

Estudo sobre a unidimensionalidade do Teste Matrizes Progressivas Coloridas de Raven

Unidimensionality Study about Raven's Coloured Standard Progressive Matrices

Fermino Fernandes Sisto*, Fabián Javier Marín Rueda, & Daniel Bartholomeu

Universidade São Francisco

Resumo

Este estudo avaliou o ajuste do modelo Rasch quanto à unidimensionalidade das Matrizes Progressivas Coloridas de Raven. Foram estudadas 441 crianças de ambos os sexos, de 1ª a 4ª série do ensino fundamental. Para avaliar a possibilidade de violação, optou-se pela análise de componentes principais de resíduos, pois eles podem ser suficientes para violar a unidimensionalidade intrínseca dos itens. Foram selecionados os itens de maior saturação em relação à dimensão positiva e negativa do primeiro fator, em relação ao teste como um todo e a cada série em particular. Esses itens foram novamente analisados pelo modelo de Rasch, separadamente, e correlacionadas as pontuações do fator positivo com o negativo. Os resultados sugeriram que o instrumento como um todo e a série B possibilitam a interpretação da existência de outra dimensão além da extraída pelo modelo de Rasch e nas outras duas séries as correlações estiveram próximas do ponto de rejeição.

Palavras-chave: Modelo Rasch; unidimensionalidade; teste Raven.

Abstract

This study assessed the adjustment of the Rasch model concerning unidimensionality of Raven's Colored Standard Progressive Matrices. Four hundred and forty-one (441) elementary school children, of both genders, attending first to fourth grade were investigated. Considering that residuals may be enough to violate the intrinsic unidimensionality of the items, residual principal component analysis was used to assess the possibility of this violation. Items with higher factorial loading relating to the positive and negative dimensions on the first factor were selected, in respect to the test as a whole and to each subset in particular. These items were re-analyzed with the Rasch model within each data group. Then the positive and negative factor measures were correlated. Results suggested the possibility of other dimensions besides those extracted by Rasch model based on the test and subset B, though the correlation coefficients of the other subsets were close to rejection.

Keywords: Rasch model; unidimensionality; Raven's test.

Historicamente, a primeira tentativa de construir um método estatístico para estudar os testes de inteligência teve como característica um modelo unidimensional de intercorrelações entre testes de inteligência (Spearman, 1904). Conhecido como a teoria dos dois fatores de Spearman, foi defendido que para explicar a manifestação inteligente observada nos testes eram necessários um fator geral, denominado de fator *g*, e fatores específicos, denominados de fatores *s*, apesar de mais tarde ter sido aceita a possibilidade de fatores de grupo, principalmente em razão do fator verbal (Spearman, 1927). O modelo podia ser testado usando a conhecida diferença tetrádica.

Spearman propôs algumas possibilidades de explicação para *g*. Ao final, acabou por definir que os raciocínios que subjazem ao fator *g* são a educação de relações e a educação de correlatos. Contudo, Spearman (1927) não rejeitou a idéia de que fatores não identificáveis poderiam influenciar na manifestação de *g*. Apenas para exemplificar, aceitou que se alguma habilidade e cuidado forem usados durante toda a realização de um teste, eles necessariamente comporiam o *g* medido, assim como o cansaço, entre outros.

Entre os vários testes construídos com base em sua proposta teórica do fator *g*, as Matrizes Progressivas Coloridas de Raven, um dos testes de inteligência mais amplamente usados no mundo, será analisado neste estudo. De certa forma, no geral, pode-se afirmar que esse teste deveria ser unidimensional, pois estaria medindo o fator *g*.

Na teoria dos testes, unidimensionalidade é uma característica de testes que medem uma única dimensão psicológica. Em outros termos, apenas uma única característica ou habilidade da pessoa estaria envolvida nas respostas. Não se deve confundir consistência interna com dimensionalidade, porque apesar de unidimensionalidade implicar em alta consistência interna, alta consistência interna não implica em unidimensionalidade. Não é apenas possível, mas também fácil obter alta consistência interna em testes que são multidimensionais.

Schmidt e Hunter (1998) revisaram 85 anos de estudos de validade, enfatizando a importância e utilidade de instrumentos bem construídos. Por sua vez, Jensen (1998) afirma que a quantidade de variância de *g* contida nas melhores medidas padronizadas, atualmente em uso, pode ser tão baixa quanto 64-75%. Usando técnicas modernas de construção de testes, avaliações muito precisas de habilidades

* Endereço: USF, Rua Alexandre Rodrigues Barbosa, 45, 13251-900, Itatiba, SP. fermino.sisto@gmail.com

estritamente definidas podem ser construídas e fornecer medidas que correlacionam a níveis mais altos. Por exemplo, as técnicas da Teoria de Resposta ao Item (TRI) poderiam ser usadas para construir uma escala unidimensional altamente precisa.

O estudo da unidimensionalidade assume um papel de grande importância quando esse constructo é considerado em sua relação com evidência de validade de um teste. No caso de evidências de validade de constructo ela informa se uma interpretação teórica pode ser dada às pontuações do teste (American Psychological Association, 1985). De certa forma, a unidimensionalidade se enquadra nos chamados estudos de estrutura interna (Cronbach & Meehl, 1955), nos quais as pontuações do item são relacionadas umas com as outras, pois considera as relações entre os itens dentro de um teste. Nesse caso específico, a 'teoria' testada é que todos os itens medem o mesmo constructo ou habilidade. Nesse contexto, a presença da unidimensionalidade significaria que as respostas das pessoas para todos os itens do teste dependeriam do mesmo traço subjacente ou habilidade, podendo denominá-lo de 'unifatorial'.

Ser unidimensional ou não acarreta conseqüências quando se está pontuando um teste e interpretando essa pontuação. No caso específico do Raven, é possível supor que ele contenha dois grupos de itens, quais sejam, itens relacionados à educação de relação e itens relacionados à educação de correlatos. Com base nessa forma de analisar é possível supor que a pontuação total do sujeito reflete uma combinação de ambos os tipos de educação ou habilidades. Uma pessoa poderá ter pontuação maior em educação de relações (Ex.: 10 pontos) e menor em educação de correlatos (Ex.: 5 pontos) e fará os mesmos 15 pontos que outra pessoa com maior pontuação em educação de correlatos (Ex.: 10) e menor em educação de relações (Ex.: 5 pontos).

Por esse exemplo é fácil observar que a pontuação total da pessoa pode refletir uma combinação desconhecida. Nesse contexto, com base apenas na pontuação total do sujeito, tal como o teste Raven sugere que a pessoa seja analisada, não permite inferir uma interpretação teórica adequada da pontuação total. Em razão dessas possibilidades a validade de constructo do teste, como definido pelos Standards (APA, 1985), estaria seriamente comprometida, e justifica avaliar se o teste Raven é unidimensional.

Nada obstante, o manual do teste Raven (Angelini, Alves, Custódio, Duarte & Duarte, 1999) apresenta uma fundamentação que facilita a interpretação de que deveria ser unidimensional, pois estaria medindo o fator g, algumas particularidades em sua construção e correção facilitam, também, a interpretação de que poderia não ser. Pelo manual do Raven (Angelini et al., 1999, p.129) há a informação de que o uso das pontuações totais não deve ser feito cegamente. Os três subconjuntos do Raven são corrigidos separadamente e se a pontuação de um subconjunto desvia muito de outros, eles não devem ser somados para produzir uma pontuação total. Essa informação permite interpretar que no manual é reconhecida a possibilidade de os subconjuntos medirem diferentes dimensões em certas pessoas.

Essa informação é indicativa de que a questão da unidimensionalidade deve ser analisada não apenas no teste como um todo, mas também no âmbito dos subconjuntos. Ao lado disso, deve-se levar em consideração que o manual traz a informação de que na construção do Raven, além dos tipos de educação há também itens de aprendizagem. Essa informação traz como conseqüência que cada subconjunto e o teste como um todo podem conter dois ou mais grupos de itens, dificultando mais ainda a interpretação teórica adequada das pontuações dos subconjuntos e do teste como um todo.

Para estudar a dimensionalidade de uma escala alguns métodos estão disponíveis para isso. Por um lado, os estudos de validação de constructo, classificados dentro da Teoria Clássica dos Testes, utilizam modelos estatísticos, cujas teorias consideram as pontuações com base em itens que são relacionados uns com os outros por meio da análise fatorial (Craighead, Smucker, Craighead, & Ilardi, 1998; Endler, Parker & Summerfeldt, 1998).

A análise fatorial, assim como outros métodos, entretanto, quando aplicada a variáveis binárias, que é o caso do Teste Raven produz quase sempre muitos fatores, porque de fato é uma análise multifatorial, alguns dos quais relacionados à dificuldade do item (Green, Lissitz, & Mulaik, 1977; Hattie, 1985; McDonald & Ahlawat, 1974), muitas vezes extraindo fatores artificiais. Deve-se considerar, que nesse caso, a média e o desvio padrão não podem causar fatores artificiais, porque a análise fatorial considera pontuações padrão nos quais os efeitos da média e desvios-padrão são parcialmente retirados. Ao lado disso, há que se considerar que os modelos de análise fatorial desenvolvidos para trabalhar com variáveis binárias (Bartholomew, 1980; Christoffersson, 1975; Muthén, 1978) não resolvem o problema do número de fatores, pois os testes estatísticos usados baseiam-se em que o traço latente subjacente é normalmente distribuído, o que pode não ser válido.

Por outro lado, os modelos de Teoria de Resposta ao Item (TRI) foram especialmente desenvolvidos para variáveis binárias e não necessitam da assunção de que o traço latente é normalmente distribuído (Allen & Yen, 1979; Lord, 1953, 1980; van der Linden & Hambleton, 1997). Dentre os modelos de TRI, o modelo de Rasch (1960) foi escolhido para este estudo por algumas razões técnicas. Em primeiro lugar, há que se considerar que a teoria estatística do modelo é bem desenvolvida e simples em comparação com outros modelos de TRI (Fischer, 1995; Glas, 1988; Glas & Verhelst, 1995). Um outro modelo poderia ser cogitado para o tipo de dado que o Raven fornece, qual seja, o modelo logístico de dois parâmetros. Esse modelo fornece muitos testes estatísticos, mas há carência de fundamentos matemáticos rigorosos, o que não é o caso de algumas estatísticas do modelo Rasch. Além disso, o modelo logístico de dois parâmetros possui teste estatístico pouco desenvolvido, baseado em freqüências bivariadas, exatamente nos aspectos em que as violações da unidimensionalidade são mais prováveis (van den Wollenberg, 1982).

No geral, aceitam-se três utilidades para análise de Rasch, quando se trata de estudar resultados de atividades. Uma delas é fornecer uma seqüência hierárquica de itens de acordo com a dificuldade; a outra, é avaliar o ajuste de itens ao modelo Rasch e determinar se os itens pertencem ao mesmo traço latente; e, finalmente, revelar grupos de itens de mesmo nível de dificuldade.

Ao lado disso, há um critério bastante claro de unidimensionalidade para verificar se os dados se ajustam ao modelo, impossibilitando a artificialidade. Nessa teoria o raciocínio básico para estudar se os dados se ajustam no modelo fundamenta-se na suposição de que a pessoa ao responder a um item do teste manifesta alguma quantidade de uma certa habilidade. Em outros termos, considera-se que cada pessoa possui uma quantidade dessa habilidade, passível de ser representada por um valor numérico ou pontuação, que indica seu lugar em uma escala dessa habilidade. É aceitável conceber que em cada nível de habilidade existe uma probabilidade, que pode ser determinada, de as pessoas desse nível fornecerem a resposta correta para o item; ou seja, a probabilidade esperada é pequena para pessoas de baixa habilidade e grande para as de alta habilidade.

No modelo de Rasch (1960) a probabilidade de uma resposta correta aumenta em razão da habilidade do sujeito, ao mesmo tempo em que diminui em razão da dificuldade do item. A fórmula na qual se baseia o modelo Rasch assume que essas características são constantes para todos os sujeitos e itens do teste, caracterizando, assim, a unidimensionalidade. Em decorrência, a habilidade do sujeito permanece constante por todo o teste ou, em outros termos, um sujeito usa a mesma habilidade para responder todos os itens; e a dificuldade relativa do item permanece constante por todos os sujeitos estudados.

A decorrência esperada é que os itens podem ser organizados por ordem de dificuldade e essa ordem independe das habilidades das pessoas, ou seja, pessoas inteligentes, cansadas, desmotivadas ou interessadas não mudarão a ordem encontrada. Há também uma outra decorrência de importância para o processo de avaliação, caracterizada de “suficiência da pontuação total”, pois a pontuação total é considerada como “medida suficiente” para a habilidade subjacente, já que ela conteria toda a habilidade da pessoa na variável que se está avaliando. Em outros termos, nenhuma informação relevante é perdida ao se relatar a pontuação total, nem ela contém outras informações além da habilidade em questão.

No caso do teste Raven, se ele estiver ajustado pelo modelo de Rasch a pontuação total da pessoa será um resumo da habilidade subjacente ao teste e essa pontuação é interpretável. Caso ele não esteja ajustado, isso indicaria que a pontuação total não é uma “medida suficiente” para indicar a quantidade da habilidade subjacente e, conseqüentemente, essa pontuação terá problemas para uma interpretação de seu significado teórico, pois estaria refletindo uma combinação das habilidades envolvidas e se desconhece como se deu essa combinação. Isso implicaria na

necessidade de se mudar a forma de pontuar o teste Raven, pois sua avaliação é feita pela pontuação total, para que ele não negligencie dimensões relevantes que estão sendo mensuradas por ele; ou encontrar uma organização que se ajuste à proposta de unidimensionalidade.

Em decorrências das questões levantadas este estudo avalia o ajuste do modelo Rasch quanto à unidimensionalidade em relação ao teste Raven. De certa forma, analisa a suposição de unidimensionalidade do modelo, focalizando o pressuposto de suficiência da pontuação total e do aumento monotônico e paralelo das funções de resposta ao item. Dentre os vários testes usados para verificar esse ajuste, tais como, os de Van den Wollenberg (1982), Glas (1988), optou-se pela análise de componentes principais de resíduos (Linacre, 1996). Essa análise parte do princípio de que o modelo Rasch extrai dos dados sua melhor dimensão possível, não podendo, assim, ser negada a possibilidade de que cada observação, em maior ou menor grau, conteria ainda informações peculiares, o que não deixa de ser uma contradição à unidimensionalidade intrínseca do item. Uma forma de se conhecer essas peculiaridades seria analisar a diferença entre o que o modelo Rasch prediz e o que é observado ou, em outros termos, os resíduos. Em razão disso, propôs uma forma de avaliá-los usando a análise de componentes principais, com vistas a encontrar indicações de estruturas secundárias ou subdimensões nos dados.

Método

Participantes

A amostra consistiu de 441 crianças de primeira à quarta série de uma escola municipal do interior do estado de São Paulo, com uma população de aproximadamente 45.000 habitantes. Trata-se de uma cidade com um poder aquisitivo regular, congregando uma boa diversidade sócio-cultural. As idades variaram de 7 a 11 anos ($m= 8,7$; $dp=1,22$). Com relação ao gênero havia 49,9% de meninas e 50,1% de meninos. A seguir é apresentado na Tabela 1 as frequências de participantes por idade e séries.

Instrumento

- *Teste Matrizes Progressivas Coloridas de Raven*. O instrumento caracteriza-se por um caderno composto por três séries (A, AB e B) cada uma contendo 12 itens, um em cada folha. Cada item constitui-se de desenhos, com um pedaço faltante, possibilitando apenas uma resposta correta. O indivíduo é solicitado completá-lo, escolhendo entre seis alternativas a que ele acredita ser correta. A aplicação desse instrumento se deu conforme as instruções contidas em seu manual em todas as crianças, sendo de aplicação coletiva quando indicado. Segundo os autores do manual, o instrumento possui validade de constructo fornecida pela diferenciação da idade cronológica, sendo evidenciado um aumento das médias com a idade, o que, conforme Anastasi (1967), pode ser considerado como evidência. Na avaliação do teste atribui-se um ponto a cada resposta certa, sendo a pontuação

Resultados

Tabela 1

Frequência de Participantes por Idade e Série

Idades	7		8		9		10		11		Total	
	m	f	m	f	m	f	m	f	m	f		
Séries	1	43	45	10	5	2	2	2	0	0	0	109
	2	6	3	34	46	9	5	5	0	0	0	108
	3	0	0	5	3	36	50	15	9	6	1	125
	4	0	0	0	0	2	1	35	34	13	14	99
Total		49	48	49	54	49	58	57	43	19	15	441
				97		103		107		100		34

Nota. m=masculino; f=feminino

Tabela 2

Média (m), Desvio-padrão (dp), Valor Mínimo (mín) e Máximo (máx) e Consistência Interna (Alfa de Cronbach) dos Subgrupos do Raven e no Geral

Subgrupo	Itens	m	mín	máx	dp	Alfa
A	3-12	6,93	0	10	3,52	0,67
AB	1-12	7,61	0	12	2,92	0,80
B	1-12	5,85	0	12	2,88	0,80
Geral	34	20,39	2	34	7,34	0,89

total dos indivíduos fornecida pelo número de acertos. Os itens A1 e A2, por serem usados como exemplos, não são computados na análise. As médias, desvios padrão e consistência interna (alfa de Cronbach) dos subgrupos A, AB, B e no geral são relatados na Tabela 2.

Procedimentos de análise dos dados

Primeiramente foram computados os erros por item, série por série, com vistas a identificar os acertos e erros para uma possível interpretação dos resultados. No caso da série A foram excluídos os itens 1 e 2 de todas as análises por terem sido usados para ensino, conforme indicação do manual, como também porque foram acertados por todos os participantes.

No estudo seguinte, foi feita a análise de componentes principais dos resíduos. Um teste com uma extensão razoável precisaria ter pelo menos três itens para afirmar a presença de uma dimensão secundária, e, caso tenha menos, é provável que ele seja unidimensional. Nessas condições, estudos de simulação indicaram que eigenvalues menores que 1,4 podem ser considerados aleatórios (Smith & Miao, 1994). Em decorrência, quando os resultados se enquadraram nesses critérios, prosseguiu-se a análise, selecionando os itens de maior saturação em relação à dimensão positiva e negativa do primeiro fator. Esses itens do fator positivo e negativo foram submetidos novamente ao modelo de Rasch, separadamente, e foram correlacionadas as pontuações do fator positivo com o fator negativo.

Para interpretar as correlações encontradas entre os fatores extraídos dos resíduos, foi assumido que a hipótese é de que se trataria de testes paralelos e considerou-se o limite mínimo de 0,51 para a correlação de Pearson, pois se aconselha precaução quando o contraste das medidas dos fatores explicar mais variância (Wright, 2000). Assim, uma correlação menor que 0,51 foi considerada indicadora da presença de outro fator, ou seja, a forte possibilidade de a escala não ser unidimensional; e acima de 0,51 foi considerada como não indicadora da existência de outra dimensão de importância.

Computadas as respostas para a série A, pode-se observar que até o item 6 as respostas corretas atingem mais de 90% de acertos. Nos itens 7 a 10 a maior parte dos sujeitos (56,3%, 65%, 61,9% e 68,6%, respectivamente) respondeu corretamente. Levando em conta a configuração dos desenhos dessa série, observou-se que ocorreu o mesmo tipo de erro nas escolhas da alternativa, pois os sujeitos mantiveram a forma do desenho, mas inverteram as cores.

Por sua vez, nos itens 11 e 12 a frequência de erros foi maior que a de acertos (26,7% e 22,9%, respectivamente). Das respostas erradas no item 11, as maiores frequências foram nas alternativas 5 (37,4%) e 2 (22,9%); por sua vez, no item 12, nas alternativas 4 e 6 (29,6% e 31,7%, respectivamente). Analisando qualitativamente as respostas dos sujeitos observou-se que eles não levaram em consideração, conjuntamente, a linha e a coluna para a solução, mas repetiram o desenho anterior correspondente ou à linha ou à coluna para completar a figura.

Em relação à série AB, a maior parte dos sujeitos respondeu corretamente os itens de 1 a 5 (entre 68-96%). Em relação aos itens de 6 a 11 as porcentagens de acerto variaram entre 47,3-65% e nas análises desses desenhos observou-se que as crianças não consideraram as informações da linha e da coluna; escolhendo como resposta em quase todos os casos uma figura igual à imediatamente acima, mas também escolheram ou a oposta ou a imediatamente ao lado, em alguns casos. Novamente observou-se, nos casos de erro, que a tendência foi a repetição de uma figura.

Finalmente, no que concerne à série B, nos itens de 1 a 5 a maior parte dos sujeitos respondeu corretamente (entre 90-60%). Por sua vez, os itens 6, 7, e 10 constituíram um segundo grupo em termos de acertos, pois 44,2%, 39,5% e 33,1%, respectivamente, responderam corretamente. Em relação aos itens 8, 9, 11 e 12, a frequência de erros foi maior que a de acertos. No item 8 as alternativas mais frequentes foram 5 e 4 (31% e 25,8% respectivamente) e somente 22,2% das crianças assinalou a resposta correta. Analisando a escolha das alternativas em razão da configuração dos desenhos observou-se que ocorreu o mesmo tipo de erro evidenciado nas séries anteriores, ou seja, nítida preponderância na escolha de uma figura igual à imediatamente acima ou imediatamente ao lado da alternativa.

Análise de componentes principais dos resíduos

A - todos os itens do teste Raven (n=34)

O fator 1 explicou 2.5 de 34 unidades de variância residual. A proporção do fator no Yardstick foi de 67,3/1. Considerando que o eigenvalue é superior a 1,4 e inferior a 3,0, optou-se pela continuidade da análise dada a possibilidade de ser necessário mais de um fator. Os resultados do fator 1, com saturação igual ou superior a 0,19, estão na Tabela 3.

Os itens com cargas positivas selecionados foram b1, b2, b3 e b4 e os com carga negativa foram os ab12, b10, b7, b11

Tabela 3
Estatísticas do Fator 1, Extraído por Análise de Componentes Principais dos Resíduos, em Relação a 34 Itens

Saturação fatorial	Measure	Infit MNSQ	Outfit MNSQ	Item
0,66	-2,25	-0,96	0,82	b1
0,57	-1,50	1,02	0,74	b2
0,56	-0,96	0,85	0,70	b3
0,49	-0,76	0,78	0,69	b4
0,19	-0,16	0,82	0,77	ab7
0,19	0,44	0,76	0,64	ab6
-0,57	2,62	1,26	2,16	ab12
-0,36	1,79	1,00	1,53	b10
-0,33	1,42	1,06	1,22	b7
-0,32	2,71	0,96	1,35	b11
-0,29	2,20	0,92	1,94	b9
-0,23	2,58	0,95	1,71	b8
-0,22	-0,07	1,27	1,45	a8
-0,20	1,31	0,87	0,99	ab9
-0,19	0,07	1,04	1,01	a9

Tabela 5
Estatísticas do Fator 1, Extraído por Análise de Componentes Principais dos Resíduos, da Série AB

Saturação fatorial	Measure	Infit MNSQ	Outfit MNSQ	Item
0,60	-0,06	0,96	0,94	Ab15
0,51	0,88	0,75	0,60	Ab6
0,30	-2,85	0,81	1,26	Ab2
0,27	-3,12	0,92	3,32	Ab3
0,18	-0,54	0,79	0,75	Ab4
0,13	0,17	0,83	0,88	Ab7
0,05	1,45	0,86	1,06	Ab8
-0,50	3,44	1,35	6,57	Ab12
-0,39	1,39	1,14	2,02	Ab10
-0,39	1,39	1,00	1,64	Ab9
-0,14	1,43	1,13	1,69	Ab11
-0,14	-3,58	0,88	0,91	Ab1

e b9. Esses dois conjuntos de itens foram submetidos ao modelo de Rasch, separadamente, para novo cálculo das pontuações dos sujeitos e as pontuações totais. As pontuações positivas forneceram uma média de 1,89 ($dp=1,86$) e as pontuações negativas forneceram uma média de -1,25 ($dp=1,70$).

Em seguida as pontuações de cada sujeito em cada um dos conjuntos foram correlacionadas e o coeficiente de correlação de Pearson encontrado foi de $r=0,18$. Esse coeficiente foi corrigido pela fórmula de profecia Spearman-Brown, tendo por base a extensão de 34 itens, evidenciando um valor de $r=0,45$. Esse coeficiente, inferior ao 0,51 sugerido como indicativo de que os resíduos não são importantes, indicou que os dois conjuntos, cujos resíduos estavam mais discrepantes, possuem uma parte da variância não explicada pela comunalidade entre eles, possibilitando interpretar que peculiaridades nas respostas desses itens não explicadas pela unidimensionalidade não são desprezíveis. Dessa forma, essa análise sugere a necessidade de um outro fator para explicar uma parte da variância que a dimensão captada como preponderante pelo modelo de Rasch não explica. Em outros termos, o Teste Matrizes Progressivas Coloridas de Raven não é unidimensional.

Tabela 4
Estatísticas do Fator 1, Extraído por Análise de Componentes Principais dos Resíduos, da Série A

Saturação fatorial	Measure	Infit MNSQ	Outfit MNSQ	Item
0,54	-2,77	0,73	0,25	A4
0,47	-3,29	1,04	0,89	A3
0,40	-1,94	0,99	0,91	A5
0,20	0,33	0,99	1,16	A10
0,17	-1,69	0,90	1,37	A6
0,09	1,21	0,83	1,10	A7
-0,54	0,65	1,22	1,31	A8
-0,54	3,20	0,99	4,19	A11
-0,20	0,80	0,89	0,89	A9
-0,08	3,51	1,08	1,15	A12

Tabela 6
Estatísticas do Fator 1, Extraído por Análise de Componentes Principais dos Resíduos, da Série B

Saturação fatorial	Measure	Infit MNSQ	Outfit MNSQ	Item
0,62	-3,59	0,92	0,70	B1
0,49	-2,07	0,78	0,68	B3
0,39	-2,70	0,87	0,86	B2
0,39	-1,84	0,67	1,04	B4
0,13	-0,81	0,87	1,09	B5
-0,54	2,18	0,97	1,42	B11
-0,50	1,07	0,93	2,11	B10
-0,49	0,64	1,24	1,83	B7
-0,35	1,56	0,88	2,44	B9
-0,28	2,02	0,93	1,53	B8
-0,14	3,27	1,07	2,11	B12
-0,02	0,26	1,21	1,63	B6

B - Série A

O fator explicou 1.4 de 10 unidades de variância residual, um por item. A proporção do fator no Yardstick foi de 39,7/1. Mesmo o eigenvalue estando no limite, optou-se por continuar a análise, dado que a extensão da série é pequena e o valor de poder do Yardstick. Os resultados do fator 1 estão na Tabela 4.

Os itens com cargas positivas selecionados foram a3, a4, a5 e a10 e os com carga negativa foram os a8, a9 e a11. Esses dois conjuntos de itens foram submetidos ao modelo de Rasch, separadamente, para novo cálculo das pontuações e as pontuações totais. O conjunto de itens com pontuações positivas forneceu uma média de 3,15 ($dp=1,70$) e o conjunto de itens com pontuações negativas forneceu uma média de 0,04 ($dp=1,95$).

O coeficiente de correlação de Pearson entre as pontuações de cada sujeito em cada um dos conjuntos foi de $r=0,33$. Esse coeficiente foi corrigido pela fórmula de profecia Spearman-Brown, tendo por base a extensão de 10 itens, evidenciando um valor de $r=0,45$. Esse coeficiente, inferior ao 0,51, pôde ser interpretado no sentido de que os dois conjuntos produziram uma parte da variância não explicada pela comunalidade entre eles, ainda que em uma região de

rejeição bastante limítrofe. Assim, as particularidades desses itens não explicadas pela unidimensionalidade podem não ser desprezíveis. Dessa forma, essa análise deixa a dúvida da necessidade de um outro fator para explicar uma parte da variância; em outros termos, não se pôde afirmar nem negar a unidimensionalidade da Série A das Matrizes Progressivas Coloridas de Raven.

C - Série AB

O fator 1 explicou 1.4 de 12 unidades de variância residual, um por item. A proporção do fator no Yardstick foi de 4,7/1. Novamente adotou-se o mesmo procedimento em relação ao eigenvalue encontrado, principalmente em razão do Yardstick. Os resultados do fator 1 estão na Tabela 5.

Os itens com cargas positivas selecionados foram ab2, ab5 e ab6 e os com carga negativa foram os ab9, ab10, ab12 e foram utilizados os mesmos procedimentos das duas análises anteriores. Assim, os itens com pontuações positivas forneceram uma média de 1,85 ($dp=2,68$) e os itens com pontuações negativas forneceram uma média de -0,58 ($dp=1,79$).

Por sua vez, o coeficiente de correlação de Pearson encontrado foi de $r=0,36$, cuja correção, considerando a extensão de 12 itens da série, evidenciou um valor de $r=0,53$. Esse coeficiente, superior ao 0,51 sugeriu a unidimensionalidade da Série AB das Matrizes Progressivas Coloridas de Raven.

D - Série B

O fator 1 explicou 1.9 de 12 unidades de variância residual, um por item. A proporção do fator no Yardstick foi de 7,3/1. Quanto ao eigenvalue, tomou-se a mesma decisão. Os resultados do fator 1 estão na Tabela 6.

Os itens com cargas positivas selecionados foram b1, b2, b3, b4 e os com carga negativa foram os b7, b9, b10 e b11, sendo que os itens com pontuações positivas forneceram uma média de 1,92 ($dp=2,15$) e os de pontuações negativas, uma média de -1,09 ($dp=1,79$). O coeficiente de correlação de Pearson encontrado foi de $r=0,30$, cuja correção resultou em um $r=0,46$. Novamente, os dois conjuntos, cujos resíduos estavam mais discrepantes, deixaram uma parte da variância não explicada pela comunalidade entre eles. Em outros termos, não se pode afirmar a unidimensionalidade da Série B das Matrizes Progressivas Coloridas de Raven.

Discussão e Conclusões

A unidimensionalidade do teste de Raven foi testada pelo modelo de Rasch, estudando o teste como um todo, e cada subconjunto (A, AB e B) separadamente. Em três das quatro situações estudadas a unidimensionalidade não pôde ser comprovada. Esse dado é mais preocupante, principalmente, no caso do teste como um todo (pontuação total). Nas três séries, há uma tendência à unidimensionalidade, ainda que em duas delas os resíduos indiquem a possibilidade de uma outra dimensão presente, interpretação essa que, aparentemente, se fortalece no instrumento como um todo.

Tomando-se o instrumento como um todo, os itens com

cargas positivas selecionados foram b1, b2, b3 e b4 e os com carga negativa foram os ab12, b10, b7, b11 e b9. No que concerne aos itens positivos, evidenciou-se por uma análise do desenho que b1 e b2 implicam em aprendizagem (nos termos de Spearman), pois a tarefa consiste simplesmente em repetir a mesma figura. Já os outros dois possuem um caráter de relação, já que evocam esse tipo de educação em sua resolução. Um fato interessante foi notado no item b4, pois as alternativas mais assinaladas depois da correta foram as 4 e 5; na primeira as crianças repetiram a figura situada no quadrante acima da que deveria ser completada, e a opção 5 representa uma réplica do desenho completo em miniatura.

Com relação aos itens negativos, todos solicitam da criança educação de correlatos, tratando-se, pois, de itens mais elaborados. Uma análise mais detida do tipo de erro apresentado por essas crianças permitiu identificar que elas tenderam a copiar o desenho imediatamente acima ou ao lado do espaço a ser completado. A exceção foi no item b7 no qual a alternativa 6 obteve uma alta frequência de respostas e não representava repetição de nenhum dos lados. Aparentemente, as crianças não conseguiram perceber como completar o raciocínio dessa situação, considerando apenas um princípio. Assim, em relação ao teste como um todo, o sistema que o modelo Rasch conseguiu produzir para colocar os itens na escala não conseguiu usar de forma mais ou menos proporcional a variância de todos os itens. A separação qualitativa mais aparente, pelas análises propostas por este estudo, parece indicar que as eduções de correlatos se diferencia das outras solicitações.

No que concerne à série A, os itens com cargas positivas selecionados foram a3, a4, a5 e a10 e os com carga negativa foram os a8, a9 e a11. Dos itens positivos, todos são referentes a aprendizagem exceto o 10 que implica educação de relação. Analisando os tipos de erros mais frequentes desse último, percebe-se que houve uma inversão de cores na escolha da alternativa errada, sugerindo que as crianças possam ter sido levadas pelas mesmas. Os itens negativos solicitaram educação de relações, com exceção do A11 que exige educação de correlatos. Nesse item, o erro mais cometido demonstra que as crianças se prenderam a um princípio e não observam os demais subjacentes. Novamente, a escala produzida pelo modelo Rasch não conseguiu usar de forma mais ou menos proporcional a variância de todos os itens da série A, ficando bastante clara a diferença entre os itens que solicitam raciocínio por educação de correlatos dos outros.

Os itens da Série AB analisados foram os desenhos ab2, ab5 e ab6 que forneceram carga positiva e ab9, ab10 e ab12, com cargas negativas. Dentre os positivos, encontrou-se que o item ab2 é caracteristicamente de aprendizagem, já ab5 de relação e ab6 requer educação de correlatos. Nesses dois últimos, as crianças que erraram tenderam a repetir o item imediatamente acima da figura. Em contrapartida, dos itens negativos, somente ab12 exigiu educação de correlatos, sendo que os outros eram problemas característicos de relação. Nos itens ab9 e ab10 os erros se deram devidos à repetição do desenho imediatamente superior ao espaço. A seu turno, no item ab12 a maior parte das crianças errou ou copiando a

figura de cima ou ao lado, ou seja, tenderam a considerar um princípio ou outro e não foram capazes de integrar os dois aspectos numa idéia correlata na solução da tarefa. De certa forma, a mescla e variedade das solicitações dos itens em relação aos grupos formados pela análise de componentes principais (negativo e positivo) pode ter possibilitado a unidimensionalidade dessa série, cujo resultado poderia ser interpretado no sentido de que o modelo Rasch retirou de forma bastante similar as variâncias de cada item.

Finalmente, os itens na série B que agruparam as cargas positivas escolhidos para análise foram b1, b2, b3, b4 e os com carga negativa foram os b7, b9, b10 e b11. Dos primeiros, verifica-se que os itens b1 e b2 implicam aprendizagem, enquanto b3 e b4 requerem educação de relações. Já dos itens negativos, todos possuem uma maior complexidade, exigindo educação de correlatos. Mais uma vez, a escala produzida pelo modelo Rasch não conseguiu usar de forma mais ou menos proporcional a variância de todos os itens da série A e os itens que solicitam educação de correlatos se diferenciam dos outros.

Algumas considerações podem ser feitas para uma maior elucidação das possíveis razões das violações à unidimensionalidade. No caso do subgrupo A, isso pôde ser explicado por meio de uma breve análise qualitativa dos itens. Deve-se considerar que para a resolução dos sete primeiros itens os sujeitos se valem, basicamente, de um processo de aprendizagem (no sentido de Spearman), na medida que todos seguem a mesma linha dos dois primeiros exemplos. No caso dos itens 8, 9 e 10, os problemas propostos parecem sugerir uma maior capacidade de educação de relações, já que existe uma variação para ser observada nas figuras. Finalmente, os itens 11 e 12 exigem capacidade de educar correlatos, dado que as crianças devem considerar dois tipos de variação ao mesmo tempo na sua solução. De fato dos 12 itens propostos, apenas dez são avaliados e, desses, metade implica em educação e metade, não. Ao lado disso, observou-se que os erros das questões de educação de correlatos se referiram a repetir a figura imediatamente acima ou do lado.

No subgrupo Ab, os itens 1, 2 e 3 exigem somente que o indivíduo reproduza as figuras colocadas (aprendizagem, no sentido de Spearman). Em contrapartida, os itens 4, 5, 7, 9, 10 e 11 exigem educação de relações e os itens 6, 8 e 12 educação de correlatos. Dessa forma, 9 itens solicitam o raciocínio por educação (seis de relação e três de correlatos) e três, não. Esse dado sugere que essa série está mais homogênea em relação à educação de relação, caracterizando uma série de mediana dificuldade. Ao lado disso, há que se considerar que a análise dos tipos de erros sugeriram que as crianças optaram por responder repetindo a figura imediatamente ou acima ou do lado, tal como foi observado na série anterior.

Finalmente, os itens que requerem aprendizagem no subgrupo B são 1 e 2, educação de relações os itens 3 e 4 e os demais, correlatos. Nesse contexto, os dados parecem sugerir que a educação de correlatos produz a separação, facilitando o aparecimento de uma outra dimensão do raciocínio, o que

vem sendo observado nas análises deste estudo. Aparentemente, trata-se de um tipo de raciocínio considerado mais difícil para a realização das crianças. Novamente, os erros cometidos seguiram o mesmo padrão das séries anteriores.

Para futuras revisões do instrumento, pode ser desejável criar diferentes escalas, ajustando-as conforme o tipo de relação implicada na sua solução. Ou talvez mudar a estrutura de cada série ou mudar a proporção dos três tipos de itens. Segundo Angelini et al. (1999), as Matrizes Progressivas Coloridas de Raven foram planejadas para medir a capacidade educativa. Dessa forma, os itens que avaliam aspectos relativos à aprendizagem poderiam ser re-examinados. Uma análise mais detida dos erros revelou que esses itens influenciaram o tipo de resposta das crianças em outros itens, pois uma boa parte das respostas erradas foram repetições seja das alternativas imediatamente acima, seja das do lado ou da diagonal.

No entanto, trata-se muito mais de sugestões para outros estudos, pois essas possibilidades não foram estudadas neste trabalho. De fato, essas análises sugerem que a unidimensionalidade do teste Matrizes Progressivas Coloridas de Raven carece de sustentação e a pontuação total, como é usada para a avaliação por esse instrumento, pode não ser uma medida suficiente das habilidades envolvidas para as crianças responderem ao instrumento.

Referências

- Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory*. Brooks/Cole, Monterey, CA.
- Anastasi, A. (1967). *Testes psicológicos*. São Paulo: Herder, Universidade de São Paulo.
- Angelini, A.L., Alves, I.C.B., Custódio, E.M., Duarte, W.F., & Duarte, J.L.M. (1999). *Manual: Matrizes Progressivas Coloridas de Raven*. São Paulo: Centro Editor de Testes e Pesquisas em Psicologia.
- American Psychological Association (1985). *Standards for educational and psychological tests*. Washington, DC: APA.
- Bartholomew, D.J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society*, 42, 293-321.
- Christofferson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40, 5-32.
- Craighead, W.E., Smucker, M.R., Craighead, L.W., & Ilardi, S.S. (1998). Factor analysis of the children's depression inventory in a community sample. *Psychological Assessment*, 10, 156-165.
- Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Endler, N.S., Parker, J.D.A., & Summerfeldt, L.J. (1998). Coping with health problems: developing a reliable and valid multidimensional measure. *Psychological Assessment*, 10, 195-205.
- Fischer, G.H. (1995). Derivations of the Rasch model. In G.H. Fischer & I.W. Molenaar (Eds.), *Rasch models: foundations, recent developments, and applications* (pp.15-38). New York: Springer Verlag.
- Glas, C.A.W. & Verhelst, N.D. (1995). Testing the Rasch model. In G.H. Fischer & I.W. Molenaar (Eds.), *Rasch models: foundations, recent developments, and applications* (pp.69-96). New York: Springer Verlag.

- Glas, C.A.W. (1988). The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika*, 53, 525–546.
- Green, S.B., Lissitz, R.W., & Mulaik, S.A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37, 827–838.
- Hattie, J.A. (1985). Methodology review: assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139–164.
- Jensen, A. (1998). *The G Factor: The Science of Mental Ability*. New Jersey: Praeger Mackintosh.
- Linacre, J.M. (1996). Structure in Rasch residuals: Why principal components analysis? *Rasch Measurement Transactions*, 10(3). Retrieved on 01/12/2003 from <http://www.rasch.org/rmt/rmt122m.htm>
- Lord, F.M. (1953). The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, 13, 517–548.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Erlbaum.
- McDonald, R.P. & Ahlwat, K.S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, 27, 82–99.
- Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551–560.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielson & Lydiche.
- Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.
- Smith, R.M. & Miao, C.Y. (1994). Assessing unidimensionality for Rasch measurement. In M. Wilson (Ed.), *Objective measurement: theory into practice* (pp.132–154). Norwood NJ: Ablex.
- Spearman, C. (1904). “General intelligence”, objectively determined and measured. *American Journal of Psychology*, 15, 201–293.
- Spearman, C. (1927). *Las habilidades del hombre: su naturaleza y medición*. Buenos Aires: Paidós, MacMillan.
- van den Wollenberg, A.L. (1982). Two new test statistics for the Rasch model. *Psychometrika*, 47, 123–140.
- van der Linden, W.J. & Hambleton, R.K. (Eds.) (1997). *Handbook of modern item response theory*. New York: Springer.
- Wright, B.D. (2000). Conventional factor analysis vs. Rasch residual factor analysis. *Rasch Measurement Transactions*, 14(2), 753. Retrieved on 01/12/2003 from <http://www.rasch.org/rmt/rmt142m.htm>.

Submissão: 13/05/2004
Última revisão: 27/10/2004
Aceite final: 15/12/2004