – 40 itens), Raciocínio Verbal (VRw – 40 itens), Raciocínio Espacial (SR – 30 itens) e Raciocínio Mecânico (MR – 44 itens), os quais sugerem um forte Fator Geral que associa a operação mental avaliada em todos eles (raciocínio) aos componentes específicos a cada conteúdo. Com os dados de uma amostra de 629 sujeitos de Brasília-DF, de ambos os sexos e nível de escolaridade superior (idade média = 27,2 e DP = 7,52), foi efetuada uma análise fatorial de cada prova e os itens submetidos à análise da Teoria de Resposta ao Item (TRI). Os Resultados da análise fatorial indicaram que cada subescala constitui um fator específico com bons índices de consistência interna (alfa de Cronbach): 0,85 para MR; 0,87 para AR; 0,92 para SR; 0,80 para VRw; 0,90 para NR. Os Resultados da estimação dos parâmetros pela Teoria de Resposta ao Item (TRI) indicam que os subtestes MR e NR avaliam os raciocínios adequadamente para sujeitos cuja aptidão se situa entre -1 DP e +2 DP, isto é, cerca de 80% dos sujeitos. O subteste SR avalia uma faixa de aptidão mais restrita, sujeitos entre -1 DP e +1,5 DP, cerca de 43%. No caso dos subtestes VR e AR, observou-se que eles avaliam uma faixa de aptidão de 80% da amostra (entre -2 DP e +1 DP), mas com escores inferiores à média. Assim, a BRD constitui-se de subtestes que apesar de se tratar de componentes específicos, possuem correlação entre si (entre 0,40 e 0,50), o que confirma ser uma medida geral de inteligência. As curvas de informações dos subtestes mostram serem os mesmos adequados para a população brasileira.

Palavras-chave: inteligência, psicometria TRI.

Differential aptitude battery: preliminary validation for Brazil

Summary

The Differential Aptitude Battery (BRD) was built by Leandro S. Almeida in 1982, based on the Belgian "Tests de Raisonnement Différentiel". The BRD is composed of five scales: Numeric Reasoning (30 items), Abstract Reasoning (35 items), Verbal Reasoning (40 items), Spatial Reasoning (30 items), and Mechanical Reasoning (44 items); they are all correlated, suggesting a second order factor, identified as the factor g. With data collected from 629 subjects from Brasília (DF, Brazil), both sexes and age averaging 27.2 years, was performed a factor analysis of each scale, showing basic unidimensionality; items were analyzed through the three parameters logistic model of IRT. Scales also showed high internal consistency (Cronbach's alpha), ranging from .80 through .92. The characteristic curves of the scales showed that they are particularly useful to measure aptitudes between -2 and +2 z.

Key-words: intelligence, psychometrics, IRT.

A Bateria de Raciocínio Diferencial (BRD) foi validada em Portugal por Leandro S. Almeida (1986), como tese do seu doutorado pela Universidade do Porto, Portugal. O autor seguiu, nesta tarefa, a teoria diferencial da inteligência. O teste é derivado

da bateria belga "Tests de Raisonnement Différentiel" de Meuris (1970), que já tinha sido anteriormente adaptada à população portuguesa (Almeida, 1982a, 1982b; Almeida & Campos, 1982), mas vinha mostrando algumas dificuldades, particularmente

1. Trabalho apresentado na Seção Coordenada "Adaptação de testes de inteligência para o Brasil". XXVIII Reunião Anual de Psicologia - SBP, Ribeirão Preto, SP, 1998.

permitindo baixa discriminabilidade, razão pela qual o autor reelaborou a bateria em 1995.

O teste consta de cinco provas, nomeado pelo autor de Bateria de Provas de Raciocínio Diferencial (BPRD); na validação brasileira, a palavra Prova foi eliminada por dar demasiada conotação escolar. A seguir, serão apresentadas brevemente as provas para a adaptação portuguesa realizada por Almeida:

Raciocínio Numérico – NR

A tarefa nesta prova consiste em continuar séries numéricas. Para elaborá-la, Almeida (1985) introduziu algumas mudanças, na prova original belga, a saber:

- aumentar o número de itens de 30 para 40;
- introduzir itens fáceis e difíceis para dar maior dispersão aos resultados do teste;
- utilizar somente números inteiros, eliminando os fracionários por induzir desânimo nos respondentes;
- exigir a resposta de dois números em lugar de um, como no original belga, para melhor perceber a compreensão, por parte do respondente, da lei que rege a série, particularmente no caso de duas séries alternadas de números no item;
- igualar o número de séries lineares (uma sequência única de números) e de séries alternadas (duas séries de números intercaladas).

O autor construiu um conjunto de 40 itens e os aplicou a alunos do 2º ano da Licenciatura em Psicologia na Universidade do Porto, utilizando a técnica da “reflexão falada” (Claparède, 1934 *apud* Almeida, 1996; Meuris, 1969). Desta aplicação resultaram 37 itens. Estes foram aplicados a 349 alunos de ambos os sexos e de vários anos de escolaridade (ensino secundário). Após a análise dos itens (dificuldade e discriminação), resultou a nova prova de 34 itens. Estes foram aplicados a 235 alunos de várias escolas e cujos dados analisados levaram à constituição da prova final de NR com 30 itens. Nesta aplicação também se mediu o tempo necessário para responder a tal prova, o qual foi posto em cerca de 20 minutos. Esta prova ficou constituída conforme da Tabela 1.

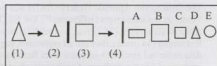
Tabela 1. Distribuição dos itens do NR por tipo de sequência

Tipo de sequência	Itens na versão final
Itens lineares Ex: 1 2 4 8 6 — —	1, 2, 3, 7, 11, 17, 18, 27, 28, 29
Itens alternados Ex: 1 10 3 20 5 — —	4, 5, 9, 10, 12, 13, 15, 16, 20, 23, 24, 25
Itens mistos Ex: 5 10 5 11 5 12 — —	6, 8, 14, 19, 21, 26
Itens com 3 séries alternadas Ex: 17 18 3 7 19 5 7 20 — —	22, 30

A prova NR da bateria brasileira BRD ficou idêntica à da versão portuguesa.

Raciocínio Abstrato – AR

A prova de Raciocínio Abstrato (AR) utilizou os 25 itens da bateria TRD de Meuris (1970), aos quais foram acrescentados mais 22 itens novos construídos nos mesmos moldes dos itens da bateria belga. Os itens deste teste constam de Figuras geométricas ou simplesmente figurativas, nos quais o sujeito tem como tarefa descobrir a relação existente entre duas figuras apresentadas e, em seguida, aplicar esta relação entre uma terceira figura dada e uma quarta que ele deve escolher entre várias alternativas dadas. Veja o exemplo a seguir:



Neste exemplo, a relação entre 1 e 2 deve se manter entre 3 e 4, resultando que a resposta correta será a alternativa C.

Os 47 itens da prova AR foram aplicados a uma amostra de 466 alunos de vários níveis escolares e de ambos os sexos do norte de Portugal. Da análise das respostas resultaram eliminados 4 itens, restando 43 que foram aplicados a uma amostra de 247 alunos da mesma região. Da análise da dificuldade e da discriminação foram retidos os 35 itens com os melho-

res índices. O tempo verificado suficiente para esta prova é de cerca de 15 minutos.

A adaptação brasileira mantém estes 35 itens refeitos através do CorelDraw para assumirem um formato mais elegante e claro.

Raciocínio Verbal - VR

É uma prova de analogias e teve que ser totalmente refeita já que a prova belga provocou várias dificuldades na aplicação em Portugal. Ademais, a prova passou de 25 (TRD belga) para 40 itens. Veja o formato da prova:

DIA está para NOITE como CLARO está para ()

- a. luz
- b. energia
- c. escuro
- d. claridade
- e. eletricidade

Sendo C a resposta correta.

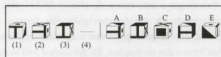
A lista inicial de itens era de 54, para cuja construção foi utilizada a diversidade de relações que se podem estabelecer entre as palavras, tais como: continuidade, oposição, sinonímia, pertença, causalidade etc. Os itens foram analisados através da "reflexão falada" com alunos da Licenciatura em Psicologia e junto a professores do ensino secundário, visando a compreensão dos mesmos. Os 53 itens resultantes dessa análise foram aplicados a 394 alunos do ensino secundário, cujos resultados serviram para a análise da dificuldade e da discriminação dos itens. Resultou disso um elenco de 43 itens que foi aplicado a 218 alunos de várias escolas secundárias. Após a análise dos itens, foram retidos 40 deles para a prova final.

Esses itens foram retidos também na adaptação brasileira, apenas reformulando expressões idiomáticas de Portugal para o nosso contexto. Ademais, como em aplicações experimentais, este teste VR se mostrava muito fácil para a população brasileira, foi decidido incluir na BRD brasileira um segundo teste de Raciocínio Verbal, este construído mais para sujeitos de nível superior. Este novo teste foi elaborado por Wagner B. Andriola (Andriola & Pasquali, 1995)

em tese de mestrado na Universidade de Brasília, que utilizou, para sua construção, dos mesmos critérios utilizados pelo Prof. Leandro em Portugal. O objetivo em utilizar duas provas de VR consiste em se poder validar, de preferência, um único teste após a aplicação nacional no Brasil, resultando na combinação dos melhores itens de ambas as provas.

Prova de Raciocínio Espacial - SR

Os itens desta prova provêm da bateria TRD (Meuris, 1970), sendo acrescidos mais 10, e são compostos de um cubo cujas faces vêm identificadas por diferentes figuras geométricas. O cubo gira duas vezes, modificando assim o desenho das faces expostas a cada movimento. A tarefa do sujeito consiste em descobrir qual a posição do cubo no próximo movimento do mesmo, escolhendo entre as alternativas dadas. Veja o exemplo:



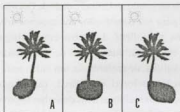
A resposta correta é B, pois o cubo se moveu linearmente para a direita.

A prova piloto de 35 itens foi aplicada a 374 alunos do ensino secundário. A análise dos índices de dificuldade e de discriminação reduziu os itens para 32, que foram aplicados a 214 alunos dos mesmos locais. Desta aplicação, após análise dos itens, foram retidos os 30 melhores itens. Esta prova exige cerca de 30 minutos para ser realizada.

A adaptação brasileira reteve os 30 itens, apenas redesenhando-os no CorelDraw para melhor qualidade estética dos cubos.

Prova de Raciocínio Mecânico - MR

Esta prova foi construída a partir de outras provas de raciocínio mecânico existentes no mercado que pretendiam avaliar a aptidão mecânica (raciocínio, compreensão e informação). Para a elaboração dos itens foi pedida ajuda de profissionais dos ramos da Engenharia e da Física. Foram, assim, elaborados 63 itens, no formato do exemplo seguinte:



Prova de Raciocínio Mecânico - MR

Exemplo D:

Qual das árvores (A, B, C) tem a sua sombra mais corretamente desenhada? Se igual, marque D.

Resposta: A árvore da figura C. Então, marque a letra C na sua Folha de Respostas.

Para a análise inicial dos itens, através da "reflexão falada", foi utilizado um grupo de estudantes de Psicologia e uma turma de alunos do 11º ano de escolaridade. A versão corrigida, em termos de dificuldades de percepção dos esquemas expressos pelos itens e das instruções, foi aplicada a 340 alunos de vários estabelecimentos de ensino secundário. Dessa aplicação resultou a eliminação de 12 itens, bem como a reformulação de outros, além da introdução de 5 novos a partir da bateria belga de Meuris (1969).

Esta versão de 56 itens foi aplicada a 323 alunos do ensino secundário de Porto e Braga. Da análise destes itens 40 deles foram retidos, mostrando índices psicométricos aceitáveis. Os itens cobrem problemas de alavancas e roldanas, resistência de materiais, esquemas geométricos, problemas de hidráulica, engrenagens, velocidade, transmissão por correias, força e dinâmica, gravidade e momentos e outros (percepção, ótica).

Segundo o autor, apesar de aceitáveis, os itens pareciam ser ainda pouco discriminativos. Assim, na adaptação brasileira procuramos melhorar a qualidade técnica dos itens, substituindo vários deles e acrescentando mais quatro. Esses itens foram discutidos com profissionais da Engenharia, que indicaram mudanças necessárias em vários deles para corresponderem à realidade da Ciência da Mecânica.

Metodologia da aplicação Brasileira

Amostra

A amostra de validação compreende 629 sujeitos adultos de Brasília-DF, de ambos os sexos – 384 homens (61%) e 245 mulheres (39%), com escolaridade de nível superior e idade média = 27,2 (DP = 7,52).

Procedimentos

Aplicação

As aplicações ocorreram em abril de 1997. A aplicação, feita por psicólogos, foi em grupo com salas que tinham não mais que 50 sujeitos cada. As subescalas do instrumento eram entregues aos sujeitos, em seguida passavam-se as instruções em grupo. Sujeitos levaram, em média, 30 minutos de duração para responder cada subescala.

Correção

A resposta dos sujeitos foi dada em folhas de leitura ótica. Essas folhas de resposta foram enviadas ao LabPAM após a leitura ótica das mesmas, a qual foi feita no Centro de Seleção e Promoção de Eventos (CESPE) da Universidade de Brasília.

Parâmetros psicométricos da amostra brasileira

Uma análise de componentes inicial mostrou que todos os itens da BRD podem ser considerados como representantes do respectivo construto que pretendem medir. Tal informação sobre a unidimensionalidade dos subtestes da bateria permitiu que se procedesse, em primeiro lugar, a uma análise dos itens individualmente, dentro de cada um dos 5 subtestes da bateria, fazendo uso do modelo logístico de 3 parâmetros de Lord (1980), dentro da visão da Teoria de Resposta ao Item, utilizando o pacote ASC (1996). Os resultados estão apresentados dentro da discussão de cada subteste a seguir. Todos os subtestes serão discutidos sob o ângulo de:

Análise dos itens

- *tabela da análise da TRI.* Esta análise permite determinar os parâmetros de dificuldade, discriminação e acerto ao acaso dos itens (Pasquali, 1996; Hambleton, Swaminathan, & Rogers, 1991).
- *curva característica do teste (ICC).* A ogiva, que permite visualizar os parâmetros dos itens individualmente, será dada somente para o conjunto dos itens do fator, a qual representa a média dos mesmos parâmetros no respectivo fator.

Validade de construto

- *dados da análise fatorial.* Inicialmente foi feita a análise dos componentes através dos componentes principais para determinar a unidimensionalidade das várias escalas ou subtestes de raciocínio, tomando como critérios o *eigenvalue*, o *screeplot* e a razão entre o *eigenvalue* do primeiro fator e do segundo fator, na qual o *eigenvalue* do primeiro fator deveria ser superior em 5 vezes ao do segundo. Por estas análises verificou-se que todos os subtestes de raciocínio, de fato, podiam ser expressos num único fator. Feita esta decisão, foi realizada a análise fatorial dos eixos principais (*Principal Axis Factoring* - PAF) com a extração de um fator.
- *curva de informação do teste.* Esta curva, que informa sobre o montante de informação que cada item traz para o traço latente medido por cada subteste de raciocínio, foi feita somente para o subteste em seu todo, a qual corresponde à média de informação dos itens do respectivo subteste.

Fidedignidade

- *alfa de Cronbach.* Reflete o grau de covariância dos itens entre si, servindo assim de indicador da consistência interna do teste.
- *coeficiente K-R21.* O índice Kuder-Richardson 21 supõe que, além de dicotômicos, os itens têm o mesmo nível de dificuldade, por isso é mais rigoroso do que os outros índices quando utilizado em um conjunto de itens com diferentes níveis de dificuldade.

I - Raciocínio Abstrato - AR

Análise dos itens

Os resultados da análise dos itens do AR estão detalhados na Tabela I-1 e na Figura I-1.

Tabela I-1. Parâmetros dos itens do AR via TRI (N = 629)

Item	a	b	c	Resid	PC	Pbs	Pbt
1	0,95	-2,05	0,20	1,30	0,92	0,57	0,52
2	1,12	-1,76	0,19	0,72	0,91	0,64	0,59
3	1,24	-1,73	0,19	0,56	0,91	0,68	0,63
4	1,18	-1,93	0,20	0,43	0,93	0,64	0,59
5	0,93	-1,74	0,20	0,83	0,90	0,54	0,52
6	1,21	-1,99	0,20	0,67	0,93	0,66	0,60
7	1,45	-1,95	0,19	0,58	0,94	0,73	0,66
8	1,30	-1,59	0,19	0,54	0,90	0,70	0,65
9	0,89	-1,73	0,20	0,89	0,89	0,54	0,50
10	1,04	-2,03	0,20	0,60	0,93	0,59	0,54
11	1,24	-1,83	0,19	0,57	0,93	0,68	0,62
12	0,89	-0,89	0,19	0,54	0,75	0,52	0,52
13	0,99	-1,25	0,20	0,26	0,85	0,55	0,54
14	1,04	-1,23	0,19	0,48	0,85	0,61	0,59
15	0,93	-1,33	0,20	0,56	0,85	0,56	0,53
16	1,09	-0,84	0,19	0,32	0,79	0,59	0,58
17	0,90	-1,16	0,20	0,75	0,82	0,54	0,52
18	0,89	-1,06	0,21	0,45	0,82	0,49	0,47
19	0,86	-1,49	0,20	0,83	0,87	0,53	0,49
20	1,05	-1,23	0,20	0,24	0,86	0,57	0,53
21	0,87	-1,12	0,20	0,74	0,82	0,52	0,49
22	0,96	-0,84	0,19	0,48	0,78	0,56	0,54
23	0,86	0,11	0,20	0,68	0,59	0,42	0,43
24	0,89	-1,54	0,20	1,09	0,87	0,56	0,52
25	0,95	-0,79	0,20	0,40	0,78	0,53	0,51
26	0,91	-1,68	0,20	0,77	0,89	0,53	0,48
27	0,92	-0,18	0,20	0,36	0,65	0,46	0,48
28	0,93	-0,60	0,19	0,73	0,73	0,55	0,53
29	1,01	-0,76	0,21	0,33	0,78	0,51	0,51
30	0,97	-0,01	0,19	0,48	0,61	0,48	0,50
31	0,95	0,04	0,19	0,46	0,59	0,48	0,50
32	0,81	0,46	0,19	0,78	0,50	0,41	0,42
33	1,10	-0,23	0,20	0,44	0,67	0,50	0,52
34	0,89	1,68	0,20	0,76	0,31	0,25	0,24
35	0,95	0,87	0,19	0,70	0,40	0,37	0,41

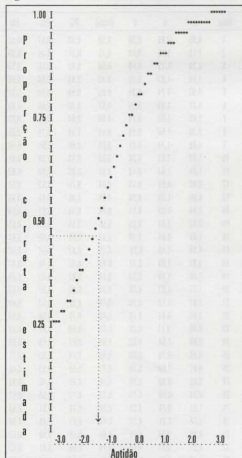
Legenda: a = discriminação do item; b = dificuldade do item; c = % de acerto aleatório; resid = resíduo;

PC = proporção de resposta correta; Pbs = correlação ponto-biserial; Pbt = correlação item - total

Somário dos Parâmetros Finais:

	Média	DP
Teta	0,00	1,00
a	1,00	0,14
b	-1,01	0,88
c	0,20	0,00

Figura I-1. Curva característica do teste AR



A Tabela I-1 e a Figura I-1 permitem os seguintes comentários:

Discriminação dos itens (parâmetro *a*): todos os itens do AR se apresentam com altos índices de discriminação, os quais vão de 0,86 a 1,24, tendo como média $a = 1,00$ ($DP = 0,14$). Sabendo-se que este índice varia entre 0 (nada discriminativo) até cerca de 4 (extremamente discriminativo) e que, na prática, esta escala varia de 0 a 2, então os itens do AR se apresentam bem discriminativos.

Dificuldade dos itens (parâmetro *b*): este parâmetro é expresso em termos de escores padrões que variam de -3 (itens extremamente fáceis) até +3 (itens extremamente difíceis), passando pelo valor 0 (itens de dificuldade mediana). Você vê que há

muitos itens fáceis, alguns de dificuldade mediana, mas muito poucos de dificuldade alta, o que torna o teste em geral fácil para a amostra de sujeitos utilizados (veja a Tabela I-2). De fato, o AR é maximamente útil para aferir sujeitos de aptidão relativamente baixa, ou seja, de -1,01 valores de teta (b médio = -1,01; $DP = 0,88$), conforme indica o sumário dos parâmetros finais e a Figura I-1.

Tabela I-2. Distribuição dos itens do AR por faixa de dificuldade

Faixa de dificuldade	Distribuição esperada		Distribuição observada	
	% prevista	Ponto z de corte	N. Itens	% Real
I	10	até - ,84	21	60
II	20	-,84 a -,26	6	17
III	40	-,26 a +,26	5	14
IV	20	,26 a ,84	1	3
V	10	,84 ou mais	2	6

Ademais, o parâmetro de dificuldade mostra que os itens não estão adequadamente distribuídos dentro do teste em ordem crescente de dificuldade, exigindo uma nova ordenação dos mesmos.

Acertos aleatórios (parâmetro *c*): este parâmetro avalia se houve respostas corretas dadas ao azar; a escala é de porcentagem. Como os itens possuíam 5 alternativas de resposta possível, sendo apenas uma delas a correta, acertos ao azar esperados são da ordem de 20%; apenas dois dos itens obtiveram um escore de 21%, concluindo-se que de fato não houve acertos devidos a uma resposta aleatória (c médio = 0,20; $DP = 0,00$).

Adequação do modelo (*Resíduo*): esta estatística avalia se o modelo utilizado de análise dos itens, a saber, o modelo logístico de 3 parâmetros, é adequado para tal análise. Em outras palavras, ele indica se há adequação (*goodness-of-fit*) entre o modelo e os dados empíricos. É aceitável um Resíduo máximo de 2,0. Nenhum dos itens obteve tal escore; pode-se, assim, concluir que o modelo de análise estatística utilizado se adequa aos dados empíricos do AR.

Conclui-se, assim, que todos os itens do AR são apropriados para avaliar o raciocínio abstrato, ressaltando apenas que a ordenação dos mesmos em termos de dificuldade dentro do teste deve ser refeita.

Validade de Construto

A validade do AR foi estabelecida através da análise fatorial dos eixos principais com extração de 01 fator e da curva de informação do teste. A Tabela I-3 apresenta os resultados da análise fatorial e a Figura I-2 a curva de informação.

Tabela I-3. Matriz fatorial do AR

Item	Carga	h ²
7	,74188	,55038
8	,72240	,52186
3	,70881	,50212
11	,68545	,46984
2	,68089	,46034
6	,63821	,40476
4	,61490	,37811
14	,54367	,29557
1	,53394	,28510
16	,51836	,26847
5	,48710	,23726
10	,48123	,23158
9	,45462	,20668
15	,45281	,20503
17	,45064	,20308
13	,44786	,20067
22	,43588	,18999
24	,43553	,18969
28	,43218	,18678
19	,42812	,17850
12	,41660	,17355
20	,40623	,16582
21	,39827	,15731
25	,37774	,14269
30	,35581	,12660
26	,34332	,11787
18	,34237	,11722
33	,34107	,11633
27	,33719	,11370
31	,32952	,10858
29	,32126	,10322
32	,28838	,08317
23	,27424	,07521
35	,25045	,06273
34	,13164	,01733
Eigenvalue	8,51159	
% Var total	24,3	
n itens	35	
Alfa	,8683	

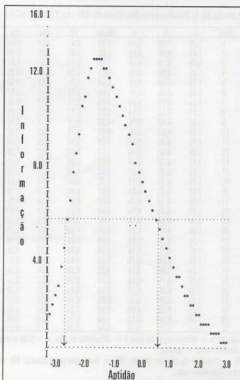
Todos os itens possuem carga no fator AR, tendo 3 deles carga inferior a 0,30, mas superior a 0,25, e apenas 1 (o item 34) com carga praticamente nula. Assim, pode-se concluir que o AR está sendo aferido adequada e validamente por mais de 30 dos 35 itens do teste. Cerca de 24% da variância total dos 35 itens é explicada pelo fator ou construto envolvido no AR.

O tamanho da carga fatorial fala da validade do item: quanto maior a carga, mais o item constitui uma representação adequada do construto medido pelo fator, no caso, o raciocínio abstrato. A coluna PBT da Tabela I-1 também fala da validade dos itens, mostran-

do a relação que eles possuem com o construto (teta), apenas que este índice é uma correlação entre o item e o teta, sendo este estimado a partir do conjunto dos itens do teste. Neste caso, a carga constitui melhor expressão da relação do item com o construto do que o PBT, visto que ela é a expressão da covariância entre o item e o construto. Contudo, os dois índices são muito congruentes e confluem para a mesma informação, a qual diz que praticamente todos os itens são bons representantes do construto raciocínio abstrato.

A validade do AR é também demonstrada e detalhada pela curva de informação do teste, conforme se observa na Figura I-2.

Figura I-2. Curva de informação do teste AR



Esta curva tem como índice esperado de informação o valor 7,427, sendo sua informação média de 5,774 (na escala expressa pela ordenada). A curva afirma que o AR é particularmente válido para aferir sujeitos que possuem uma aptidão que se situa entre cerca de $-2,4 \pm 0,7$ níveis de teta, ou seja, da aptidão para racio-

cinio abstrato. Isso significa que o teste é apto para avaliar cerca de 75% dos sujeitos da população.

Fidedignidade

A fidedignidade do AR é estimada através do índice de consistência dado pelo alfa de Cronbach, o qual é de 0,87, muito adequado. Também o teste de Kuder-Richardson (1937) confirma este dado ($K-R_{21} = 0,919$).

II - Raciocínio Espacial – SR

Análise dos itens

Os resultados da análise dos itens do SR estão detalhados na Tabela II-1 e Figura II-1.

Tabela II-1. Parâmetros dos itens do SR via TRI (N=629)

Item	a	b	c	Resid	PC	Pbt	Pbt
1	1,29	-1,06	0,16	0,89	0,83	0,57	0,59
2	1,44	-0,95	0,15	0,85	0,81	0,62	0,64
3	0,95	-0,18	0,15	1,19	0,63	0,52	0,51
4	1,39	-0,80	0,15	0,81	0,78	0,65	0,65
5	1,28	-0,50	0,15	0,68	0,71	0,63	0,62
6	1,25	-0,53	0,15	1,08	0,71	0,62	0,62
7	1,55	-0,73	0,14	0,85	0,77	0,68	0,69
8	1,35	-0,53	0,14	0,92	0,72	0,65	0,65
9	1,14	-0,46	0,15	0,89	0,70	0,59	0,59
10	1,26	-0,61	0,15	0,68	0,73	0,62	0,62
11	1,16	-0,46	0,15	1,12	0,70	0,60	0,59
12	1,30	-0,38	0,14	1,73	0,68	0,67	0,66
13	1,16	-0,26	0,14	1,50	0,64	0,61	0,60
14	1,63	-0,44	0,14	1,32	0,70	0,71	0,70
15	1,89	-0,33	0,14	1,26	0,66	0,59	0,58
16	1,32	-0,37	0,15	0,61	0,69	0,62	0,60
17	1,45	-0,32	0,15	0,85	0,67	0,66	0,65
18	1,35	-0,33	0,15	0,69	0,67	0,64	0,63
19	1,30	-0,06	0,14	0,99	0,60	0,65	0,63
20	1,04	0,27	0,15	0,80	0,53	0,51	0,49
21	1,09	0,43	0,15	0,53	0,48	0,51	0,48
22	1,32	0,51	0,14	0,70	0,44	0,56	0,55
23	1,80	0,54	0,15	0,75	0,43	0,57	0,57
24	1,74	0,66	0,14	0,53	0,39	0,54	0,54
25	1,78	0,71	0,15	0,71	0,39	0,51	0,51
26	1,79	0,61	0,15	0,75	0,42	0,53	0,52
27	1,85	0,67	0,13	0,89	0,38	0,56	0,56
28	1,97	0,66	0,13	0,66	0,38	0,56	0,56
29	1,94	0,66	0,13	1,08	0,37	0,58	0,58
30	1,43	1,29	0,14	0,54	0,26	0,37	0,36

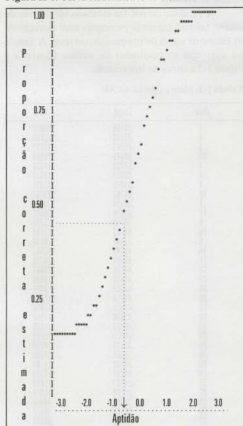
Legenda: a = discriminação do item; b = dificuldade do item; c = % de acerto aleatório; resid = residuo;

PC = proporção de resposta correta; Pbt = correlação ponto-biserial; Pbt = correlação item - teta

Sumário dos parâmetros finais

	Média	DP
Teta	0,00	1,00
a	1,42	0,28
b	-0,08	0,60
c	0,14	0,01

Figura II-1. Curva característica do teste SR



A Tabela II-1 e a Figura II-1 permitem os seguintes comentários:

Discriminação dos itens (parâmetro a): todos os itens do SR se apresentam com altos índices de discriminação, os quais vão de 0,95 a 1,97, tendo como média $a = 1,42$ ($DP = 0,28$). Sabendo-se que esse índice varia entre 0 (nada discriminativo) até cerca de 4 (extremamente discriminativo) e que, na prática, esta escala varia de 0 a 2, então os itens do SR se apresentam bem discriminativos.

Dificuldade dos itens (parâmetro b): este parâmetro é expresso em termos de escores padrões que variam de -3 (itens extremamente fáceis) até +3 (itens extremamente difíceis), passando pelo valor 0 (itens de dificuldade mediana). Pode-se ver que há muitos itens de dificuldade mediana, mas relativa-

mente poucos de dificuldade baixa ou alta, o que torna o teste em geral adequado para a amostra de sujeitos utilizados (veja Tabela II-2). De fato, o SR é maximamente útil para aferir sujeitos de aptidão mediana, ou seja, de $-0,08$ valores de teta (b médio = $-0,08$; $DP = 0,60$), isto é, na prática o nível de aptidão necessário para responder ao teste corresponde exatamente com a aptidão média dos sujeitos.

Tabela II-2. Distribuição dos itens do SR por faixa de dificuldade

Faixa de Dificuldade	Distribuição ideal		Distribuição real	
	% prevista	Ponto z	N. Itens	% Real
I	10	até $-0,84$	2	7
II	20	$-0,84$ a $-0,26$	14	47
III	40	$-0,26$ a $+0,26$	3	10
IV	20	$+0,26$ a $+0,84$	10	33
V	10	$+0,84$ ou mais	1	3

Ademais, o parâmetro de dificuldade mostra que os itens estão razoavelmente bem distribuídos dentro do teste em termos de ordem crescente de dificuldade, exigindo apenas pequenas alterações na ordenação dos mesmos.

Acertos aleatórios (parâmetro c): este parâmetro avalia se houve respostas corretas dadas ao azar; a escala é de porcentagem. Como os itens possuíam 5 alternativas de resposta possível, sendo apenas uma delas a correta, acertos ao azar esperados são da ordem de 20%; nenhum dos itens atinge tal percentual, concluindo-se que, de fato, não houve acertos devidos a uma resposta aleatória (c médio = $0,14$; $DP = 0,01$).

Adequação do modelo (Resíduo): esta estatística avalia se o modelo utilizado de análise dos itens, a saber, o modelo logístico de 3 parâmetros, é adequado para tal análise. Em outras palavras, ele indica se há adequação (*goodness-of-fit*) entre o modelo e os dados empíricos. É aceitável um Resíduo máximo de 2,0. Nenhum dos itens obteve tal escore; pode-se, assim, concluir que o modelo de análise estatística utilizado se adequa aos dados empíricos do SR.

Conclui-se, assim, que todos os itens do SR são apropriados para avaliar o raciocínio espacial, ressaltando apenas que a ordenação dos mesmos em termos de dificuldade dentro do teste deve ser refeita.

Validade de Construto

A validade do SR foi estabelecida através da análise fatorial dos eixos principais e da curva de informação do teste. A Tabela II-3 apresenta os resultados da análise fatorial e a Figura II-2 a curva de informação.

Tabela II-3. Matriz fatorial do SR

Item	Carga	h ²
14	,66842	,44412
7	,65797	,43292
12	,62898	,39562
2	,60385	,36464
19	,59535	,35445
6	,59439	,35330
4	,59417	,35304
8	,59417	,35304
17	,58536	,34294
5	,56555	,31994
13	,55662	,30982
18	,55278	,30556
1	,55203	,30473
11	,54709	,29931
9	,53542	,28667
29	,53225	,28329
10	,53056	,28149
23	,50966	,25976
27	,50753	,25758
22	,50732	,25737
15	,50714	,25719
16	,50581	,25504
28	,48163	,23197
24	,46131	,21281
3	,44767	,20041
26	,42334	,17921
25	,41494	,17218
21	,39819	,15855
20	,37801	,14289
30	,28311	,08922
Eigenvalue	9,13247	
% Var total	30,4	
n itens	30	
alfa	,9195	

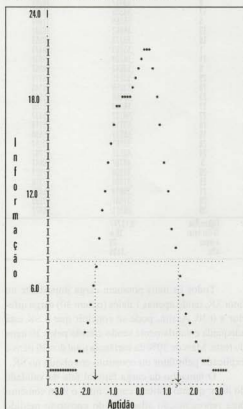
Todos os itens possuem carga importante no fator SR, tendo apenas 1 deles (o item 30) carga inferior a 0,30. Assim, pode-se concluir que o SR está adequada e validamente sendo aferido pelos 30 itens do teste. Mais de 30% da variância total dos 30 itens é explicada pelo fator ou construto envolvido no SR.

O tamanho da carga fatorial fala da validade do item: quanto maior a carga, mais o item constitui uma representação adequada do construto medido pelo fator, no caso, o raciocínio espacial. A coluna PBt da Figura II-1 também fala da validade dos itens, mostrando a relação que eles possuem com o construto (o teta), apenas que este índice é uma correlação entre o item e o teta, sendo este estimado a

partir do conjunto dos itens do teste. Neste caso, a carga fatorial constitui melhor expressão da relação do item com o construto do que o PBT, visto que ela é a expressão da covariância entre o item e o construto. Contudo, os dois índices são muito congruentes e confluem para a mesma informação, a qual diz que todos os itens são bons representantes do construto raciocínio espacial.

A validade do SR é também demonstrada e detalhada pela curva de informação do teste, conforme se observa na Figura II-2.

Figura II-2. Curva de informação do SR



Esta curva tem como índice esperado de informação o valor 14,336, sendo sua informação

média de 7,962 (na escala expressa pela ordenada). A curva afirma que o SR é particularmente válido para aferir sujeitos que possuem uma aptidão que se situa entre aproximadamente -1,2 e +1,5 níveis de teta, isto é, de aptidão para raciocínio espacial. O teste permite avaliar adequadamente 80% da população.

Fidedignidade

A fidedignidade do SR é estimada através do índice de consistência dado pelo alfa de Cronbach, o qual é de 0,92, muito adequado. Também o teste de Kuder-Richardson (1937) confirma este dado ($K-R_{21} = 0,935$).

III - Raciocínio Verbal – VR

Análise dos itens

Os resultados da análise dos itens do VR estão detalhados na Tabela III-1 e Figura III-1.

Tabela III-1. Parâmetros dos itens do VR via TRI (N = 629)

Item	a	b	c	Resid	PC	PBs	PBT
1	0,78	-2,36	0,21	0,26	0,93	0,42	0,39
2	0,75	-1,96	0,21	0,90	0,90	0,44	0,42
3	0,66	-1,14	0,21	0,95	0,80	0,40	0,39
4	0,71	-1,43	0,21	0,47	0,84	0,43	0,42
5	0,89	-1,69	0,21	0,45	0,89	0,54	0,51
6	1,10	-1,21	0,20	0,50	0,85	0,60	0,60
7	0,96	-1,69	0,21	0,46	0,90	0,56	0,53
8	0,61	-1,56	0,21	1,03	0,84	0,37	0,35
9	0,83	-1,86	0,21	0,84	0,90	0,50	0,48
10	0,98	-2,41	0,21	0,36	0,95	0,54	0,49
11	0,79	-1,17	0,21	0,57	0,82	0,47	0,47
12	0,82	-1,10	0,20	0,71	0,81	0,51	0,50
13	0,77	-1,10	0,21	0,39	0,80	0,48	0,47
14	0,99	-1,15	0,20	0,39	0,83	0,59	0,57
15	0,87	-1,12	0,20	0,60	0,82	0,53	0,53
16	0,76	-1,26	0,21	0,27	0,83	0,46	0,46
17	0,69	-0,61	0,20	0,68	0,71	0,45	0,44
18	0,67	-0,85	0,21	0,65	0,75	0,43	0,42
19	1,12	-1,11	0,20	0,64	0,84	0,60	0,60
20	0,64	-1,13	0,21	0,65	0,79	0,41	0,39
21	0,73	-1,72	0,21	0,65	0,87	0,46	0,43
22	0,80	-1,27	0,21	0,36	0,83	0,49	0,48
23	0,98	3,00	0,17	1,25	0,15	0,00	-0,05
24	0,91	3,00	0,24	1,51	0,27	-0,00	-0,06
25	0,73	-0,45	0,21	0,53	0,69	0,44	0,45
26	0,70	-1,22	0,21	0,89	0,81	0,47	0,44
27	0,91	-1,25	0,20	0,30	0,84	0,55	0,53
28	0,86	0,09	0,20	0,31	0,50	0,46	0,49
29	0,62	-0,26	0,21	0,67	0,65	0,39	0,38
30	0,79	-0,22	0,20	0,36	0,65	0,47	0,47

31	0,80	0,11	0,20	0,72	0,59	0,44	0,44
32	0,74	0,17	0,20	0,61	0,57	0,44	0,44
33	0,85	0,05	0,20	0,52	0,59	0,48	0,50
34	0,62	-0,15	0,21	0,77	0,63	0,39	0,37
35	0,58	-0,36	0,22	0,78	0,67	0,35	0,33
36	0,96	-0,43	0,21	0,58	0,70	0,51	0,52
37	0,65	1,29	0,24	0,82	0,44	0,20	0,18
38	0,71	0,10	0,20	0,51	0,58	0,42	0,42
39	0,82	-0,15	0,20	0,29	0,63	0,48	0,48
40	0,91	0,26	0,19	0,63	0,53	0,69	0,52

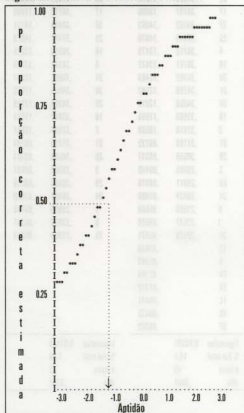
Legenda: a = discriminação do item; b = dificuldade do item; c = % de acerto aleatório; resid = resíduos;

PC = proporção de resposta correta; PBs = correlação ponto-biserial; PBt = correlação item - teta

Sumário dos Parâmetros finais

	Média	DP
Teta	0,80	1,80
a	0,80	0,13
b	-0,68	1,15
c	0,21	0,61

Figura III-1. Curva característica do teste VR



A Tabela III-1 e a Figura III-1 permitem os seguintes comentários:

Discriminação dos itens (parâmetro a): todos os itens do VR se apresentam com índices razoáveis de discriminação, os quais vão de 0,58 a 1,12, tendo como média $a = 0,80$ ($DP = 0,13$). Sabendo-se que este índice varia entre 0 (nada discriminativo) até cerca de 4 (extremamente discriminativo) e que, na prática, esta escala varia de 0 a 2, então os itens do VR se apresentam aceitavelmente discriminativos.

Dificuldade dos itens (parâmetro b): esse parâmetro é expresso em termos de escores padrões que variam de -3 (itens extremamente fáceis) até +3 (itens extremamente difíceis), passando pelo valor 0 (itens de dificuldade mediana). Você vê que há muitos itens fáceis e de dificuldade mediana, mas muito poucos de dificuldade alta, o que torna o teste em geral fácil para a amostra de sujeitos utilizados (veja Tabela III-2). De fato, o VR é maximamente útil para aferir sujeitos de aptidão relativamente baixa, ou seja, de -0,68 valores de teta (b médio = -0,68, com um enorme $DP = 1,15$, devido a dois itens extremamente difíceis, os itens 23 e 24).

Tabela III-2. Distribuição dos itens do VR por faixa de dificuldade

Faixa de dificuldade	Distribuição ideal		Distribuição real	
	% prevista	Ponto z	N. Itens (n = 40)	% Real
I	10	até -0,84	23	57
II	20	-0,84 a -0,26	4	10
III	40	-0,26 a +0,26	10	25
IV	20	+0,26 a +0,84	0	0
V	10	+0,84 ou mais	3	8

Ademais, o parâmetro de dificuldade mostra que os itens não estão adequadamente distribuídos dentro do teste em ordem crescente de dificuldade, exigindo uma nova ordenação dos mesmos.

Acertos aleatórios (parâmetro *c*): este parâmetro avalia se houve respostas corretas dadas ao azar; a escala é de porcentagem. Como os itens possuíam 5 alternativas de resposta possível, sendo apenas uma delas a correta, acertos ao azar esperados são da ordem de 20%. Houve uma pequena percentagem de acertos aleatórios no VR, pois a média do *c* ficou em 21%.

Adequação do modelo (Resíduo): esta estatística avalia se o modelo utilizado de análise dos itens, a saber, o modelo logístico de 3 parâmetros, é adequado para tal análise. Em outras palavras, ele indica se há adequação (*goodness-of-fit*) entre o modelo e os dados empíricos. É aceitável um Resíduo máximo de 2,0. Nenhum dos itens obteve tal escore; pode-se, assim, concluir que o modelo de análise estatística utilizado se adequa aos dados empíricos do VR.

Conclui-se, assim, que todos os itens do VR são apropriados para avaliar o raciocínio verbal. Entretanto dois itens, o 23 e o 24, apresentam correlações nulas com o restante do teste (veja seus PBs e PBT na Tabela III-1), embora tal evento ocorra normalmente com itens muito difíceis ou muito fáceis, uma vez que a estatística que faz esta análise se baseia em grupos críticos, que, no caso de tais itens, onde os sujeitos ou acertam ou erram, um dos grupos extremos praticamente não existe. De fato, o valor PBs reproduz a informação sobre parâmetro *a* apenas baseado no escore total e nos grupos critério, enquanto o valor do "parâmetro *a*" faz esta análise diretamente em relação ao teta; resultando que os dois itens se apresentam razoavelmente discriminativos pela análise da TRI, enquanto falham na análise tradicional da correlação ponto-biserial. Além disso, a ordenação dos itens do VR, em termos de dificuldade dentro do teste, deve ser refeita.

A validade do VR foi estabelecida através da análise fatorial dos eixos principais e da curva de informação do teste. A Tabela III-3 apresenta os resultados da análise fatorial e a Figura III-2 a curva de informação.

Tabela III-3. Matriz fatorial do VR

Primeira análise			Segunda análise (sem os itens fracos)		
40 itens			34 itens		
Item	Carga	H2	Item	Carga	H2
6	,58077	,33729	6	,57974	,33610
19	,55207	,30478	19	,55223	,30496
14	,51559	,26584	14	,51921	,26958
5	,48724	,23740	5	,48711	,23727
27	,47213	,22291	27	,47507	,22569
40	,46492	,21815	40	,46438	,21563
15	,46317	,21453	15	,46274	,21413
36	,45959	,21123	36	,45895	,21064
7	,45098	,20339	7	,44698	,19879
28	,44176	,19515	28	,44320	,19642
33	,43533	,18951	33	,43387	,18825
30	,40908	,16735	30	,41251	,17016
13	,39897	,15918	39	,39857	,15886
39	,38727	,15782	13	,39854	,15884
22	,38823	,15072	22	,39326	,15466
17	,38737	,15005	17	,38711	,14886
32	,38682	,14963	32	,38459	,14791
25	,37510	,14070	25	,37442	,14019
4	,36712	,13478	16	,36363	,13215
16	,36711	,13477	4	,36311	,13185
38	,35302	,12463	31	,35081	,12307
31	,34788	,12102	38	,34865	,12226
26	,34658	,12012	26	,34884	,12169
18	,33893	,11555	18	,33477	,11207
2	,32416	,10508	2	,32506	,10567
21	,31196	,09732	21	,31279	,09784
20	,30556	,09337	20	,30568	,09344
3	,29065	,08448	3	,29388	,08637
29	,29011	,08416	29	,29803	,08458
34	,28624	,07808	34	,28457	,07800
8	,23598	,05568	1	,23950	,05736
1	,23587	,05563	8	,23823	,05301
35	,22528	,05075	35	,22809	,05203
12		,03630			
9		,01946			
24		,01394			
23		,01312			
11		,00441			
10		,00435			
37		,00282			
Eigenvalue	6,03331		Eigenvalue	5,61460	
% Var total	15,1		% Var total	17,0	
n itens	40		n itens	33	
alfa	,8007		alfa	,8337	

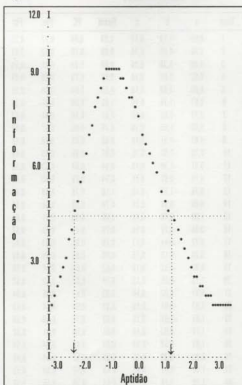
Eliminando os 7 piores itens do VR (os de número 9, 10, 11, 12, 23, 24, 37), o teste se torna mais válido (todos os 33 itens têm cargas aceitáveis) e mais preciso, pois a consistência interna sobe de 0,80 para 0,83, além de explicar maior percentagem de variância.

O tamanho da carga fatorial fala da validade do item: quanto maior a carga, mais o item constitui uma representação adequada do construto medido pelo fator, no caso, o raciocínio verbal. A coluna PBt da Tabela III-1 também fala da validade dos itens, mostrando a relação que eles possuem com o construto (o teta), este índice é uma correlação entre o item e o teta, sendo este estimado a partir do conjunto dos itens do teste. Neste caso, a carga fatorial constitui melhor expressão da relação do item com o construto do que o PBt, visto que ela é a expressão da covariância entre o item e o construto. Os dois índices são normalmente congruentes e confluem para a mesma informação; no nosso caso, isto não é verdade para os itens 23 e 24, visto que as análises PBts e PBts se baseiam no escore total do teste e trabalham com grupos critérios, o que no caso de itens extremamente fáceis ou difíceis estes grupos critérios não funcionam, uma vez que, se todos os sujeitos acertam ou erram o item, um dos grupos extremos nem existe ou é muito mal representado.

A validade do VR é também demonstrada e detalhada pela curva de informação do teste, conforme se observa na Figura III-2.

Esta curva tem como índice esperado de informação o valor 6,658, sendo sua informação média de 4,871 (na escala expressa pela ordenada). A curva afirma que o VR é particularmente válido para aferir sujeitos que possuem uma aptidão que se situa entre aproximadamente -2,2 e +0,90 níveis de teta, isto é, de aptidão para raciocínio verbal. O teste avalia adequadamente cerca de 80% da população.

Figura III-2. Curva característica do teste VR



Fidedignidade

A fidedignidade do VR é estimada através do índice de consistência dado pelo alfa de Cronbach, o qual é de 0,83, adequado. O teste de Kuder-Richardson (1937) também concorre para demonstrar a boa qualidade da fidedignidade do VR ($K-R_{21} = 0,887$).

IV - Raciocínio Numérico - NR

Análise dos itens

Os resultados da análise dos itens do NR estão detalhados na Tabela IV-1 e Figura IV-1.

Tabela IV-1. Parâmetros dos itens do NR via TRI (N = 629)

Item	a	b	c	Resid	PC	Pbs	Pbt
1	0,55	-1,78	0,17	1,23	0,84	0,27	0,25
2	0,85	-0,86	0,16	0,60	0,76	0,52	0,51
3	0,88	-1,28	0,16	0,60	0,83	0,51	0,50
4	0,82	-1,04	0,16	0,88	0,79	0,50	0,40
5	0,65	-1,59	0,16	0,90	0,84	0,36	0,35
6	0,67	-1,75	0,16	1,05	0,86	0,38	0,36
7	0,77	-1,62	0,16	0,81	0,86	0,44	0,43
8	0,80	-1,55	0,16	0,78	0,86	0,47	0,46
9	0,83	-1,31	0,16	0,49	0,83	0,48	0,48
10	0,78	-1,65	0,16	0,96	0,86	0,43	0,43
11	0,91	-1,39	0,16	0,84	0,84	0,53	0,52
12	0,79	-0,92	0,16	0,54	0,76	0,50	0,40
13	0,75	-1,13	0,16	1,10	0,79	0,48	0,47
14	0,88	-0,46	0,15	0,78	0,68	0,56	0,54
15	1,08	-0,09	0,14	0,86	0,59	0,61	0,61
16	0,98	-0,12	0,15	0,76	0,60	0,58	0,57
17	0,72	-0,44	0,17	0,76	0,68	0,43	0,42
18	0,78	0,13	0,15	0,90	0,55	0,50	0,48
19	0,93	-0,33	0,15	0,82	0,65	0,57	0,56
20	1,13	-0,35	0,15	0,74	0,67	0,62	0,62
21	0,92	0,02	0,15	0,93	0,57	0,55	0,54
22	0,93	0,01	0,15	0,37	0,58	0,54	0,53
23	1,01	-0,62	0,15	0,63	0,72	0,59	0,58
24	1,12	0,03	0,14	0,81	0,56	0,61	0,62
25	1,02	0,52	0,14	0,94	0,44	0,54	0,54
26	0,92	0,30	0,15	0,35	0,52	0,50	0,49
27	1,15	0,31	0,15	0,55	0,50	0,56	0,56
28	1,29	0,64	0,13	0,95	0,39	0,56	0,58
29	1,16	1,10	0,13	0,89	0,29	0,47	0,48
30	1,30	0,72	0,13	1,03	0,36	0,56	0,58

Legenda: a = discriminação do item; b = dificuldade do item; c = % de acerto aleatório; resid = resíduo;

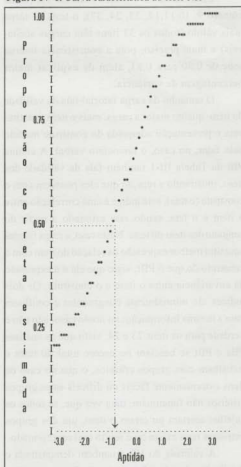
PC = proporção de resposta correta; Pbs = correlação ponto-biserial; Pbt = correlação item - teta

Sumário dos Parâmetros Finais:

	Média	DP
Teta	0,00	1,00
a	0,91	0,18
b	-0,55	0,82
c	0,15	0,01

A Tabela IV-1 e a Figura IV-1 permitem os seguintes comentários:

Figura IV-1. Curva característica do teste NR



Discriminação dos itens (parâmetro a): todos os itens do NR se apresentam com altos índices de discriminação, os quais vão de 0,55 a 1,30, tendo uma média de $a = 0,91$ ($DP = 0,18$). Sabendo-se que este índice varia entre 0 (nada discriminativo) até cerca de 4 (extremamente discriminativo) e que, na prática, esta escala varia de 0 a 2, então os itens do NR se apresentam razoavelmente discriminativos.

Dificuldade dos itens (parâmetro b): este parâmetro é expresso em termos de escores padrões que variam de -3 (itens extremamente fáceis) até +3 (itens extremamente difíceis), passando pelo valor 0

(itens de dificuldade mediana). Vê-se que há muitos itens fáceis e de dificuldade mediana, mas relativamente poucos de dificuldade alta, tendo como média um $b = -0,55$ ($DP = 0,82$, veja Figura IV-1).

Tabela IV-2. Distribuição dos itens do NR por faixa de dificuldade

Faixa de dificuldade	Distribuição ideal		Distribuição real	
	% prevista	Ponto z	N. Itens (n = 30)	% Real
I	10	até -0,84	13	43
II	20	-0,84 a -0,26	5	17
III	40	-0,26 a +0,26	6	20
IV	20	+0,26 a +0,84	5	17
V	10	+0,84 ou mais	1	3

Ademais, o parâmetro de dificuldade mostra que os itens não estão adequadamente distribuídos dentro do teste em ordem crescente de dificuldade, exigindo uma nova ordenação dos mesmos.

Acertos aleatórios (parâmetro c): este parâmetro avalia se houve respostas corretas dadas ao azar; a escala é de porcentagem. Como a resposta neste teste é construída, as alternativas são infinitas.

Adequação do modelo (Resíduo): esta estatística avalia se o modelo utilizado de análise dos itens, a saber, o modelo logístico de 3 parâmetros, é adequado para tal análise. Em outras palavras, ele indica se há adequação (*goodness-of-fit*) entre o modelo e os dados empíricos. É aceitável um Resíduo máximo de 2,5. Nenhum dos itens obteve tal escore; pode-se, assim, concluir que o modelo de análise estatística utilizado se adequa aos dados empíricos do NR.

Conclui-se, assim, que todos os itens do NR são apropriados para avaliar o raciocínio numérico, ressalvando apenas que a ordenação dos mesmos em termos de dificuldade dentro do teste deve ser refeita.

Validade de Construto

A validade do NR foi estabelecida através da análise fatorial dos eixos principais e da curva de informação do teste. A Tabela IV-3 apresenta os resultados da análise fatorial e a Figura IV-2 a curva de informação.

Tabela IV-3. Matriz fatorial do NR

Item	Carga	h ²
20	,61125	,37362
24	,59514	,35420
15	,59010	,34822
23	,57736	,33334
16	,56838	,31178
19	,54969	,30216
14	,53370	,28484
27	,53041	,28133
30	,52784	,27862
28	,52762	,27838
21	,52681	,27753
11	,51355	,26374
25	,51309	,26326
22	,51061	,26072
2	,50393	,25385
3	,48686	,23703
4	,48175	,23208
12	,47238	,22371
9	,46687	,21796
18	,46462	,21587
26	,46007	,21167
13	,45954	,21110
8	,45079	,20321
29	,43131	,18603
7	,41584	,17292
10	,41090	,16884
17	,39315	,15456
6	,33343	,11118
5	,31637	,10009
1	,22591	,05104
Eigenvalue	7,16296	
% Var total	23,9	
N itens	30	
alfa	,9816	

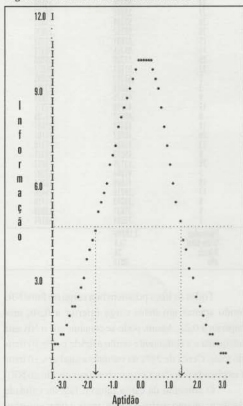
Todos os itens possuem boa carga no fator NR, tendo apenas um deles carga inferior a 0,30, mas superior a 0,22. Assim, pode-se concluir que o NR está adequada e validamente sendo aferido pelos 30 itens do teste. Cerca de 24% da variância total dos 30 itens é explicada pelo fator ou construto envolvido no NR.

O tamanho da carga fatorial fala da validade do item: quanto maior a carga, mais o item constitui uma representação adequada do construto medido pelo fator, no caso, o raciocínio numérico. A coluna PBt da Tabela IV-1 também fala da validade dos itens, mostrando a relação que eles possuem com o construto (o teta), este índice é uma correlação entre o item e o teta, sendo este estimado a partir do conjunto dos itens do teste. Neste caso, a carga constitui melhor expressão da relação do item com o construto do que o PBt, visto que ela é a expressão da covariância entre o item e o construto. Contudo, os dois índices são muito congruentes e confluem para a mesma informação, a

qual diz que os itens, praticamente todos, são bons representantes do construto raciocínio numérico.

A validade do NR é também demonstrada e detalhada pela curva de informação do teste, conforme se observa na Figura IV-2.

Figura IV-2. Curva característica do teste NR



Esta curva tem como índice esperado de informação o valor 7,730, sendo sua informação média de 4,965 (na escala expressa pela ordenada). A curva afirma que o NR é particularmente válido para aferir sujeitos que possuem uma aptidão que se situa entre cerca de -1,6 e +1,5 níveis de teta, isto é, de raciocínio numérico, aparecendo, portanto, perfeitamente adequada para a população da amostra de sujeitos, avaliando cerca de 90% da população.

Fidedignidade

A fidedignidade do NR é estimada através do índice de consistência dado pelo alfa de Cronbach, o qual é de 0,902, muito adequado. Também o teste de Kuder-Richardson (1937) confirma este dado ($K-R_{21} = 0,902$).

V - Raciocínio Mecânico – MR

Análise dos itens

Os resultados da análise dos itens do MR estão detalhados na Tabela V-1 e na Figura V-1.

Tabela V-1. Parâmetros dos itens do MR via TRI (N = 629)

Item	a	b	c	Resid	PC	Pbs	Pbt
1	0,72	-2,27	0,20	0,70	0,92	0,32	0,33
2	0,94	-2,50	0,20	0,30	0,98	0,37	0,39
3	0,78	-1,44	0,20	0,52	0,85	0,42	0,43
4	0,78	1,24	0,20	0,37	0,38	0,35	0,34
5	0,70	-2,64	0,20	1,17	0,94	0,26	0,27
6	0,52	-1,34	0,21	1,25	0,80	0,27	0,25
7	0,68	-1,80	0,20	0,56	0,88	0,35	0,35
8	0,99	-1,53	0,20	0,49	0,88	0,47	0,50
9	0,50	-0,17	0,20	0,85	0,63	0,35	0,35
10	1,00	1,35	0,17	0,70	0,30	0,40	0,40
11	0,64	-0,14	0,21	0,46	0,63	0,38	0,37
12	0,88	-0,92	0,20	0,33	0,78	0,50	0,51
13	0,85	0,01	0,20	0,47	0,60	0,48	0,49
14	0,63	-0,28	0,20	1,15	0,65	0,38	0,39
15	1,00	-1,16	0,20	0,33	0,83	0,52	0,54
16	0,43	-0,91	0,21	1,66	0,73	0,20	0,17
17	0,86	0,77	0,18	0,50	0,43	0,45	0,45
18	0,83	-0,33	0,20	0,30	0,67	0,47	0,48
19	0,73	-0,18	0,20	0,45	0,64	0,43	0,42
20	0,63	-1,05	0,20	0,82	0,77	0,38	0,38
21	0,82	-0,29	0,20	0,65	0,66	0,48	0,48
22	0,89	-0,33	0,20	0,32	0,67	0,50	0,51
23	0,72	-0,18	0,20	0,42	0,63	0,44	0,43
24	0,84	0,64	0,21	1,05	0,49	0,39	0,38
25	0,88	0,00	0,20	0,91	0,21	0,86	0,80
26	0,52	0,28	0,21	0,87	0,57	0,29	0,27
27	0,91	0,25	0,19	0,41	0,54	0,50	0,50
28	1,11	0,07	0,18	0,69	0,57	0,58	0,59
29	1,14	0,29	0,17	0,92	0,58	0,58	0,60
30	0,81	-0,13	0,20	0,47	0,63	0,46	0,46
31	0,87	0,45	0,20	0,72	0,51	0,45	0,45
32	0,71	0,49	0,20	0,26	0,51	0,40	0,39
33	0,76	2,57	0,24	1,17	0,31	0,10	0,06
34	0,79	1,16	0,18	0,97	0,37	0,39	0,38
35	0,94	0,74	0,17	0,82	0,42	0,49	0,49
36	0,74	1,29	0,19	0,76	0,38	0,36	0,35
37	0,64	2,18	0,21	1,21	0,31	0,23	0,18
38	0,76	0,07	0,19	0,60	0,58	0,46	0,46

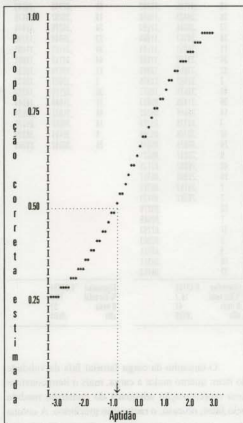
39	0,81	0,42	0,20	0,33	0,52	0,43	0,42
40	0,83	1,53	0,18	0,40	0,31	0,34	0,33
41	0,89	0,73	0,19	0,33	0,44	0,45	0,45
42	0,90	-0,24	0,20	0,35	0,65	0,52	0,51
43	0,60	-0,61	0,21	0,40	0,71	0,41	0,40
44	0,84	1,03	0,19	0,22	0,40	0,39	0,38

Legenda: a = discriminação do item; b = dificuldade do item; c = % de acerto aleatório; resid = resíduos;
PC = proporção de resposta correta; PBs = correlação ponto-biserial; Pbt = correlação item - tota

Sumário dos Parâmetros Finais:

	Média	DP
Tota	0,00	1,00
a	0,80	0,15
b	0,00	1,22
c	0,20	0,01

Figura V-1. Curva característica do teste MR



A Tabela V-1 e a Figura V-1 permitem os seguintes comentários:

Discriminação dos itens (parâmetro a): todos os itens do MR se apresentam com índices de discriminação, que vão de razoáveis (menor que $a = 0,43$) até altos (maior que $a = 1,14$), tendo como média $a = 0,80$ ($DP = 0,15$). Sabendo-se que este índice varia entre 0 (nada discriminativo) até cerca de 4 (extremamente discriminativo) e que, na prática, esta escala varia de 0 a 2, então os itens do MR se apresentam razoavelmente discriminativos.

Dificuldade dos itens (parâmetro b): este parâmetro é expresso em termos de escores padrões que variam de -3 (itens extremamente fáceis) até +3 (itens extremamente difíceis), passando pelo valor 0 (itens de dificuldade mediana). Você vê que há muitos de dificuldade mediana, mas também há vários itens de dificuldade muito baixa e outros de dificuldade muito alta, tendo uma média de $b = 0,00$ ($DP = 1,12$), o que torna o teste razoavelmente adequado para a amostra de sujeitos utilizados (veja Figura V-1). De fato, o MR é maximamente útil para aferir sujeitos de aptidão.

Tabela V-2. Distribuição dos itens do MR por faixa de dificuldade

Faixa de dificuldade	Distribuição ideal		Distribuição real	
	% prevista	Ponto z	N. Itens (n = 44)	% Real
I	10	até -0,84	11	25
II	20	-0,84 a -0,26	5	11
III	40	-0,26 a +0,26	10	23
IV	20	+0,26 a +0,84	9	20
V	10	+0,84 ou mais	9	20

Ademais, o parâmetro de dificuldade mostra que os itens não estão adequadamente distribuídos dentro do teste em ordem crescente de dificuldade, exigindo uma nova ordenação dos mesmos.

Acertos aleatórios (parâmetro c): este parâmetro avalia se houve respostas corretas dadas ao azar; a escala é de porcentagem. Como os itens possuem 3 ou 4 alternativas de resposta possível, sendo apenas uma delas a correta, acertos ao azar esperados

são da ordem de 33% e 25%. Nenhum item obteve tais escores, concluindo-se que de fato não houve acertos devidos a uma resposta aleatória (c médio = 0,20; DP = 0,01).

Adequação do modelo (Resíduo): esta estatística avalia se o modelo utilizado de análise dos itens, a saber, o modelo logístico de 3 parâmetros, é adequado para tal análise. Em outras palavras, ele indica se há adequação (*goodness-of-fit*) entre o modelo e os dados empíricos. É aceitável um Resíduo máximo de 2,0. Nenhum dos itens obteve tal escore; pode-se, assim, concluir que o modelo de análise estatística utilizado se adequa aos dados empíricos do MR. O item 25, por ser extremamente difícil, não pôde ser avaliado corretamente pelas estatísticas PBs e Pbt, porque estas trabalham com o escore total do teste e com grupos critérios; sabendo-se que com itens muito fáceis ou muito difíceis, os quais são acertados ou errados por praticamente todos os sujeitos, a formação de grupos critérios fica comprometida.

Conclui-se, assim, que todos os itens do MR são apropriados para avaliar o raciocínio mecânico, ressaltando apenas que a ordenação dos mesmos em termos de dificuldade dentro do teste deve ser refeita.

Validade de Construto

A validade do MR foi estabelecida através da análise fatorial dos eixos principais e da curva de informação do teste. A Tabela V-3 apresenta os resultados da análise fatorial e a Figura V-2 a curva de informação.

Eliminando os 12 piores itens do MR (os de número 33, 16, 5, 6, 37, 1, 26, 2, 7, 20, 43, 9), o teste se torna mais válido (todos os 32 itens têm cargas aceitáveis) e mais preciso, pois, apesar da diminuição substantiva no número de itens, a consistência interna sobe de 0,85 para 0,86, além de explicar maior percentagem de variância. Assim, pode-se concluir que o MR está sendo adequada e validamente sendo aferido, particularmente por 32 itens.

Tabela V-3. Matriz fatorial do MR

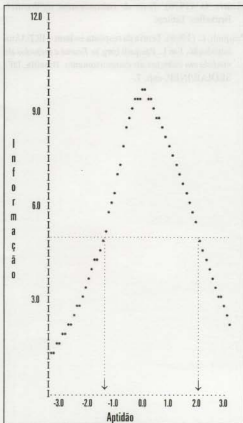
Primeira análise			Segunda análise (sem os itens fracos)		
44 itens			32 itens		
Item	Carga	h^2	Item	Carga	h^2
29	,58952	,35823	29	,60389	,36468
28	,58346	,35228	28	,58855	,35587
35	,48408	,23433	35	,48566	,23586
27	,47578	,22637	27	,48407	,23433
22	,46638	,21751	22	,46234	,21376
12	,46415	,21544	12	,45859	,21031
15	,45855	,20300	15	,44349	,19668
13	,44745	,20022	13	,44123	,19468
42	,43464	,18891	42	,44054	,19407
17	,42421	,17995	30	,42956	,18452
18	,42831	,17666	21	,42380	,17893
21	,41766	,17444	17	,42280	,17859
8	,41180	,16958	18	,41645	,17343
30	,40948	,16767	38	,41479	,17205
38	,40804	,16650	41	,38776	,15822
41	,40195	,16158	8	,38596	,15678
10	,38134	,14542	10	,39146	,15324
31	,37893	,14359	39	,38006	,14444
19	,37135	,13780	31	,37739	,14243
39	,36823	,13560	19	,36297	,13175
23	,36846	,12893	24	,35276	,12444
24	,34357	,11804	23	,35089	,12288
11	,34851	,11595	34	,33168	,11001
34	,33179	,11088	44	,33153	,10991
32	,33053	,10925	11	,32614	,10637
3	,32962	,10865	3	,32607	,10632
44	,32678	,10677	36	,32128	,10322
36	,31406	,89863	32	,31686	,10040
14	,30466	,89282	40	,31022	,89624
4	,30379	,89229	14	,30618	,89375
40	,30346	,89209	4	,30494	,89299
25	,30075	,89045	25	,30332	,89200
9	,28847	,88321			
43	,28682	,87125			
20	,25328	,86411			
7	,25143	,86322			
2	,20307	,84124			
26		,83279			
1		,82585			
37		,82283			
6		,82063			
5		,80956			
16		,80317			
33		,80163			
Eigenvalue	6,48136		Eigenvalue	6,04398	
% Var total	14,7		% Var total	18,9	
N itens	44		N itens	32	
alfa	,8526		alfa	,8594	

O tamanho da carga fatorial fala da validade do item: quanto maior a carga, mais o item constitui uma representação adequada do construto medido pelo fator, no caso, o raciocínio mecânico. A coluna

Pbt da Figura V-1 também fala da validade dos itens, mostrando a relação que eles possuem com o construto (o teta), este índice é uma correlação entre o item e o teta, sendo este estimado a partir do conjunto dos itens do teste. Neste caso, a carga constitui melhor expressão da relação do item com o construto do que o Pbt, visto que ela é a expressão da covariância entre o item e o construto. Contudo, os dois índices são muito congruentes (exceção feita para o item 25, o item mais difícil do teste) e confluem para a mesma informação, a qual diz que os itens, praticamente todos, são bons representantes do construto raciocínio mecânico.

A validade do MR é também demonstrada e detalhada pela curva de informação do teste, conforme se observa na Figura V-2.

Figura V-2. Curva de informação do teste MR



Esta curva tem como índice esperado de informação o valor 7,427, sendo sua informação média de 5,357 (na escala expressa pela ordenada). A curva afirma que o MR é particularmente válido para aferir sujeitos que possuem uma aptidão que se situa entre cerca de $-1,1$ e $+2,0$ níveis de teta, isto é, de raciocínio mecânico. O teste serve para avaliar cerca de 85% da população.

Fidedignidade

A fidedignidade do MR é estimada através do índice de consistência dado pelo alfa de Cronbach, o qual é de 0,86, muito adequado. Também o teste de Kuder-Richardson (1937) confirma este dado ($K-R_{21} = 0,875$).

Considerações finais

A Bateria de Raciocínio Diferencial - BRD -, embora ainda em fase de finalização de sua adaptação, já se apresenta altamente promissora para a população brasileira, demonstrando índices elevados em sua qualidade psicométrica. A amostra de 600 casos é ainda muito pequena para um trabalho desta natureza. A expectativa é de se coletar uma amostra nacional de cerca de 5.000 casos, pesquisa esta ora em andamento no LabPAM.

A BRD avalia aspectos da inteligência, especificamente a habilidade de raciocínio, trabalhando com materiais verbal, abstrato, numérico, espacial e mecânico. A disponibilidade de testes validados no país para avaliar aspectos diferenciados do raciocínio é ainda muito reduzida, o que torna o presente empreendimento algo de importância para os pesquisadores e profissionais de Psicologia no Brasil. Para os pesquisadores, a BRD será de utilidade para a pesquisa das habilidades de raciocínio na nossa cultura e para os profissionais ela será útil na área da seleção e da orientação vocacional, em especial no nível de segundo grau e universitário.

Um objetivo mais pragmático da adaptação da BRD visa mostrar o uso e utilidade da teoria moderna da Psicometria (TRI) no país no trabalho científico de validação de instrumentos psicológicos, uma vez que esta tecnologia é ainda incipiente no Brasil.

Referências bibliográficas

- Almeida, L. S. e Campos, B. P. (1982). Aferição dos testes de raciocínio diferencial de G. Meuris para a população do Ensino Secundário do Grande Porto. *Psicologia*, 3, (3-4), 195-204.
- Almeida, L. S. (1982a). *Testes de raciocínio diferencial (Caderno dos testes)*. Porto: Faculdade de Psicologia e de Ciências da Educação, Serviço de Consultoria Psicológica e Orientação Vocacional.
- Almeida, L. S. (1982b). *Aferição dos testes de raciocínio diferencial de G. Meuris junto dos alunos do Ensino Secundário do Grande Porto*. Porto: Faculdade de Psicologia e de Ciências da Educação, Serviço de Consultoria Psicológica e Orientação Vocacional.
- Almeida, L. S. (1985). *Bateria de provas de raciocínio diferencial*. Faculdade de Psicologia e de Ciências da Educação, Porto.
- Almeida, L. S. (1986). *O Raciocínio Diferencial dos jovens portugueses: sua avaliação, desenvolvimento e diferenciação*. Porto: Faculdade de Psicologia e de Ciências da Educação, (Tese de Doutoramento).
- Andriola, W. B. e Pasquali, L. (1995). A construção de um teste de Raciocínio Verbal (RV). *Psicologia: Reflexão e Crítica*, 8, (1), 75-83. Porto Alegre, RS.
- ASC (1996). *Item and test analysis package*. St. Paul, MN: Assessment Systems Corporation.
- Hambleton, R. K.; Swaminathan, H. e Rogers, H. J. (1991). *Fundamentals of item response theory*. London: Sage.
- Kuder, G. F. e Richardson, M. W. (1937). The theory of estimation of test reliability. *Psychometrika*, 28, 221-238.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ, Erlbaum.
- Meuris, G. (1969). La mesure du raisonnement dans l'enseignement secondaire. *Informat. Psychologique*, 36, 3-25.
- Meuris, G. (1970). *Tests de raisonnement différentiel*. Bruxelles: Editest.
- Pasquali, L. (1996). Teoria da resposta ao item - IRT: Uma introdução. Em L. Pasquali (org.) *Teoria e métodos de medida em ciências do comportamento*. Brasília, DF: SEDIAE/INEP, cap. 7.